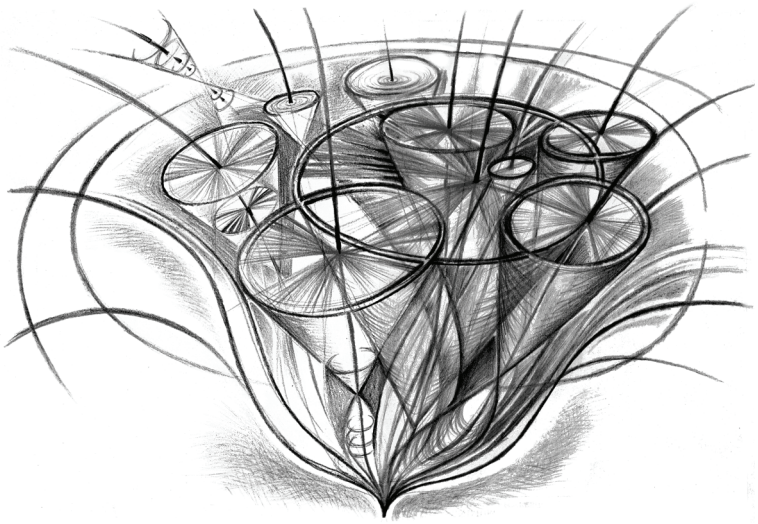
SELECTED TOPICS IN THEORETICAL PHYSICS II



Petr Kulhánek

Statistical Physics
Electromagnetic Field
Relativity

AGA 2026

Text © Petr Kulhánek
ISBN: 978-80-906638-3-1

Contents

FOREWORD 8

1. STATISTICAL PHYSICS 11

1.1 Essentials of Thermodynamics 12
1.1.1 First and Second Thermodynamic Law 13
1.1.2 Thermodynamic Potentials 14

1.2 Basic Concepts of Statistical Physics 18
1.2.1 Glossary of Terms 18
1.2.2 Ergodicity Problem 22
1.2.3 Liouville Theorem 23

1.3 Gibbs Canonical Ensemble 26
1.3.1 Derivation of the Distribution 26
1.3.2 Distribution Constants 27
1.3.3 Partition Function and Its Interpretation 30

1.4 Simple Examples 33
1.4.1 Ideal Gas 33
1.4.2 Particle in External Field 35
1.4.3 Classical Oscillator 40

1.5 Further Examples 42
1.5.1 Quantum Oscillator (Vibrator) 42
1.5.2 Quantum Rotator 47
1.5.3 Diatomic Gas 50
1.5.4 Anharmonic Oscillator 52
1.5.5 Two-Level System 54

1.6 Grandcanonical Ensemble 56
1.6.1 Derivation of the Distribution 56
1.6.2 Distribution Constants 57
1.6.3 Partition Function 58

1.7 Fermions and Bosons 61
1.7.1 Fermi-Dirac and Bose-Einstein Distributions 62
1.7.2 White Dwarf and Neutron Star as Fermion Gas 65
1.7.3 Planck Radiation Law for Photons 68

1.8 Fluctuations and Entropy 77
1.8.1 Fluctuations 77
1.8.2 Entropy 80

1.9 MAGNETICALLY ACTIVE SYSTEMS 83
1.9.1 Fundamental Concepts 83
1.9.2 Magnetically Active Materials 87
1.9.3 Lattice Models 93

2. ELECTROMAGNETIC FIELD 99

2.1 MAXWELL EQUATIONS 100
2.1.1 Charge Conservation Law 100
2.1.2 Lorentz Equation of Motion 101
2.1.3 Electric and Magnetic Dipole 102
2.1.4 Maxwell Equations 105
2.1.5 System Completeness 111
2.1.6 Constitutive Relations 112
2.1.7 Boundary Conditions 115

2.2 ELECTROSTATICS 119
2.2.1 Poisson Equation and Its Solution 119
2.2.2 Interesting Examples 122
2.2.3 Electric Field Energy 126
2.2.4 Multipole Expansion 128
2.2.5 Electric Dipole in External Field 132

2.3 MAGNETOSTATICS 133
2.3.1 Vector Potential 133
2.3.2 Poisson Equation and Its Solution 134
2.3.3 Biot-Savart Law 136
2.3.4 Multipole Expansion 136
2.3.5 Magnetic Dipole 138

2.4 CONSERVATION LAWS 141
2.4.1 Conservation of Energy 141
2.4.2 Conservation of Momentum 142
2.4.3 Field Pressure 146
2.4.4 Conservation of Angular Momentum 148
2.4.5 Field as a Real Entity 149

2.5 ELECTROMAGNETIC WAVES 150
2.5.1 Basic Concepts 150
2.5.2 Waves in Vacuum 156
2.5.3 Waves in Anisotropic Media 159
2.5.4 Waves in Conducting Media 162
2.5.5 Light at Interface of Two Media 163
2.5.6 Raytracing 169

2.6 ELECTROMAGNETIC POTENTIALS 171
2.6.1 Definition of Potentials 171
2.6.2 Potentials and Electromagnetic Field Tensor 172
2.6.3 Maxwell Equations in Potentials 173
2.6.4 Potentials in Free Space 175
2.6.5 Retarded and Advanced Potentials 176

2.7 Radiation of Charged Particles 178
2.7.1 Radiative Multipole Expansion 178
2.7.2 Electric Dipole Radiation 181
2.7.3 Thomson Scattering 184
2.7.4 Rayleigh Scattering 186
2.7.5 Bremsstrahlung of Electrons on Ions 187
2.7.6 Cyclotron Radiation 189
2.7.7 Higher Multipole Radiation 191
2.7.8 Liénard-Wiechert Potentials 192
2.7.9 Relativistic Bremsstrahlung and Synchrotron Radiation 195

2.8 Field in Cavity 199
2.8.1 Cavity Potentials 199
2.8.2 Natural Field Oscillations 200
2.8.3 Energy and Momentum of Cavity Field 201
2.8.4 Hamilton Equations 203
2.8.5 Thermal Radiation 205

3. RELATIVITY 209

3.1 Special Relativity 210
3.1.1 Lorentz Transformation 210
3.1.2 Minkowski Metric 219
3.1.3 Relativistic Dynamics 230
3.1.4 Relativistic Paradoxes 233

3.2 Introduction to General Relativity 236
3.2.1 Equivalence Principle 236
3.2.3 Covariant Derivative 240
3.2.4 Geodesic Equation 244
3.2.5 Newtonian Limit 248

3.3 Time Dilation and Gravitational Redshift 252
3.3.1 Different Views on Time Dilation 254
3.3.2 Pound-Rebka Experiment 256
3.3.3 Hafele-Keating Experiment 257
3.3.4 Gravity Probe A 259
3.3.5 Berkeley Experiment 260

3.4 Einstein Field Equations 262
3.4.1 Riemann Curvature Tensor 262
3.4.2 Continuity Equation in Classical Physics 268
3.4.3 Stress-Energy Tensor 271
3.4.4 Einstein Field Equations 273

3.5 Gravitational Waves 278
3.5.1 Wave Equation 278
3.5.2 Properties of Gravitational Waves 279
3.5.3 Indirect Observations of Gravitational Waves 283
3.5.4 Detection of Gravitational Waves 284
3.5.5 Future Detectors 288

3.6 Spherical Source of Gravitation 292
3.6.1 Schwarzschild Solution 292
3.6.2 Embedding Diagrams 298
3.6.3 Black Holes 301
3.6.4 Motion in the Schwarzschild Geometry 304
3.6.5 Gravitational Lenses 312
3.6.6 Frame-Dragging by a Rotating Body 315

3.7 Brief Introduction to Cosmology 322
3.7.1 Friedmann Metric 322
3.7.2 Properties of the Friedmann Metric 326
3.7.3 Cosmological Redshift 328
3.7.4 Scale Factor 330
3.7.5 Cosmological Constant 332

LIST OF SYMBOLS 335

BIOGRAPHICAL INDEX 341
Statistical Physics 342
Electromagnetic Field 349
Relativity 358

SUBJECT INDEX 371

REFERENCES 374

INSIDE OTHER VOLUMES 379
Selected Topics in Theoretical Physics I 381
Selected Topics in Theoretical Physics III 383



Foreword

Dear Readers,

You are holding the second volume of "Selected Topics." Its structure remains identical to that of the first volume. The only difference is that all the mathematics required for study has been intentionally consolidated into the first volume. As a result, you no longer need to search for the mathematical minimum scattered across various appendices, as was the case in the previous edition; instead, you will find it concentrated in one single place. I hope that you will find this approach significantly more practical.

The second volume of "Selected Topics" is dedicated to statistical physics, the electromagnetic field, and relativity. In the statistical physics section, the reader will be introduced to the description of systems comprising a large number of particles that form a **statistical ensemble**. Due to the immense number of individual constituents of the same kind, a complete description of all trajectories at the level of theoretical mechanics is impossible, even on the world's largest computers. Therefore, the description relies on a statistical approximation, where we infer the properties of a large ensemble based on the behavior of a single constituent. The central tool in the statistical description of natural processes is the **partition function**. The fundamental statistical distributions are covered both in classical physics and, for quantum mechanics, the Fermi-Dirac and Bose-Einstein distributions. This knowledge is then applied, for instance, to electron and photon gases, resulting in, among other things, a complete derivation of **Planck law of blackbody radiation**. The application of the statistical approach is also demonstrated on the rotational and vibrational states of molecules, and the fundamentals of lattice models of ferromagnets are discussed as well. The student will encounter non-equilibrium statistics later, in the third volume of "Selected Topics."

The second part of this textbook is devoted to the electromagnetic field. Electricity and magnetism are two sides of the same coin. A changing magnetic field induces an electric field, and vice versa. Much time has passed since 1873, when **James Clerk Maxwell** published his monumental work, "*A Treatise on Electricity and Magnetism*." Maxwell electrodynamics predicted the existence of many phenomena that, in just a single century, have transformed our civilization beyond recognition. Today, we can hardly imagine our lives without electrical and electronic devices. Electromagnetic phenomena also play a vital role in cosmic processes – dust in nebulae electrostatically coalesces into the seeds from which planets form, stars are giant spheres of plasma with pronounced electromagnetic manifestations, and magnetic fields permeate the universe across all scales. Furthermore, our technologies do not utilize electricity and magnetism solely in electrical machines and electronic devices. The rapid rise of plasma physics has brought entirely unprecedented applications – the deposition of ultra-hard coatings, fabric treatments, plasma cutting and machining, and **plasma medicine**, which originated in recent years. A substantial portion of the text on electromagnetism is dedicated to the radiation of charged particles, which is associated with changes in the momentum of a charged particle, whether in the direction of motion (e.g., **bremsstrahlung**) or perpendicular to the trajectory (e.g., cyclotron and synchrotron radiation).

However, relatively commonplace forms of radiation-matter interaction are also covered, such as Thomson scattering or Rayleigh scattering, the latter being responsible for

the blue color of the sky. The culmination of the electrodynamics section is the description of the electromagnetic field in a resonant cavity as a system of harmonic oscillators that satisfy the classical Hamilton equations.

The final part of this textbook is dedicated to relativity. Maxwell's electrodynamics demonstrated that velocities transform differently than they do in classical mechanics. Modifying mechanics to satisfy the requirements of electrodynamics led to special relativity. Albert Einstein further generalized it to arbitrary coordinate systems, thereby giving rise to general relativity, which is our current, highly successful theory of gravitational interaction. General relativity has yielded a whole range of previously unimaginable phenomena – the confinement of spacetime into a black hole, the bending of light by gravitational lensing, gravitational waves, the **frame-dragging** effect caused by a rotating body, and the existence of many other processes that, even more than a century after the birth of this theory, still take one's breath away. Although we will only cover the basics of general relativity, even these may captivate the reader enough to embark on a deeper study of this beautiful physical theory.

Variables across all three volumes are strictly represented in **italic type**. Functions, abbreviations, numerals, and various mathematical operations are displayed in regular (roman) type. Vectors, tensors, and composite objects are typeset in **bold type**. In exceptional cases where ambiguity or confusion might arise, arrows are placed above the vectors and tensors. Latin indices denote the sequence of a quantity, coordinate axes, etc. Greek letters are used to label the components of four-vectors, for example A_α, where $\alpha = 0$ (the timelike component) and 1, 2, 3 (the spacelike components).

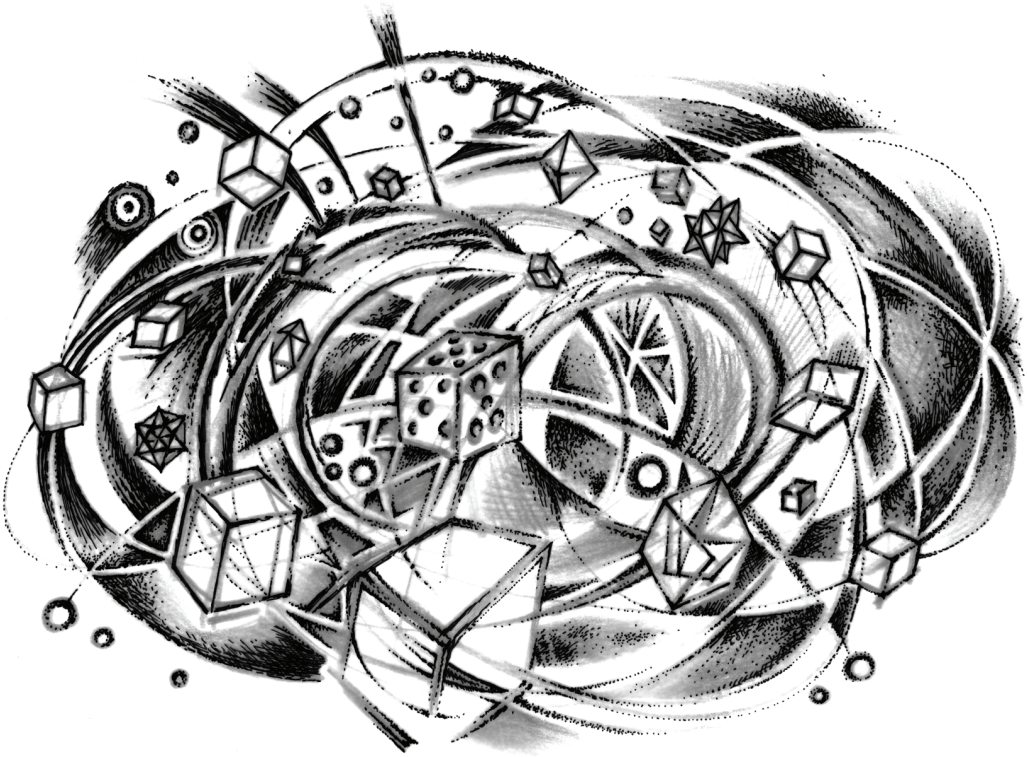
Since the number of letters in the alphabet is limited, some quantities are designated by the same symbol. However, their meaning can easily be deduced from the context. The list of symbols included at the end of the book can also assist with this. When reading, do not skip the footnotes and remarks, as they often contain important insights necessary for understanding the phenomenon under discussion; in the book, you will find them on a gray background. Illustrative examples are separated from the rest of the text by a black half-circle at the beginning and the end. Important relations are marked on the left side with a black triangle. Hopefully, this notation will help the reader navigate this demanding textbook more easily.

What to say in conclusion? My thanks go to the many students of mine who studied from the texts that formed the basis of this book and who meticulously spotted typos and ambiguities. My thanks also go to Ing. arch. Ivan Havlíček, who took care of the introductory graphics for individual chapters, the cover, and several illustrations. I would like to wish future students that they experience the unique and incommunicable vertigo that comes from grasping and understanding the laws describing the processes occurring in the nature around us.

Petr Kulhánek, 2026, Prague

On the **aldebaran.cz** website, under the Study section, you can download an electronic version of this textbook and other supplementary materials for the lectures.

1. Statistical Physics



1.1 Essentials of Thermodynamics

To describe the behavior of a large ensemble of many identical systems (a classic example is a gas composed of many identical molecules), essentially only four approaches can be used:

1) The behavior is declared a "divine miracle" and is not investigated further.
2) On the basis of simple experimental results, laws governing the ensemble are sought. For instance, pressure in a closed container increases with increasing temperature or an increasing number of particles, whereas gas pressure decreases if the dimensions of the container are increased. A combination of these relations yields the *state equation*. However, it must always be kept in mind that this derivation is based on experiments without investigating the nature of the phenomena themselves, and the exact limits of validity for these derived laws may not always be fully understood. This approach is the subject of *thermodynamics* and is called descriptive, empirical, or *phenomenological*.
3) An attempt is made to calculate the trajectory of each individual particle constituting the ensemble from fundamental laws, such as Hamilton equations. This procedure necessarily fails for ensembles of many particles, where such a description is beyond human capability. However, various numerical methods can be used, such as not describing all systems in the ensemble, and so on.
4) Laws describing the behavior of the ensemble as a whole are derived theoretically from knowledge of the behavior of individual system components using statistical methods. The obtained results are probabilistic in nature, but for ensembles of many particles, this is by no means a disadvantage – rather the contrary. This approach is the subject of *statistical physics*.

The ensemble of systems described by statistical physics methods can be highly diverse. It can be a simple monoatomic gas, a neutron star composed of neutrons, or a ferromagnet composed of many elementary magnets (spins). The systems of the described ensemble can be both classical and quantum. The extraordinary importance of the harmonic oscillator was already mentioned in quantum theory, and therefore, an ensemble of quantum harmonic oscillators is also examined in this section.

Simple ideal gas laws and the state equation are derived, and the concept of entropy is introduced from a statistical perspective. Likewise, quantum systems of fermions and bosons or many elementary quantum rotators (for example, rotating molecules) are studied. Statistical physics represents an exceptionally beautiful and elegant part of physics, with the origins of which names of such giants as Ludwig Boltzmann or Josiah Gibbs, or in quantum statistics, Enrico Fermi, Paul Dirac, Satyendra Bose, and Albert Einstein, are associated.

Before proceeding to the actual construction of statistical physics, familiarity with some basic concepts of thermodynamics is required. This section is by no means a systematic exposition of thermodynamics. It is merely an overview of certain concepts from this field needed to compare the results of thermodynamics and statistical physics. In thermodynamics, the term "*described system*" is used. In statistical physics, a strict

distinction is made between a *system* (a single described entity) and an *ensemble* (a large number of these entities).

For the study of this section, knowledge of Pfaffian differential forms is required. Their fundamentals can be found in "*Selected Topics I*", Section 3.9.

1.1.1 First and Second Thermodynamic Law

The first thermodynamic law is nothing other than the law of conservation of system energy: *The internal energy of a system can increase by supplied heat or by adding further particles, and decrease by work performed by the system.*

$$\mathrm{d}U = \mathrm{d}Q - \mathrm{d}A + \mathrm{d}U_N \,. \tag{1.1}$$

The internal energy of a system itself is an exact differential. However, the terms on the right-hand side are not exact differentials. This fact follows from numerous experiments. The supplied heat $\int \mathrm{d}Q$ varies along different paths in the phase space (p, V, T) and thus depends on the path. A similar situation applies to the other terms on the right-hand side. The individual terms are examined below in more detail:

Work Performed by System (dA)

Work performed by a system can be of various natures: mechanical, electrical, magnetic, polarization, elastic, etc., and the resulting expression is the sum of many terms. For now, however, only the expression for mechanical work is considered; electrical and magnetic terms are discussed later

$$\mathrm{d}A = F\,\mathrm{d}l + \cdots = pS\,\mathrm{d}l + \cdots = p\,\mathrm{d}V + \cdots \tag{1.2}$$

Internal Energy Caused by Change in Particle Number (dU_N)

If additional particles enter the system from the outside, the internal energy of the system increases proportionally to the particle increment:

$$\mathrm{d}U_N = \mu\,\mathrm{d}N \,. \tag{1.3}$$

The proportionality coefficient μ is called the *chemical potential* of the system and depends on the type of substance of which the system consists. From a mathematical point of view, this is not a true potential, and the name has merely a historical origin. If the system contains multiple particle species, the increment of internal energy associated with a change in the particle number is given by the sum of similar terms over all particle species (the summation convention is applied):

$$\mathrm{d}U_N = \mu_k\,\mathrm{d}N_k \,. \tag{1.4}$$

Heat (dQ)

Heat is one of the central concepts of thermodynamics; therefore, it is particularly inconvenient that it does not take the form of an exact differential. Fortunately, it can be shown that an integrating factor always exists, transforming heat into a differential form that is an exact differential. This statement represents the core of the *second thermodynamic law*, which appears in many formulations. For the present purposes, the most

important form is stated as follows: *The differential of heat possesses an integrating factor, which is the reciprocal value of the absolute temperature. The newly formed exact differential form is called entropy and is denoted as* dS:

$$\mathrm{d}S \equiv \frac{1}{T}\,\mathrm{d}Q\,. \tag{1.5}$$

- Other formulations of the second thermodynamic law exist and carry profound significance for thermodynamics, for instance: *A perpetual motion machine of the second kind (a machine continuously and cyclically performing work by cooling a thermal bath) cannot exist*. The relationship between these two formulations can be found in any thermodynamics textbook.
- The correct integrating factor can be reached, for example, by analyzing a Carnot cycle, where the integration of dQ is shown to *depend* on the integration path, whereas the integration of the quantity dQ/T is *path-independent* and therefore constitutes an exact differential.
- The fact that the reciprocal value of temperature is the correct integrating factor for the differential of heat can also be demonstrated by comparing the cross-derivatives of the entropy differential.
- In the thermodynamics of equilibrium processes, entropy has the meaning of a heat differential corrected by an integrating factor to become an exact differential. For entropy, all statements of the theorem of five equivalences hold: the integral of entropy is path-independent, the integral over a closed curve (a cyclic process) is zero, etc. Therefore, preference is always given to entropy, and instead of the differential of heat, the expression is written as:

► $$\mathrm{d}Q = T\,\mathrm{d}S\,. \tag{1.6}$$

After expressing all quantities on the right-hand side of the first thermodynamic law (1.1), the form to be utilized is obtained as:

► $$\mathrm{d}U = T\,\mathrm{d}S - p\,\mathrm{d}V + \mu_k\,\mathrm{d}N_k\,. \tag{1.7}$$

In the last term, the summation convention is applied; the summation is performed over all particle species whose number can change (for example, electrons, neutral particles, and ions in a plasma).

1.1.2 Thermodynamic Potentials

In the previous section, two quantities forming exact differential forms were introduced: internal energy and entropy. However, a procedure exists by which a whole series of other, highly useful exact differential forms can be constructed. Enthalpy, Helmholtz free energy, Gibbs free energy (potential), and grandcanonical potential are discussed below. The significance of these quantities for statistical physics is highly important.

Enthalpy *H*

The right-hand side of the internal energy differential (1.7) is made exact in the $p\,\mathrm{d}V$ term. It is written as $\mathrm{d}(pV) - V\mathrm{d}p$, and the first term is moved to the left-hand side. A new quantity in the form of an exact differential, termed enthalpy, is thus obtained:

$$\mathrm{d}U = T\,\mathrm{d}S - p\,\mathrm{d}V + \mu_k\,\mathrm{d}N_k\,,$$

$$\mathrm{d}U = T\,\mathrm{d}S - \mathrm{d}(pV) + V\,\mathrm{d}p + \mu_k\,\mathrm{d}N_k\,,$$

$$\mathrm{d}(U + pV) = T\,\mathrm{d}S + V\,\mathrm{d}p + \mu_k\,\mathrm{d}N_k\,.$$

Many relations can be written for the newly introduced quantity. Its definition (in the parentheses on the left) and its first differential (the right-hand side of the equality) are known. From the form of the first differential, the variables on which enthalpy depends can be identified. Furthermore, as it is an exact differential, its partial derivatives yield the coefficients of the differential form on the right-hand side:

$$H \equiv U + pV\,,$$

$$\mathrm{d}H = T\,\mathrm{d}S + V\,\mathrm{d}p + \mu_k\,\mathrm{d}N_k\,,$$

$$H = H(S, p, N_k)\,, \tag{1.8}$$

$$T = \left(\frac{\partial H}{\partial S}\right)_{p,N_k}, \qquad V = \left(\frac{\partial H}{\partial p}\right)_{S,N_k}, \qquad \mu_k = \left(\frac{\partial H}{\partial N_k}\right)_{S,p,N_{i\neq k}}.$$

From knowledge of enthalpy, the temperature, volume, and chemical potentials of the system can be determined. From the first relation (1.8), the internal energy of the system can also be determined using enthalpy:

$$U = H - pV = H - p\left(\frac{\partial H}{\partial p}\right). \tag{1.9}$$

This relation is called a *Gibbs-Helmholtz equation of the first kind*. From the internal energy, heat capacities of the system and other quantities can then be calculated.

Note 1: The subscripts in the partial derivatives indicate that the respective quantities are kept constant during differentiation, which is common practice. In thermodynamics, where processes depend on the path, this path is usually explicitly indicated. In this text, it is automatically understood that the quantities with respect to which the derivation is not performed are constant, and subscripts are omitted hereafter.

Note 2: The letter H stands for the Greek capital letter eta (associated with the word *enthalpy*).

Helmholtz Free Energy *F*

The procedure is analogous to that for enthalpy, but now the $T\,\mathrm{d}S$ term in the internal energy expression (1.7) is made exact:

$$\mathrm{d}U = T\,\mathrm{d}S - p\,\mathrm{d}V + \mu_k\,\mathrm{d}N_k\,,$$

$$\mathrm{d}(U - TS) = -S\,\mathrm{d}T - p\,\mathrm{d}V + \mu_k\,\mathrm{d}N_k\,.$$

Again, a new exact differential form, termed *Helmholtz free energy*, is obtained. Just as with enthalpy, a series of relations can immediately be written alongside the definition:

$$F \equiv U - TS\,,$$

$$\mathrm{d}F = -S\,\mathrm{d}T - p\,\mathrm{d}V + \mu_k\,\mathrm{d}N_k\,,$$

$$F = F(T, V, N_k)\,, \tag{1.10}$$

$$S = -\left(\frac{\partial F}{\partial T}\right), \qquad p = -\left(\frac{\partial F}{\partial V}\right), \qquad \mu_k = \left(\frac{\partial F}{\partial N_k}\right).$$

Free energy carries extraordinary importance in statistical physics. It will be demonstrated that this quantity can be determined on the basis of statistical considerations. Free energy is a function of easily visualizable quantities: temperature, system volume, and the number of particles of various species. If this function is known, the entropy of the system – the intuitive understanding of which often poses difficulties – is easily determined by simple differentiation. Differentiating free energy with respect to volume yields the pressure in the system, hence the state equation, and differentiating with respect to the particle number determines the chemical potentials of the system. Nothing more could be desired. Perhaps the internal energy, but even that presents no issue. From the definition of free energy $F = U - TS$, the relation $U = F + TS$ is obtained, and the already calculated value of entropy is substituted:

$$U = F - T\left(\frac{\partial F}{\partial T}\right). \tag{1.11}$$

This relation is called a *Gibbs-Helmholtz equation of the second kind*, serving as the starting point for many other derived quantities calculated from internal energy.

◖ **Example 1.1:** Prove that the Gibbs-Helmholtz equation (1.11) can be written in the simple form $U = \partial(\beta F)/\partial\beta$, where $\beta \equiv 1/T$. ◗

Gibbs Potential *G*

By exactly the same procedure as in the previous cases, the internal energy differential is made exact in both the $p\,\mathrm{d}V$ and $T\,\mathrm{d}S$ terms. The result is a new quantity, the Gibbs potential (frequently termed as Gibbs free energy), for which the following holds:

$$G \equiv U - TS + pV\,,$$

$$\mathrm{d}G = -S\,\mathrm{d}T + V\,\mathrm{d}p + \mu_k\,\mathrm{d}N_k\,,$$

$$G = G(T,\, p, N_k), \tag{1.12}$$

$$S = -\left(\frac{\partial G}{\partial T}\right), \qquad V = \left(\frac{\partial G}{\partial p}\right), \qquad \mu_k = \left(\frac{\partial G}{\partial N_k}\right).$$

The Gibbs potential (or Gibbs free energy) is of great importance for chemistry and thermodynamics; it is not utilized in this statistical physics text. The parentheses mean that all other quantities (except the one with respect to which the derivative is taken) are held constant; subscripts are no longer written.

Grandcanonical Potential Ω

The last potential carrying great significance for the present purposes is the grandcanonical potential. The internal energy differential is made exact in the $T\,\mathrm{d}S$ and $\mu\,\mathrm{d}N$ terms. The result is:

$$\Omega \equiv U - T\,S - \mu_k N_k\,,$$

$$\mathrm{d}\Omega = -\,S\,\mathrm{d}T - p\,\mathrm{d}V - N_k\,\mathrm{d}\mu_k\,,$$

$$\Omega = \Omega(T, V, \mu_k), \tag{1.13}$$

$$S = -\left(\frac{\partial \Omega}{\partial T}\right), \qquad p = -\left(\frac{\partial \Omega}{\partial V}\right), \qquad N_k = -\left(\frac{\partial \Omega}{\partial \mu_k}\right).$$

In the statistical physics of ensembles with a variable number of particles, the grandcanonical potential can be determined, at least theoretically. From it, the system entropy, pressure (the state equation), and the numbers of individual particles are easily found. From the definition of the grandcanonical potential, the internal energy is then calculated

$$U = \Omega + TS + \mu_k N_k = \Omega - T\left(\frac{\partial \Omega}{\partial T}\right) - \mu_k\left(\frac{\partial \Omega}{\partial \mu_k}\right), \tag{1.14}$$

which serves as the basis for calculating many other quantities. For a deeper study of thermodynamics, the classic and comprehensive textbook by Herbert Callen [3] can be consulted. Highly detailed textbooks on statistical physics by Mehran Kardar [4] or Pathria and Beale [6] also exist. Further information on statistical physics can be found in publications [5], [7], and [8].

1.2 Basic Concepts of Statistical Physics

1.2.1 Glossary of Terms

System

A system is defined as any described entity. A set of independent parameters necessary for the description is called generalized coordinates. Standard procedures of theoretical mechanics assign a generalized momentum to each generalized coordinate. If the initial values of coordinates and momenta are known, the trajectory of the system in classical physics can be predicted using Hamilton equations. In the case of a quantum system, the state is given by a vector in a Hilbert space, and the time evolution is determined by the action of an evolution operator. In this case, the trajectory is probabilistic in nature and is quantum mechanically "blurred."

Phase Space

Phase space is defined as the space of generalized coordinates and momenta (q, p). If subscripts are omitted, the entire sets of all generalized coordinates and momenta are automatically implied. Several reasons justify the use of momenta instead of velocities:

- To track time evolution, Hamilton equations for positions and momenta are used.
- The world is quantum mechanically blurred at the elementary level due to Heisenberg uncertainty relations. These relations hold between a generalized coordinate and its corresponding generalized momentum.
- If a situation is symmetric with respect to a displacement in a certain generalized coordinate, the corresponding generalized momentum is conserved. These two quantities inherently belong together.
- Poisson brackets of generalized coordinates and corresponding generalized momenta equal one; other brackets are zero.
- In statistical physics, volume is conserved during the statistical evolution of many systems in a phase space with (q, p) axes, and the ensemble of systems behaves like an incompressible fluid (Liouville theorem), which is highly advantageous for the description.

Ensemble

An ensemble is defined as a large number of identical systems with the same phase space. The systems can have different initial conditions and are represented in the phase space at a given moment by a set of many points. During time evolution, the individual points move along their phase trajectories given by Hamilton equations.

Phase Volume

In phase space, an elementary volume can be introduced in a standard way as a $2f$-dimensional differential (where f is the number of degrees of freedom):

$$\mathrm{d}\phi = \mathrm{d}q_1 \cdots \mathrm{d}q_f \, \mathrm{d}p_1 \cdots \mathrm{d}p_f = \prod_k \mathrm{d}q_k \prod_k \mathrm{d}p_k \;=\; \mathrm{d}^f q \; \mathrm{d}^f p \,. \tag{1.15}$$

This expression naturally generalizes the "ordinary" volume in three dimensions, where $\mathrm{d}V = \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z$. A finite region Ω of the phase space then possesses a volume:

► $$\Delta\phi = \int_{\Omega} \mathrm{d}\phi = \int_{\Omega} \mathrm{d}^f q \; \mathrm{d}^f p \;. \tag{1.16}$$

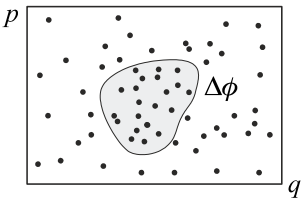


Fig. 1.1: Phase volume with one spatial and one momentum coordinate.

◖ Example 1.2: Phase Volume of Oscillator

Determine the phase volume occupied in the phase space by harmonic oscillators whose maximum energy is E. Also, determine the volume occupied by a single energy quantum of a harmonic oscillator.

Solution: Theoretical mechanics shows that the phase trajectory of a harmonic oscillator is an ellipse. Since a harmonic oscillator conserves energy during its motion, the equation of the ellipse can be written as the equation of an energy "hypersurface":

$$H \equiv \frac{p^2}{2m} + \frac{m\omega^2 x^2}{2} = E \,. \tag{1.17}$$

Oscillators possessing an energy exactly equal to E are located on this hypersurface (in this case, an ordinary ellipse). Lower energy trajectories lie inside the ellipse, whereas higher energy ones lie outside. The energy hypersurface defines a set (including the boundary) in the phase space where oscillators with energy less than or equal to E are located. The phase volume of this set is given by:

$$\phi = \int_{\Omega} \mathrm{d}x\,\mathrm{d}p \,; \qquad \Omega = \left\{ (x,p) : \; \frac{m\omega^2 x^2}{2} + \frac{p^2}{2m} \le E \right\}. \tag{1.18}$$

If substitutions $\xi = (m\omega^2/2)^{1/2} x$ and $\eta = p/(2m)^{1/2}$ are performed, the integration region simplifies to a circle (a two-dimensional ball):

$$\phi = \frac{2}{\omega} \int_{\Omega} \mathrm{d}\xi \, \mathrm{d}\eta \,; \qquad \Omega = \left\{ (\xi,\eta) : \; \xi^2 + \eta^2 \le E \right\}. \tag{1.19}$$

The result of the integral is the volume of a 2D ball $V_2(E^{1/2})$, which in this case represents the ordinary area of a circle with radius $E^{1/2}$:

$$\phi = 2\pi E / \omega \,. \tag{1.20}$$

The energy spectrum of a harmonic oscillator jumps by quanta $\Delta E = \hbar\omega$; therefore, the size of the phase space corresponding to one quantum is given by:

$$\Delta\phi = 2\pi\,\Delta E/\omega = 2\pi\hbar\omega/\omega = 2\pi\hbar\,. \tag{1.21}$$

The value $2\pi\hbar$ represents the size of a single (1D) state in the phase space. Simultaneously, it constitutes the "pixel" of blurring of the phase trajectory in quantum theory.

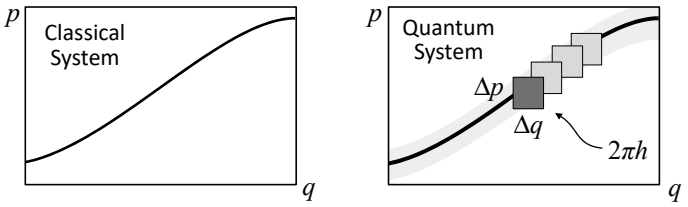


Fig. 1.2: Volume of state in phase space.

Example 1.3: Phase Volume of Molecule

Calculate the phase volume bounded by the energy hypersurface of a molecule behaving as a rigid dumbbell.

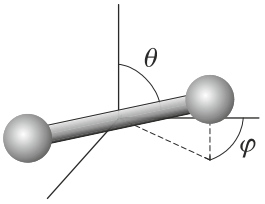


Fig. 1.3: Molecule as rigid dumbbell.

Solution: The molecule possesses 5 degrees of freedom. It can move as a whole (described by center-of-mass coordinates x, y, z) and rotate in two independent angles θ and φ of spherical coordinates. The phase space thus has 5 coordinate and 5 momentum axes, making it ten-dimensional. A molecule of mass m is always located on the energy hypersurface (each atom has a mass of $m/2$ at a distance of $l/2$ from the center of mass):

$$E = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2 + \dot{z}^2\right) + \frac{1}{2}m\left(\frac{l}{2}\right)^2\left(\dot{\theta}^2 + \sin^2\theta\,\dot{\varphi}^2\right) = \text{const.}$$

The Hamiltonian is determined by a standard procedure:

$$H = \frac{p_x^2}{2m} + \frac{p_y^2}{2m} + \frac{p_z^2}{2m} + \frac{2}{ml^2}\left(p_\theta^2 + \frac{p_\varphi^2}{\sin^2\theta}\right).$$

The calculation of the volume of the phase space region enclosed by the energy hypersurface can now be initiated:

$$\phi = \int\limits_{H \le E} \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}\varphi\,\mathrm{d}\theta\,\mathrm{d}p_x\,\mathrm{d}p_y\,\mathrm{d}p_z\,\mathrm{d}p_\varphi\,\mathrm{d}p_\theta \;\Rightarrow$$

$$\phi = \int_V \mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \int_0^{2\pi} \mathrm{d}\varphi \int_0^{\pi} \sin\theta\,\mathrm{d}\theta \int_{p_1^2+\cdots p_5^2 \le E} \sqrt{2} m^{5/2} l^2\ \mathrm{d}^5 p \ \Rightarrow$$

$$\phi = 2^{5/2}\pi V m^{5/2} l^2 E^{5/2} V_5(1)\,.$$

In the second line, standard substitutions were performed, analogously to the previous example. The quantity $V_5(1)$ is the volume of a unit five-dimensional ball.

◗

Weight Factor

Each degree of freedom is blurred in quantum theory with a "pixel" value of $2\pi\hbar$ occupied by one quantum state (this value applies generally, not only to the quantum oscillator for which it was derived). In multiple dimensions, the size of this fundamental "pixel," or the size of one quantum state, is $(2\pi\hbar)^f$, representing a simple product of the individual dimensional "pixels." A dimensionless phase volume can be highly advantageously introduced by the expression:

► $$\mathrm{d}\Gamma \equiv \frac{\mathrm{d}\phi}{(2\pi\hbar)^f}\ ; \qquad \Delta\Gamma = \int_\phi \frac{\mathrm{d}^f q\ \mathrm{d}^f p}{(2\pi\hbar)^f}\,. \tag{1.22}$$

This dimensionless quantity is called a weight factor and is utilized primarily for quantum systems. It represents the phase volume divided by the size of one quantum state; thus, Γ has the meaning of the number of quantum states contained within the phase volume ϕ.

Probability Density

If an ensemble contains a large number of systems, a *system number density* in an element of phase volume can be introduced as $\Delta N/\Delta\phi$. The higher this number, the more systems are located at that position, and the higher the probability of finding a system in that region. The *probability density* is proportional to the system number density in the phase space. For normalization reasons, the probability density is sometimes divided by the total number of particles. It is also possible to utilize the dimensionless phase volume (weight factor); hence, a total of 4 options exist for introducing and normalizing the probability density:

$$\rho \equiv \frac{\mathrm{d}N}{\mathrm{d}\phi}, \qquad \mathrm{d}w = \rho\,\mathrm{d}\phi, \qquad \int \mathrm{d}w = \int \frac{\mathrm{d}N}{\mathrm{d}\phi}\mathrm{d}\phi = N\,; \tag{1.23}$$

► $$\rho \equiv \frac{1}{N}\frac{\mathrm{d}N}{\mathrm{d}\phi}, \qquad \mathrm{d}w = \rho\mathrm{d}\phi, \qquad \int \mathrm{d}w = \int \frac{1}{N}\frac{\mathrm{d}N}{\mathrm{d}\phi}\mathrm{d}\phi = 1\,; \tag{1.24}$$

$$\rho \equiv \frac{\mathrm{d}N}{\mathrm{d}\Gamma}, \qquad \mathrm{d}w = \rho\mathrm{d}\Gamma, \qquad \int \mathrm{d}w = \int \frac{\mathrm{d}N}{\mathrm{d}\Gamma}\mathrm{d}\Gamma = N\,; \tag{1.25}$$

► $$\rho \equiv \frac{1}{N}\frac{\mathrm{d}N}{\mathrm{d}\Gamma}, \qquad \mathrm{d}w = \rho\mathrm{d}\Gamma, \qquad \int \mathrm{d}w = \int \frac{1}{N}\frac{\mathrm{d}N}{\mathrm{d}\Gamma}\mathrm{d}\Gamma = 1\,. \tag{1.26}$$

The total probability is normalized either to one (as is most common in mathematics) or to the total number of particles. In this textbook, the second and fourth options of probability density normalization (to one) are utilized.

Phase Space Averaging

If the probability density is known, dynamic variables can be averaged over the phase space. The values of the dynamic variable A for all systems are summed with a weight given by the probability density ρ:

$$\langle A \rangle \equiv \int A(q,p)\,\mathrm{d}w = \int A(q,p)\,\rho\,\mathrm{d}\phi;$$

$$\langle A \rangle \equiv \int A(q,p)\,\mathrm{d}w = \int A(q,p)\,\rho\,\mathrm{d}\Gamma; \tag{1.27}$$

$$\langle A \rangle \equiv \sum A_n w_n .$$

The first case applies if the phase space is used, the second if the weight factor is used, and the third for a system with discrete, quantized states, where summation is simply performed over the corresponding probabilities of individual states. The probabilities satisfy the normalization condition:

$$\int \mathrm{d}w = 1; \qquad \sum_n w_n = 1. \tag{1.28}$$

1.2.2 Ergodicity Problem

The mean value of a dynamic variable $A(q,p)$ in an ensemble can in principle be determined in two ways. The first option is averaging over the ensemble using the introduced probability density:

$$\langle A \rangle \equiv \int A(q,p)\mathrm{d}w = \int A(q,p)\,\rho\,\mathrm{d}\Gamma . \tag{1.29}$$

The second option is to choose one of the systems of the ensemble and average the quantity A over a sufficiently long time interval τ:

$$\overline{A} \equiv \lim_{\tau\to\infty} \frac{1}{\tau} \int_{t_0}^{t_0+\tau} A\big(q(t),p(t)\big)\,\mathrm{d}t . \tag{1.30}$$

Naturally, the result of the limit should not depend on the initial time t_0. A highly debated and very old question is whether both methods yield the same result:

$$\langle A \rangle \overset{?}{=} \overline{A} . \tag{1.31}$$

This problem is called the ergodicity problem; it is resolved positively in many individual cases, but a general solution for mechanical systems is not known. In this text, variables are averaged using relation (1.29), assuming that averaging via (1.30) would lead to the same result.

1.2.3 Liouville Theorem

Flow is generally described by the density and flux of some additive quantity A. This quantity can be the flux of mass, charge, heat, energy, or similar. The density and flux are defined by the relations:

$$\rho = \lim_{\Delta V \to 0} \frac{\Delta A}{\Delta V}, \qquad \mathbf{j} = \rho \mathbf{u} \tag{1.32}$$

and form a relativistic four-vector (ρ, $\mathbf{j}$) transforming via a Lorentz matrix. The quantity $\mathbf{u}(t, \mathbf{x})$ represents the velocity field. If the quantity A is conserved during the flow, the continuity equation holds:

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \rho \mathbf{u} = 0 . \tag{1.33}$$

This expression sums both the time derivative and all spatial derivatives:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho u_k}{\partial x_k} = 0 .$$

In the present case, the situation is only slightly more complex. The additive quantity is the number of systems in the ensemble, and the density is the probability density. However, the flux must be taken in the phase space of all coordinates and momenta:

$$\mathbf{j} = (\rho \dot{q}_1, \cdots, \rho \dot{q}_f, \rho \dot{p}_1, \cdots, \rho \dot{p}_f),$$

just as the divergence in the continuity equation will contain derivatives over all axes of the phase space:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho \dot{q}_k}{\partial q_k} + \frac{\partial \rho \dot{p}_k}{\partial p_k} = 0 .$$

Clearly, if systems are not lost in the phase space, such a law of conservation of the system number must hold. The derivatives of the products are performed below:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial q_k} \dot{q}_k + \frac{\partial \dot{q}_k}{\partial q_k} \rho + \frac{\partial \rho}{\partial p_k} \dot{p}_k + \frac{\partial \dot{p}_k}{\partial p_k} \rho = 0 .$$

In the third and fifth terms, Hamilton equations are utilized:

$$\dot{q}_k = \frac{\partial H}{\partial p_k}; \qquad \dot{p}_k = -\frac{\partial H}{\partial q_k}$$

yielding:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial q_k} \dot{q}_k + \frac{\partial^2 H}{\partial q_k \partial p_k} \rho + \frac{\partial \rho}{\partial p_k} \dot{p}_k - \frac{\partial^2 H}{\partial p_k \partial q_k} \rho = 0 .$$

Due to the commutativity of mixed partial derivatives, both mentioned terms eventually cancel out. The continuity equation is thus obtained in the form:

$$\frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial q_k}\dot{q}_k + \frac{\partial \rho}{\partial p_k}\dot{p}_k = 0 .$$

Since $\rho = \rho(t, q_k, p_k)$, the following relation is obtained:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = 0 . \tag{1.34}$$

The probability density remains constant, and the probability of system occurrence in the phase space behaves like an incompressible fluid. Equation (1.34) is called the *Liouville theorem* and carries fundamental importance in statistical physics.

Note: The continuity equation can be similarly modified for the flow of an ordinary fluid described only by a density field and a velocity field:

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \rho \mathbf{u} = 0 \quad \Rightarrow \quad \frac{\partial \rho}{\partial t} + \partial_k (\rho u_k) = 0 \quad \Rightarrow \quad \frac{\partial \rho}{\partial t} + \frac{\partial \rho}{\partial x_k} u_k + \frac{\partial u_k}{\partial x_k} \rho = 0 .$$

The first two terms yield the total time derivative of the density, and the last term can be modified using the divergence:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + \rho \operatorname{div} \mathbf{u} = 0$$

Evidently, the flow of a normal fluid is incompressible, i.e., $\mathrm{d}\rho/\mathrm{d}t = 0$, if $\operatorname{div} \mathbf{u} = 0$.

According to relation (1.34), the probability density is constant. The number of constants available in the calculation is considered below. If the generalized coordinates and generalized momenta are calculated from Hamilton equations, the solution depends on the initial state (f positions and f momenta), i.e., it contains $2f$ integration constants of motion, more precisely $2f - 1$, because one constant is consumed by the choice of the time axis origin t_0:

$$q_k = q_k(t, \alpha_1, \ldots, \alpha_{2f-1}) ,$$
$$p_k = p_k(t, \alpha_1, \ldots, \alpha_{2f-1}) .$$

If the obtained solution could be completely inverted to calculate these integration constants directly from the system evolution:

$$\alpha_k = \alpha_k(q_1, \ldots, q_f , p_1, \ldots, p_f) ,$$

All conservation laws of the ensemble would be obtained in this way. Unfortunately, these constants are not always defined on the entire domain, and from analytical mechanics, the existence of only seven fundamental conservation laws is guaranteed: energy, momentum, and angular momentum. Since the total derivative of the probability density must be constant according to the Liouville theorem, it is reasonable to assume that the density is a function of these known integrals of motion:

$$\rho = \rho(E, \mathbf{p}, \mathbf{L}) .$$

If a coordinate system moving and rotating with the ensemble center of mass is selected, the only remaining quantity on which the probability density can depend is energy:

► $$\rho = \rho(E)\,. \tag{1.35}$$

This statement represents the most serious consequence of the Liouville theorem.

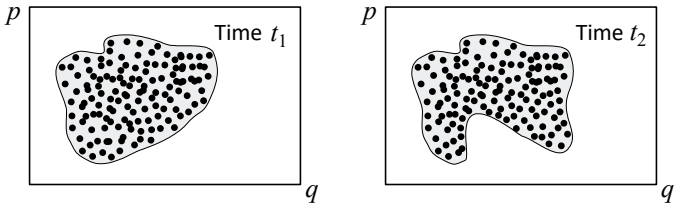


Fig. 1.4: Liouville theorem. The phase volume can deform, but its magnitude remains constant.

Due to the incompressibility of the "flowing probability", the phase volume occupied by a selected macroscopic part of the systems does not change. If the systems initially occupied a certain volume ϕ in the phase space, this volume will deform in various ways during the time evolution, but its magnitude will remain constant. This volume (or weight factor) will therefore again be merely a function of the system energy:

► $$\phi = \phi(E)\,; \qquad \Gamma = \Gamma(E)\,. \tag{1.36}$$

This statement represents an alternative formulation of the Liouville theorem. Further information can be found in publication [8].

Energy Density of States

Due to the Liouville theorem, the probability element can be written as:

$$\mathrm{d}w = \rho\,\mathrm{d}\Gamma = \rho(E)\,\mathrm{d}\Gamma(E) = \rho\frac{\mathrm{d}\Gamma}{\mathrm{d}E}\,\mathrm{d}E = \rho(E)\,\gamma(E)\,\mathrm{d}E\,, \tag{1.37}$$

where the following notation was introduced:

$$\gamma(E) \equiv \frac{\mathrm{d}\Gamma}{\mathrm{d}E} \tag{1.38}$$

termed the density of energy states (recall that Γ represents the number of quantum states in the considered phase volume). For continuous problems, the density of energy states is a continuous function; however, it often contains a discrete part as well:

$$\gamma(E) = g(E) + \sum_n g_n\,\delta(E - E_n)\,. \tag{1.39}$$

The symbol δ denotes the Dirac distribution (an analog to the Kronecker delta in l_2 spaces for L_2 spaces). The coefficients g_n are called the degeneracy degree of the state n.

1.3 Gibbs Canonical Ensemble

1.3.1 Derivation of the Distribution

The derivation is based on the following assumptions:

- A system of N particles can exchange energy with the surroundings. This implies, for example, the possibility of thermal energy exchange through the system walls.
- The number of system particles is constant. The system does not exchange particles with the surroundings; particles are neither created nor destroyed within it.
- The only external parameter is the system volume (determined by external factors, such as the shape of the container). Work can generally depend on many external factors: $\mathrm{d}A = A_\mathrm{k}\,\mathrm{d}a_\mathrm{k}$. The quantities A_k are termed generalized forces, and the quantities a_k are called external parameters. In this simple example, only a single term exists: $\mathrm{d}A = p\mathrm{d}V$.
- The energy associated with the system surface $E_{\partial V}$ is negligible compared to the total system energy E_V, i.e., surface phenomena are neglected.
- According to the Liouville theorem, the probability density and the phase volume are assumed to depend only on the system energy.

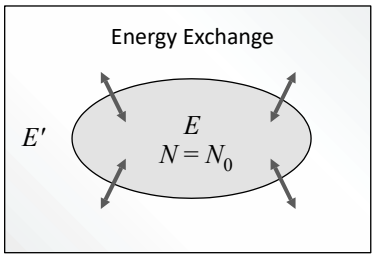


Fig. 1.5: Gibbs canonical ensemble.

The stated assumptions can be simply expressed by mathematical relations:

$$\begin{aligned} &(1) \quad E \neq \text{const}\,, \\ &(2) \quad N = \text{const}\,, \\ &(3) \quad \mathrm{d}A = p\mathrm{d}V\,, \\ &(4) \quad E_{\partial V} \ll E_V\,, \\ &(5) \quad \phi = \phi(E)\,; \qquad \rho = \rho(E)\,. \end{aligned} \tag{1.40}$$

The first thermodynamic law for this system takes a simple form:

$$\mathrm{d}U = T\,\mathrm{d}S - p\,\mathrm{d}V\,. \tag{1.41}$$

Since the number of system particles remains unchanged, the last term in (1.7) is zero. The variables of the surroundings are denoted with a prime. The probability of finding the system and its surroundings in a state with a specific energy is given by the additivity of energy and the multiplicativity of phase volume:

$$\mathrm{d}w_{\mathrm{tot}} = \rho(E_{\mathrm{tot}})\,\mathrm{d}\Gamma_{\mathrm{tot}} = \rho(E+E')\,\mathrm{d}\Gamma\,\mathrm{d}\Gamma' . \tag{1.42}$$

For independent subsystems, probabilities multiply, and this relation must hold:

$$\mathrm{d}w_{\mathrm{tot}} = \mathrm{d}w(E)\,\mathrm{d}w(E') = \rho(E)\mathrm{d}\Gamma(E)\;\rho(E')\mathrm{d}\Gamma(E') = \rho(E)\rho(E')\,\mathrm{d}\Gamma\mathrm{d}\Gamma' . \tag{1.43}$$

Comparing both options reveals that the probability density must satisfy the relation:

$$\rho(E+E') = \rho(E)\,\rho(E') . \tag{1.44}$$

In mathematics, the exponential function in the form $\rho(E) = \exp[cE]$ is shown to be the only function with this property. Adding a normalization constant to this distribution yields the canonical distribution in the form:

$$\rho(E) = K\,\mathrm{e}^{cE} = \mathrm{e}^{c_1+c_2E} . \tag{1.45}$$

In the exponential, the constants are denoted by α and $-\beta$. The choice of sign is immaterial; if incorrect, β would be negative. As system energy increases, the probability of its occurrence decreases, meaning the minus sign is correct. The expression is given for both continuous and discrete cases:

$$\rho(E) = \mathrm{e}^{\alpha-\beta E} ; \qquad w_n = \mathrm{e}^{\alpha-\beta E_n} . \tag{1.46}$$

The constants α and β are derived from the condition that the statistical results must asymptotically transition into the known laws of thermodynamics. In the discrete case, the possible values of the energy spectrum depend on the external parameters of the system – in this case, on the volume occupied by the system, i.e., $E_n = E_n(V)$.

1.3.2 Distribution Constants

The constants α and β are determined below by comparing the differential of the mean energy value with the first thermodynamic law. The derivation can be performed continuously or discretely; for instance, the continuous expression is:

$$U = \int E\,\mathrm{d}w = \int E\,\rho(E)\,\mathrm{d}\Gamma = \int E\,\rho(E)\,\gamma(E)\,\mathrm{d}E \tag{1.47}$$

and in the discrete case:

$$U = \sum E_n w_n = \sum E_n(V)\,w_n . \tag{1.48}$$

This derivation is based on the discrete case, whereas some future derivations will be conducted continuously for variety. The differential of expression (1.48) is found as:

$$\mathrm{d}U = \sum\left[\left(\frac{\partial E_n}{\partial V}\,\mathrm{d}V\right)w_n\right] + \sum E_n\,\mathrm{d}w_n .$$

The energy density of the n-th state corresponds to the pressure (up to a sign change). In mechanics, pressure is energy density: $p = \Delta F/\Delta S = \Delta F\,\Delta l/(\Delta S\,\Delta l) = -\Delta E/\Delta V$. A nega-

tive sign is chosen because force is the negative gradient of energy. In the second expression, an ingenious trick is used, expressing the energy E_n via the probability (1.46):

$$\mathrm{d}U = -\sum\left(p_n w_n \mathrm{d}V\right) + \sum\left(\frac{\alpha}{\beta} - \frac{1}{\beta}\ln w_n\right)\mathrm{d}w_n \,.$$

In the sums, only expressions over which summation actually occurs are retained; other terms are factored out:

$$\mathrm{d}U = -\sum\left(p_n w_n\right)\mathrm{d}V + \frac{\alpha}{\beta}\,\mathrm{d}\sum w_n - \frac{1}{\beta}\sum\left(\ln w_n\,\mathrm{d}w_n\right).$$

In the first term, the mean value of partial pressures equals the total pressure. In the second term, the sum of all probabilities equals one, and the differential of unity is zero. The third expression is modified according to the relation $f\,\mathrm{d}g = \mathrm{d}\,(fg) - g\,\mathrm{d}f$:

$$\mathrm{d}U = -\,p\,\mathrm{d}V - \frac{1}{\beta}\,\mathrm{d}\sum\left(w_n \ln w_n\right) + \frac{1}{\beta}\sum\left(w_n\ \mathrm{d}\ln w_n\right).$$

The last expression is shown to be zero below:

$$\sum\left(w_n\,\mathrm{d}\ln w_n\right) = \sum\left(w_n\frac{1}{w_n}\mathrm{d}w_n\right) = \sum\mathrm{d}w_n = \mathrm{d}\sum w_n = \mathrm{d}\,1 = 0\,.$$

From statistical considerations, the expression for the energy differential is finally obtained, allowing comparison with the first thermodynamic law $\mathrm{d}U = -p\mathrm{d}V + T\mathrm{d}S$:

$$\mathrm{d}U = -\,p\,\mathrm{d}V - \frac{1}{\beta}\,\mathrm{d}\sum\left(w_n \ln w_n\right). \tag{1.49}$$

Evidently, the coefficient β must be proportional to the reciprocal value of the absolute temperature, and the sum in the second expression to the entropy. This relation holds up to an arbitrary multiplicative coefficient, which must be determined experimentally:

► $$\beta = 1/\left(k_{\mathrm{B}}T\right), \tag{1.50}$$

► $$S = -k_{\mathrm{B}}\sum\left(w_n \ln w_n\right). \tag{1.51}$$

Note 1: As in any physical theory, a single selectable constant, k_{B}, exists in statistics. This Boltzmann constant allows creating various statistical theories, but only one corresponds to real nature. This situation is identical to introducing the Planck constant in quantum theory. After comparing the first derived relations (for instance, the ideal gas state equation) with reality, its value is found to be:

$$k_{\mathrm{B}} = 1{,}38\times10^{-23}\ \mathrm{J\,K^{-1}}. \tag{1.52}$$

Note 2: A statistical expression for entropy (1.51) is obtained, which averages the logarithm of the probability of the n-th state. From the statistical perspective, entropy equals the mean value of the logarithm of probability up to constants:

$$S = -k_{\mathrm{B}}\left\langle \ln w \right\rangle. \tag{1.53}$$

Note 3: The relationship between entropy and system realization probability was derived by Ludwig Boltzmann and is known as the Boltzmann equation:

$$S = -k_{\mathrm{B}} \ln P \,. \tag{1.54}$$

The minus sign in the entropy relation may be surprising. Since the probability is normalized to one, i.e., each partial probability is less than one, the logarithm of probability is negative. The minus sign before the relation thus ensures a non-negative entropy value. By comparison with the first thermodynamic law, the meaning of the coefficient β in the probability distribution was identified. Further modifications of the statistical definition of entropy and subsequent comparison with thermodynamics yield the meaning of the coefficient α:

$$S = -k_{\mathrm{B}} \sum \left[w_n \ln w_n \right].$$

Substituting for $\ln w_n$ from the relation $w_n = \exp[\alpha - \beta E_n] \Rightarrow \ln w_n = \alpha - \beta E_n$ yields:

$$S = -k_{\mathrm{B}} \sum \left[w_n (\alpha - \beta E_n) \right] =$$

$$= -k_{\mathrm{B}} \alpha \sum w_n \; + \; k_{\mathrm{B}} \beta \sum E_n \, w_n =$$

$$= -k_{\mathrm{B}} \, \alpha + k_{\mathrm{B}} \beta U \,.$$

The unknown coefficient α is now easily determined:

$$\alpha = \frac{-S + k_{\mathrm{B}} \beta U}{k_{\mathrm{B}}} = \frac{-S + U/T}{k_{\mathrm{B}}} =$$

$$= \frac{U - TS}{k_{\mathrm{B}} T} = \frac{F}{k_{\mathrm{B}} T} \,.$$

The values of both coefficients are thus obtained as:

► $$\alpha = \frac{F}{k_{\mathrm{B}} T}; \quad \beta = \frac{1}{k_{\mathrm{B}} T} \,. \tag{1.55}$$

The probability distribution is therefore (for both continuous and discrete cases):

► $$\rho(E) = \mathrm{e}^{\frac{F-E}{k_{\mathrm{B}} T}}; \quad w_n(E) = \mathrm{e}^{\frac{F-E_n}{k_{\mathrm{B}} T}} \,. \tag{1.56}$$

Expressions are frequently abbreviated using the coefficient $\beta = 1/k_{\mathrm{B}} T$:

► $$\rho(E) = \mathrm{e}^{\beta(F-E)}; \quad w_n(E) = \mathrm{e}^{\beta(F-E_n)} \,. \tag{1.57}$$

The derived relations are called the Gibbs canonical distribution. It is named after the prominent American physicist Josiah Gibbs, who worked on thermodynamics and statistical physics. Among other achievements, he formulated the well-known Gibbs phase rule valid for phase transitions. In quantum theory, a density operator is used instead of probability density:

$$\hat{\rho} = \exp[\beta(F - \hat{H})] \,. \tag{1.58}$$

Note: In the expression for probability $\mathrm{d}w = \rho \mathrm{d}\Gamma = \rho\gamma\ \mathrm{d}E$, the probability density ρ is an exponentially decreasing function of energy. Conversely, the density of energy states γ increases with increasing energy. The resulting probability density therefore possesses a maximum near a certain characteristic energy, which is represented in the system with the highest probability.

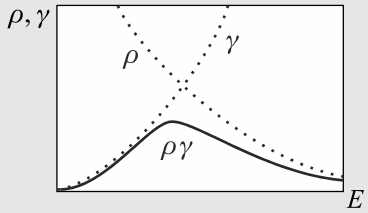


Fig. 1.6: Probability density and density of states.

1.3.3 Partition Function and Its Interpretation

Since both distribution constants are now known, Helmholtz free energy can be determined from the normalization condition. This step is the key to success. If free energy is known, differentiating it yields extensive information about the system, such as the state equation. The free energy calculation is performed in parallel for the discrete and continuous cases to allow comparison of both procedures. The left side presents the discrete calculation, while the right side shows the continuous one:

$$\sum w_n = 1 \quad \Rightarrow$$
$$\sum \mathrm{e}^{\beta(F-E_n)} = 1 \quad \Rightarrow$$
$$\sum \mathrm{e}^{-\beta E_n} = \mathrm{e}^{-\beta F} \quad \Rightarrow$$
$$\ln\left(\sum \mathrm{e}^{-\beta E_n}\right) = -\beta F \quad \Rightarrow$$
$$F = -k_\mathrm{B} T \ln\left(\sum \mathrm{e}^{-\beta E_n}\right);$$

$$\int \rho\ \mathrm{d}\Gamma = 1 \quad \Rightarrow$$
$$\int \mathrm{e}^{\beta(F-E)}\ \mathrm{d}\Gamma = 1 \quad \Rightarrow$$
$$\int \mathrm{e}^{-\beta E}\ \mathrm{d}\Gamma = \mathrm{e}^{-\beta F} \quad \Rightarrow$$
$$\ln\left(\int \mathrm{e}^{-\beta E}\ \mathrm{d}\Gamma\right) = -\beta F \quad \Rightarrow$$
$$F = -k_\mathrm{B} T \ln\left(\int \mathrm{e}^{-\beta E}\ \mathrm{d}\Gamma\right).$$

The quantity inside the parentheses is called the partition function (partition sum, state sum) and represents the central quantity of statistical physics, denoted as Z. Since the logarithm argument must be dimensionless, using the weight factor instead of the phase volume is preferable. Every statistical calculation begins by determining the partition function. The basic structure of a statistical calculation is described below:

Statistical Calculation Scheme

1) The energies E_n that the system can occupy are determined. In the classical case, these always include all energy values that can occur in the system. In the quantum case, the spectrum of the Hamiltonian operator must be determined (for instance, by solving a Schrödinger equation).

2) The partition function Z is found as a sum of Boltzmann factors $\exp[-\beta E]$over the entire domain of the energy spectrum:

$$Z = \sum \mathrm{e}^{-\beta E_n} \; ; \qquad \text{resp.} \qquad Z = \int \mathrm{e}^{-\beta E} \, \mathrm{d}\Gamma \; . \tag{1.59}$$

This specific step can be complicated, often resolved via graphical or numerical methods. All actual possibilities must be summed without omission.

3) Logarithm evaluation yields the free energy F, the pivotal statistical quantity:

$$F = -k_\mathrm{B} T \;\; \ln Z \; . \tag{1.60}$$

4) From the knowledge of the Helmholtz free energy, system entropy and pressure (the state equation) are determined according to relation (1.10):

$$S = -\left(\frac{\partial F}{\partial T} \right), \qquad p = -\left(\frac{\partial F}{\partial V} \right). \tag{1.61}$$

5) Other derived quantities are determined, i.e., internal energy and its derivatives (such as specific heats, susceptibility, etc.). The starting point can be a Gibbs-Helmholtz equation (1.11) for calculating internal energy from free energy:

$$U = F + T S = F - T \left(\frac{\partial F}{\partial T} \right). \tag{1.62}$$

Partition Function and Its Interpretation

- The partition function is the sum of all Boltzmann factors over possible energy values (discrete or continuous):

$$Z = \sum \mathrm{e}^{-\beta E_n} \; ; \qquad \text{resp.} \qquad Z = \int \mathrm{e}^{-\beta E} \, \mathrm{d}\Gamma \; .$$

- In quantum theory, the partition function can be written using the energy operator:

$$Z = \sum \mathrm{e}^{-\beta E_n} = \sum_n \langle n | \mathrm{e}^{-\beta \hat{H}} | n \rangle = \mathrm{Tr}\left(\mathrm{e}^{-\beta \hat{H}} \right),$$

where the symbol Tr() denotes the trace (the sum of diagonal elements in a representation) of the operator function in parentheses. The resulting expression is known as a Slater equation, named after the prominent American physicist John Clark Slater, who worked primarily on quantum theory.

- The partition function possesses a unique relationship to free energy and can be determined from experimental free energy measurements:

$$F = -k_\mathrm{B} T \;\; \ln Z \qquad \Rightarrow \qquad Z = \mathrm{e}^{-\beta F} \; .$$

- The partition function is the reciprocal value of the normalization constant in the probability distribution, which follows by substituting F from the previous relation:

$$\rho(E) = \mathrm{e}^{\beta (F - E)} = \frac{1}{Z} \mathrm{e}^{-\beta E} \; ; \qquad w_n(E) = \mathrm{e}^{\beta (F - E_n)} = \frac{1}{Z} \mathrm{e}^{-\beta E_n} \; .$$

- The partition function is a Laplace transform of the energy states density $\gamma(E)$:

$$Z = \int \mathrm{e}^{-\beta E}\, \mathrm{d}\Gamma = \int_0^\infty \mathrm{e}^{-\beta E}\, \gamma(E)\, \mathrm{d}E \quad .$$

  Conversely, if the partition function is known (for example, from an experimental measurement of free energy), the energy states density is obtained by performing an inverse Laplace transform:

$$\gamma(E) = \frac{1}{2\pi \mathrm{i}} \int_{\sigma-\mathrm{i}\infty}^{\sigma+\mathrm{i}\infty} \mathrm{e}^{\beta E}\, Z(\beta)\, \mathrm{d}\beta \, .$$

The individual meanings of the partition function are summarized below:

$$\begin{aligned} Z &= \sum \mathrm{e}^{-\beta E_n} \; ; \\ Z &= \mathrm{Tr}\, \mathrm{e}^{-\beta \hat{H}} \; ; \\ Z &= \mathrm{e}^{-\beta F} \; ; \\ Z &= \frac{1}{K} \; ; \\ Z &= \mathscr{L}_\beta(\gamma) \, . \end{aligned} \qquad (1.63)$$

All prerequisites to initiate and successfully complete a statistical calculation are now established. In the next chapter, this procedure is introduced using simple examples.

◖ Example 1.4: Canonical Distribution Maximum

Prove that the canonical distribution possesses a maximum at $E_n = \langle E \rangle$.

Solution: Consider a discrete canonical distribution $w_n(\beta) = A \exp[-\beta E_n]$. From the normalization condition, the normalization constant A is found::

$$\sum w_k = 1 \quad \Rightarrow \quad A = \frac{1}{\sum_k \mathrm{e}^{-\beta E_k}} \quad \Rightarrow \quad w_n = \frac{\mathrm{e}^{-\beta E_n}}{\sum_k \mathrm{e}^{-\beta E_k}} \, .$$

As established previously, the normalization constant equals the reciprocal of the partition function. The condition for the maximum with respect to parameter β is found:

$$\frac{\partial w_n}{\partial \beta} = 0 \quad \Rightarrow \quad E_n = \sum_k E_k\, w_k \quad \Rightarrow \quad E_n = \langle E \rangle \, .$$

The state corresponding to the mean energy value is represented most prominently within the entire canonical ensemble of the system. ◗

1.4 Simple Examples

1.4.1 Ideal Gas

For the first time, a complete statistical calculation is performed according to the procedure from Section 1.3.3. The system consists of N identical classical particles that interact neither with each other nor with the surroundings (the potential energy is zero). An ensemble would comprise many such systems (for example, the system is a container filled with gas, and the ensemble is many such containers). The choice was made to describe the container as a whole, which allows, for instance, determining the pressure inside this container. The procedure follows the previously outlined scheme exactly:

1) Energy spectrum. The system energy can assume any non-negative value and is given solely by the sum of the kinetic energies of all particles:

$$E = \sum_{a=1}^{N} \frac{\mathbf{p}_a^2}{2m}.$$

2) Partition function. The partition function is found as a sum (in this case, continuous, i.e., an integral) of all Boltzmann factors:

$$Z = \int \mathrm{e}^{-\beta E}\,\mathrm{d}\Gamma = \int \exp\left[-\sum_{a=1}^{N}\frac{\mathbf{p}_a^2}{2mk_\mathrm{B}T}\right]\frac{\mathrm{d}^{3N}x\,\mathrm{d}^{3N}p}{(2\pi\hbar)^{3N}} =$$

$$= \frac{1}{(2\pi\hbar)^{3N}}\,V^N \int \exp\left[-\sum_{a=1}^{N}\frac{\mathbf{p}_a^2}{2mk_\mathrm{B}T}\right]\mathrm{d}^{3N}p =$$

$$= \frac{1}{(2\pi\hbar)^{3N}}\,V^N \left(\int_{-\infty}^{\infty}\exp\left[-\frac{\xi^2}{2mk_\mathrm{B}T}\right]\mathrm{d}\xi\right)^{3N} =$$

$$= \frac{1}{(2\pi\hbar)^{3N}}\,V^N \left(\sqrt{2\pi mk_\mathrm{B}T}\right)^{3N}.$$

where the individual momenta are expanded into components, and the $3N$ sums in the exponential argument are converted into a product of exponentials. The integral thus becomes a product of $3N$ identical Gaussian integrals (I.3.576). The result is:

$$Z(T,V,N) = A^N\,V^N T^{3N/2};$$
$$A \equiv (2\pi mk_\mathrm{B})^{3/2}/(2\pi\hbar)^3. \tag{1.64}$$

3) Free energy. Free energy is easily determined from relation (1.60):

$$F(T,V,N) = -k_\mathrm{B}T\ln Z = -Nk_\mathrm{B}T\ln\left(A\,VT^{3/2}\right).$$

4) Thermodynamic quantities. Entropy and pressure (the state equation) are determined as partial derivatives (1.10) of the free energy. The chemical potential is unnecessary due to the constant number of particles.

$$S = -\left(\frac{\partial F}{\partial T}\right) = +Nk_{\mathrm{B}} \ln\left(AVT^{3/2}\right) + \frac{3Nk_{\mathrm{B}}}{2}, \tag{1.65}$$

$$p = -\left(\frac{\partial F}{\partial V}\right) = \frac{N k_{\mathrm{B}} T}{V}. \tag{1.66}$$

If a limit transition to absolute zero is performed, entropy relation (1.65) diverges to $-\infty$. This occurs because the classical calculation fails at low temperatures, and a quantum calculation is required. In a quantum approach, the expression approaches zero with decreasing temperature. Through coefficient A, relation (1.65) depends on the Planck constant. This is because the Planck constant determines the state size in phase space, and entropy relates to the probability of a specific state. The second part has the character of an integration constant $S_0 = 3Nk_{\mathrm{B}}/2$, representing the entropy value at absolute zero (its non-zero value relates to the third thermodynamic law, which was not mentioned). Notably, entropy is proportional to the particle number. This is logical; entropy represents thermal energy multiplied by an integrating factor, and energy is additive. A separate section is dedicated to entropy later.

The second relation (1.66) constitutes the ideal gas state equation ($pV = Nk_{\mathrm{B}}T$) and naturally cannot depend on the Planck constant. Pressure in this relation represents a non-quantum gas. An alternative formulation for the state equation is provided below:

$$p = \frac{Nk_{\mathrm{B}}T}{V}; \tag{1.67}$$

$$p = nk_{\mathrm{B}}T; \qquad n \equiv \frac{N}{V}. \tag{1.68}$$

The quantity n is called particle number density. The state equation is sometimes expressed using the universal gas constant; however, such an expression is unnecessary for the purposes of this textbook..

5) Internal energy. Internal energy could be determined by direct integration from its definition, i.e., $U = \langle E \rangle = \int E \, \mathrm{d}w$, but utilizing the free energy definition $F = U - TS$ as a thermodynamic potential is faster:

$$U = F + TS =$$

$$= -Nk_{\mathrm{B}}T \ln\left(AVT^{3/2}\right) + \frac{3Nk_{\mathrm{B}}T}{2} + Nk_{\mathrm{B}}T \ln\left(AVT^{3/2}\right) = \frac{3}{2} Nk_{\mathrm{B}}T$$

The result is:

$$U = \frac{3}{2} Nk_{\mathrm{B}}T; \qquad u \equiv \frac{U}{N} = \frac{3}{2} k_{\mathrm{B}}T, \tag{1.69}$$

where u represents the energy of a single particle. The internal energy again does not depend on the Planck constant; notably, each degree of freedom of the system contributes a value of $k_{\mathrm{B}}T/2$ to the internal energy. This statement is known as the *equipartition theorem* of energy in classical statistical mechanics.

Finally, all derived relations for an ideal gas are listed below:

$$Z = A^N\, V^N T^{3N/2}; \qquad A = \left(\frac{mk_\mathrm{B}}{2\pi\hbar^2}\right)^{3/2},$$

$$F = -Nk_\mathrm{B}T \ln\left(A\, VT^{3/2}\right),$$

$$S = \frac{3Nk_\mathrm{B}}{2} + Nk_\mathrm{B} \ln\left(A\, V\, T^{3/2}\right), \tag{1.70}$$

$$p = \frac{Nk_\mathrm{B}T}{V},$$

$$U = \frac{3}{2}\, Nk_\mathrm{B}T\,.$$

Note: The partition function calculation is usually performed only for a system with a single particle. This partition function is denoted by the lowercase letter z. As seen from the calculation, for N independent particles, the total partition function represents a product of the integrals of the individual particle partition functions. For identical non-quantum particles, the following relation holds:

$$Z = z_1\, z_2 \ldots z_N = z^N\,. \tag{1.71}$$

The partition function is multiplicative with respect to the particle number, whereas free energy, entropy, pressure, and internal energy constitute additive quantities.

1.4.2 Particle in External Field

Properties of the probability density of a particle in an external potential field $V(x, y, z)$ are examined below. The ensemble comprises many such particles. Phase space consists of coordinates and momenta ($x, y, z\,, p_x, p_y, p_z$). The system energy takes the form:

$$E = \frac{p_x^2 + p_y^2 + p_z^2}{2m} + V(x, y, z)\,.$$

The probability element is given by:

$$\mathrm{d}w = \rho\, \mathrm{d}\phi = \mathrm{e}^{\beta(F-E)}\, \mathrm{d}^3x\, \mathrm{d}^3p = K\, \mathrm{e}^{-\beta E}\, \mathrm{d}^3x\, \mathrm{d}^3p\,.$$

If the weight space were used instead of the phase space, the expression would be:

$$\mathrm{d}w = \rho\, \mathrm{d}\Gamma = \mathrm{e}^{\beta(F-E)}\, \frac{\mathrm{d}^3x\, \mathrm{d}^3p}{(2\pi\hbar)^3} = \tilde{K}\, \mathrm{e}^{-\beta E}\, \mathrm{d}^3x\, \mathrm{d}^3p\,.$$

Evidently, both expressions lead to the same final relation; the normalization constant must be determined from the condition $\int \mathrm{d}w = 1$ in both cases. Substituting for the energy yields the resulting probability distribution:

$$\mathrm{d}w(\mathbf{x}, \mathbf{p}) = K \exp\left[-\beta\left(\frac{p_x^2 + p_y^2 + p_z^2}{2m} + V(x, y, z)\right)\right] \mathrm{d}^3x\, \mathrm{d}^3p\,. \tag{1.72}$$

Due to the properties of the exponential function, the result can be written as a product of the probabilities for the coordinates and for the momenta, meaning that the distribution of coordinates and momenta is independent:

$$\mathrm{d}w = \mathrm{d}w_1(\mathbf{x})\,\mathrm{d}w_2(\mathbf{p})\,;$$

$$\mathrm{d}w = K_1 \exp\left[-\frac{V(x,y,z)}{k_\mathrm{B}T}\right]\mathrm{d}^3x \times K_2 \exp\left[-\frac{p_x^2 + p_y^2 + p_z^2}{2mk_\mathrm{B}T}\right]\mathrm{d}^3p\,.$$

In the case of a suitable form of potential energy, the distribution can even decompose into products of probabilities in individual axes. For momenta, this decomposition always occurs automatically. In the subsequent text, both probabilities are discussed separately and are thus denoted by a single symbol dw.

$$\mathrm{d}w(\mathbf{p}) = \mathrm{d}w(p_x)\;\;\mathrm{d}w(p_y)\;\;\mathrm{d}w(p_z)\,;$$

$$\mathrm{d}w(\mathbf{p}) = K_x \exp\left[-\frac{p_x^2}{2mk_\mathrm{B}T}\right]\mathrm{d}p_x \times K_y \exp\left[-\frac{p_y^2}{2mk_\mathrm{B}T}\right]\mathrm{d}p_y \times K_z \exp\left[-\frac{p_z^2}{2mk_\mathrm{B}T}\right]\mathrm{d}p_z.$$

All three probability elements are given by the same exponential function; therefore, the probabilities of all three momentum projections are identical. The distribution constants can be easily determined from the normalization condition $\int \mathrm{d}w = 1$, which holds for each part of the distribution separately.

Barometric Formula

The distribution of particle positions in a gravitational field $V = mgy$ is analyzed below. The mass of a single particle is m, and the height above the surface is y. The particle occurrence probability is given by:

$$\mathrm{d}w(x,y,z) = K \exp\left[-\frac{mgy}{k_\mathrm{B}T(y)}\right]\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,.$$

Evidently, the particle occurrence probability does not depend on the coordinates x and z, which can be integrated out, with the integration result absorbed into the normalization constant. Only the particle occurrence probability in the vertical direction remains:

$$\mathrm{d}w(y) = C \exp\left[-\frac{mgy}{k_\mathrm{B}T(y)}\right]\mathrm{d}y\,. \tag{1.73}$$

The probability density of particle occurrence dw/dy must be proportional to the number of individuals in a given region of the phase space (in this case, only in the volume), hence to the particle number density $n(y)$ above the ground:

$$n(y) = n_0 \exp\left[-\frac{mgy}{k_\mathrm{B}T(y)}\right]. \tag{1.74}$$

The pressure $p = nk_\mathrm{B}T$: is now easily determined:

$$p(y) = n_0\, k_\mathrm{B}\, T(y) \exp\left[-\frac{mgy}{k_\mathrm{B}T(y)}\right]. \tag{1.75}$$

This relation represents the well-known *Boltzmann barometric formula*, describing the decrease in particle pressure with height in the atmosphere. If the temperature profile of the atmosphere is known, the height profile of pressure is easily calculated.

◖ **Example 1.5: Rotating Cylinder**
Determine the gas density in a cylinder of radius R and length L rotating around its axis with an angular velocity ω.

Solution: The potential energy of a single rotating particle is determined as the negative rotational energy ($T + V = \text{const}$):

$$V(r) = -\frac{J\omega^2}{2} = -\frac{mr^2\omega^2}{2}.$$

From the barometric formula, the particle density is immediately obtained as:

$$n(r) = n_0 \exp\left[\frac{m r^2\omega^2}{2k_B T}\right]. \qquad \blacksquare$$

Boltzmann Probability Distribution

The distribution of a single component of momentum or velocity is analyzed below. For concreteness, the projection of momentum or velocity into the x-axis is considered; however, any axis could be chosen:

$$\mathrm{d}w(p_x) = K \exp\left[-\frac{p_x^2}{2mk_B T}\right] \mathrm{d}p_x .$$

First, the distribution constant is determined from the normalization condition. The integration is performed using relation (I.3.576):

$$\int \mathrm{d}w(p_x) = 1 \qquad \Rightarrow$$

$$\int_{-\infty}^{+\infty} K \exp\left[-\frac{p_x^2}{2mk_B T}\right] \mathrm{d}p_x = 1 \qquad \Rightarrow$$

$$K\sqrt{2\pi m k_B T} = 1 \qquad \Rightarrow$$

$$K = \frac{1}{\sqrt{2\pi m k_B T}}.$$

The Boltzmann distribution in momenta or velocities thus has the character of a Gaussian package $y = A \exp(-\alpha x^2)$:

$$\blacktriangleright \qquad \mathrm{d}w(p_x) = \frac{1}{\sqrt{2\pi m k_B T}} \exp\left[-\frac{p_x^2}{2mk_B T}\right] \mathrm{d}p_x ; \qquad (1.76)$$

$$\blacktriangleright \qquad \mathrm{d}w(v_x) = \frac{m}{\sqrt{2\pi m k_B T}} \exp\left[-\frac{m v_x^2}{2k_B T}\right] \mathrm{d}v_x . \qquad (1.77)$$

The most probable projection of momentum or velocity is zero, which is due to the chaotic nature of the motion. The higher the temperature, the higher the fraction of particles with high velocities. Positive and negative projections are equally represented; the distribution is symmetric. The area under the distribution curve always equals one, i.e., the total probability of particle occurrence.

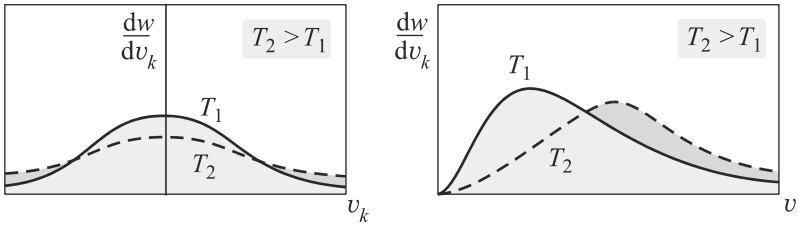


Fig. 1.7: Boltzmann distribution (left) and Maxwell distribution (right).

Maxwell Probability Distribution

This section covers the distribution of the total particle velocity magnitude, termed the Maxwell distribution. First, the distribution in all three velocities is written, which is a product of the distributions in the individual axes:

$$\mathrm{d}w(\mathbf{v}) = \mathrm{d}w(v_x)\,\mathrm{d}w(v_y)\,\mathrm{d}w(v_z) = \frac{m^3}{(2\pi m k_\mathrm{B} T)^{3/2}}\,\exp\left[-\frac{m\left(v_x^2+v_y^2+v_z^2\right)}{2k_\mathrm{B}T}\right]\mathrm{d}v_x\mathrm{d}v_y\mathrm{d}v_z\,.$$

If a transition to spherical coordinates is performed in the (x, y, z) space and integration over the angular variables is conducted, a single variable remains – the distance from the origin r, and the volume element is $\mathrm{d}V = 4\pi r^2\,\mathrm{d}r$. Now, the same operation is performed in the velocity space, i.e., with axes labeled (v_x, v_y, v_z). The result is analogous: the volume element has the value $\mathrm{d}v_x\,\mathrm{d}v_y\,\mathrm{d}v_z = 4\pi v^2\mathrm{d}v$.

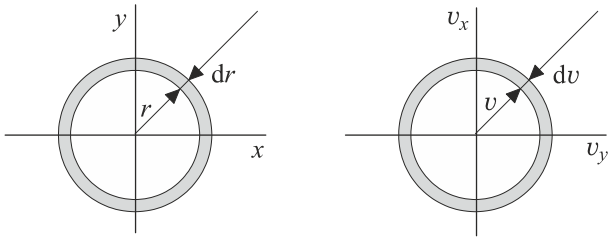


Fig. 1.8: Spherical element in velocity space.

The distribution in velocity magnitudes (Maxwell distribution) can now be written as:

$$\mathrm{d}w(v) = \frac{4\pi m^3}{(2\pi m k_\mathrm{B} T)^{3/2}}\,v^2\,\exp\left[-\frac{mv^2}{2k_\mathrm{B}T}\right]\mathrm{d}v\,. \qquad (1.78)$$

This function is of the type $y = x^2\exp[-x^2]$; the profile is shown in Fig. 1.7 on the right. The probability of finding a particle with a specific velocity has a maximum dependent

on temperature. Finding a particle with either low or high velocity is unlikely. At high velocities, the probability of finding a particle with that velocity decreases exponentially. High-velocity particles from the tail of the Maxwell distribution can reach the escape velocity from the Earth, and the Earth atmosphere loses them. The probability density dw/dv is shown in the Fig. 1.9.

Typical Velocities

The distribution yields the most probable velocity (where the distribution has a maximum), the mean velocity magnitude (which divides the distribution into two equal halves), and the root-mean-square velocity (important for root-mean-square fluctuations, discussed in a separate section). The most probable velocity is determined as the maximum of the probability density:

$$\frac{\mathrm{d}}{\mathrm{d}v} v^2 \exp\left[-\frac{m v^2}{2k_\mathrm{B}T}\right] = 0 \qquad \Rightarrow \qquad v_0 = \sqrt{\frac{2k_\mathrm{B}T}{m}} \,.$$

The mean velocity magnitude value is calculated from the definition, which leads to an integral of type (I.3.575):

$$v_\mathrm{s} = \langle v \rangle = \int_0^\infty v \,\mathrm{d}w(v) = \frac{4\pi m^3}{(2\pi m k_\mathrm{B}T)^{3/2}} \int_0^\infty v^3 \exp\left[-\frac{m v^2}{2k_\mathrm{B}T}\right] \mathrm{d}v = \sqrt{\frac{8k_\mathrm{B}T}{\pi m}} \,.$$

The root-mean-square value leads to a simple integral of type (I.3.574):

$$v_\mathrm{kv} = \sqrt{\langle v^2 \rangle} = \sqrt{\int_0^\infty v^2 \mathrm{d}w(v)} = \sqrt{\frac{4\pi m^3}{(2\pi m k_\mathrm{B}T)^{3/2}} \int_0^\infty v^4 \exp\left[-\frac{m v^2}{2k_\mathrm{B}T}\right] \mathrm{d}v} = \sqrt{\frac{3k_\mathrm{B}T}{m}} \,.$$

All three characteristic velocities differ slightly and are of the same order of magnitude:

► $$v_0 = \sqrt{\frac{2\,k_\mathrm{B}T}{m}} \quad ; \qquad v_\mathrm{s} = \sqrt{\frac{8\,k_\mathrm{B}T}{\pi m}} \quad ; \qquad v_\mathrm{kv} = \sqrt{\frac{3\,k_\mathrm{B}T}{m}} \,. \tag{1.79}$$

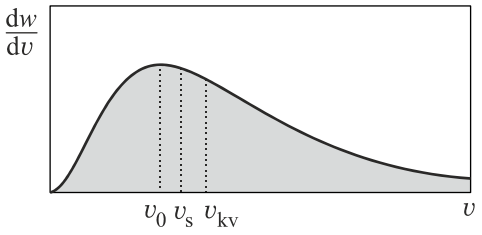


Fig. 1.9: Most probable, mean, and root-mean-square velocity.

As the temperature increases, the mean velocity value of the particles increases. Using the most probable velocity, which is sometimes called the thermal velocity, the Boltzmann and Maxwell distributions can be simply written as:

$$\mathrm{d}w(\mathbf{v}) = C \exp\left[-\frac{\mathbf{v}^2}{v_0^2}\right] \mathrm{d}^3\mathbf{v} \,; \qquad \mathrm{d}w(v) = A v^2 \exp\left[-\left(\frac{v}{v_0}\right)^2\right] \mathrm{d}v \,. \tag{1.80}$$

1.4.3 Classical Oscillator

Thermodynamic Quantities

A system of N identical independent harmonic oscillators is considered below. Let ε denote the energy of a single oscillator and z represent the partition function of a single oscillator. These quantities are determined as:

$$\varepsilon = \frac{1}{2} m\omega^2 x^2 + \frac{p^2}{2m} \ .$$

$$z = \int \exp\left[-\frac{m\omega^2 x^2}{2k_\mathrm{B}T} - \frac{p^2}{2mk_\mathrm{B}T} \right] \frac{\mathrm{d}x\,\mathrm{d}p}{2\pi\hbar} \qquad \Rightarrow$$

$$z = \frac{1}{2\pi\hbar} \left(\int_{-\infty}^{\infty} \exp\left[-\frac{m\omega^2 x^2}{2k_\mathrm{B}T} \right] \mathrm{d}x \right) \left(\int_{-\infty}^{\infty} \exp\left[-\frac{p^2}{2mk_\mathrm{B}T} \right] \mathrm{d}p \right) = \frac{1}{2\pi\hbar} \sqrt{\frac{2\pi k_\mathrm{B}T}{m\omega^2}} \sqrt{2\pi m k_\mathrm{B}T} \quad \Rightarrow$$

$$z = \frac{k_\mathrm{B}T}{\hbar\omega}; \qquad Z = \left(\frac{k_\mathrm{B}T}{\hbar\omega} \right)^N .$$

Notably, the partition function is dimensionless, representing the ratio of thermal energy to the elementary quantum of the oscillator. The weight space introduces the Planck constant even into non-quantum calculations. In classical quantities, it naturally vanishes. The individual thermodynamic quantities are determined below:

$$F = -k_\mathrm{B}T \ln Z = -Nk_\mathrm{B}T \ln \frac{k_\mathrm{B}T}{\hbar\omega},$$

$$S = -\frac{\partial F}{\partial T} = Nk_\mathrm{B} + Nk_\mathrm{B} \ln \frac{k_\mathrm{B}T}{\hbar\omega},$$

$$U = F + TS = Nk_\mathrm{B}T \ .$$

The partition function is multiplicative with respect to the particle number, whereas other quantities are additive. Entropy possesses a constant term and a logarithmic term that diverges in the limit $T \to 0$ because the classical calculation cannot be applied near absolute zero. The results obtained for a classical oscillator are summarized below:

► $$Z = \left(\frac{k_\mathrm{B}T}{\hbar\omega} \right)^N ;$$

$$F = -Nk_\mathrm{B}T \ln \frac{k_\mathrm{B}T}{\hbar\omega};$$

$$S = Nk_\mathrm{B} + Nk_\mathrm{B} \ln \frac{k_\mathrm{B}T}{\hbar\omega};$$

$$U = Nk_\mathrm{B}T \ .$$

(1.81)

Probability Distribution

Finally, the probability distribution in positions and momenta of a classical oscillator is written with the determined normalization constants (the normalization constants are the reciprocal values of the corresponding parts of the partition function):

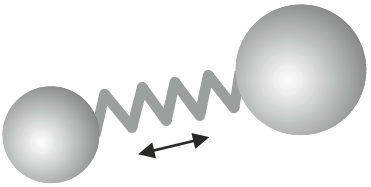
$$\begin{aligned} \mathrm{d}w(x) &= \sqrt{\frac{m\omega^2}{2\pi k_\mathrm{B}T}}\,\exp\left[-\frac{m\omega^2 x^2}{2k_\mathrm{B}T}\right]\mathrm{d}x\,, \\ \mathrm{d}w(p) &= \sqrt{\frac{1}{2\pi m k_\mathrm{B}T}}\,\exp\left[-\frac{p^2}{2m k_\mathrm{B}T}\right]\mathrm{d}p\ . \end{aligned} \tag{1.82}$$

Both distributions have the character of a Gaussian package, and in principle, very large displacements from the equilibrium position and large momenta are possible for an ensemble of many oscillators at a given temperature. However, these are highly improbable. Notably, both the amplitude and the half-width of the Gaussian package depend on the temperature.



1.5 Further Examples

1.5.1 Quantum Oscillator (Vibrator)



Thermodynamic Quantities

The problem of a single harmonic oscillator is defined by relations (I.2.50)

$$\hat{\mathbf{H}}\,|n\rangle = \varepsilon_n\,|n\rangle\,; \qquad \hat{\mathbf{H}} \equiv \frac{\hat{\mathbf{P}}^2}{2m} + \frac{1}{2}\; m\omega^2\hat{\mathbf{X}}^2.$$

The harmonic oscillator energy spectrum (I.2.63) was derived in several ways (Schrödinger, Dirac, and Heisenberg representations):

$$\varepsilon_n = \left(n+\frac{1}{2}\right)\hbar\omega\,, \qquad |n\rangle = \psi_n(\xi) = \alpha_n H_n(\xi)\,\mathrm{e}^{-\xi^2/2}\;; \qquad n = 0, 1, 2, \ldots$$

The independent variable and normalization coefficients are given by the relations:

$$\xi \equiv \sqrt{\frac{m\omega}{\hbar}}\;\; x\,; \qquad \alpha_n = \frac{1}{\sqrt{\pi^{1/2}\;\; n!\;\; 2^n}}.$$

The calculation of the partition function can now be initiated, first for a single oscillator:

$$z = \sum_{n=0}^{\infty} \exp\left[-\frac{(n+\frac{1}{2})\hbar\omega}{k_{\mathrm{B}}T}\right] =$$

$$= \exp\left[-\frac{\hbar\omega}{2\,k_{\mathrm{B}}T}\right] \sum_{n=0}^{\infty} \exp\left[-\frac{n\,\hbar\omega}{k_{\mathrm{B}}T}\right] =$$

$$= \exp\left[-\frac{\hbar\omega}{2\,k_{\mathrm{B}}T}\right] \sum_{n=0}^{\infty} \left[\exp\left(-\frac{\hbar\omega}{k_{\mathrm{B}}T}\right)\right]^n.$$

The remaining series is a geometric series that can be easily summed using relation (I.3.584), with the quantum number n as the exponent:

$$z = \frac{\mathrm{e}^{-\frac{\hbar\omega}{2k_\mathrm{B}T}}}{1-\mathrm{e}^{-\frac{\hbar\omega}{k_\mathrm{B}T}}} = \frac{1}{\mathrm{e}^{+\frac{\hbar\omega}{2k_\mathrm{B}T}} - \mathrm{e}^{-\frac{\hbar\omega}{2k_\mathrm{B}T}}} = \frac{1}{2\sinh\left(\dfrac{\hbar\omega}{2k_\mathrm{B}T}\right)}.$$

For N independent oscillators, the partition function is the corresponding power:

$$Z = 2^{-N}\,\sinh^{-N}\left(\frac{\hbar\omega}{2k_\mathrm{B}T}\right).$$

Free energy, entropy, and internal energy of the system are found below by the standard procedure:

$$F = -k_\mathrm{B}T\ln Z = Nk_\mathrm{B}T\,\ln\left[2\sinh\left(\frac{\hbar\omega}{2k_\mathrm{B}T}\right)\right];$$

$$S = -\frac{\partial F}{\partial T} = -Nk_\mathrm{B}\ln\left[2\sinh\left(\frac{\hbar\omega}{2k_\mathrm{B}T}\right)\right] + \frac{N\hbar\omega}{2T}\coth\frac{\hbar\omega}{2k_\mathrm{B}T};$$

$$U = F + TS = \frac{N\hbar\omega}{2}\coth\frac{\hbar\omega}{2k_\mathrm{B}T}.$$

The results are summarized below (coefficient β is used instead of temperature):

$$\begin{aligned}
&Z = 2^{-N}\,\sinh^{-N}\left(\beta\hbar\omega/2\right);\\
&F = Nk_\mathrm{B}T\,\ln\left[2\sinh\left(\beta\hbar\omega/2\right)\right];\\
&S = -Nk_\mathrm{B}\ln\left[2\sinh\left(\beta\hbar\omega/2\right)\right] + \tfrac{1}{2}N\beta\hbar\omega k_\mathrm{B}\coth\left(\beta\hbar\omega/2\right);\\
&U = \frac{N\hbar\omega}{2}\coth\left(\beta\hbar\omega/2\right).
\end{aligned} \qquad (1.83)$$

The last relation for internal energy was derived by Albert Einstein in 1906. In this quantum case, entropy no longer diverges for $T \to 0$. The mean energy of the harmonic oscillator system in the limits of low and high temperatures is found below:

1) $T \to 0$: $\quad U = \lim\limits_{T\to 0}\dfrac{N\hbar\omega}{2}\coth\left(\beta\hbar\omega/2\right) = N\dfrac{\hbar\omega}{2}.$

2) $k_\mathrm{B}T \gg \hbar\omega$: $\quad U = \dfrac{N\hbar\omega}{2}\coth\left(\beta\hbar\omega/2\right) \to \dfrac{N\hbar\omega}{2}\dfrac{2}{\beta\hbar\omega} = Nk_\mathrm{B}T.$

The low-temperature case is purely quantum mechanical. At absolute zero, all oscillators are in the ground state, and the internal energy equals the number of oscillators multiplied by the ground-state energy. Conversely, the high-temperature case is purely classical. The mean thermal energy is substantially larger than the fundamental energy quantum, and the internal energy is given by classical relation (1.81). The temperature at which the mean thermal energy equals the vibrational quantum (the argument of the

exponentials equals one, $\hbar\omega \approx k_B T$) is called the vibrational temperature, denoted as T_V. The low-temperature limit implies $T \ll T_V$, whereas the high-temperature limit means $T \gg T_V$. The vibrational temperature is given by the relation:

► $$T_V = \frac{\hbar\omega}{k_B} \ . \tag{1.84}$$

For various quantum vibrators (such as vibrating molecules or a crystal lattice), this quantity constitutes a characteristic parameter. Both oscillation frequency and vibrational temperature decrease with entity mass. For instance, this temperature is 6,100 K for a hydrogen molecule, but "only" 3,340 K for heavier nitrogen.

◖ Example 1.6: Philip Morse Potential

Determine the vibrational temperature for a pair of atoms whose interaction is described by the potential named after the American physicist Philip Morse [10]:

$$\begin{aligned} V(r) &= V_0 \left[1 - \mathrm{e}^{-\alpha(r-r_0)}\right]^2 = \\ &= V_0 - 2V_0\, \mathrm{e}^{-\alpha(r-r_0)} + V_0\, \mathrm{e}^{-2\alpha(r-r_0)} . \end{aligned} \tag{1.85}$$

Solution: First, the potential energy minimum is determined by setting the first derivative to zero. The minimum is easily found at $r = r_0$. For small displacements from the equilibrium, the oscillation stiffness equals the second derivative of the potential energy at the minimum (I.1.26). The frequency is determined from the oscillator stiffness:

$$\omega = \sqrt{V''(r_0)/m} \ .$$

A general relation for the vibrational temperature is now easily derived:

$$T_V = \frac{\hbar\omega}{k_B} = \frac{\hbar}{k_B}\sqrt{V''(r_0)/m} \ .$$

The vibrational temperature can be determined from the characteristic profile of the potential energy. In this case, performing the derivatives yields the resulting relation:

$$T_V = \frac{\hbar\omega}{k_B} = \frac{\hbar\alpha}{k_B}\sqrt{\frac{2V_0}{m}} \ .$$

◗

The heat capacity of the harmonic oscillator system is determined below:

$$C \equiv \frac{\partial U}{\partial T} = N k_B \left(\frac{\hbar\omega}{2k_B T}\right)^2 \sinh^{-2}\left(\frac{\hbar\omega}{2k_B T}\right). \tag{1.86}$$

The calculated profiles of internal energy and heat capacity are plotted below. At low temperatures, internal energy is given by zero-point oscillations ($N\hbar\omega/2$), while at high temperatures, it is a linear function of temperature. The transition occurs around the vibrational temperature. Well below this temperature, vibrational degrees of freedom contribute nothing to the heat capacity and are said to be "*frozen*." At high temperatures, they contribute a constant value to the specific heat.

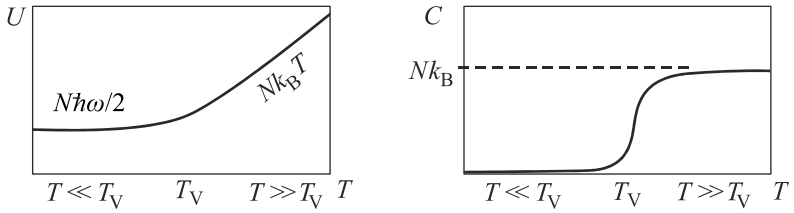


Fig. 1.10: Internal energy and heat capacity profiles for a quantum oscillator.

Each vibrational degree of freedom adds a value of k_B to the system heat capacity at high temperatures. In crystalline solids, the heat capacity relative to the number of vibrating entities is close to $3\,k_B$. This relationship was experimentally discovered in crystals by French physicists Pierre Louis Dulong and Alexis Thérèse Petit in 1819. The Boltzmann constant can be interpreted as the heat capacity of a single one-dimensional harmonic oscillator. Notably, for crystalline substances, room temperature is above the vibrational temperature (unlike in diatomic gases).

Probability of Finding Vibrator in Energy State

The probability is given by the Boltzmann factor:

$$w_n = C\,\exp\left[-\frac{(n+1/2)\hbar\omega}{k_B T}\right].$$

The normalization constant is determined either from the condition that the sum of all probabilities equals one, or by recognizing that it represents the reciprocal of the partition function. The resulting relation is:

► $$w_n = 2\sinh\left(\frac{\beta\hbar\omega}{2}\right)\mathrm{e}^{-\beta(n+1/2)\hbar\omega}\quad. \tag{1.87}$$

The analysis is performed below in the low- and high-temperature limits:

1) $T \ll T_V$: $$w_n = \lim_{T\to 0} 2\sinh\left(\frac{\hbar\omega}{2k_B T}\right)\exp\left(-\frac{(n+1/2)\hbar\omega}{k_B T}\right) = \begin{cases}1 & \text{pro } n=0,\\ 0 & \text{pro } n\neq 0;\end{cases}$$

2) $T \gg T_V$: $$w_n = 2\sinh\left(\frac{\hbar\omega}{2k_B T}\right)\exp\left(-\frac{(n+1/2)\hbar\omega}{k_B T}\right) \to 2\cdot\frac{\hbar\omega}{2k_B T}\cdot 1 = \frac{\hbar\omega}{k_B T}.$$

The first case is again purely quantum mechanical, showing that only the ground energy state is occupied at absolute zero. Conversely, the second case is classical. At high temperatures, all states are represented equally.

Probability of Finding Vibrator in Position

The probability density can be determined directly (the wave functions are real):

$$\rho(x) = \langle x|\hat{\rho}|x\rangle = \langle x|\mathrm{e}^{-\beta\hat{H}}|x\rangle = \langle x|n\rangle\langle n|\mathrm{e}^{-\beta\hat{H}}|m\rangle\langle m|x\rangle = \sum_{n,m}\psi_n^* w_n \delta_{nm}\psi_m = \sum_n \psi_n^2 w_n.$$

The respective Hermite polynomials are substituted for the wave functions, and the probability distribution in the energy representation (1.87) is applied. Only the respective series of Hermite polynomials needs to be summed, which is not easy. Another option is an indirect calculation using a trick devised by Swiss-American nuclear physicist, one of the directors of the CERN complex and Nobel laureate in physics, Felix Bloch. The actions of operators $\hat{p}\hat{\rho}$ and $\hat{x}\hat{\rho}$ are utilized in the *x*-representation and in the Heisenberg matrix representation:

$$\frac{\mathrm{d}\rho}{\mathrm{d}x} = \sum_n 2\psi_n \frac{\mathrm{d}\psi_n}{\mathrm{d}x} w_n = \frac{2}{-\mathrm{i}\hbar} \sum_n \psi_n (-\mathrm{i}\hbar\partial\psi_n/\partial x) w_n = \frac{2\mathrm{i}}{\hbar} \sum_n \psi_n (\hat{P}\psi_n) w_n =$$

$$= \frac{2\mathrm{i}}{\hbar} \sum_{n,k} \psi_n P_{nk} \psi_k w_n = -\sum_{k=0}^{\infty} \sqrt{\frac{2m\omega}{\hbar}} \left(\sqrt{k+1}\,\psi_{k+1}\psi_k - \sqrt{k}\,\psi_{k-1}\psi_k \right) w_k =$$

$$= -\sqrt{\frac{2m\omega}{\hbar}} \sum_{k=0}^{\infty} \sqrt{k+1} \left(w_k - w_{k+1} \right) \psi_{k+1}\psi_k =$$

$$= -\sqrt{\frac{2m\omega}{\hbar}} \left(1 - \mathrm{e}^{-\beta\hbar\omega} \right) \sum_{k=0}^{\infty} \sqrt{k+1}\, w_k\, \psi_{k+1}\psi_k \,.$$

The calculation utilizes the matrix expansion of the momentum operator from Section I.2.3.3 and the reality of the wave functions. Analogously, operator $\hat{x}\hat{\rho}$ is sought:

$$x\rho = \langle x | \hat{x}\hat{\rho} | x \rangle = \langle x | \hat{x}\, \mathrm{e}^{-\beta\hat{H}} | x \rangle = \sum_{n,k} \psi_n \psi_k X_{nk} w_k =$$

$$= \sum_{k=0}^{\infty} \sqrt{\frac{\hbar}{2m\omega}} \left(\sqrt{k+1}\,\psi_{k+1}\psi_k + \sqrt{k}\,\psi_{k-1}\psi_k \right) w_k =$$

$$= \sqrt{\frac{\hbar}{2m\omega}} \sum_{k=0}^{\infty} \sqrt{k+1} \left(w_k + w_{k+1} \right) \psi_{k+1}\psi_k =$$

$$= \sqrt{\frac{\hbar}{2m\omega}} \left(1 + \mathrm{e}^{-\beta\hbar\omega} \right) \sum_{k=0}^{\infty} \sqrt{k+1}\, w_k\, \psi_{k+1}\psi_k \,.$$

The actions of both operators lead to the same series. Therein lies the ingenious Bloch trick. Dividing both obtained equations mathematically eliminates the inconvenient series, yielding a simple equation:

$$\frac{\dfrac{\mathrm{d}\rho}{\mathrm{d}x}}{x\rho} = -\frac{2m\omega}{\hbar} \frac{\left(1 - \mathrm{e}^{-\beta\hbar\omega} \right)}{\left(1 + \mathrm{e}^{-\beta\hbar\omega} \right)},$$

which leads to a differential equation in a separated form:

$$\frac{\mathrm{d}\rho}{\rho} = -\frac{2m\,\omega}{\hbar} \tanh\left(\frac{\beta\hbar\omega}{2} \right) x\,\mathrm{d}x \,.$$

The solution is simple; the integration constant is determined from the normalization:

► $$\rho(x) = \sqrt{\frac{\lambda}{\pi}}\,\exp\left[-\lambda x^2\right]; \qquad \lambda(T) \equiv \frac{m\omega}{\hbar}\,\mathrm{th}\left(\frac{\hbar\omega}{2k_\mathrm{B}T}\right). \tag{1.88}$$

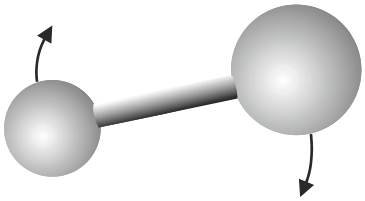
This now-famous formula was derived by Felix Bloch in 1932. The formula carries great importance in the theory of crystal lattice vibrations. As in previous cases, the low- and high-temperature limits are derived below:

1) $T \ll T_\mathrm{V}$: $$\rho(x) \to \sqrt{\frac{m\omega}{\pi\hbar}}\,\exp\left[-\frac{m\omega x^2}{\hbar}\right] = \psi_0^2(x).$$

2) $T \gg T_\mathrm{V}$: $$\rho(x) \to \sqrt{\frac{m\omega^2}{2\pi k_\mathrm{B}T}}\,\exp\left[-\frac{m\omega^2 x^2}{2k_\mathrm{B}T}\right].$$

The first case corresponds again to a purely quantum mechanical solution, representing the oscillator probability density in the quantum ground state. The high-temperature case yields classical result (1.82).

1.5.2 Quantum Rotator



Properties of a rotating particle with a non-zero angular momentum **L** and a non-zero moment of inertia J are examined below. This system can represent, for example, a rotating diatomic molecule or a similar structure. First, the partition function for a system consisting of a single molecule is derived. The standard translational relation $p^2/2m$ for rotational motions transforms into an analogous relation $L^2/2J$:

$$\varepsilon = \frac{L^2}{2J} = \frac{l(l+1)\hbar^2}{2J}; \qquad l = 0,1,2,...$$

Relation (I.2.149) for the quantization of the angular momentum magnitude was utilized. Each such energy state is degenerate, occurring $2l+1$ times. Individual states with identical energy differ by the magnetic quantum number (see *Selected Topics I*), assuming values $m = -l, -l+1, \ldots, 0, \ldots, l-1, l$. Each Boltzmann factor must be taken into account in the partition function as many times as the given state is degenerate:

$$z = \sum_{l=0}^{\infty} g_l \exp\left[-\frac{l(l+1)\hbar^2}{2J\,k_\mathrm{B}T}\right] = \sum_{l=0}^{\infty}(2l+1)\exp\left[-\frac{l(l+1)\hbar^2}{2J\,k_\mathrm{B}T}\right].$$

This expression represents a series that cannot be solved analytically. This series can only be summed numerically or in the low- and high-temperature limits. These regions are determined by the exponential argument. If the argument equals one, a characteristic

temperature is obtained at which the thermal energy equals the rotational energy. Based on the relation $\hbar^2 \approx 2J\,k_B T$, the *rotational temperature* is given by:

$$T_R = \frac{\hbar^2}{2k_B J} \ . \tag{1.89}$$

Similarly to the vibrational temperature, the rotational temperature represents a highly characteristic parameter for a given system. The values of the rotational and vibrational temperatures for selected gases are listed below:

Gas	Rotational Temperature	Vibrational Temperature
H_2	85 K	6100 K
HCl	15 K	4140 K
N_2	3 K	3340 K
O_2	2 K	2230 K

The series is summed below in the low- and high-temperature limits:

Low Temperatures

At low temperatures ($T \ll T_R$), exponentials decrease sharply with increasing l. Because terms converge very rapidly, taking into account only the first two is sufficient:

$$z = 1 + 3\exp\left[-\frac{\hbar^2}{Jk_B T}\right] = 1 + 3\exp\left(-2T_R/T\right).$$

Thermodynamic quantities in the low-temperature limit are determined standardly:

$$Z = \left[1 + 3\exp\left(-2T_R/T\right)\right]^N ,$$

$$F = -k_B T \ln Z = -N k_B T \ln\left[1 + 3\exp\left(-2T_R/T\right)\right].$$

The second term is small, allowing a first-order Taylor expansion $\mathrm{n}(1+x) \sim x$. For free energy, a simpler formula is obtained:

$$F \approx -3N k_B T \exp\left(-2T_R/T\right).$$

Entropy and internal energy are found below:

$$S = -\frac{\partial F}{\partial T} = 3Nk_B \exp\left(-2T_R/T\right) + 6Nk_B \frac{T_R}{T}\exp\left(-2T_R/T\right),$$

$$U = F + TS = 6N k_B T_R \exp\left(-2T_R/T\right).$$

High Temperatures

At high temperatures ($T \gg T_R$), many states with large l are occupied, and the summation can be replaced by integration:

$$z = \sum_{l=0}^{\infty} (2l+1)\exp\left[-l(l+1)\frac{T_R}{T}\right] \approx \int_0^{\infty} (2x+1)\exp\left[-x(x+1)\frac{T_R}{T}\right]\mathrm{d}x \,.$$

The substitution $\xi = x(x+1)$is performed in the integral below:

$$z \approx \int_0^\infty \exp\left[-\frac{T_\mathrm{R}}{T}\,\xi\right]\mathrm{d}\xi \;= \frac{T}{T_\mathrm{R}}\,.$$

Thermodynamic quantities in the high-temperature limit are determined standardly:

$$Z = \left(T / T_\mathrm{R}\right)^N ,$$

$$F = -k_\mathrm{B}T \ln Z = -Nk_\mathrm{B}T \ln\left(T/T_\mathrm{R}\right),$$

$$S = -\frac{\partial F}{\partial T} = Nk_\mathrm{B}\left[1+\ln\left(T/T_\mathrm{R}\right)\right],$$

$$U = F + TS = N k_\mathrm{B}T\ .$$

As expected, classical results are obtained in the high-temperature limit. The results obtained for both the low- and high-temperature limits are summarized below:

$T \ll T_\mathrm{R}$	$T \gg T_\mathrm{R}$
$Z = \left[1+3\exp\left(-2T_\mathrm{R}/T\right)\right]^N$	$Z = \left(T / T_R\right)^N$
$F \doteq -3Nk_\mathrm{B}T \exp\left(-2T_\mathrm{R}/T\right)$	$F = -N k_\mathrm{B}T \ln\left(T/T_\mathrm{R}\right)$
$S = 3Nk_\mathrm{B} \exp\left(-2T_R / T\right) + +6Nk_\mathrm{B}\dfrac{T_\mathrm{R}}{T}\exp\left(-2T_\mathrm{R}/T\right)$	$S = Nk_\mathrm{B}\left[1+\ln\left(T/T_\mathrm{R}\right)\right]$
$U = 6N k_\mathrm{B}T_\mathrm{R} \exp\left(-2T_\mathrm{R}/T\right)$	$U = N k_\mathrm{B}T$

(1.90)

The situation is analogous to the oscillator case. Up to the rotational temperature, the system cannot absorb heat and its degrees of freedom are "*frozen*." Above this temperature, each rotator contributes the Boltzmann constant to the heat capacity.

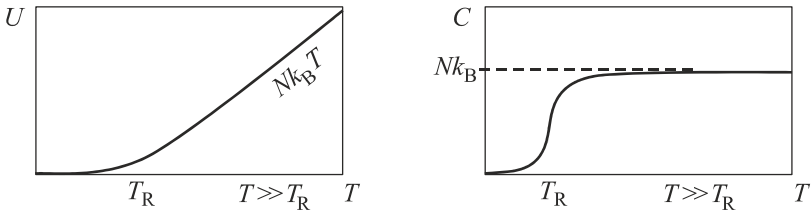


Fig. 1.11: Internal energy and heat capacity profiles for a quantum rotator.

For diatomic molecules, rotational temperatures are substantially lower than vibrational temperatures. During gradual gas heating, the rotational degrees of freedom are released first, followed later by the vibrational degrees of freedom.

◖ **Example 1.7:** Determine the most probable rotational quantum number for a quantum rotator (the state with the highest representation).

Solution: The probability of the system being in a state with the azimuthal quantum number l is given by the formula derived above:

$$w_l = A(2l+1)\exp\left[-\frac{l(l+1)\hbar^2}{2Jk_\mathrm{B}T}\right] = A(2l+1)\exp\left[-l(l+1)\frac{T_\mathrm{R}}{T}\right].$$

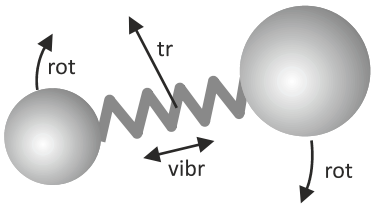
At low temperatures, the system does not rotate, and the probability is nearly zero. At high temperatures, the maximum is found by the standard procedure, treating the variable l as a continuous variable:

$$\frac{\partial w_l}{\partial l} = 0 \qquad \Rightarrow \qquad l_\mathrm{max} = \sqrt{\frac{T}{2T_\mathrm{R}}} - \frac{1}{2} \approx \sqrt{\frac{T}{2T_\mathrm{R}}}\,.$$

Typical azimuthal quantum numbers of rotating molecules at a given temperature can be determined from the calculated relationship. ◗

1.5.3 Diatomic Gas

A system composed of N diatomic molecules with distinguishable atoms is considered below (otherwise, the wave function symmetry would have to be analyzed). Many gases behave in this manner. Single molecule energy consists of translational, vibrational, rotational, and other (for instance, nuclear) degrees of freedom.



The partition function for a single molecule represents a *product* of the partition functions of the individual degrees of freedom, and the thermodynamic quantities constitute a *sum* of the corresponding terms:

$$\varepsilon = \varepsilon_\mathrm{tr} + \varepsilon_\mathrm{vib} + \varepsilon_\mathrm{rot} + \varepsilon_\mathrm{nuc} + \cdots \quad .$$

$$z = \sum \mathrm{e}^{-\beta(\varepsilon_\mathrm{tr} + \varepsilon_\mathrm{vib} + \varepsilon_\mathrm{rot} + \cdots)} =$$

$$= \int \mathrm{e}^{-\beta\varepsilon_\mathrm{tr}}\,\mathrm{d}\Gamma \cdot \sum \mathrm{e}^{-\beta\varepsilon_\mathrm{vib}} \cdot \sum \mathrm{e}^{-\beta\varepsilon_\mathrm{rot}} \cdots =$$

$$= z_\mathrm{tr} \cdot z_\mathrm{vib} \cdot z_\mathrm{rot} \cdots$$

The total partition function for N particles is given by:

$$Z = z_\mathrm{tr}^N \cdot z_\mathrm{vib}^N \cdot z_\mathrm{rot}^N \cdots$$

By definition, the fundamental thermodynamic quantities are additive, satisfying:

$$F = -k_B T \ln Z = F_{tr} + F_{vib} + F_{rot} + \cdots ,$$
$$S = -\frac{\partial F}{\partial T} = S_{tr} + S_{vib} + S_{rot} + \cdots ,$$
$$U = F + TS = U_{tr} + U_{vib} + U_{rot} + \cdots ,$$
$$C_V = \left(\frac{\partial U}{\partial T}\right)_V = C_{tr} + C_{vib} + C_{rot} + \cdots .$$

Heat capacity contributions from individual degrees of freedom are examined below:

Translational Degrees of Freedom

$$U_{tr} = \frac{3}{2} N k_B T \quad \Rightarrow \quad c \equiv \frac{1}{N}\left(\frac{\partial U}{\partial T}\right)_V = \frac{3}{2} k_B .$$

The translational degrees of freedom contribute a constant value to the specific heat capacity of the gas (heat capacity relative to the particle number).

Vibrational Degrees of Freedom

$$U_{vib} = \frac{N\hbar\omega}{2} \coth\left(\frac{\hbar\omega}{2k_B T}\right) \quad \Rightarrow \quad c \equiv \frac{1}{N}\left(\frac{\partial U}{\partial T}\right)_V = k_B \left(\frac{\hbar\omega}{2k_B T}\right)^2 \sinh^{-2}\left(\frac{\hbar\omega}{2k_B T}\right).$$

The calculated profiles are plotted below. At low temperatures, internal energy is determined by zero-point oscillations ($N\hbar\omega/2$), whereas at high temperatures, it is a linear function of temperature. The transition occurs around the vibrational temperature. Below this temperature, vibrational degrees of freedom contribute nothing to the specific heat and are said to be "*frozen*." At high temperatures, they contribute a constant value. Taking the low- and high-temperature limits yields:

$$T \ll T_V \quad \Rightarrow \quad U = N\hbar\omega/2 , \quad C = 0 , \quad c = 0 ;$$
$$T \gg T_V \quad \Rightarrow \quad U = N k_B T , \quad C = N k_B , \quad c = k_B .$$

At high temperatures, each vibrational degree of freedom adds a value of k_B to the gas heat capacity. As established previously, the Boltzmann constant can be interpreted as the heat capacity of a single vibrating molecule.

Rotational Degrees of Freedom

For rotational degrees of freedom, the analytical profile of internal energy and heat capacity at constant volume is unknown. However, the values in the low- and high-temperature limits relative to the rotational temperature are known, see (1.90):

$$T \ll T_R \quad \Rightarrow \quad U = 6N k_B T_R \exp(-2T_R / T) , \quad C \simeq 0 , \quad c \simeq 0 ;$$
$$T \gg T_R \quad \Rightarrow \quad U = N k_B T , \quad C = N k_B , \quad c = k_B .$$

Evidently, the rotational states contribute to the specific heat in the same manner as the vibrational states; the contribution manifests at temperatures higher than the rotational temperature. At lower temperatures, the rotational states are again "*frozen*." Each rota-

tional state contributes a value of the Boltzmann constant to the heat capacity. The resulting specific heat capacity profile exhibits a stepwise character:

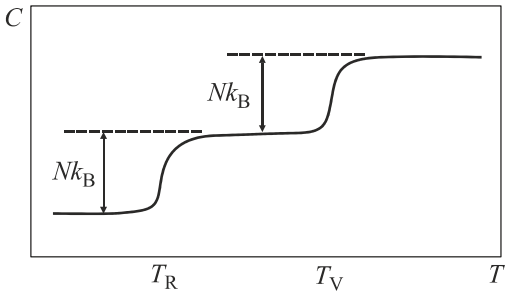


Fig. 1.12: Heat capacity of a diatomic gas.

As the temperature increases, additional degrees of freedom are activated; each "*unfrozen*." degree of freedom contributes a value of k_B to the specific heat capacity. Each translational degree of freedom contributes a value of $k_B/2$ independently of the temperature, totaling $k_B/2$. Further details regarding rotational and vibrational spectra can be found in textbook [10].

Note: For cyanogen HCN, the transition between the second and first rotational levels corresponds to a wavelength of 1.3 mm, correlating with the cosmic microwave background (CMB) radiation maximum. Therefore CMB radiation causes rotational excitation of the interstellar gas, with a rotational temperature of 15 K.

1.5.4 Anharmonic Oscillator

A interesting situation occurs if the third term (minimum asymmetry) or even the fourth term in the potential energy Taylor expansion near the minimum is significant. This no longer describes harmonic oscillations, but an anharmonic oscillator with energy:

$$E = \frac{p^2}{2m} + \frac{m\omega^2 x^2}{2} + a_3 x^3 + a_4 x^4 . \tag{1.91}$$

Under the assumption of high temperature ($k_B T \gg \hbar\omega$), the classical partition function for N independent oscillators can be calculated. At low temperatures, the problem would require a quantum mechanical solution. For a single oscillator

$$\begin{aligned} z &\equiv \int \mathrm{e}^{[-\beta E]} \frac{\mathrm{d}p\,\mathrm{d}x}{2\pi\hbar} = \\ &= \frac{1}{2\pi\hbar} \int \exp\left[-\left(\frac{p^2}{2mk_B T} + \frac{m\omega^2 x^2}{2k_B T} + \frac{a_3 x^3}{k_B T} + \frac{a_4 x^4}{k_B T} \right) \right] \mathrm{d}p\,\mathrm{d}x = \\ &= \frac{\sqrt{2\pi m k_B T}}{2\pi\hbar} \int_{-\infty}^{+\infty} \exp\left[-\frac{m\omega^2 x^2}{2k_B T} \right] \exp\left[-\frac{a_3 x^3}{k_B T} - \frac{a_4 x^4}{k_B T} \right] \mathrm{d}x \end{aligned} \tag{1.92}$$

Notably, any small third-order anharmonic term leads to a non-convergent integral (1.92). Convergence is ensured by the presence of the fourth-order term with $a_4 > 0$. It is assumed below that the anharmonic terms are small compared to the harmonic one, allowing an expansion of the second exponential up to the second order in the argument:

$$z \approx \frac{\sqrt{2\pi m k_\mathrm{B} T}}{2\pi\hbar} \int\limits_{-\infty}^{+\infty} \exp\left[-\frac{m\omega^2 x^2}{2k_\mathrm{B}T}\right]\left(1-\frac{a_3 x^3}{k_\mathrm{B}T}-\frac{a_4 x^4}{k_\mathrm{B}T}+\frac{1}{2k_\mathrm{B}^2T^2}\left(a_3 x^3+a_4 x^4\right)^2\right)\mathrm{d}x\,.$$

This expression represents a sum of Gaussian integrals with various powers of x multiplying the fundamental exponential. Integrals with odd powers are zero; therefore, only even terms up to the sixth order in x are retained:

$$z \approx \frac{\sqrt{2\pi m k_\mathrm{B} T}}{2\pi\hbar} \int\limits_{-\infty}^{+\infty} \exp\left[-\frac{m\omega^2 x^2}{2k_\mathrm{B}T}\right]\left(1-\frac{a_4 x^4}{k_\mathrm{B}T}+\frac{a_3^2 x^6}{2k_\mathrm{B}^2T^2}+\cdots\right)\mathrm{d}x$$

Following trivial integration using relation (I.3.574), or via software tools such as MATLAB or Mathematica, the following expression is obtained:

$$z \approx \frac{\sqrt{2\pi m k_\mathrm{B} T}}{2\pi\hbar}\sqrt{\frac{2\pi k_\mathrm{B} T}{m\omega^2}}\left(1-3\frac{a_4 k_\mathrm{B} T}{m^2\omega^4}+\frac{15}{2}\frac{a_3^2 k_\mathrm{B} T}{m^3\omega^6}+\cdots\right) \Rightarrow$$

$$z \approx \frac{k_\mathrm{B} T}{\hbar\omega}-3\frac{a_4 k_\mathrm{B}^2 T^2}{\hbar m^2\omega^5}+\frac{15}{2}\frac{a_3^2 k_\mathrm{B}^2 T^2}{\hbar m^3\omega^7}+\cdots \Rightarrow$$

$$Z \approx \left(\frac{k_\mathrm{B} T}{\hbar\omega}-3\frac{a_4 k_\mathrm{B}^2 T^2}{\hbar m^2\omega^5}+\frac{15}{2}\frac{a_3^2 k_\mathrm{B}^2 T^2}{\hbar m^3\omega^7}+\cdots\right)^N . \tag{1.93}$$

The thermodynamic quantities are determined by the standard procedure below:

$$F = -k_\mathrm{B} T \ln Z = -N k_\mathrm{B} T \ln\left(\frac{k_\mathrm{B} T}{\hbar\omega}-3\frac{a_4 k_\mathrm{B}^2 T^2}{\hbar m^2\omega^5}+\frac{15}{2}\frac{a_3^2 k_\mathrm{B}^2 T^2}{\hbar m^3\omega^7}+\cdots\right), \tag{1.94}$$

$$S = -\frac{\partial F}{\partial T} \Rightarrow$$

$$S \approx N k_\mathrm{B} \ln\left(\frac{k_\mathrm{B} T}{\hbar\omega}-3\frac{a_4 k^2 T_\mathrm{B}^2}{\hbar m^2\omega^5}+\frac{15}{2}\frac{a_3^2 k_\mathrm{B}^2 T^2}{\hbar m^3\omega^7}+\cdots\right)+N k_\mathrm{B}\frac{1-6\dfrac{a_4 k_\mathrm{B} T}{m^2\omega^4}+15\dfrac{a_3^2 k_\mathrm{B} T}{m^3\omega^6}+\cdots}{\left(1-3\dfrac{a_4 k_\mathrm{B} T}{m^2\omega^4}+\dfrac{15}{2}\dfrac{a_3^2 k_\mathrm{B} T}{m^3\omega^6}+\cdots\right)},$$

$$U = F + TS = N k_\mathrm{B} T\frac{1-6\dfrac{a_4 k_\mathrm{B} T}{m^2\omega^4}+15\dfrac{a_3^2 k_\mathrm{B} T}{m^3\omega^6}+\cdots}{\left(1-3\dfrac{a_4 k_\mathrm{B} T}{m^2\omega^4}+\dfrac{15}{2}\dfrac{a_3^2 k_\mathrm{B} T}{m^3\omega^6}+\cdots\right)} \Rightarrow,$$

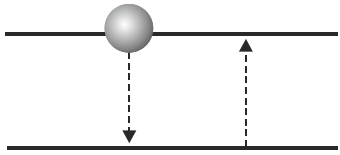
$$U \approx Nk_{\mathrm{B}}T\left(1-3\frac{a_4 k_{\mathrm{B}}T}{m^2\omega^4}+\frac{15}{2}\frac{a_3^2 k_{\mathrm{B}}T}{m^3\omega^6}+\cdots\right), \tag{1.95}$$

$$C=\frac{\partial U}{\partial T}\approx Nk_{\mathrm{B}}+Nk_{\mathrm{B}}^2\left(\frac{15\,a_3^2}{m^3\omega^6}-\frac{6a_4}{m^2\omega^4}\right)T+\cdots. \tag{1.96}$$

Evidently, the anharmonicity of vibrations in crystals or molecules leads to a violation of the Dulong-Petit law. A linear and possibly a quadratic term in temperature appears in the heat capacity.

Notably, if the fourth power of x in energy expansion (1.91) is set exactly to zero, integral (1.92) does not converge and the system becomes unstable; with increasing displacement from equilibrium, the potential energy approaches infinity. Assuming that the fourth power of x in the expansion is even minutely small ensures both integral convergence and system stability.

1.5.5 Two-Level System



The behavior of a quantum system with two close energy levels $\varepsilon_0 = 0$ and $\varepsilon_1 = \varepsilon$, characterized by degeneracy factors g_0 and g_1, is analyzed below. The partition function for a single particle contains only two terms:

$$z = g_0 + g_1 \exp(-\beta\varepsilon). \tag{1.97}$$

The procedure follows the standard method below:

$$Z=\left[g_0+g_1\exp\left(-\frac{\varepsilon}{k_{\mathrm{B}}T}\right)\right]^N, \tag{1.98}$$

$$F=-k_{\mathrm{B}}T\ln Z=-Nk_{\mathrm{B}}T\ln\left[g_0+g_1\exp\left(-\frac{\varepsilon}{k_{\mathrm{B}}T}\right)\right], \tag{1.99}$$

$$S=-\frac{\partial F}{\partial T}=Nk_{\mathrm{B}}\ln\left[g_0+g_1\exp\left(-\frac{\varepsilon}{k_{\mathrm{B}}T}\right)\right]+\frac{N\varepsilon/T}{\left[1+g\exp\left(\frac{\varepsilon}{k_BT}\right)\right]};\quad g\equiv\frac{g_0}{g_1}, \tag{1.100}$$

$$U=F+TS=\frac{N\varepsilon}{\left[1+g\exp\left(\frac{\varepsilon}{k_{\mathrm{B}}T}\right)\right]}, \tag{1.101}$$

$$C_V = \frac{\partial U}{\partial T} = \frac{Ng\,\varepsilon^2}{k_\mathrm{B}T^2} \frac{\exp\left(\dfrac{\varepsilon}{k_\mathrm{B}T}\right)}{\left[1 + g\,\exp\left(\dfrac{\varepsilon}{k_\mathrm{B}T}\right)\right]^2}, \tag{1.102}$$

► $$c_V = \frac{C_V}{N} = \frac{g\,\varepsilon^2}{k_\mathrm{B}T^2} \frac{\exp\left(\dfrac{\varepsilon}{k_\mathrm{B}T}\right)}{\left[1 + g\,\exp\left(\dfrac{\varepsilon}{k_\mathrm{B}T}\right)\right]^2}. \tag{1.103}$$

A well-known relation for the two-level system contribution to the specific heat is obtained, often referred to as the *Schottky anomaly*. This contribution converges to zero in both the low- and high-temperature regions, implying the existence of a temperature at which the specific heat contribution is maximal. The maximum can be determined numerically; for $g = 1$, the result is $c_\mathrm{max} \sim 0.34\ k_\mathrm{B}$.

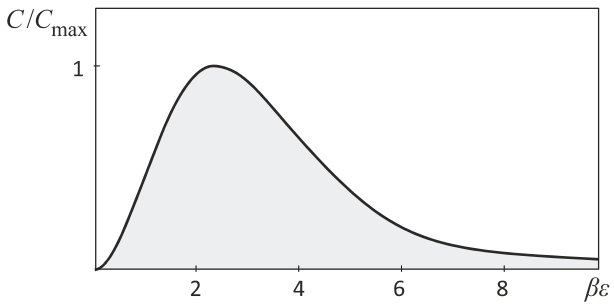


Fig. 1.13: Heat capacity of a two-level system.

Further information regarding heat capacities, rotational and vibrational spectra, and the relationship between thermodynamics, statistical physics, and chemistry can be found in textbook [9]. Interesting details about the spectroscopy of rotational and vibrational states are also provided in publication [10].

1.6 Grandcanonical Ensemble

1.6.1 Derivation of the Distribution

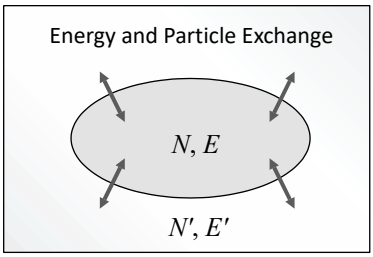


Fig. 1.14: Gibbs grandcanonical ensemble.

In contrast to the canonical distribution, particle exchange with the surroundings is permitted in the grandcanonical distribution. Other assumptions remain identical. The system can exchange energy, surface phenomena are neglected, and volume is the only external parameter for now. With the Liouville theorem assumed, the first thermodynamic law is written in the standard form for a single particle species:

$$\mathrm{d}U = T\,\mathrm{d}S - p\,\mathrm{d}V + \mu\,\mathrm{d}\bar{N}. \tag{1.104}$$

Since N in statistics represents the instantaneous number of particles in the system, the mean number in the first thermodynamic law must be denoted with a bar. The probability distribution in the discrete and continuous cases is denoted as:

$$w_{nN} = w_{nN}(E); \qquad \rho_N = \rho_N(E). \tag{1.105}$$

Subscript n numbers quantum states, and subscript N represents the particle number. Additivity in energy and particle number of the system and its surroundings is required:

$$E_{\mathrm{tot}} = E + E'; \qquad N_{\mathrm{tot}} = N + N'. \tag{1.106}$$

Similarly to the canonical case, independence (multiplicativity) of the probability distributions of the system and the surroundings is required:

$$w_{n\,N+N'}(E+E') = w_{nN}(E)\,w_{nN'}(E'). \tag{1.107}$$

The only solution is an exponential function of the type:

$$w_{nN}(E_{nN}) = \mathrm{e}^{c_1 + c_2 E_{nN} + c_3 N} = \mathrm{e}^{\alpha - \beta E_{nN} + \gamma N}. \tag{1.108}$$

For the continuous case, the probability density takes the form:

$$\rho_N(E_N) = \mathrm{e}^{\alpha - \beta E_N + \gamma N}. \tag{1.109}$$

The constants are denoted as α, $-\beta$, and γ, and are determined in the subsequent section.

1.6.2 Distribution Constants

The constants are determined similarly to the canonical distribution by comparing the differential of internal energy with the thermodynamic relation. The procedure is somewhat laborious; readers not interested in the derivation may skip directly to the result at the end of this section. In principle, three equations must be utilized to determine the three constants: probability normalization, energy averaging, and particle number averaging (the last equation was unnecessary for the canonical distribution). The problem can be solved both discretely and continuously:

$$\begin{aligned} &\sum_{n,N} w_{nN} = 1, && \sum_{N} \int \rho_N(E_N)\,\mathrm{d}\Gamma_N = 1, \\ &\sum_{n,N} E_{nN} w_{nN} = U, && \sum_{N} \int E\,\rho_N(E_N)\,\mathrm{d}\Gamma_N = U, \\ &\sum_{n,N} N\, w_{nN} = \bar{N}; && \sum_{N} N \int \rho_N(E_N)\,\mathrm{d}\Gamma_N = \bar{N}. \end{aligned} \tag{1.110}$$

Even in the continuous case, summation over all possible particle numbers must be performed. The derivation is conducted below for the discrete case. Similarly to the canonical distribution, the differential of internal energy is found first:

$$\mathrm{d}U = \mathrm{d}\sum_{n,N} E_{nN}(V)\, w_{nN} = \sum_{n,N}\left[\left(\frac{\partial E_{nN}}{\partial V}\,\mathrm{d}V + \frac{\partial E_{nN}}{\partial N}\mathrm{d}N\right) w_{nN}\right] + \sum_{n,N} E_{nN}\,\mathrm{d}w_{nN}\,.$$

The derivative of energy with respect to volume is again interpreted as partial pressure, the change in energy with the particle number represents the partial chemical potential, and E_{nN} in the last term is expressed from distribution (1.108):

$$\mathrm{d}U = \sum_{n,N}\left(-p_{nN} w_{nN}\mathrm{d}V\right) + \sum_{n,N}\left(\mu_{nN}\, w_{nN}\mathrm{d}N\right) + \sum_{n,N}\left(\frac{\alpha}{\beta} + \frac{\gamma}{\beta}N - \frac{1}{\beta}\ln w_{nN}\right)\mathrm{d}w_{nN}\,.$$

The first term is interpreted as mechanical work, identical to the canonical distribution. The last term is expanded, and the constants are factored out:

$$\mathrm{d}U = -p\,\mathrm{d}V + \sum_{n,N}\mu_{nN}\, w_{nN}\mathrm{d}N + \frac{\gamma}{\beta}\sum_{n,N} N\,\mathrm{d}w_{nN} + \frac{\alpha}{\beta}\mathrm{d}\sum_{n,N} w_{nN} - \frac{1}{\beta}\sum_{n,N}\ln w_{nN}\,\mathrm{d}w_{nN}.$$

The third and fifth terms are modified according to the standard product derivative relation $f\,\mathrm{d}g = \mathrm{d}(fg) - g\mathrm{d}f$; the fourth term is zero since the sum of all probabilities equals one and the differential of unity is zero:

$$\mathrm{d}U = -p\mathrm{d}V + \sum_{n,N}\mu_{nN}\, w_{nN}\mathrm{d}N + \frac{\gamma}{\beta}\mathrm{d}\sum_{n,N} N\, w_{nN} - \frac{\gamma}{\beta}\sum_{n,N} w_{nN}\,\mathrm{d}N - \frac{1}{\beta}\mathrm{d}\sum_{n,N}(\ln w_{nN})\, w_{nN}.$$

As one of the final steps, the second and fourth terms are combined:

$$\mathrm{d}U = -p\,\mathrm{d}V \;+\; \sum_{n,N}\left(\mu_{nN} - \frac{\gamma}{\beta}\right) w_{nN}\mathrm{d}N \;+ \frac{\gamma}{\beta}\,\mathrm{d}\bar{N} \;-\; \frac{1}{\beta}\mathrm{d}\sum_{n,N} w_{nN}\ln w_{nN}.$$

If this expression is to correspond to the first thermodynamic law (1.104), the following relations must hold:

$$\beta = \frac{1}{k_\mathrm{B}T}, \qquad \gamma = \frac{\mu}{k_\mathrm{B}T}, \qquad S = -k_\mathrm{B}\sum_{n,N} w_{nN}\ln w_{nN}\,. \tag{1.111}$$

The second condition ($\gamma/\beta = \mu$) ensures the correspondence of the third term with the matching term of the first thermodynamic law, canceling the second term. Other conditions are identical to the canonical distribution. Two constants are now determined, and a single undetermined distribution constant remains – the normalization constant α. This constant is determined, for instance, from the relation for entropy:

$$S = -k_\mathrm{B}\sum_{n,N} w_{nN}\ln w_{nN} = -k_\mathrm{B}\sum_{n,N} w_{nN}\left(\alpha - \beta E_{nN} + \gamma N\right) = -k_\mathrm{B}\alpha + k_\mathrm{B}\beta U - k_\mathrm{B}\gamma\bar{N}\,.$$

Following trivial modifications, the sought constant is calculated:

$$\alpha = \frac{U - TS - \mu\bar{N}}{k_\mathrm{B}T} = \frac{\Omega}{k_\mathrm{B}T}\,. \tag{1.112}$$

Since all constants are now known, the resulting distribution can be written for both the discrete and continuous cases:

$$w_{nN} = \mathrm{e}^{\beta(\Omega - E_{nN} + \mu N)}\,; \qquad \rho_N = \mathrm{e}^{\beta(\Omega - E_N + \mu N)}\,. \tag{1.113}$$

The grandcanonical potential is clearly related to the probability normalization; if this constant is explicitly written, the distribution assumes a frequently used form:

$$w_{nN} = K\exp\left[\frac{-E_{nN} + \mu N}{k_\mathrm{B}T}\right]; \qquad \rho_N = K\exp\left[\frac{-E_N + \mu N}{k_\mathrm{B}T}\right]. \tag{1.114}$$

1.6.3 Partition Function

The grandcanonical potential has a similar significance for systems with a variable number of particles as free energy has for the canonical distribution. The grandcanonical potential is calculated from the distribution normalization condition. The calculation is performed in parallel for the discrete (left) and continuous (right) cases:

$$\sum_{n,N} w_{nN} = 1 \;\Rightarrow$$
$$\sum_{n,N} \mathrm{e}^{\beta(\Omega - E_{nN} + \mu N)} = 1 \;\Rightarrow$$
$$\sum_{n,N} \mathrm{e}^{-\beta E_{nN} + \beta\mu N} = \mathrm{e}^{-\beta\Omega} \;\Rightarrow$$

$$\sum_{N}\int \rho_N\,\mathrm{d}\Gamma_N = 1 \;\Rightarrow$$
$$\sum_{N}\int \mathrm{e}^{\beta(\Omega - E_N + \mu N)}\,\mathrm{d}\Gamma_N = 1 \;\Rightarrow$$
$$\sum_{N}\int \mathrm{e}^{-\beta E_N + \beta\mu N}\,\mathrm{d}\Gamma_N = \mathrm{e}^{-\beta\Omega} \;\Rightarrow$$

The grandcanonical potential Ω is now easily calculated:

$$\Omega = -k_\mathrm{B}T \begin{cases} \ln\left(\sum\limits_{n,N} \mathrm{e}^{-\beta E_{nN} + \beta\mu N}\right); & \text{Discrete case,} \\ \ln\left(\sum\limits_{N} \int \mathrm{e}^{-\beta E_N + \beta\mu N}\, \mathrm{d}\Gamma_N\right); & \text{Continuous case.} \end{cases} \tag{1.115}$$

The quantity inside the logarithm parentheses is called the *grandcanonical partition function* and represents the central quantity of statistical physics with a variable number of particles, denoted as Ξ (capital Xi). Since the logarithm argument must be dimensionless, using the weight factor instead of the phase volume is preferable.

Statistical Calculation with Variable Particle Number

1) The energies E_{nN} that the system can occupy are determined first. In the classical case, these include all energy values that can occur in the system. In the quantum case, the spectrum of the Hamiltonian operator must be determined (for instance, by solving a Schrödinger equation).
2) The partition function Ξ is found as a sum of all terms $\exp[-\beta E + \beta\mu N]$ over the entire domain of the energy spectrum:

$$\Xi = \sum_{n,N} \mathrm{e}^{-\beta E_{nN} + \beta\mu N}; \quad \text{resp.} \quad \Xi = \sum_{N} \int \mathrm{e}^{-\beta E_N + \beta\mu N}\, \mathrm{d}\Gamma_N. \tag{1.116}$$

3) The grandcanonical potential is determined as:

$$\Omega = -k_\mathrm{B}T \ln \Xi\,. \tag{1.117}$$

4) he fundamental thermodynamic quantities S, p, and N are determined via:

$$S = -\left(\frac{\partial\Omega}{\partial T}\right), \qquad p = -\left(\frac{\partial\Omega}{\partial V}\right), \qquad \bar{N} = -\left(\frac{\partial\Omega}{\partial\mu}\right). \tag{1.118}$$

5) he internal energy U and its derivatives are determined by:

$$U = \Omega + TS + \mu\bar{N}\,. \tag{1.119}$$

6) Since the grandcanonical potential is a function of the chemical potential μ, it must be eliminated from all derived thermodynamic relations. Theoretically, this can be performed independently of (1.118) from the relation:

$$\bar{N} = \sum_{n,N} N\, w_{nN} \quad \Rightarrow \quad \bar{N} = \bar{N}(T,V,\mu) \quad \Rightarrow \quad \mu = \mu(\bar{N},T,V)\,. \tag{1.120}$$

Practically, eliminating the chemical potential from the equations can cause difficulties. However, broader thermodynamic framework, the chemical potential can also be understood as a parameter in a parametric specification of curves. For instance, if the dependencies $p = p(T, V, \mu)$ and $\bar{N} = \bar{N}(T, V, \mu)$ are known from (1.118), coordinates $(p, \bar{N})$ can be plotted in a graph with axes $(p, \bar{N})$ for a specific interval of μ values, thereby graphically constructing the dependence of pressure on the average particle number.

Grandcanonical and Canonical Partition Functions Relationship

The relationship for the grandcanonical partition function is modified below:

$$\Xi = \sum_{n,N} \mathrm{e}^{-\beta E_{nN} + \beta \mu N} = \sum_{N} \left[\left(\sum_{n} \mathrm{e}^{-\beta E_{nN}} \right) \cdot \left(\mathrm{e}^{\beta \mu} \right)^{N} \right].$$

Denoting the canonical partition function for N particles as Z_N and introducing the *fugacity* (denoted by the lowercase Greek letter zeta ζ):

$$\varsigma \equiv \mathrm{e}^{\beta \mu}, \tag{1.121}$$

a clear relation is obtained:

$$\Xi = \sum_{N} Z_N \varsigma^N. \tag{1.122}$$

This relation holds in classical physics, where individual particles in a system of N particles are distinguishable in principle. In quantum theory, particles are indistinguishable, and each of the $N!$ possible particle permutations represents the identical state. In relation (1.122), each permutation is counted in the sum, meaning that a single state is counted multiple times ($N!$ times) instead of once. Therefore, the correct relation in quantum theory is the expression:

$$\Xi = \sum_{N} \frac{Z_N}{N!} \varsigma^N, \tag{1.123}$$

which strongly resembles a Taylor expansion in fugacity, where the coefficients are the canonical partition functions for N particles.

1.7 Fermions and Bosons

Indistinguishable Particles

In quantum theory, only the probability of a particle occurring at a certain position and time can be predicted. This probability possesses a maximum at the position of the classical trajectory, decreasing exponentially with distance; far from the trajectory, the probability is very small but non-zero. If two identical particles are present, determining which particle is which is impossible. The probability of one particle occurring at the position of the other is non-zero. Consequently, identical particles are considered *indistinguishable* in quantum theory. This characteristic leads to the classification of all particles into two fundamental types, *fermions* and *bosons*, which differ in their properties and behavior (see Section I.2.8). Fermions possess a half-integer spin, and two fermions cannot occupy the identical quantum state (satisfying the Pauli exclusion principle); their wave function is antisymmetric, and their creation and annihilation operators satisfy anticommutation relations. Conversely, bosons possess an integer spin, any quantity of bosons can occupy the identical quantum state, their wave function is symmetric, and their creation and annihilation operators satisfy commutation relations. All field particles mediating interactions (photons, gluons, W^+, W^-, Z^0) are bosons with spin 1. Bosons also include particles composed of two quarks (mesons), which possess a spin equal to 0 (scalar mesons) or 1 (vector mesons). Conversely, all fundamental building blocks of matter (quarks and leptons) are fermions with spin ½. Fermions also include particles composed of three quarks (baryons), such as the neutron and proton. From a statistical perspective, particularly at low temperatures, fermions and bosons exhibit completely different behavior, and their probability distributions differ. Fermions are subject to the Fermi-Dirac distribution, and bosons follow the Bose-Einstein distribution.

Representation of Occupation Numbers

The number of particles in a given energy state is called the occupation number:

Energy State	Occupation Number
ε_1	N_1
ε_2	N_2
ε_3	N_3
⋮	⋮

For fermions, the occupation number can assume only the values 0 and 1 by the Pauli exclusion principle. For bosons, it can be any non-negative integer 0, 1, 2, ... In this representation, the particle number and total energy are written as:

$$N = \sum_i N_i \,; \qquad E_N = \sum_i \varepsilon_i N_i \,. \tag{1.124}$$

The grandcanonical partition function can be modified in the representation of occupation numbers as follows:

$$\Xi = \sum_{N,n}{}' \, \mathrm{e}^{-\beta E_{nN} + \beta\mu N} = \sum_{N_1, N_2, \ldots} \exp\left[-\beta \sum_i (\varepsilon_i N_i - \mu N_i) \right] =$$

$$= \sum_{N_1, N_2, \ldots} \prod_i \exp\left[-\beta (\varepsilon_i N_i - \mu N_i) \right] = \prod_i \sum_{N_i} \exp\left[-\beta N_i (\varepsilon_i - \mu) \right].$$

The prime symbol on the summation indicates that all particle permutations are considered a single state in quantum theory, and these permutations contribute a single term to the total sum. Evidently, the total grandcanonical sum decomposes into a product of partial grandcanonical sums of individual states:

$$\Xi = \prod_i \Xi_i \,; \qquad \Xi_i = \sum_{N_i} \exp\left[-\beta N_i (\varepsilon_i - \mu) \right]. \tag{1.125}$$

Consequently, thermodynamic quantities are given by sums of individual contributions:

$$\begin{aligned}
&\Omega = \sum_i \Omega_i \,; && \Omega_i = -k_\mathrm{B} T \ln \Xi_i \,, \\
&S = \sum_i S_i \,; && S_i = -\left(\frac{\partial \Omega_i}{\partial T} \right), \\
&p = \sum_i p_i \,; && p_i = -\left(\frac{\partial \Omega_i}{\partial V} \right), \\
&\bar{N} = \sum_i \bar{N}_i \,; && \bar{N}_i = -\left(\frac{\partial \Omega_i}{\partial \mu} \right).
\end{aligned} \tag{1.126}$$

1.7.1 Fermi-Dirac and Bose-Einstein Distributions

Fermi-Dirac Distribution

Fermions are examined first. According to the Pauli exclusion principle, two fermions cannot occupy the identical quantum state. Therefore, either no fermion is present in a given state, or a single one is present:

$$N_i = 0, 1. \tag{1.127}$$

The partition function of the *i*-th state in the representation of occupation numbers contains only two terms:

$$\Xi_i = \sum_{N_i} \exp\left[-\beta N_i (\varepsilon_i - \mu) \right] \; = \; 1 + \; \exp\left[-\beta (\varepsilon_i - \mu) \right].$$

The grandcanonical potential of the *i*-th state is easily calculated as:

$$\Omega_i = -k_\mathrm{B} T \ln \Xi_i = -k_\mathrm{B} T \ln\left(1 + \; \exp\left[-\beta (\varepsilon_i - \mu) \right] \right)$$

and the mean particle number in the *i*-th state is obtained as:

▶
$$\bar{N}_i = -\frac{\partial \Omega_i}{\partial \mu} = \frac{1}{\exp\left[\beta(\varepsilon_i - \mu)\right] + 1}. \tag{1.128}$$

This expression is called the Fermi-Dirac distribution. Its profile in the low-temperature limit ($T \to 0, \beta \to \infty$) is found below:

$$\lim_{\beta\to\infty} \bar{N}_i = \lim_{\beta\to\infty} \frac{1}{\exp\left[\beta(\varepsilon_i - \mu)\right] + 1} = \begin{cases} 1 & \text{pro } \varepsilon_i < \mu, \\ 0 & \text{pro } \varepsilon_i > \mu. \end{cases}$$

All states are filled by one particle each up to the Fermi limit $\varepsilon_F = \mu$. Above the Fermi limit, the states are unoccupied. The Fermi limit is thus the last occupied energy level at zero temperature. At absolute zero, the chemical potential equals the Fermi limit. Fermions exhibit exclusive behavior. If a state is occupied by a particle, another particle cannot occupy this state. At absolute zero, particles tend to occupy the state with the lowest possible energy. If that state is already occupied, the nearest free state is taken. Consequently, at absolute zero, all states up to the Fermi limit are occupied. The distribution is named after the Italian physicist Enrico Fermi and the English physicist Paul Adrien Maurice Dirac.

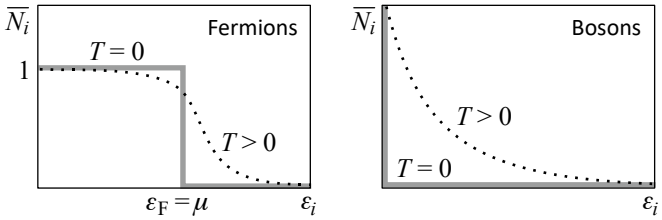


Fig. 1.15: Distribution of fermions (left) and bosons (right).

Bose-Einstein Distribution

The distribution of a boson ensemble is analyzed below. Bosons do not satisfy the Pauli exclusion principle, and any quantity of bosons can occupy a given state. The occupation numbers are therefore given by:

$$N_i = 0, 1, 2, \ldots \tag{1.129}$$

The grandcanonical partition function of the *i*-th state bocomes an infinite series:

$$\Xi_i = \sum_{N_i=0}^{\infty} \exp\left[-\beta N_i (\varepsilon_i - \mu)\right] =$$

$$= \sum_{N_i=0}^{\infty} \left(\exp\left[-\beta(\varepsilon_i - \mu)\right]\right)^{N_i}.$$

This expression constitutes a geometric series that is easily summed:

$$\Xi_i = \frac{1}{1 - \exp\left[-\beta(\varepsilon_i - \mu)\right]}. \tag{1.130}$$

The grandcanonical potential of the *i*-th state is easily calculated as:

$$\Omega_i = -k_B T \ln \Xi_i =$$
$$= +k_B T \ln\left(1 - \exp\left[-\beta(\varepsilon_i - \mu)\right]\right)$$

and the mean particle number in the i-th state is obtained as:

► $$\bar{N}_i = -\frac{\partial \Omega_i}{\partial \mu} = \frac{1}{\exp\left[\beta(\varepsilon_i - \mu)\right] - 1}. \tag{1.131}$$

This expression is called the Bose-Einstein distribution, named after the Indian physicist Satyendra Bose and the German-American physicist Albert Einstein. Notably, this distribution differs from the Fermi-Dirac distribution only by the sign in the denominator. This sign variation is typical for relations describing fermions and bosons (symmetric and antisymmetric wave functions, commutator and anticommutator). From the convergence condition of the geometric series, the chemical potential of a boson ensemble must satisfy the condition:

$$\mu \leq \varepsilon_0 \ . \tag{1.132}$$

The distribution profile in the low-temperature limit ($T \to 0, \beta \to \infty$) is found below:

$$\lim_{\beta\to\infty} \bar{N}_i = \lim_{\beta\to\infty} \frac{1}{\exp\left[\beta(\varepsilon_i - \mu)\right] - 1} = \begin{cases} \infty \ \text{pro} \ \ \varepsilon_i = \varepsilon_0 = \mu \ , \\ 0 \ \text{pro} \ \ \varepsilon_i > \varepsilon_0 \ . \end{cases}$$

In summing the geometric series, an infinite number of particles was implicitly assumed. At absolute zero, all particles occupy the ground energy state. In real systems, the particle number is finite. The Bose-Einstein distribution at both zero and non-zero temperatures is illustrated in Fig. 1.15 on the right. A state of matter in which particles accumulate in the ground state is called a *Bose-Einstein condensate* (BEC). A typical example is represented by Cooper pairs of electrons, which exhibit superconducting and superfluid properties as a Bose-Einstein condensate at low temperatures.

Both distributions can be written collectively as:

► $$\bar{N}_i = \frac{1}{\exp\left[\beta(\varepsilon_i - \mu)\right] \pm 1}. \tag{1.133}$$

The "+" sign applies to fermions, and the "–" sign applies to bosons. The conditions under which both distributions merge are examined below. Evidently, this merging occurs when the unity in the denominator can be neglected, i.e., the exponential term dominates and $\bar{N}_i << 1$. Then:

$$\bar{N}_i \ll 1 \qquad \Rightarrow \qquad \bar{N}_i \sim \frac{1}{\exp\left[\beta(\varepsilon_i - \mu)\right]} \sim K \exp\left[-\beta\varepsilon_i\right]$$

and both distributions transition into the Boltzmann distribution, where quantum states are mostly empty. This situation occurs under the following conditions:

- A high number of quantum states
- A dilute gas (small number of particles)
- High temperatures (particles excited to high energy states)

The properties of fermions and bosons are summarized in a table below:

	Fermions	Bosons
Spin	1/2, 3/2, ...	0, 1, 2, ...
Particle Examples	electrons, neutrinos, quarks	all field particles
Hamiltonian	$\hat{H}(1,2) = \hat{H}(2,1)$	$\hat{H}(1,2) = \hat{H}(2,1)$
Wave Function	$\lvert 1,2 \rangle = -\lvert 2,1 \rangle$	$\lvert 1,2 \rangle = +\lvert 2,1 \rangle$
Creation and Annihilation Operators	$\left[a_k^+ , a_l \right]_+ = \delta_{kl}$	$\left[a_k^+ , a_l \right]_- = \delta_{kl}$
Statistics ($\bar{N}_i$)	$\dfrac{1}{\exp\left[\beta(\varepsilon_i - \mu)\right] + 1}$	$\dfrac{1}{\exp\left[\beta(\varepsilon_i - \mu)\right] - 1}$

1.7.2 White Dwarf and Neutron Star as Fermion Gas

In the final stages of stellar evolution, the gravitational collapse of a star into a black hole can be halted by the pressure of a degenerate electron gas (white dwarf) or a degenerate neutron gas (neutron star). Therefore, an ensemble of non-zero mass fermions is examined below. The fundamental characteristics of both particles are:

$$s = 1/2\,; \qquad \mu \neq 0\,; \qquad m_0 \neq 0\,. \tag{1.134}$$

A non-zero chemical potential requires a calculation according to relation (1.120); a non-zero rest mass implies a relationship between momentum and energy for the non-relativistic case:

$$\varepsilon = \frac{p^2}{2m}\,; \qquad p = \sqrt{2m\varepsilon}\,; \qquad \mathrm{d}\,p = \frac{1}{2}\sqrt{2m}\;\varepsilon^{-1/2}\,\mathrm{d}\varepsilon\,. \tag{1.135}$$

The weight factor element is determined below:

$$\mathrm{d}\Gamma_\varepsilon = g\,\frac{\mathrm{d}^3x\,\mathrm{d}^3p}{(2\pi\hbar)^3} \rightarrow \frac{4\pi g V}{(2\pi\hbar)^3}\,p^2\mathrm{d}p = \frac{4\pi g V}{(2\pi\hbar)^3}\cdot 2m\varepsilon\cdot\frac{1}{2}\sqrt{2m}\;\varepsilon^{-1/2}\,\mathrm{d}\varepsilon\,.$$

For the weight factor of non-relativistic fermions, the following relation holds:

► $$\begin{aligned} \mathrm{d}\Gamma_\varepsilon &= \gamma(\varepsilon)\,\mathrm{d}\varepsilon\,; \\ \gamma(\varepsilon) &= \alpha V\,\varepsilon^{1/2}\,; \\ \alpha &\equiv \frac{g\,m^{3/2}}{\sqrt{2}\pi^2\hbar^3}\,. \end{aligned} \tag{1.136}$$

In the next section, it is shown that the number of quantum states does not grow as rapidly as in the case of photons, but it still exhibits an increasing trend, and for higher energies, the density of energy states can be considered nearly continuous.

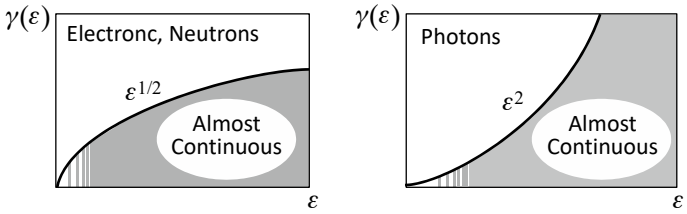


Fig. 1.16: Density of energy states for fermions (left). For comparison, the density of energy states for photons is shown on the right.

The grandcanonical partition function corresponding to a single energy state is:

$$\Xi_\varepsilon = \sum_{N_\varepsilon = 0}^{1} \mathrm{e}^{-\beta N_\varepsilon(\varepsilon-\mu)} = 1 + \mathrm{e}^{-\beta(\varepsilon-\mu)}; \tag{1.137}$$

The thermodynamic quantities of interest for this state yield:

$$\Omega_\varepsilon = -k_\mathrm{B}T \ln \Xi_\varepsilon = -k_\mathrm{B}T \ln\left[1 + \mathrm{e}^{-\beta(\varepsilon-\mu)}\right]; \tag{1.138}$$

$$\bar{N}_\varepsilon = -\frac{\partial \Omega_\varepsilon}{\partial \mu} = \frac{1}{\mathrm{e}^{\beta(\varepsilon-\mu)} + 1}; \tag{1.139}$$

$$U_\varepsilon = \varepsilon \bar{N}_\varepsilon = \frac{\varepsilon}{\mathrm{e}^{\beta(\varepsilon-\mu)} + 1}. \tag{1.140}$$

Integral quantities, representing sums over all states, are found below. The logarithmic dependence in the grandcanonical potential is eliminated via integration by parts. The remaining calculations are straightforward:

$$\Omega = \int \Omega_\varepsilon \,\mathrm{d}\Gamma_\varepsilon = \int_0^\infty -k_\mathrm{B}T \ln\left[1 + \mathrm{e}^{-\beta(\varepsilon-\mu)}\right] \alpha V \varepsilon^{1/2} \,\mathrm{d}\varepsilon =$$

$$-k_\mathrm{B}T\alpha V \int_0^\infty \varepsilon^{1/2} \ln\left[1 + \mathrm{e}^{-\beta(\varepsilon-\mu)}\right] \mathrm{d}\varepsilon =$$

$$-k_\mathrm{B}T\alpha V \left\{ \left[\frac{2}{3}\varepsilon^{3/2} \ln\left[1 + \mathrm{e}^{-\beta(\varepsilon-\mu)}\right]\right]_0^\infty - \int_0^\infty \frac{2}{3}\varepsilon^{3/2} (-\beta) \frac{\mathrm{e}^{-\beta(\varepsilon-\mu)}}{1 + \mathrm{e}^{-\beta(\varepsilon-\mu)}} \right\}.$$

The first expression from the integration by parts is zero – the upper limit is nullified by the logarithmic dependence, and the lower limit by the power dependence. The second term containing the integral could be solved numerically, but for the present purposes, it is sufficient to retain it in a general form:

$$\Omega = -\frac{2}{3} \alpha V \int_0^\infty \frac{\varepsilon^{3/2}}{\mathrm{e}^{\beta(\varepsilon-\mu)} + 1} \,\mathrm{d}\varepsilon. \tag{1.141}$$

Pressure, or the state equation, is easily determined from the grandcanonical potential:

$$p = -\frac{\partial \Omega}{\partial V} = \frac{2}{3}\alpha \int_0^\infty \frac{\varepsilon^{3/2}}{\mathrm{e}^{\beta(\varepsilon-\mu)}+1}\,\mathrm{d}\varepsilon. \tag{1.142}$$

Additional thermodynamic quantities are obtained again by integrating the quantities corresponding to a single state:

$$\bar{N} = \int \bar{N}_\varepsilon \,\mathrm{d}\Gamma_\varepsilon = \alpha V \int_0^\infty \frac{\varepsilon^{1/2}}{\mathrm{e}^{\beta(\varepsilon-\mu)}+1}\,\mathrm{d}\varepsilon\,; \tag{1.143}$$

$$U = \int U_\varepsilon \,\mathrm{d}\Gamma_\varepsilon = \alpha V \int_0^\infty \frac{\varepsilon^{3/2}}{\mathrm{e}^{\beta(\varepsilon-\mu)}+1}\,\mathrm{d}\varepsilon\,. \tag{1.144}$$

The integrals in the relations must be found numerically. Dividing the last two relations by the volume yields a set of three intensive quantities describing the fermion gas:

$$p = \frac{2}{3}\alpha \int_0^\infty \frac{\varepsilon^{3/2}}{\mathrm{e}^{\beta(\varepsilon-\mu)}+1}\,\mathrm{d}\varepsilon\,; \tag{1.145}$$

$$n \equiv \frac{\bar{N}}{V} = \alpha \int_0^\infty \frac{\varepsilon^{1/2}}{\mathrm{e}^{\beta(\varepsilon-\mu)}+1}\,\mathrm{d}\varepsilon\,; \tag{1.146}$$

$$u \equiv \frac{U}{V} = \alpha \int_0^\infty \frac{\varepsilon^{3/2}}{\mathrm{e}^{\beta(\varepsilon-\mu)}+1}\,\mathrm{d}\varepsilon\,. \tag{1.147}$$

Even without calculating the integrals, the relationship between pressure and internal energy density for non-zero mass fermions is evident:

► $$p = \frac{2}{3}u\,. \tag{1.148}$$

Stars in the white dwarf or neutron star phases are in the final stages of their evolution. Thermonuclear fusion does not occur in the core, and although the interior temperature is substantial from a human perspective, it is negligible relative to the active life of a star, essentially representing zero temperature. In the low-temperature limit ($\beta \to \infty$), the integrals can be easily calculated since:

$$\bar{N}_\varepsilon = \frac{1}{\mathrm{e}^{\beta(\varepsilon-\mu)}+1} \to \begin{cases} 1 & \text{pro } \varepsilon \le \mu_0; \\ 0 & \text{pro } \varepsilon > \mu_0, \end{cases}$$

where μ_0 is the chemical potential in the absolute zero limit. Integration over the step-wise distribution becomes trivial:

$$p = \frac{4}{15}\alpha \mu_0^{5/2}; \qquad n = \frac{2}{3}\alpha \mu_0^{3/2}; \qquad u = \frac{2}{5}\alpha \mu_0^{5/2}. \tag{1.149}$$

The chemical potential represents a parameter of the equations and can be easily eliminated from the relation for the particle density:

► $$\mu_0 = \text{const}\, n^{2/3} \quad \Rightarrow \quad p = \text{const}\, n^{5/3}\,; \quad u = \frac{3}{2}\, p\,. \tag{1.150}$$

The fermion gas exhibits polytropic behavior with a coefficient of 5/3. Even at zero temperature, a massive non-zero pressure caused by quantum processes (the exclusion behavior of fermions) exists. A gas with a mean thermal energy lower than the Fermi limit ($k_B T < \mu_0$) is called degenerate. In the case of a relativistic calculation, the polytropic coefficient yields 4/3, which is precisely at the boundary of stability and instability for a polytropic star (see the final volume *Selected Topics III*, Section III.1.5).

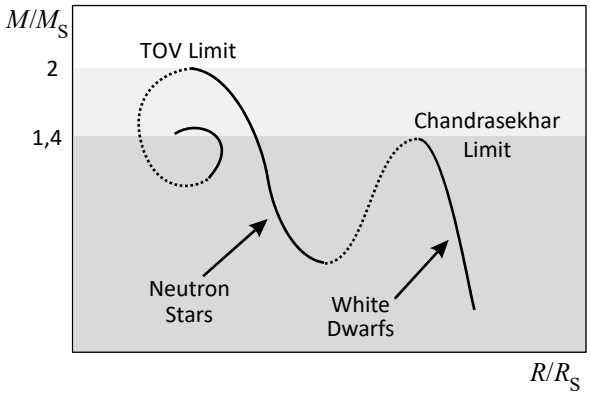


Fig. 1.17: Stable configurations (solid line) of a star without ongoing thermonuclear fusion.

1.7.3 Planck Radiation Law for Photons

Photons are particles of electromagnetic interaction propagating at the speed of light. Their characteristics (spin, chemical potential, rest mass, and electric charge) are:

$$s = 1\,; \qquad \mu = 0\,; \qquad m_0 = 0\,; \qquad Q_e = 0\ . \tag{1.151}$$

In special relativity, a simple relation holds for the energy of a particle:

$$E = \sqrt{m_0^2 c^4 + p^2 c^2}\ . \tag{1.152}$$

For small momenta, this relationship transitions into the non-relativistic expression (1.135). Currently, however, photons propagating at the speed of light are considered, whose rest mass is zero (only such particles can propagate at the speed of light). Substituting $m_0 = 0$ yields a simple relation for their energy:

$$\varepsilon = pc\,. \tag{1.153}$$

An ensemble of photons enclosed in a region of volume V at temperature T is analyzed below. To a first approximation, photons are treated as a continuous system, and the weight space element of a single photon is determined. The degeneracy factor $g = 2$ because electromagnetic radiation is transverse, with two independent polarization modes. All feasible integrations are performed automatically for the weight factor element, the momentum space element is transformed into spherical coordinates, and momentum is converted to energy according to relation (1.153):

$$\mathrm{d}\Gamma_\varepsilon = g\,\frac{\mathrm{d}\phi}{(2\pi\hbar)^3} = 2\,\frac{\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}p_x\,\mathrm{d}p_y\,\mathrm{d}p_z}{(2\pi\hbar)^3} \to \frac{2V}{(2\pi\hbar)^3}\,4\pi p^2\mathrm{d}p = \frac{V\varepsilon^2}{\pi^2\hbar^3c^3}\,\mathrm{d}\varepsilon\,.$$

The density of energy states is thus obtained, increasing quadratically with energy:

$$\mathrm{d}\Gamma_\varepsilon = \gamma(\varepsilon)\,\mathrm{d}\varepsilon\,; \qquad \gamma(\varepsilon) = \frac{V\varepsilon^2}{\pi^2\hbar^3c^3}\,. \tag{1.154}$$

The density of photon energy states increases substantially faster than in the case of fermion states and can be considered a continuous quantity. This behavior is clearly visible in Fig. 1.16. The fundamental statistical and thermodynamic quantities for a single state are summarized below:

$$\Xi_\varepsilon = \sum_{N_\varepsilon=0}^{\infty} \mathrm{e}^{-\beta N_\varepsilon \varepsilon} = \frac{1}{1-\mathrm{e}^{-\beta\varepsilon}}\,, \tag{1.155}$$

$$\Omega_\varepsilon = -k_\mathrm{B}T\ln\Xi_\varepsilon = k_\mathrm{B}T\ln\left[1-\mathrm{e}^{-\beta\varepsilon}\right], \tag{1.156}$$

$$\bar{N}_\varepsilon = -\lim_{\mu\to 0}\frac{\partial\Omega_\varepsilon}{\partial\mu} = \frac{1}{\mathrm{e}^{\beta\varepsilon}-1}\,. \tag{1.157}$$

$$U_\varepsilon = \varepsilon\bar{N}_\varepsilon = \frac{\varepsilon}{\mathrm{e}^{\beta\varepsilon}-1}\,. \tag{1.158}$$

All these relations were derived previously, see e.g. Fig. 1.15; setting $\mu = 0$ is sufficient. The calculation of the total thermodynamic quantities is performed below:

$$\Omega = \int_0^\infty \Omega_\varepsilon\,\mathrm{d}\Gamma_\varepsilon = \frac{V}{\pi^2\hbar^3c^3}\,k_\mathrm{B}T\int_0^\infty \varepsilon^2\ln\left[1-\mathrm{e}^{-\beta\varepsilon}\right]\mathrm{d}\varepsilon\,.$$

Integration by parts (similarly to the fermion gas) yields the following expression:

$$\Omega = -\frac{V}{\pi^2\hbar^3c^3}\,\frac{1}{3}\int_0^\infty \frac{\varepsilon^3}{\mathrm{e}^{\beta\varepsilon}-1}\,\mathrm{d}\varepsilon\,. \tag{1.159}$$

In this case, the integral can be solved analytically, see (I.3.591), but for the present purposes, it can be retained in a general form. The pressure, mean photon number, and internal energy are determined below:

$$p = -\frac{\partial\Omega}{\partial V} = \frac{1}{3}\,\frac{1}{\pi^2\hbar^3c^3}\int_0^\infty \frac{\varepsilon^3}{\mathrm{e}^{\beta\varepsilon}-1}\,\mathrm{d}\varepsilon\,, \tag{1.160}$$

$$\bar{N} = \int \bar{N}_\varepsilon\,\mathrm{d}\Gamma_\varepsilon = \frac{V}{\pi^2\hbar^3c^3}\int_0^\infty \frac{\varepsilon^2}{\mathrm{e}^{\beta\varepsilon}-1}\,\mathrm{d}\varepsilon\,, \tag{1.161}$$

$$U = \int_0^\infty \varepsilon\bar{N}_\varepsilon\,\mathrm{d}\Gamma_\varepsilon = \frac{V}{\pi^2\hbar^3c^3}\int_0^\infty \frac{\varepsilon^3}{\mathrm{e}^{\beta\varepsilon}-1}\,\mathrm{d}\varepsilon\,. \tag{1.162}$$

Introducing intensive quantities (last two terms divided by the volume) yields:

$$p = \frac{1}{3}\frac{1}{\pi^2\hbar^3c^3}\int_0^\infty \frac{\varepsilon^3}{\mathrm{e}^{\beta\varepsilon}-1}\,\mathrm{d}\varepsilon\,, \tag{1.163}$$

$$n = \frac{1}{\pi^2\hbar^3c^3}\int_0^\infty \frac{\varepsilon^2}{\mathrm{e}^{\beta\varepsilon}-1}\,\mathrm{d}\varepsilon\,, \tag{1.164}$$

$$u = \frac{1}{\pi^2\hbar^3c^3}\int_0^\infty \frac{\varepsilon^3}{\mathrm{e}^{\beta\varepsilon}-1}\,\mathrm{d}\varepsilon\,. \tag{1.165}$$

Notably, even without performing the integration, a comparison of equations (1.163) and (1.165) reveals the relationship between radiation pressure and energy density:

► $$p = \frac{1}{3}u\,. \tag{1.166}$$

The differential of the energy density is easily found from relation (1.165):

$$\mathrm{d}u = \frac{1}{\pi^2\hbar^3c^3}\frac{\varepsilon^3}{\mathrm{e}^{\beta\varepsilon}-1}\mathrm{d}\varepsilon\ . \tag{1.167}$$

The energy flux, or radiation intensity ($I = uc$), can now be written as a function of the state energy ε, angular frequency ($\varepsilon = \hbar\omega$), or wavelength ($\omega = 2\pi c/\lambda$):

$$\mathrm{d}I(\varepsilon) = \frac{1}{\pi^2\hbar^3c^2}\frac{\varepsilon^3}{\exp\left(\dfrac{\varepsilon}{k_\mathrm{B}T}\right)-1}\,\mathrm{d}\varepsilon\,; \tag{1.168}$$

$$\mathrm{d}I(\omega) = \frac{\hbar}{\pi^2c^2}\frac{\omega^3}{\exp\left(\dfrac{\hbar\omega}{k_\mathrm{B}T}\right)-1}\,\mathrm{d}\omega\,; \tag{1.169}$$

$$\mathrm{d}I(\lambda) = -16\pi^2\hbar c^2\frac{\lambda^{-5}}{\exp\left(\dfrac{2\pi\hbar c}{\lambda k_\mathrm{B}T}\right)-1}\,\mathrm{d}\lambda\,. \tag{1.170}$$

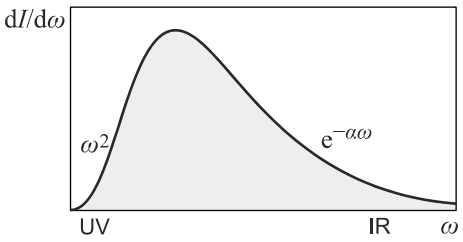


Fig. 1.18: Planck radiation law.

This expression represents the famous Planck radiation law derived in 1901. Most frequently, the quantity $\mathrm{d}I/\mathrm{d}\omega$ is utilized. The profile is illustrated in Fig. 1.18.

For low frequencies, $\mathrm{d}I/\mathrm{d}\omega \approx \omega^2$ (in the denominator, $\exp[x] - 1 \approx x$). For high frequencies, the exponential dominates, and $\mathrm{d}I/\mathrm{d}\omega \approx \exp[-\hbar\omega/k_\mathrm{B}T]$ holds. The relationship for low frequencies is called the Rayleigh-Jeans law, which was known prior to the discovery of the Planck law. This relationship diverges for high frequencies, a phenomenon termed the ultraviolet catastrophe.

Wien Displacement Law

The radiation maximum corresponding to the Planck law is examined below. This maximum depends on temperature and is determined by differentiating relation (1.169):

$$\frac{\mathrm{d}}{\mathrm{d}\omega}\left(\frac{\omega^3}{\exp\left(\hbar\omega/k_\mathrm{B}T\right)-1}\right) = 0 \qquad \Rightarrow \qquad \frac{\hbar\omega}{k_\mathrm{B}T}\exp\left(\frac{\hbar\omega}{k_\mathrm{B}T}\right) = 3\left[\exp\left(\frac{\hbar\omega}{k_\mathrm{B}T}\right)-1\right].$$

The resulting equation is a transcendental equation of the type:

$$x\,\mathrm{e}^x = 3\left(\mathrm{e}^x - 1\right); \qquad x \equiv \frac{\hbar\omega}{k_\mathrm{B}T}.$$

Both sides are divided by the exponential function, and the polynomial dependencies are transferred to the left:

$$3 - x = 3\,\mathrm{e}^{-x}.$$

In this form, the equation can be easily solved graphically by plotting the left and right sides to find the intersection of both curves:

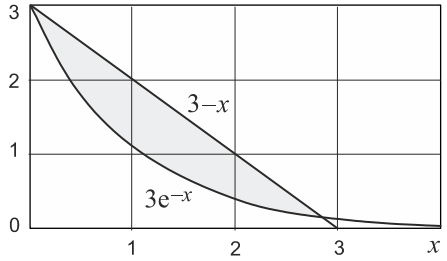


Fig. 1.19: Graphical solution of the equation $3 - x = 3\mathrm{e}^{-x}$.

Both curves intersect at the point $x_0 \simeq 2.822$; therefore:

$$\omega_\mathrm{max} = \frac{2.822\,k_\mathrm{B}T}{\hbar}.$$

Distribution (3.170) holds for the frequency. A similar procedure yields the maximum in terms of wavelength, known as the Wien displacement law:

► $$\lambda_\mathrm{max} = \frac{b}{T}; \qquad b = 0.00289\ \mathrm{K\cdot m}. \tag{1.171}$$

Higher temperature bodies radiate at shorter wavelengths. Examples include:

- Cosmic microwave background radiation ($T \sim 3$ K) has a maximum at about 1 mm.
- The human body ($T \sim 300$ K) has a maximum at about 10 μm.
- Cool stars ($T \sim 3{,}000$ K) radiate in the infrared region at about 1,000 nm.
- Stars such as the Sun ($T \sim 6{,}000$ K) radiate in the visible spectrum at about 500 nm.
- Highly hot stars ($T \sim 30{,}000$ K) radiate in the ultraviolet region at about 100 nm.

Stefan-Boltzmann Law

The total radiated energy per unit time from a unit area of a body can also be determined from the Planck radiation law.

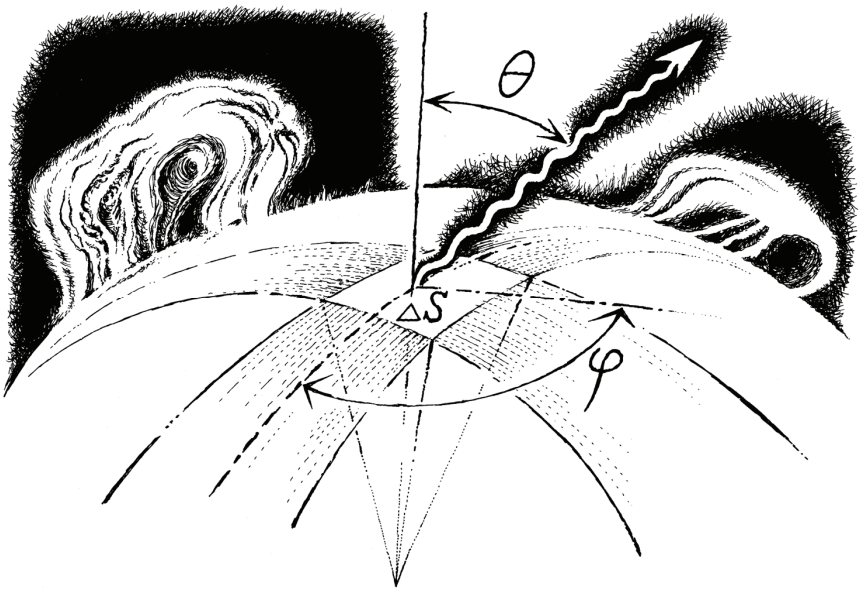


Fig. 1.20: Integration of the Planck law over angular dependencies.

The intensity radiated in a given direction (φ, θ) per frequency interval dω is equal to d$I(\omega)$; over the entire frequency range, the value is:

$$I_{\varphi,\theta} = \int_0^\infty \cos\theta \, \mathrm{d}I(\omega) \ . \tag{1.172}$$

This value is averaged over the entire solid angle Ω since rays from various surface elements reach the observer at different angles:

$$I = \left\langle I_{\varphi,\theta} \right\rangle_\Omega = \frac{1}{4\pi} \int I_{\varphi,\theta} \, \mathrm{d}\Omega \tag{1.173}$$

The total solid angle equals 4π. The solid angle element in spherical coordinates is:

$$\mathrm{d}\,\Omega = \frac{\mathrm{d}S}{r^2} = \frac{\mathrm{d}l_\varphi \, \mathrm{d}l_\theta}{r^2} = \frac{r\mathrm{d}\varphi \cdot r\sin\theta \, \mathrm{d}\theta}{r^2} = \sin\theta \, \mathrm{d}\varphi \, \mathrm{d}\theta \, ; \tag{1.174}$$

$$I = \frac{1}{4\pi}\int\left[\int_0^\infty \frac{\hbar}{\pi^2 c^2}\,\frac{\omega^3}{\exp\left(\hbar\omega/k_\mathrm{B}T\right)-1}\cos\theta\,\mathrm{d}\omega\right]\mathrm{d}\Omega\,.$$

The relation is integrated over the entire frequency spectrum and over the outer spatial hemisphere:

$$I = \frac{1}{4\pi}\int_0^\infty\left(\int_0^{\pi/2}\left(\int_0^{2\pi}\frac{\hbar}{\pi^2 c^2}\,\frac{\omega^3}{\exp\left(\hbar\omega/k_\mathrm{B}T\right)-1}\cos\theta\sin\theta\,\mathrm{d}\varphi\right)\mathrm{d}\theta\right)\mathrm{d}\omega\,.$$

Integration over the angles is trivial, yielding:

$$I = \frac{\hbar}{4\pi^2 c^2}\int_0^\infty \frac{\omega^3}{\exp\left(\hbar\omega/k_\mathrm{B}T\right)-1}\,\mathrm{d}\omega\,.$$

As a final step, the substitution $x = \hbar\omega/k_\mathrm{B}T$ is performed in the integral below:

$$I = \frac{k_\mathrm{B}^4}{4\hbar^3\pi^2 c^2}T^4\int_0^\infty\frac{x^3}{\exp\left(x\right)-1}\,\mathrm{d}x = \frac{k_\mathrm{B}^4}{4\hbar^3\pi^2 c^2}T^4\cdot\frac{\pi^4}{15} = \frac{\pi^2 k_\mathrm{B}^4}{60\hbar^3 c^2}T^4\,.$$

The calculation of the integral can be found in *Selected Topics I*, Section 3.10.4. The result is the well-known Stefan-Boltzmann law:

► $$I = \sigma T^4\,;\qquad \sigma = \frac{\pi^2 k_\mathrm{B}^4}{60\hbar^3 c^2} = 5.8\times10^{-8}\ \mathrm{W\,m^{-2}\,K^{-4}}\,. \tag{1.175}$$

Note: In the derivation of the Planck law, the weight factor was derived continuously. However, photons could also be modeled as standing waves inside a rectangular box, with wave vectors along the coordinate axes and nodes at the boundaries. The result would be identical to the continuous derivation.

◖ Example 1.8: Temperature of Sun

Determine the surface temperature of the Sun, given that the radiation maximum is at a wavelength of 500 nm.

Solution: According to the Wien law, the surface temperature is equal to:

$$T = \frac{b}{\lambda_\mathrm{max}} \sim 5{,}800\ \mathrm{K}\,.$$

Hot stars generally radiate at shorter wavelengths. Typical blue stars possess a surface temperature exceeding 9,000 K, yellow and green stars around 6,000 K, and red stars only about 3,000 K. The Wien law can also be applied to substantially cooler bodies. For instance, a human with a surface temperature of approximately 310 K radiates nearly as a blackbody with a radiation maximum at a wavelength of 10 µm. Therefore, sensors for person detection must exhibit maximum sensitivity in this spectral region.

◗

◖ Example 1.9: Power Output of Sun

Find the total radiant power output of the Sun, given its surface temperature $T = 5{,}800\,\mathrm{K}$.

Solution: The radiant power output of the Sun is determined from the Stefan-Boltzmann law:

$$P_\mathrm{S} = IS = \sigma T^4 4\pi R_\mathrm{S}^2 = 4\times 10^{26}\ \mathrm{W}\,. \tag{1.176}$$

The enormous value of the solar radiant power output is due to its large mass. On average, one kilogram of solar matter produces a very small power output. With a solar mass of 2×10^{30} kg, this value corresponds to only 0.0002 W/kg. ◗

◖ Example 1.10: Solar Constant

Determine the solar radiation intensity in the vicinity of Earth.

Solution: The solar constant represents the intensity of solar radiation (energy perpendicularly incident on a unit area per unit time) above the atmosphere of Earth. This quantity can be calculated as the ratio of the total power output of the Sun to the total surface area of a sphere passing through Earth with its center in the Sun:

$$I_\mathrm{Z} = \frac{P_\mathrm{S}}{4\pi R_\mathrm{ZS}^2} = 1.4\ \mathrm{kW\,m^{-2}}\,. \tag{1.177}$$

Near Earth, a power of 1.4 kW is incident on each square meter of a surface positioned perpendicularly to the solar radiation. This immense power is directly utilized in solar panels of space probes and in solar power plants. At the surface of Earth, this power is reduced by atmospheric scattering. Apart from nuclear energy, all commonly available energy on Earth originates from solar energy. The incident solar radiation power is, for instance, partially absorbed by plants and stored via photosynthesis into chemical bond energy. After many years, this energy is recovered during the combustion of coal, oil, or gasoline. The incident radiation also causes water evaporation from the surface of Earth, enabling the water cycle. Therefore, the energy utilized in hydroelectric power plants also has its primal origin in solar energy. ◗

◖ Example 1.11: Temperature of Planet

Determine the equilibrium temperature of a planet at distance d from a host star whose temperature is T_* and radius is R_*. The planet possesses a surface reflectivity (albedo) A.

Solution: First, the radiant power output of the star is determined, which is given by the Stefan-Boltzmann law:

$$P_* = I_* S = \sigma T_*^4\ 4\pi R_*^2\,.$$

The stellar radiation intensity at the orbital radius of the planet, i.e., at distance d, is calculated below (the star radiates into the entire space, meaning the surface of a sphere with radius d is considered):

$$I_\mathrm{P} = \frac{P_*}{4\pi d^2} = \frac{\sigma T_*^4\ 4\pi R_*^2}{4\pi d^2} = \frac{\sigma T_*^4 R_*^2}{d^2}\,.$$

The planet absorbs this radiation approximately across the projection of its surface facing the star; from the total incident radiation, fraction A is reflected and the fraction $(1–A)$ is absorbed. The power absorbed by the planet is therefore given by:

$$P_\mathrm{P} = (1-A) I_\mathrm{P}\ \pi R_\mathrm{P}^2 = (1-A)\frac{\sigma T_*^4 R_*^2}{d^2}\pi R_\mathrm{P}^2 .$$

The total energy balance of the planet is formulated below, assuming no other energy source exists. The absorbed power is radiated across the entire surface of the planet according to the Planck or Stefan-Boltzmann law:

$$P_\mathrm{P} = \sigma T_\mathrm{P}^4\ 4\pi R_\mathrm{P}^2 ,$$

$$(1-A)\frac{\sigma T_*^4 R_*^2}{d^2}\pi R_\mathrm{P}^2 = \sigma T_\mathrm{P}^4\ 4\pi R_\mathrm{P}^2 ,$$

$$T_\mathrm{P} = \sqrt[4]{\frac{(1-A)}{4}}\ \sqrt{\frac{R_*}{d}}\ T_* . \tag{1.178}$$

Earth has an albedo of around 0.3 (absorbing 70 % of radiation), and the equilibrium temperature yields approximately −18 °C. The greenhouse effect shifts this value to a pleasant +15 °C. ◗

◖ Example 1.12: Radiation as Dense as Water

Determine during which phase of cosmic expansion (at what temperature) radiation possessed a density identical to that of water.

Solution: A simple relation, originating from the Einstein mass-energy formula, holds between matter density and energy density:

$$\rho_\mathrm{E} = \rho_\mathrm{M}\, c^2 . \tag{1.179}$$

The matter density ρ_M corresponds to the density of water. The radiation energy density is determined from the energy flux, which is given by the Stefan-Boltzmann law:

$$\rho_\mathrm{E} = \frac{I}{c} = \frac{\sigma T^4}{c} . \tag{1.180}$$

By comparing both relations, the temperature of the universe at which electromagnetic radiation possessed a density identical to water is determined:

$$T = \sqrt[4]{\frac{\rho_\mathrm{M}\, c^3}{\sigma}} = 8\times 10^8\ \mathrm{K} .$$

The universe reached this temperature approximately 4 minutes after the Big Bang, precisely when the formation of the first light elements was initiated. ◗

◖ Example 1.13: Photon Energy During Expansion

Determine the dependence of photon energy density on the cosmic scale factor.

Solution: Between pressure and energy density, non-relativistic fermions satisfy the relation, $p = ⅔\, u$, a photon gas yields $p = ⅓\, u$, and classical mechanics gives $p = u$. Therefore, a general relation is sometimes introduced as:

$$p = wu\,, \tag{1.181}$$

where w is a parameter characteristic of a given entity. The decrease in radiation energy density ($w = 1/3$) with increasing dimensions of a thermally isolated system (for instance, the entire universe, which has no surroundings) is easily determined. For $\mathrm{d}Q = 0$, only the corresponding term from the first thermodynamic law remains:

$$\mathrm{d}U = -p\,\mathrm{d}V,$$

$$\mathrm{d}(uV) = -\frac{1}{3}u\,\mathrm{d}V.$$

The volume is assumed to be proportional to the cube of the dimensions (the scale factor for the universe), i.e.:

$$\mathrm{d}(uR^3) = -\frac{1}{3}u\,\mathrm{d}R^3,$$

$$R^3\,\mathrm{d}u + 3R^2u\,\mathrm{d}R + u\,R^2\,\mathrm{d}R = 0\,,$$

$$\frac{\mathrm{d}u}{u} + 4\frac{\mathrm{d}R}{R} = 0\,,$$

$$\ln uR^4 = K\,,$$

$$u \sim \frac{1}{R^4}\,. \tag{1.182}$$

The radiation energy density (or density of particles with zero rest mass) decreases as $u \sim 1/R^4$. This behavior is due to the fact that with increasing volume, the density decreases as $1/R^3$, but additionally, the photon wavelength elongates during expansion, "following" the cosmic expansion. This further reduces the photon energy (which is proportional to $1/\lambda$), and the resulting decrease is equal to $1/R^4$.

Further information regarding the physics of fermions and bosons can be found in publications [11], [12].

1.8 Fluctuations and Entropy

1.8.1 Fluctuations

In the previous sections, the mean value of a dynamical variable A averaged over an ensemble was introduced:

$$\langle A \rangle \equiv \sum_n A_n w_n \, .$$

The deviation of the quantity A from the mean value is defined below:

$$\Delta A \equiv A - \langle A \rangle \, . \tag{1.183}$$

The mean value of the deviations is naturally zero (approximately equal numbers of positive and negative deviations from the mean value occur):

$$\langle \Delta A \rangle = 0 \, . \tag{1.184}$$

To determine the average magnitude of deviations, either the absolute values of deviations or the squares of deviations must be averaged (the average of the squares must naturally be square-rooted at the end to ensure the result possesses the identical dimension as the original quantity):

$$\Delta A_{\text{abs}} \equiv \langle |\Delta A| \rangle ; \quad \Delta A_{\text{kv}} \equiv \sqrt{\langle (\Delta A)^2 \rangle} \, . \tag{1.185}$$

The second quantity is called the *root-mean-square fluctuation* of the quantity A. In statistics, the un-rooted quantity itself (the mean of the squared deviations, i.e., *variance* or the *second central moment* of the quantity A) also carries great significance:

$$\text{var}\, A \equiv \left\langle (\Delta A)^2 \right\rangle = \left\langle \left(A - \langle A \rangle \right)^2 \right\rangle = \left\langle A^2 - 2A\langle A \rangle + \langle A \rangle^2 \right\rangle$$

$$= \left\langle A^2 \right\rangle - 2\langle A \rangle^2 + \langle A \rangle^2 = \left\langle A^2 \right\rangle - \langle A \rangle^2 \, .$$

Thus, the following holds:

$$\text{var}\, A \equiv \langle (\Delta A)^2 \rangle = \langle A^2 \rangle - \langle A \rangle^2 \, . \tag{1.186}$$

Fluctuations of various physical quantities are closely related to important characteristics of the system. For instance, velocity magnitude fluctuations are related to temperature, energy fluctuations relate to heat capacity, particle number fluctuations relate to system compressibility ($\partial V/\partial p$), and magnetization fluctuations (dipole moment density) relate to susceptibility ($\partial M/\partial H$). These system characteristics are relatively straightforward to measure experimentally. Conversely, in computer simulations, tracking the fluctuations of simulated quantities is straightforward, allowing the determination of specific heats, compressibility, and system susceptibility. The variance of the velocity magnitude is determined below.

Velocity Fluctuations

Relations for the variances of the velocity magnitude and the square of velocity are derived below. The values of the root-mean-square velocity and the mean velocity from relation (1.79) are utilized, whereas higher powers of velocity are obtained by direct integration over the Maxwell distribution:

$$\operatorname{var} v = \langle v^2 \rangle - \langle v \rangle^2 = v_{\mathrm{kv}}^2 - v_{\mathrm{s}}^2 = \frac{3k_{\mathrm{B}}T}{m} - \frac{8k_{\mathrm{B}}T}{\pi m} = (3 - 8/\pi)\frac{k_{\mathrm{B}}T}{m}.$$

$$\operatorname{var} v^2 = \langle v^4 \rangle - \langle v^2 \rangle^2 = 15\left(\frac{k_{\mathrm{B}}T}{m}\right)^2 - 9\left(\frac{k_{\mathrm{B}}T}{m}\right)^2 = 6\left(\frac{k_{\mathrm{B}}T}{m}\right)^2.$$

Both variances are closely related to temperature:

► $$\operatorname{var} v = (3 - 8/\pi)\frac{k_{\mathrm{B}}T}{m}; \tag{1.187}$$

► $$\operatorname{var} v^2 = 6\left(\frac{k_{\mathrm{B}}T}{m}\right)^2. \tag{1.188}$$

Energy Fluctuations

The heat capacity at constant volume is calculated from the definition below:

$$C_V = \frac{\partial U}{\partial T} = \frac{\partial}{\partial T}\left(\sum_n E_n w_n\right) = \frac{\partial}{\partial T}\left(\sum_n E_n(V)\exp\left[\frac{F(T,V) - E_n(V)}{k_{\mathrm{B}}T}\right]\right) \quad \Rightarrow$$

$$C_V = \sum_n E_n \left(\frac{\frac{\partial F}{\partial T}k_{\mathrm{B}}T - (F - E_n)k_{\mathrm{B}}}{(k_{\mathrm{B}}T)^2}\right)\exp\left[\frac{F - E_n}{k_{\mathrm{B}}T}\right] \quad \Rightarrow$$

$$C_V = \sum_n E_n \left(\frac{-Sk_{\mathrm{B}}T - (F - E_n)k_{\mathrm{B}}}{(k_{\mathrm{B}}T)^2}\right)w_n.$$

The obtained expression is divided into individual terms, and quantities over which the summation does not perform are factored out:

$$C_V = -\frac{S}{k_{\mathrm{B}}T}\sum_n E_n w_n - \frac{F}{k_{\mathrm{B}}T^2}\sum_n E_n w_n + \frac{1}{k_{\mathrm{B}}T^2}\sum_n E_n^2 w_n.$$

The subsequent modifications are straightforward:

$$C_V = -\frac{S}{k_{\mathrm{B}}T}U - \frac{F}{k_{\mathrm{B}}T^2}U + \frac{1}{k_{\mathrm{B}}T^2}\langle E^2 \rangle \quad \Rightarrow$$

$$C_V = \frac{-(F + TS)U + \langle E^2 \rangle}{k_{\mathrm{B}}T^2} = \frac{-UU + \langle E^2 \rangle}{k_{\mathrm{B}}T^2} = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_{\mathrm{B}}T^2}.$$

The energy variance is therefore given by:

► $$\operatorname{var} E = k_B T^2 C_V \ . \tag{1.189}$$

Following relations (1.187) and (1.188), another fluctuation relation is proven. The heat capacity is determined by system energy fluctuations. If the energy did not fluctuate, the system would possess a zero heat capacity at constant volume.

Dispersion of Dynamical Variable *A*

Dispersion represents a dimensionless characteristic of fluctuations (the so-called *relative fluctuation*), whose square is defined by the relation:

$$(\delta_A)^2 \equiv \frac{\langle A^2 \rangle - \langle A \rangle^2}{\langle A \rangle^2} \ . \tag{1.190}$$

The numerator contains the variance of A, whereas the denominator contains the square of the mean value. For energy dispersion, the relation is easily obtained as:

$$(\delta_E)^2 \equiv \frac{\langle E^2 \rangle - \langle E \rangle^2}{\langle E \rangle^2} = \frac{\operatorname{var} E}{U^2} = \frac{k_B T^2 C_V}{U^2} \ . \tag{1.191}$$

The energy dispersion for an ideal gas, where $U = fNk_BT/2$ and $C_V = fNk_B/2$, yields:

► $$\delta_E \sim \frac{1}{\sqrt{N}} \ . \tag{1.192}$$

The larger the number of particles the system contains, the smaller the fluctuations it exhibits. Systems with a large number of particles possess minimal relative fluctuations of total energy, whereas systems with a small number of particles exhibit large relative energy fluctuations.

Fluctuations of physical quantities carry extraordinary significance in physics, as nature forces them at the most elementary level. According to the Heisenberg uncertainty relations, a small fluctuation of any generalized coordinate leads to a large variation of the corresponding momentum, and vice versa. Quantum fields can be understood as coordinates with corresponding momenta, where fluctuations cannot be suppressed. They are responsible for the creation of virtual pairs in a vacuum and for the complex spacetime structure at small scales. Fluctuations are also significant in error theory, causing predictable measurement errors. Another application lies in numerical simulations, where tracking fluctuations allows the determination of the specific heat, susceptibility, or compressibility of a system. Although a relatively small portion of this textbook is dedicated to fluctuations, they possess extraordinary significance for statistical physics.

◖ Example 1.14: Spectral Line

Determine the spectral line profile caused by the chaotic motion of source atoms.

Solution: If a radiation source moves relative to an observer, the received frequency changes due to the Doppler effect are:

$$\omega = \omega_0 \left(1 + \frac{v}{c} \cos\alpha \right) = \omega_0 \left(1 + \frac{v_z}{c} \right) .$$

If the radiation source is, for example, a heated gas or a stellar envelope, individual atoms move at various velocities in accordance with the Maxwell-Boltzmann distribution. An observer detects individual emission events frequency-shifted, and the overall result constitutes the Doppler broadening of a spectral line. The radiation frequency fluctuation is determined from the relation for the Doppler effect:

$$\blacktriangleright \qquad \omega - \omega_0 = \omega_0 \frac{v_z}{c} \quad \Rightarrow \quad \Delta\omega^2 = \omega_0^2 \frac{v_z^2}{c^2} \quad \Rightarrow \quad \left\langle \Delta\omega^2 \right\rangle = \omega_0^2 \frac{\left\langle v_z^2 \right\rangle}{c^2} = \omega_0^2 \frac{k_\mathrm{B} T}{mc^2} .$$

The root-mean-square fluctuation (variance) of the frequency is thus obtained. It is proportional to the square of the fundamental frequency and the ratio of thermal energy to the rest energy of a single radiating particle. The cooler the radiating gas, the narrower the spectral line. The exact line profile is determined below. The line intensity is given by the radiation reaching the observer from all particles. The frequency dependence is caused by the Doppler effect. Particles with a specific velocity projection v_z in the direction toward the observer contribute to the intensity at the corresponding frequency. The quantity of particles with a given velocity (and thus the intensity at the given frequency) is determined by the canonical distribution:

$$\mathrm{d}I(v_z) = K \exp\left[-m v_z^2 / 2 k_\mathrm{B} T \right] \mathrm{d}v_z \ .$$

Velocity v_z is converted to frequency using the Doppler relation:

$$\omega = \omega_0 \left(1 + v_z / c\right) \qquad \Rightarrow \qquad v_z = c\left(\omega - \omega_0\right) / \omega_0$$

and substituted into the intensity relation:

$$\blacktriangleright \qquad \mathrm{d}I(\omega) = I_0 \exp\left[-\frac{m c^2}{2 k_\mathrm{B} T} \left(\frac{\omega - \omega_0}{\omega_0} \right)^2 \right] \mathrm{d}\omega \ . \tag{1.193}$$

Due to the Doppler effect, the spectral line acquires a Gaussian packet profile.

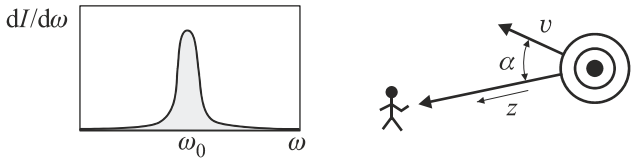


Fig. 1.21: Line profile caused by chaotic motion of sources.

◗

1.8.2 Entropy

Entropy represents a quantity introduced in the thermodynamics of equilibrium processes as a heat differential multiplied by an integrating factor. Entropy constitutes an exact differential form. Statistical physics provides statistical definition of entropy as the mean value of the logarithm of probability (up to a negligible constant):

$$\mathrm{d}S = \frac{\mathrm{d}Q}{T} ; \qquad S = -k_\mathrm{B} \sum_n w_n \ln w_n \ .$$

Thereby, entropy relates to the realization probability of a given system state. A highly interesting statement holds for entropy:

> The entropy of an equilibrium state is extremal ⇔ the system obeys a canonical or grandcanonical distribution ⇔ the system is in thermodynamic equilibrium.

Thermodynamic equilibrium is defined as a state in which the system no longer evolves and macroscopic fluxes do not exist (as these would evidence processes leading to future thermodynamic equilibrium). In equilibrium, the canonical (grandcanonical) distribution holds, and according to the previous statement, entropy reaches an extremum (maximum). Non-equilibrium processes in a closed isolated system tend toward this maximum, increasing the disorder of the system. Conversely, if the extremality of entropy in the equilibrium state is accepted as a primary principle, the canonical (grandcanonical) distribution must hold. This implication is proven below for the grandcanonical distribution:

$$S = -k_{\mathrm{B}} \sum_n w_n \ln w_n ,$$

$$\sum_n w_n = 1 ,$$

$$\sum_n E_n w_n = U ,$$

$$\sum_n N_n w_n = \bar{N} .$$

This problem involves finding the entropy extremum under three constraint conditions. Extremes with constraint conditions are found using the Lagrange multipliers method. The following function is introduced:

$$F(w_l, \lambda_1, \lambda_2, \lambda_3) =$$

$$= -k_{\mathrm{B}} \sum_n w_n \ln w_n + \lambda_1 \left(\sum_n w_n - 1 \right) + \lambda_2 \left(\sum_n E_n w_n - U \right) + \lambda_3 \left(\sum_n N_n w_n - \bar{N} \right).$$

This function constitutes entropy with added constraints (null expressions) containing multipliers λ. The extremality condition of the newly introduced function of three variables is examined below:

$$\frac{\partial F}{\partial w_l} = 0 \quad \Rightarrow \quad -k_{\mathrm{B}} \ln w_l - k_{\mathrm{B}} + \lambda_1 + \lambda_2 E_l + \lambda_3 N_l \quad \Rightarrow \quad w_l = \mathrm{e}^{c_1 + c_2 E_l + c_3 N_l} .$$

The derived condition represents the grandcanonical distribution. Thus, the grandcanonical distribution is naturally linked to the extremality of entropy. In the case of the canonical distribution, the particle number is constant, and the extremality of entropy is examined under only two constraints. The result of a similar consideration for a constant particle number is the canonical distribution.

Entropic Force

If a material particle in the system surroundings moves, its motion can cause an increase in system entropy. The opposite situation can also occur: a change in system entropy causes the motion of an external particle, representing a situation where a force acts on the particle. The energy balance is written as:

$$T\Delta S = F\Delta x\,, \tag{1.194}$$

or:

$$\mathbf{F} = T\nabla S\,. \tag{1.195}$$

Such a force is called an entropic force. The category of entropic forces includes, for instance, the elasticity of a rubber band, which relates to the entropy change during stretching. In 2010, the Dutch physicist Erik Verlinde attempted to explain gravitational interaction as an entropic force caused by quantum manifestations of the microworld. His theory relied on both entropy growth and the holographic principle (the assertion that all information about a system can be encoded on a lower-dimensional entity – for example, the entropy of a black hole relates to its surface area). The extent to which Verlinde's considerations represent a utopia or a promising hypothesis remains unknown.

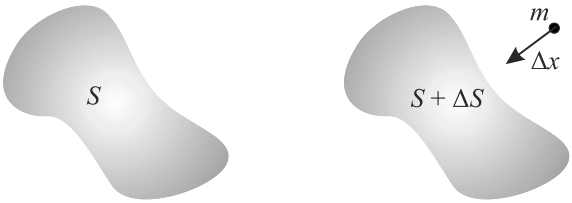


Fig. 1.22: Principle of entropic force.

Further information regarding entropy can be found in publication [13]. Verlinde's works on gravity as an entropic force can be examined in publications [14], [15], [16].

1.9 Magnetically Active Systems

1.9.1 Fundamental Concepts

System of Charged Particles

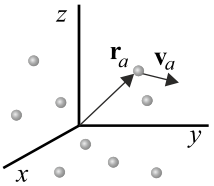


Fig. 1.23: System of charges near the coordinate origin.

A system composed of N charged particles capable of responding to external electric and magnetic fields, creating their own fields, and interacting with each other through these fields is considered below. To solve the problem comprehensively, *fields* must be calculated from the Maxwell equations supplemented by constitutive relations (describing how the system responds to the presence of fields as a whole), and *particle motions* must be calculated from the Lorentz motion equation. The position and velocity of the a-th particle are denoted as $\mathbf{r}_a$ and $\mathbf{v}_a$, respectively. As a whole, the system of particles exhibits electric and magnetic dipole moments. Certain concepts from electricity and magnetism required for the statistical description are reviewed in this section.

Electric Dipole

The *electric dipole moment* $\mathbf{p}_\mathrm{E}$ of a particle system is defined by the relation:

► $$\mathbf{p}_\mathrm{E} = \sum_a Q_a \mathbf{r}_a \ . \tag{1.196}$$

For a pair of particles with opposite charges and a mutual position vector $\mathbf{d}$, this definition yields the expression (for small d, this is referred to as an elementary dipole):

$$\mathbf{p}_\mathrm{E} = +Q\mathbf{r}_+ - Q\mathbf{r}_- = Q(\mathbf{r}_+ - \mathbf{r}_-) = Q\mathbf{d} \ .$$

Dipole interaction with an external electric field is given by the energy expression:

► $$W_\mathrm{int} = -\mathbf{p}_\mathrm{E} \cdot \mathbf{E} \ . \tag{1.197}$$

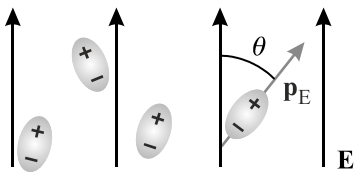


Fig. 1.24: Electric dipole in an electric field.

The lowest interaction energy occurs when a dipole points in the direction of the electric field. At low temperatures or in strong fields, the system of dipoles is aligned in the field direction. Conversely, at high temperatures and in weak fields, fluctuations occur, and the dipole position is described by the corresponding statistical distribution.

◖ Example 1.15: Electric Dipoles

Determine the distribution of an electric dipole ensemble in an external field.

$$W_{\text{int}} = -\mathbf{p}_{\text{E}} \cdot \mathbf{E} = -p_{\text{E}} E \cos\theta \qquad \Rightarrow$$

$$\mathrm{d}w(\varphi,\theta) = K \cdot \exp\left[\frac{p_{\text{E}} E \cos\theta}{k_{\text{B}} T}\right] \mathrm{d}\Omega \qquad \Rightarrow$$

$$\mathrm{d}w(\varphi,\theta) = K \cdot \exp\left[\frac{p_{\text{E}} E \cos\theta}{k_{\text{B}} T}\right] \sin\theta \, \mathrm{d}\varphi \, \mathrm{d}\theta \qquad \Rightarrow$$

$$\mathrm{d}w(\theta) = C \cdot \exp\left[\frac{p_{\text{E}} E \cos\theta}{k_{\text{B}} T}\right] \sin\theta \, \mathrm{d}\theta.$$

◗

Polarization is defined as the electric dipole moment density:

► $$\mathbf{P} = \lim_{\Delta V \to 0} \frac{\Delta \mathbf{p}_{\text{E}}}{\Delta V}. \tag{1.198}$$

While the electric dipole moment $\mathbf{p}_{\text{E}}$ represents an additive quantity, the polarization vector $\mathbf{P}$ is an intensive quantity (a density independent of the particle number). The constitutive relation between the electric displacement and the electric field can be written via polarization (polarization describes the system response to an electric field):

► $$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}(\mathbf{E}). \tag{1.199}$$

For weak fields, the system response can be considered linear, allowing a first-order Taylor expansion of the polarization vector (the summation convention is utilized):

$$P_k = \frac{\partial P_k}{\partial E_l} E_l \equiv \varepsilon_0 \kappa_{kl} \, E_l \, ; \tag{1.200}$$

$$\kappa_{kl} \equiv \frac{1}{\varepsilon_0} \frac{\partial P_k}{\partial E_l}. \tag{1.201}$$

The corresponding matrix of coefficients is called the electric susceptibility tensor. The vacuum permittivity in the definition ensures the dimensionless nature of the susceptibility. In vector notation, this can be written as:

$$\mathbf{P} = \varepsilon_0 \, \ddot{\boldsymbol{\kappa}} \cdot \mathbf{E}. \tag{1.202}$$

The relationship between the electric displacement and the electric field in the weak field limit yields:

$$D_k = \varepsilon_{kl} \, E_l \, ; \qquad \varepsilon_{kl} \equiv \varepsilon_0 (\delta_{kl} + \kappa_{kl}), \tag{1.203}$$

$$\mathbf{D} = \ddot{\boldsymbol{\varepsilon}} \cdot \mathbf{E} \, ; \qquad \ddot{\boldsymbol{\varepsilon}} = \varepsilon_0 \left(\ddot{\mathbf{1}} + \ddot{\boldsymbol{\kappa}} \right). \tag{1.204}$$

The matrix ε_{kl} is called the *permittivity tensor*. The fields **D** and **E** do not necessarily point in the identical direction. In the subsequent section, the susceptibility tensor, and consequently the permittivity tensor, is shown to be symmetric. In a homogeneous isotropic medium, the susceptibility matrix is even diagonal:

$$\mathbf{D} = \varepsilon \mathbf{E} \,. \tag{1.205}$$

The internal energy density of an electric dipole system in an electric field is given by the relation (originating from the Maxwell equations; in foundational physics courses, this corresponds to the energy density of a capacitor):

$$u_{\mathrm{E}} = \tfrac{1}{2}\, \mathbf{E} \cdot \mathbf{D} \,. \tag{1.206}$$

In the first thermodynamic law, the differential of the energy density is required; therefore, the linear expansion of the polarization vector is sufficient:

$$u_{\mathrm{E}} = \frac{1}{2} \mathbf{E} \cdot \left(\varepsilon_0 \mathbf{E} + \varepsilon_0 \ddot{\boldsymbol{\kappa}} \cdot \mathbf{E} \right) = \frac{\varepsilon_0 \mathbf{E}^2}{2} + \frac{\varepsilon_0}{2} \mathbf{E} \cdot \ddot{\boldsymbol{\kappa}} \cdot \mathbf{E} \,, \tag{1.207}$$

$$\mathrm{d}u_{\mathrm{E}} = \varepsilon_0 \mathbf{E} \cdot \mathrm{d}\mathbf{E} + \varepsilon_0 \mathbf{E} \cdot \ddot{\boldsymbol{\kappa}} \cdot \mathrm{d}\mathbf{E} = \varepsilon_0 \mathbf{E} \cdot \mathrm{d}\mathbf{E} + \mathbf{E} \cdot \mathrm{d}\mathbf{P} \,. \tag{1.208}$$

The first term relates to the field energy density, whereas the second corresponds to the material response (in this case, the dipole system) to the applied electric field. This response relates to polarization and matches the internal energy of the system:

► $$\mathrm{d}u = \mathbf{E} \cdot \mathrm{d}\mathbf{P} \,. \tag{1.209}$$

Analogously to the electric dipole, the fundamental relations for the interaction of a magnetic field with a magnetic dipole system are defined below.

Magnetic Dipole

The *magnetic dipole* moment of a particle system is defined by the relation:

► $$\mathbf{p}_{\mathrm{M}} = \frac{1}{2} \sum_a Q_a \, \mathbf{r}_a \times \mathbf{v}_a \,. \tag{1.210}$$

The simplest situation – a single charged particle orbiting along a small circle (the so-called elementary magnetic dipole) – is considered below.

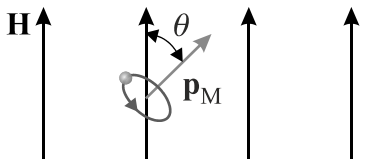


Fig. 1.25: Magnetic dipole in a magnetic field.

According to (1.210), the magnitude of the magnetic dipole moment (where T denotes the orbital period and I represents the current caused by the charge orbit) satisfies the expression:

$$p_{\mathrm{M}} = \frac{Q r v}{2} = \frac{Q r}{2} \frac{2\pi r}{T} = \frac{Q}{T} \pi r^2 = I S \,.$$

The magnetic dipole moment of a system of identical particles is related to the total angular momentum **L** (the particle mass is denoted as m_0 to prevent confusion with the magnetic quantum number):

$$\mathbf{p}_{\mathrm{M}} = \frac{Q}{2}\sum_a \mathbf{r}_a \times \mathbf{v}_a = \frac{Q}{2m_0}\sum_a m_0\, \mathbf{r}_a \times \mathbf{v}_a = \frac{Q}{2m_0}\mathbf{L}\,. \tag{1.211}$$

Since the angular momentum is quantized in the microworld, the magnetic moment is quantized correspondingly. The interaction of a dipole with an external magnetic field is given by the energy expression:

► $$W_{\mathrm{int}} = -\mathbf{p}_{\mathrm{M}} \cdot \mathbf{B} = -\mu_0\, \mathbf{p}_{\mathrm{M}} \cdot \mathbf{H}\,. \tag{1.212}$$

Analogously to electric dipoles, magnetic dipoles tend to align in the direction of the magnetic field, where they possess the lowest possible energy. The magnetic dipole moment density

► $$\mathbf{M} = \lim_{\Delta V \to 0} \frac{\Delta \mathbf{p}_{\mathrm{M}}}{\Delta V} \tag{1.213}$$

is called *magnetization*. While the magnetic dipole moment $\mathbf{p}_{\mathrm{M}}$ represents an additive quantity, the magnetization vector **M** is an intensive quantity (a density independent of the particle number). The constitutive relation between the magnetic induction and the magnetic field intensity can be written using the magnetization vector (magnetization describes the system response to a magnetic field):

► $$\mathbf{B} = \mu_0 \left[\mathbf{H} + \mathbf{M}(\mathbf{H})\right]. \tag{1.214}$$

For weak fields, the system response can be considered linear, allowing a first-order Taylor expansion of the magnetization vector (the summation convention is utilized):

$$M_k = \frac{\partial M_k}{\partial H_l} H_l = \chi_{kl}\, H_l\,; \tag{1.215}$$

► $$\chi_{kl} \equiv \frac{\partial M_k}{\partial H_l}\,. \tag{1.216}$$

The corresponding matrix of coefficients is called the *magnetic susceptibility tensor*. In vector notation, this can be written as:

$$\mathbf{M} = \ddot{\boldsymbol{\chi}} \cdot \mathbf{H}\,. \tag{1.217}$$

The relationship between the two magnetic vectors in the weak field limit yields:

$$B_k = \mu_{kl}\, H_l\,; \qquad \mu_{kl} \equiv \mu_0(\delta_{kl} + \chi_{kl})\,. \tag{1.218}$$

$$\mathbf{B} = \ddot{\boldsymbol{\mu}} \cdot \mathbf{H}\,; \qquad \ddot{\boldsymbol{\mu}} = \mu_0 \left(\ddot{\mathbf{I}} + \ddot{\boldsymbol{\chi}}\right). \tag{1.219}$$

The matrix μ_{kl} is called the *permeability tensor*. The fields **B** and **H** do not necessarily point in the identical direction. In the subsequent section, the susceptibility tensor, and consequently the permeability tensor, is shown to be symmetric. In a homogeneous isotropic medium, the susceptibility matrix is even diagonal with identical diagonal elements. In the weak field limit, an even simpler relation holds in this case:

$$\mathbf{B} = \mu\,\mathbf{H}\,. \tag{1.220}$$

The internal energy density of a magnetic dipole system in a magnetic field is given by the relation (originating from the Maxwell equations; in foundational physics courses, this corresponds to the energy density of an inductor):

$$u_{\mathrm{M}} = ½\,\mathbf{H}\cdot\mathbf{B}\,. \tag{1.221}$$

In the first thermodynamic law, only the differential of the energy density is required; therefore, the linear expansion of the magnetization vector is sufficient:

$$u_{\mathrm{M}} = \frac{\mu_0}{2}\mathbf{H}\cdot\left(\mathbf{H}+\tilde{\boldsymbol{\chi}}\cdot\mathbf{H}\right) = \frac{\mu_0\mathbf{H}^2}{2}+\frac{\mu_0}{2}\mathbf{H}\cdot\tilde{\boldsymbol{\chi}}\cdot\mathbf{H}\,, \tag{1.222}$$

$$\mathrm{d}u_{\mathrm{M}} = \mu_0\mathbf{H}\cdot\mathrm{d}\mathbf{H}+\mu_0\mathbf{H}\cdot\tilde{\boldsymbol{\chi}}\cdot\mathrm{d}\mathbf{H} = \mu_0\mathbf{H}\cdot\mathrm{d}\mathbf{H}+\mu_0\,\mathbf{H}\cdot\mathrm{d}\mathbf{M}\,. \tag{1.223}$$

The first term relates to the field energy density, whereas the second corresponds to the material response (in this case, the dipole system) to the applied magnetic field. This response relates to magnetization and matches system internal energy:

► $$\mathrm{d}u = \mu_0\,\mathbf{H}\cdot\mathrm{d}\mathbf{M}\,. \tag{1.224}$$

1.9.2 Magnetically Active Materials

First Thermodynamic Law and Free Energy

For the interaction of electric and magnetic fields with a system, only the internal energy density is known. Therefore, the first thermodynamic law is also rewritten in terms of densities. The following quantities are introduced:

$$u = \frac{U}{V}; \qquad s = \frac{S}{V}; \qquad f = \frac{F}{V}; \qquad n = \frac{N}{V}\,. \tag{1.225}$$

These quantities represent the internal energy density, entropy density, free energy density, and particle density, respectively. Symbol collisions can occur due to the limited alphabet; signs are reused where the context keeps meanings obvious (e.g., magnetic quantum number vs. mass, electric dipole moment vs. momentum, free energy density vs. degrees of freedom, mass density vs. probability density). The first thermodynamic law (1.7) for a constant particle number is converted into densities below:

$$\mathrm{d}u = T\,\mathrm{d}s - \frac{1}{V}\,p\,\mathrm{d}V\,.$$

Since volume is not an intensive quantity, its appearance in the relation for energy density is unacceptable; therefore, volume is expressed using mass density:

$$\rho V = M \qquad \Rightarrow \qquad \rho\,\mathrm{d}V + V\,\mathrm{d}\rho = 0 \qquad \Rightarrow \qquad -\frac{\mathrm{d}V}{V} = \frac{\mathrm{d}\rho}{\rho}\,.$$

The first thermodynamic law written in densities now takes the form:

► $$\mathrm{d}u = T\,\mathrm{d}s + \frac{p}{\rho}\mathrm{d}\rho\,. \tag{1.226}$$

Electrically and magnetically active systems are considered below. The internal energy terms in the electric and magnetic fields (1.209) and (1.224) are added to the right side:

► $$\mathrm{d}u = T\,\mathrm{d}s + \frac{p}{\rho}\mathrm{d}\rho + \mathbf{E}\cdot\mathrm{d}\mathbf{P} + \mu_0\mathbf{H}\cdot\mathrm{d}\mathbf{M}\,. \tag{1.227}$$

Additional terms appear in the energy conservation law, expressing that the system can perform electric and magnetic work. The final terms represent scalar products, meaning each constitutes a sum of three components. The completion of the differentials for the first, third, and fourth terms on the right side is performed below:

$$\mathrm{d}(u - Ts - \mathbf{E}\cdot\mathbf{P} - \mu_0\mathbf{H}\cdot\mathbf{M}) \;=\; -s\,\mathrm{d}T + \frac{p}{\rho}\mathrm{d}\rho - \mathbf{P}\cdot\mathrm{d}\mathbf{E} - \mu_0\mathbf{M}\cdot\mathrm{d}\mathbf{H}\,.$$

The relationship for the free energy density and its differential is thus obtained:

$$f = u - Ts - \mathbf{E}\cdot\mathbf{P} - \mu_0\mathbf{H}\cdot\mathbf{M}\,;$$

$$\mathrm{d}f = -s\,\mathrm{d}T + \frac{p}{\rho}\mathrm{d}\rho - \mathbf{P}\cdot\mathrm{d}\mathbf{E} - \mu_0\mathbf{M}\cdot\mathrm{d}\mathbf{H}\,; \tag{1.228}$$

$$f = f(T, \rho, \mathbf{E}, \mathbf{H})\,.$$

The derivatives of the free energy density with respect to its variables yield the individual differential coefficients:

$$s = -\frac{\partial f}{\partial T}\,; \qquad p = \rho\frac{\partial f}{\partial \rho}\,; \qquad \mathbf{P} = -\frac{\partial f}{\partial \mathbf{E}}\,; \qquad \mathbf{M} = -\frac{1}{\mu_0}\frac{\partial f}{\partial \mathbf{H}}\,. \tag{1.229}$$

Knowledge of the free energy density allows the determination of the entropy density, equation of state, polarization vector, and magnetization vector, i.e. the responses induced in the system by external fields. The last two relations are vector equations, written in components as:

$$P_k = -\frac{\partial f}{\partial E_k}\,; \qquad M_k = -\frac{1}{\mu_0}\frac{\partial f}{\partial H_k}\,.$$

Using relations (1.201) and (1.216), the electric and magnetic susceptibility tensors can be determined from the free energy density

$$\kappa_{kl} = -\frac{1}{\varepsilon_0}\frac{\partial^2 f}{\partial E_k\,\partial E_l}\,; \qquad \chi_{kl} = -\frac{1}{\mu_0}\frac{\partial^2 f}{\partial H_k\,\partial H_l} \tag{1.230}$$

and consequently, according to relations (1.203) and (1.218), the permittivity and permeability tensors are obtained:

$$\varepsilon_{kl} \equiv \varepsilon_0\delta_{kl} - \frac{\partial^2 f}{\partial E_k\,\partial E_l}\,; \qquad \mu_{kl} \equiv \mu_0\delta_{kl} - \frac{\partial^2 f}{\partial H_k\,\partial H_l}\,. \tag{1.231}$$

Evidently, both susceptibilities, permittivity, and permeability represent symmetric tensors. Knowledge of the free energy density enables the calculation of fundamental thermodynamic, electric, and magnetic system characteristics.

Root-Mean-Square Fluctuation of Magnetization

Only a magnetically active material is considered below. The derivation is conducted discretely, since magnetization is related to the angular momentum of particles, which is quantized. The possible states of magnetization are indexed by the subscript n.

$$w_n = \exp\left[\beta(F - W_{\text{int}})\right] = \exp\left[\beta(F + \mu_0 V\, \mathbf{M}_n \cdot \mathbf{H})\right]. \tag{1.232}$$

The mean value of magnetization is given by the relation:

$$\langle\, \mathbf{M} \,\rangle \equiv \sum_n \mathbf{M}_n w_n \,. \tag{1.233}$$

The susceptibility is found below in a one-dimensional problem for simplicity:

$$\chi = \frac{\partial \langle\, M \,\rangle}{\partial H} = \frac{\partial}{\partial H}\left(\sum_n M_n \exp\left[\beta f(T,H) V + \beta \mu_0 M_n V H\right] \right);$$

$$\chi = \sum_n \left(M_n \beta \left(\frac{\partial f}{\partial H} V + \mu_0 V M_n \right) \exp[\, \cdots \,] \right).$$

Similarly to the derivation of energy fluctuations, individual terms are separated, and quantities over which the summation does not perform are factored out:

$$\chi = \sum_n \left(M_n \beta \left(-\mu_0 \langle\, M \,\rangle V + \mu_0 M_n V \right) w_n \right);$$

$$\chi = -\beta\mu_0 \langle\, M \,\rangle V \sum_n M_n w_n \;+\; \beta\mu_0 V \sum_n M_n^2 w_n \,;$$

$$\chi = -\beta\mu_0 V \langle\, M \,\rangle^2 \;+\; \beta\mu_0 V \langle\, M^2 \,\rangle \,;$$

$$\chi = \beta\mu_0 V \left(\langle\, M^2 \,\rangle - \langle\, M \,\rangle^2 \right).$$

A simple relation between susceptibility and magnetization fluctuation is thus obtained:

► $$\chi = \beta\mu_0 V \left(\langle\, M^2 \,\rangle - \langle\, M \,\rangle^2 \right) = \beta\mu_0 V \operatorname{var} M \,. \tag{1.234}$$

Another highly important fluctuation relation is proven.

Sources of Magnetic Moment and Lande Factor

The source of a particle magnetic moment can be either the orbital (**L**) or spin (**S**) angular momentum of the particle. The combination of two momenta in quantum theory is reviewed below, providing a summary of results for two momenta denoted as **L** and **S**, regardless of their physical meaning:

$\mathbf{L}$	$L = \sqrt{l(l+1)}\,\hbar$	$L_3 = m_L \hbar \,;$	$m_L = -l, -l+1, \ldots, l-1, l$
$\mathbf{S}$	$S = \sqrt{s(s+1)}\,\hbar$	$S_3 = m_S \hbar \,;$	$m_S = -s, -s+1, \ldots, s-1, s$
$\mathbf{J} = \mathbf{L} + \mathbf{S}$	$J = \sqrt{j(j+1)}\,\hbar \,;$	$J_3 = m_J \hbar \,;$	$m_J = -j, -j+1, \ldots, j-1, j$

The resulting quantum number j can assume the values $j = |\, l - s \,|, \ldots, l + s$. Even a neutral particle (such as a neutron) can possess a magnetic dipole moment caused by a non-zero spin. The magnetic dipole moment (1.211) is modified as follows:

$$\mathbf{p}_{\mathrm{M}} = g \frac{Q}{2m_0} \mathbf{J} . \tag{1.235}$$

The quantity **J** represents the total angular momentum of the particle, and the factor g is characteristic of a specific particle species:

$$\begin{aligned} g &= -2 ; && \text{electron} \\ g &= +5.68 ; && \text{proton} \\ g &= -3.86 ; && \text{neutron} \end{aligned}$$

The situation is somewhat more complex in a system such as an entire atom. Here, *LS* coupling occurs, representing the interaction of collective orbital and spin degrees of freedom of the entire system. The result is identical to relation (1.235), except the factor g is given by a more complex formula:

$$g = 1 + \frac{j(j+1) + s(s+1) - l(l+1)}{2j(j+1)} \tag{1.236}$$

and is called the *Lande g-factor*, whose derivation is outlined below. In the quantum mechanical Hamiltonian, the atomic interaction with an external field is described by:

$$\hat{H}_{\mathrm{int}} = \mathrm{const}\, (\hat{\mathbf{L}} + 2\hat{\mathbf{S}}) \cdot \mathbf{H} .$$

The coefficient 2 for the spin momentum is given by the anomalous magnetic moment of the electron. In the time average, only components parallel to the total momentum **J** from both orbital and spin parts contribute to the interaction. Therefore, the projection of the vector combination **L**+2**S** onto the total momentum **J** is determined below:

$$P_{\mathbf{J}}(\mathbf{L} + 2\mathbf{S}) = \frac{|\mathbf{J}\rangle\langle\mathbf{J}|\mathbf{L} + 2\mathbf{S}\rangle}{\langle\mathbf{J}|\mathbf{J}\rangle} = \frac{\mathbf{J} \cdot (\mathbf{L} + 2\mathbf{S})}{J^2} \mathbf{J} .$$

This projection is transformed into the squares of momenta J^2, L^2, and S^2:

$$P_{\mathbf{J}}(\mathbf{L} + 2\mathbf{S}) = \frac{\mathbf{J} \cdot (\mathbf{J} + \mathbf{S})}{J^2} \mathbf{J} = \frac{J^2 + \mathbf{J} \cdot \mathbf{S}}{J^2} \mathbf{J} = \frac{J^2 + (\mathbf{L} + \mathbf{S}) \cdot \mathbf{S}}{J^2} \mathbf{J} = \frac{J^2 + S^2 + \mathbf{L} \cdot \mathbf{S}}{J^2} \mathbf{J} .$$

The term $\mathbf{L} \cdot \mathbf{S}$ is determined from the identity $J^2 = (\mathbf{L}+\mathbf{S})^2 = L^2 + 2\mathbf{L} \cdot \mathbf{S} + S^2$:

$$P_{\mathbf{J}}(\mathbf{L} + 2\mathbf{S}) = \frac{J^2 + S^2 + (J^2 - L^2 - S^2)/2}{J^2} \mathbf{J} = \frac{3J^2 + S^2 - L^2}{2J^2} \mathbf{J} = \left(1 + \frac{J^2 + S^2 - L^2}{2J^2} \right) \mathbf{J} .$$

Utilizing the quantization of these quantities, the expression becomes:

$$P_{\mathbf{J}}(\mathbf{L} + 2\mathbf{S}) = \left(1 + \frac{j(j+1) + s(s+1) - l(l+1)}{2j(j+1)} \right) \mathbf{J} .$$

The factor in the parentheses represents the *Lande factor*, which multiplicatively modifies the total angular momentum **J** due to the *LS* coupling.

Interaction of Magnetic Moments

Magnetic moment interaction with an external field is given by the relation:

$$W_{\text{int}} = -\,\mathbf{p}_{\text{M}} \cdot \mathbf{B} = -\,g\,\frac{Q}{2m_0}\,\mathbf{J} \cdot \mathbf{B}\,. \tag{1.237}$$

The momentum **J** tends to align either in the direction of the external field or opposite to it, depending on the sign of g. The entire expression is frequently written in the form:

$$W_{\text{int}} = -\,g\,\frac{Q}{2m_0}\,J_z B = -\,g\,\frac{Q}{2m_0}\,m_J\,\hbar\,B = -\,m_J\,g\,\frac{Q\hbar}{2m_0}\,B\,.$$

The fraction in the expression is called the *Bohr magneton* μ_{B} (an elementary charge $Q = e$ is considered) – the sign is negligible because the magnetic quantum number m_J ranges through both negative and positive values:

► $$W_{\text{int}} = m_J\,g\,\mu_{\text{B}}\,B\,; \quad \mu_{\text{B}} \equiv \frac{e\hbar}{2m_0}, \quad m_J = -j, -j+1, \ldots, j-1, j\,. \tag{1.238}$$

For a system of magnetic moments, the mutual interaction of the moments must also be considered, and instead of magnetic induction, using the magnetic field intensity is advantageous, as it relates to the free energy of the system

$$W_{\text{int}} = \sum_{a,b} W_0(\mathbf{p}_a, \mathbf{p}_b) - g\mu_0\,\frac{Q}{2m_0}\sum_a \mathbf{J}_a \cdot \mathbf{H}\,. \tag{1.239}$$

Zeeman Effect

In an external magnetic field, an originally single energy level splits into $2j+1$ sublevels according to relation (1.238). This behavior leads to the splitting of spectral lines, termed the Zeeman effect. The energy difference between two adjacent sublevels is given by: $\Delta E = g\mu_{\text{B}}B$ and the distance between the split lines can be easily determined.

Magnetic Resonance

A system of magnetic dipoles in an external field can absorb electromagnetic radiation quanta corresponding to the dipole energy level difference. The passing field induces transitions between states with different m_J. The energy spectrum (1.238) is equidistant, and resonant photons must satisfy the relation:

$$\hbar\omega_{\text{rez}} = g\mu_{\text{B}}B\,.$$

The resonant frequency is easily calculated as:

$$f_{\text{rez}} = K\,B\,; \qquad K = \begin{cases} 1{,}44\ \text{MHz T}^{-1} & \text{for atom} \\ 0{,}76\ \text{kHz T}^{-1} & \text{for nucleus} \end{cases} \tag{1.240}$$

Magnetic resonance is utilized as a successful imaging method. An object is immersed in a field B and irradiated with electromagnetic radiation at the resonant frequency. Resonance on whole atoms (or their electron shells) is called *electron magnetic resonance* (EMR), whereas resonance on nuclear particles is termed *nuclear magnetic resonance* (NMR).

Curie Law

The simplest possible system composed of N elementary magnetic dipoles is considered below, which can possess only two projections of the magnetic dipole moment, $\pm\mu_B/2$ (for instance, formed by particles with spin ½). The interaction of a single dipole with an external magnetic field possesses only two possible energy values:

$$W_{\text{int}} = -\mathbf{p}_M \cdot \mathbf{B} = \pm\frac{\mu_B}{2} B = \pm\frac{\mu_0 \mu_B H}{2}.$$

In the scalar product, only the projection of the magnetic dipole moment in the field direction is applied, representing one of the two possible values. The partition function of a single dipole contains only two terms:

$$z = \exp\left[+\frac{\mu_0 \mu_B H}{2k_B T}\right] + \exp\left[-\frac{\mu_0 \mu_B H}{2k_B T}\right] = 2\cosh\left[\frac{\mu_0 \mu_B H}{2k_B T}\right].$$

The partition function of a system of N mutually non-interacting dipoles yields:

$$Z_N = z^N = \left(2\cosh\left[\frac{\mu_0 \mu_B H}{2k_B T}\right]\right)^N.$$

The free energy density, and consequently magnetization, susceptibility, and permeability, are determined in a standard manner:

$$f = -\frac{1}{V} k_B T \ln Z_N = -\frac{N k_B T}{V} \ln\left[2\cosh\left(\frac{\mu_0 \mu_B H}{2k_B T}\right)\right],$$

$$M = -\frac{1}{\mu_0}\frac{\partial f}{\partial H} = n\frac{\mu_B}{2}\tanh\left[\frac{\mu_0 \mu_B H}{2k_B T}\right], \tag{1.241}$$

$$\chi = -\frac{1}{\mu_0}\frac{\partial^2 f}{\partial H^2} = n\frac{\mu_0 \mu_B^2}{4k_B T}\cosh^{-2}\left[\frac{\mu_0 \mu_B H}{2k_B T}\right], \tag{1.242}$$

$$\mu \equiv \mu_0 - \frac{\partial^2 f}{\partial H^2} = \mu_0 + n\frac{\mu_0^2 \mu_B^2}{4k_B T}\cosh^{-2}\left[\frac{\mu_0 \mu_B H}{2k_B T}\right]. \tag{1.243}$$

Even such a simple system (two orientation possibilities of elementary magnets) exhibits non-linear behavior. Magnetization is not a linear function of the field intensity but a general function $M = M(H)$. At high field intensities and low temperatures, magnetization exhibits saturation (attaining a maximum value independent of the field intensity magnitude). Under these conditions, all spins are aligned in the field direction. The saturation value is given by:

$$M_S = n\mu_B/2.$$

Linearizing the relation for weak fields yields:

$$M = \chi H\,; \qquad \chi \equiv \left.\frac{\partial M}{\partial H}\right|_{H=0} = \frac{n\mu_0\mu_\mathrm{B}^2}{4k_\mathrm{B}T}\ . \tag{1.244}$$

The susceptibility is inversely proportional to the temperature of the magnetic material. This relationship is called the *Curie law* and holds only in the weak field limit. It is named after the French physicist Pierre Curie.

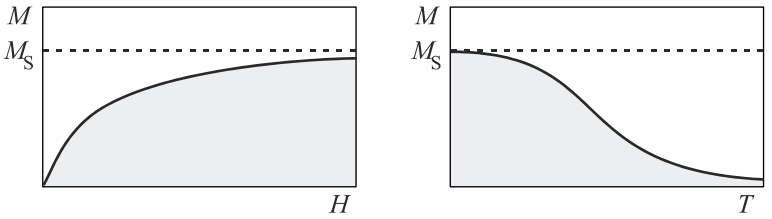
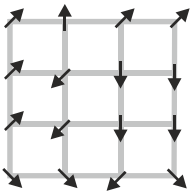


Fig. 1.26: Dependence of magnetization on the magnetic field and temperature.

1.9.3 Lattice Models



If multiple spin values and their mutual interactions are admitted, calculating the partition function is generally an insurmountable problem. Therefore, a relatively simple model of spins localized at rectangular lattice vertices is frequently utilized, closely corresponding to real materials. Generally, it is assumed that only the nearest spins at adjacent lattice vertices interact; such adjacent vertices are denoted as $\langle a, b\rangle$. The interaction rule is analogous to expression (1.239), except the summation over the mutual interaction performs only over the nearest neighbors:

$$W_\mathrm{int} = \sum_{\langle a,b\rangle} W_0(\boldsymbol{\sigma}_a, \boldsymbol{\sigma}_b) \ - \ K\sum_a \boldsymbol{\sigma}_a \cdot \mathbf{H}. \tag{1.245}$$

The quantities $\boldsymbol{\sigma}$ characterize the spin momentum and are proportional to the magnetic moment. Individual models are distinguished by the type of mutual interaction W_0.

Ising Model

This represents the simplest possible model. Spins can assume only two values (for instance, "up" and "down"), and the interaction energy is given by the rule (all coefficients and constants are absorbed into the interaction constants J and K):

$$W_\mathrm{int} = -J\sum_{\langle a,b\rangle} \sigma_a\sigma_b - K\sum_a \sigma_a H; \qquad \sigma \in \{-1,+1\}\,. \tag{1.246}$$

The value +1 denotes a spin pointing upward, while the value –1 represents a spin pointing downward. For $J > 0$), two adjacent identically oriented spins contribute $-J$ to the energy, whereas two oppositely oriented spins contribute $+J$. At low temperatures, the system tends to occupy the lowest possible energy, leading to the formation of domains of identically oriented spins (*Weiss domains*). Their theory was developed by the French physicist Pierre-Ernest Weiss. At high temperatures, the spins are oriented randomly. Even this simple model exhibits the fundamental properties of *ferromagnets* for $J > 0$. An ordered phase exists at low temperatures, while a chaotic phase appears at high temperatures. A phase transition between both phases occurs at the Curie temperature T_C. For $J < 0$, oppositely oriented spins are preferred at low temperatures, and the system exhibits *antiferromagnetic* properties. In a one-dimensional problem, finding the partition function even in the presence of external fields is a trivial matter. The problem was solved analytically within a doctoral thesis by Ernst Ising in 1925. It was shown that in a linear spin chain, the situation is sufficiently simple that no phase transition occurs. In two dimensions, the problem was solved using refined graphical methods by Lars Onsager in 1944. A Curie phase transition already exists in the 2D Ising model. A general analytical solution in three dimensions is unknown, and the partition function, magnetization, and susceptibility are determined numerically.

One-Dimensional Ising Calculation (1925)

A one-dimensional chain of N spins on a lattice with periodic boundary conditions is assumed below; thus, the following holds:

$$\sigma_{N+1} = \sigma_1 \,. \tag{1.247}$$

For the partition function, the relation applies:

$$Z_N = \sum_{\sigma} \mathrm{e}^{\left[\beta J \sum_{a=1}^{N} \sigma_a \sigma_{a+1} \; + \; \beta K \sum_{a=1}^{N} \sigma_a H \right]} . \tag{1.248}$$

The symmetrization of the second part of the expression is performed below:

$$Z = \sum_{\sigma} \mathrm{e}^{\left[\beta J \sum_{a=1}^{N} \sigma_a \sigma_{a+1} + \beta K \frac{1}{2} \sum_{a=1}^{N} (\sigma_a + \sigma_{a+1}) H \right]} = \sum_{\sigma} \prod_{a=1}^{N} \mathrm{e}^{\left[\beta J \sigma_a \sigma_{a+1} + \frac{1}{2} \beta K (\sigma_a + \sigma_{a+1}) H \right]} . \tag{1.249}$$

Introducing the matrix ($\sigma_a = \pm 1$)

$$T_{\sigma_a \sigma_b} = \begin{pmatrix} \mathrm{e}^{\beta J + \beta K H}, & \mathrm{e}^{-\beta J} \\ \mathrm{e}^{-\beta J}, & \mathrm{e}^{\beta J - \beta K H} \end{pmatrix} . \tag{1.250}$$

the partition function can be rewritten in the form:

$$Z_N = \sum_{\sigma} \prod_{a=1}^{N} T_{\sigma_a \sigma_{a+1}} = \sum_{\sigma_k = \pm 1} T_{\sigma_1 \sigma_2} T_{\sigma_2 \sigma_3} T_{\sigma_3 \sigma_4} \cdots T_{\sigma_N \sigma_1} = \mathrm{Tr}\left(T^N \right) . \tag{1.251}$$

Two identical adjacent subscripts indicate matrix multiplication. The result is the trace T^N, which represents an invariant and remains identical in any basis. In the basis of eigenvectors, the matrix T possesses eigenvalues on the diagonal, while the remaining elements are zero, i.e.:

$$T = \begin{pmatrix} \lambda_1, & 0 \\ 0, & \lambda_2 \end{pmatrix} \quad \Rightarrow \tag{1.252}$$

$$Z_N = \mathrm{Tr}\, T^N = \lambda_1^N + \lambda_2^N \,. \tag{1.253}$$

The eigenvalues are assumed to be ordered such that the first one is larger. Then, for large N, the following holds:

$$Z_N = \mathrm{Tr}\, T^N = \lambda_1^N \left[1 + (\lambda_2/\lambda_1)^N \right] \approx \lambda_1^N \,. \tag{1.254}$$

The key quantity is again the free energy, which is easily determined below:

$$\begin{aligned} F &= -k_\mathrm{B} T \ln \lambda_1^N = -N k_\mathrm{B} T \ln \lambda_1 \,; \\ f(n,T,H) &= -n k_\mathrm{B} T \ln \lambda_1(T,H) \,. \end{aligned} \tag{1.255}$$

Subsequently, additional quantities are determined using a standard procedure, particularly the magnetization $M = -\partial f/\partial H$, susceptibility $\chi = -\partial^2 f/\partial H^2$, heat capacity, etc. The magnetization increases from a zero value to a maximum (saturated) value, and its profile corresponds to a classical ferromagnet. The calculation of the matrix eigenvalues (1.250) is straightforward, yielding:

$$\lambda_{1,2} = \mathrm{e}^{\beta J} \cosh(\beta K H) \pm \sqrt{\mathrm{e}^{2\beta J} \cosh^2(\beta K H) - 2\sinh(2\beta J)} \,. \tag{1.256}$$

Potts Model

The Potts model represents a Q-state model with an interaction rule similar to the Ising model:

$$W_\mathrm{int} = -\sum_{\langle a,b \rangle} J \delta_{\sigma_a \sigma_b} \,; \qquad \sigma \in \{1, \ldots, Q\} \,. \tag{1.257}$$

Spins can orient into Q directions; adjacent identically oriented spins contribute $-J$ to the energy, while other configurations do not contribute. The model exhibits an ordered low-temperature phase featuring domains of identically oriented spins, and a high-temperature chaotic phase. Both phases are separated by a Curie phase transition occurring at the Curie temperature (T_C). The model is named after the Australian physicist Renfrey Potts.

Z_Q Model

The Z_Q model constitutes a Q-state model with an interaction given by the cosine of the mutual angle between spins:

$$W_\mathrm{int} = -\sum_{\langle a,b \rangle} J \cos(\alpha_a - \alpha_b) \,; \qquad \alpha = \frac{2\pi}{Q}\sigma \,; \qquad \sigma \in \{1, \ldots, Q\} \,. \tag{1.258}$$

For large values of Q, the model exhibits three phases: 1) a low-temperature ordered phase with typical domains; 2) a "soft" phase at intermediate temperatures, where adjacent spins differ very little in their orientation, forming characteristic vortices or spin waves; 3) a high-temperature chaotic phase. The phase transition from the low-temperature phase to the "soft" phase is called the Curie transition (T_C), whereas the phase

transition from the "soft" phase to the high-temperature phase is termed the Kosterlitz-Thouless transition (T_K). This transition features a continuous susceptibility. The loss of order during the transition from the "soft" to the chaotic phase occurs due to transverse fluctuations (Goldstone modes). The transition in the opposite direction (from the disordered to the "soft" phase) can be understood as a broken rotational symmetry, where Goldstone fluctuations appear. The transition is named after the American physicist John Michael Kosterlitz and the Scottish physicist David Thouless.

Heisenberg Model

The Heisenberg model represents a continuous model admitting all possible spin orientations. The energetic rule of the interaction is written as:

$$W_{\text{int}} = -\sum_{\langle a,b \rangle} J \cos(\alpha_a - \alpha_b); \qquad \alpha \in <0, 2\pi). \tag{1.259}$$

The three-dimensional variant is called the Heisenberg model, whereas the two-dimensional variant is termed the XY model. These models possess only two phases: the low-temperature "soft" phase is characterized by vortex pairs and spin waves, while the high-temperature phase features randomly oriented spins. Both phases are separated by the Kosterlitz-Thouless transition. In materials of this type, a phase with domains does not exist. Heisenberg magnets include, for example, the material designated as GSO with the chemical composition $Gd_2Sn_2O_7$.

Spin Glasses

Spin glasses represent materials where the coupling constant J varies from one spin pair to another, often exhibiting a random character. The interaction energy rule is given by:

$$W_{\text{int}} = -\sum_{\langle a,b \rangle} J_{ab}\, f(\sigma_a, \sigma_b) \ . \tag{1.260}$$

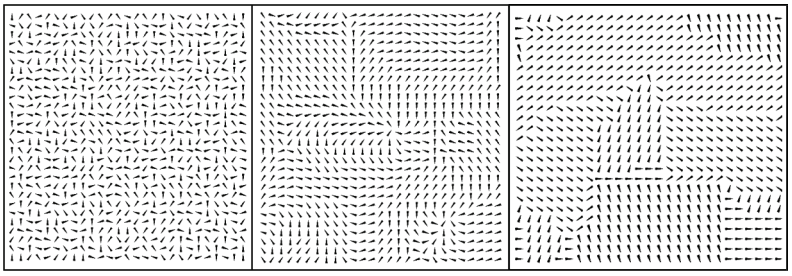


Fig. 1.27: Monte Carlo simulation of the Z_Q model using the Metropolis method. The high-temperature chaotic phase is shown on the left, the "soft" phase in the middle, and the low-temperature domains on the right.

In such materials, long-range correlations do not exist. Spin glasses are characterized by the existence of numerous metastable states, whose experimental investigation is extraordinarily difficult. Numerical simulations allow at least a framework study of these interesting materials. Spin glasses include, for example, a material with the chemical composition $LiHo_xY_{1-x}F_4$. For a high proportion of holmium, the material represents a ferromagnet, while for $x < 0.25$, it behaves as a spin glass. The magnetic moments of holmium interact almost exclusively via dipole interaction; the energy rule is analogous to the Ising model, but with a chaotic behavior of the interaction constant. Another

highly interesting material is the compound $PrAu_2Si_2$. Despite possessing a well-ordered crystalline structure, it exhibits the properties of spin glasses at low temperatures. The mechanism of origin is the so-called dynamic frustration – a phenomenon where dynamic fluctuations in the crystal prevent long-range correlations of magnetic moments. Frustration occurs because multiple incompatible ground states (with the lowest energy) exist, and individual force interactions compete to achieve one of them. The result is the freezing of the material into a spin glass state with chaotic coupling constants.

Hubbard Model

Even rapidly developing computer technology does not permit simulations of electron systems with many (e.g., 10^{23}) electrons. In such situations, the examined system is simplified to ensure it still describes the properties of the actual material under investigation. Electrons are assumed to be localized only at specific locations, such as vertices of a regular lattice. This scenario occurs, for instance, in crystalline materials where an electron is in the vicinity of a specific ion. Nevertheless, models of electrons on a lattice possess a much broader application and can be used generally to model systems with strongly correlated electrons, such as high-temperature superconductors.

The simplest models of this type include the Hubbard model. Electrons are created and annihilated at lattice vertices so that their behavior corresponds to the system energy at a given temperature. The system energy consists of two distinct terms:

$$W_{\text{int}} = -t \sum_{\langle a,b \rangle,\sigma} \left(\hat{\mathbf{a}}^{\dagger}_{a\sigma} \hat{\mathbf{a}}_{b\sigma} + \hat{\mathbf{a}}^{\dagger}_{b\sigma} \hat{\mathbf{a}}_{a\sigma} \right) + U \sum_{a} n_{a\uparrow} n_{a\downarrow} \; ; \quad n_{a\sigma} \equiv \hat{\mathbf{a}}^{\dagger}_{a\sigma} \hat{\mathbf{a}}_{a\sigma} \, . \tag{1.261}$$

The first term is given by the interaction of the nearest neighbors (the summation over all pairs of the nearest vertices). The coupling constant is denoted as t, the symbol $\mathbf{a}^{\dagger}$ denotes the creation operator of an electron with spin σ at a given vertex, and $\mathbf{a}$ represents the annihilation operator. This term enables the "hopping" or "tunneling" of electrons from one lattice vertex to another. The second part of the energy is a summation over all vertices, where the symbol $n_{a\sigma}$ denotes the number of individuals with spin σ at a given vertex. A positive coupling constant U implies electron repulsion at small distances. If one electron with spin ↑ and one electron with spin ↓ are present at a vertex, they contribute a positive value U to the total energy balance; if a single electron is present at a vertex, this vertex contributes a zero value. At low temperatures, states with the lowest possible energy are preferred, representing a single electron at a lattice vertex. Both energy terms imply a pair interaction in a certain sense. The first term concerns the interaction of electrons at the two nearest lattice vertices. The second concerns the interaction of a pair of electrons at a single vertex (Coulomb repulsion). In real materials, the ratio of coupling constants U/t ranges between 10 and 50.

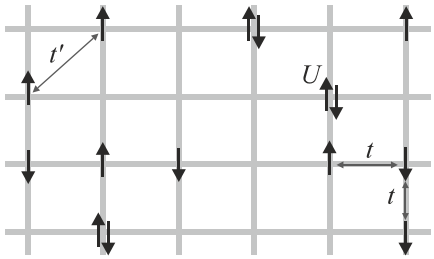


Fig. 1.28: Hubbard model.

The model was proposed by the English physicist John Hubbard in 1963 to describe electron behavior in solids. Using the Hubbard model, the transition of a material between conducting and non-conducting states is easily simulated. Currently, the model is utilized to describe the behavior of ultracold atoms trapped in an optical lattice (regularly alternating minima and maxima of electric potential created by the interference of two or more laser beams). The original model was proposed for two fermions; later, a boson variant of the Hubbard model and various other useful modifications appeared. Models do not need to be restricted to a rectangular lattice; interactions do not have to occur only between the nearest neighbors, but can also take place, for example, between diagonal neighbors with a coupling constant t', models in multiple dimensions are considered, etc.

t-J Model

In 1977, the Polish physicist Józef Spałek modified the Hubbard model for large values of the interaction constant U into a form with two pair interactions between the nearest lattice vertices:

$$W_{\text{int}} = -t \sum_{\langle a,b \rangle,\sigma} \left(\hat{\mathbf{c}}^{\dagger}_{a\sigma} \hat{\mathbf{c}}_{b\sigma} + \hat{\mathbf{c}}^{\dagger}_{b\sigma} \hat{\mathbf{c}}_{a\sigma} \right) + J \sum_{a} \left(\mathbf{S}_a \cdot \mathbf{S}_b - n_a n_b / 4 \right);$$

$$\hat{\mathbf{c}}_{a\sigma} \equiv \hat{\mathbf{a}}_{a\sigma} (1 - n_{a\sigma}); \qquad n_a = n_{a\uparrow} + n_{a\downarrow}; \qquad J = \frac{2t^2}{U}; \tag{1.262}$$

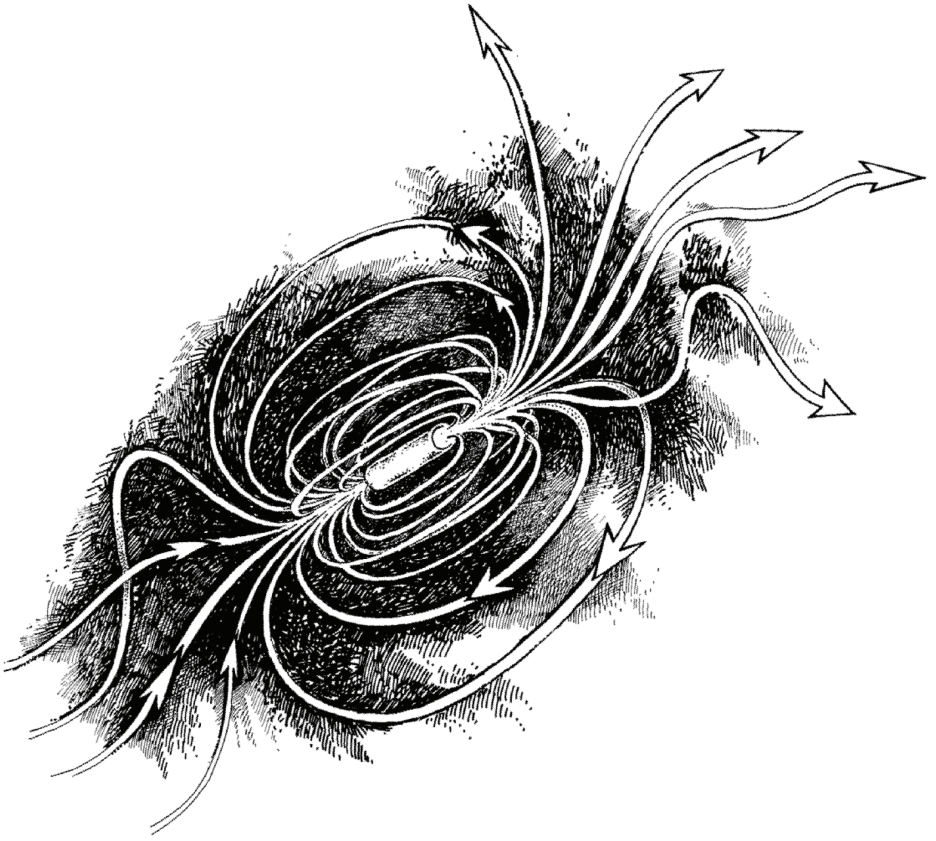
$$\mathbf{S}_a \equiv \left(\hat{\mathbf{a}}^{\dagger}_{a\uparrow} \hat{\mathbf{a}}_{b\downarrow}, \hat{\mathbf{a}}^{\dagger}_{a\downarrow} \hat{\mathbf{a}}_{b\uparrow}, (n_{a\uparrow} - n_{a\downarrow})/2 \right)$$

The model was named the t-J model according to the designation of the coupling constants of these interactions. The term with the coupling constant t is similar to the first term of the Hubbard model. The second term with the coupling constant $J = 2t^2/U$ contains the scalar product of two adjacent spins, analogously to the Heisenberg model. The operators $\mathbf{c}$ and $\mathbf{c}^{\dagger}$ appearing in the model are called projections of the annihilation and creation operators. Using this model, the behavior of Mott insulators, including their ferromagnetism, was successfully explained. Mott insulators represent non-conducting materials that should be conductors according to standard band theory. The non-conductivity is caused by pair electron-electron interactions, which band theory does not consider. This phenomenon occurs at low temperatures, for example, in NiO or CoO. Subsequently, the t-J model also became successful in explaining the high-temperature superconductivity of ceramic materials, such as $La_{2-x}Sr_xCuO_4$, discovered by Karl Alexander Müller and Johannes Georg Bednorz in 1986.

Further information regarding models of magnetic systems can be found in publication. Statistical physics represents an extraordinarily extensive area of physics, and its numerous applications are beyond the scope of this textbook.



2. Electromagnetic Field



2.1 Maxwell Equations

Each interaction possesses a characteristic charge determining the magnitude of the mutual interaction between bodies. For gravity, this "charge" is mass; for the electromagnetic interaction, it is electric charge; for the strong interaction, color charge; and for the weak interaction, flavor. Consequently, focus is first directed toward the electric charge of particles, and the differential form of the charge conservation law, known as the continuity equation, is expressed. This procedure applies to any conserved extensive quantity (extensive meaning increasing with the number of entities, such as mass, energy, momentum, angular momentum, heat, heat capacity, etc.). Prior to reading this chapter, a thorough understanding of field flux, circulation, and related vector operations – gradient, divergence, and curl – is required. Details are provided in the "*Selected Topics I*," Section I.3.5.

2.1.1 Charge Conservation Law

First, charge density and charge flux are introduced by the relations:

► $$\rho_Q \equiv \lim_{\Delta V \to 0} \frac{\Delta Q}{\Delta V} . \tag{2.1}$$

► $$\mathbf{j}_Q \equiv \rho_Q \mathbf{u} . \tag{2.2}$$

In the SI system, charge density has the unit C/m^3, and charge flux has the unit A/m^2. Charge flux is a vector quantity possessing a magnitude and a direction. The direction corresponds to the velocity field describing the charge flow. The magnitude is determined by monitoring the charge flow through a perpendicular surface. During a time interval dt, particles move a distance d$l = u\,\mathrm{d}t$ toward another imaginary surface:

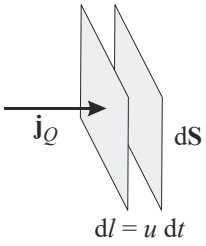


Fig. 2.1: Flux of an extensive quantity through a surface.

The magnitude of the flux is given by:

$$j_Q = \rho_Q u = \frac{\mathrm{d}Q}{\mathrm{d}V}\frac{\mathrm{d}l}{\mathrm{d}t} = \frac{\mathrm{d}Q}{\mathrm{d}S\,\mathrm{d}l}\frac{\mathrm{d}l}{\mathrm{d}t} = \frac{\mathrm{d}Q}{\mathrm{d}S\,\mathrm{d}t} . \tag{2.3}$$

Charge flux therefore represents the amount of charge flowing through a perpendicular unit area per unit time. An alternative name for this quantity is *current density*. Charge

is neither lost nor created during the flow; its temporal decrease within any arbitrary volume must equal the flux of the quantity across the surface enclosing this volume:

$$-\frac{\mathrm{d}}{\mathrm{d}t}\iiint_V \rho_Q \,\mathrm{d}V = \oiint_{\partial V} \mathbf{j}_Q \cdot \mathrm{d}\mathbf{S}\,. \tag{2.4}$$

The boundary of the volume V is denoted by ∂V. On the left-hand side, the time derivative is moved inside the integral; however, since ρ_Q is a function of both time and space, the derivative becomes a partial derivative. On the right-hand side, the Gauss theorem is applied, transforming the surface integral into a volume integral:

$$-\iiint_V \frac{\partial \rho_Q}{\partial t}\mathrm{d}V = \iiint_V \operatorname{div}\mathbf{j}_Q \,\mathrm{d}V\,. \tag{2.5}$$

Both integrals are now combined:

$$\iiint_V \left(\frac{\partial \rho_Q}{\partial t} + \operatorname{div}\mathbf{j}_Q\right)\mathrm{d}V = 0. \tag{2.6}$$

This relation must hold for any arbitrary volume during flow, which is possible only if the integrand is zero (in principle, it could be non-zero only at certain points or surfaces, generally on a set of a lower dimension than the integration domain):

► $$\frac{\partial \rho_Q}{\partial t} + \operatorname{div}\mathbf{j}_Q = 0\,. \tag{2.7}$$

The derived equation constitutes the law of conservation of charge in differential form and is called the *continuity equation for charge*. It demonstrates that the set of four quantities describing the flow of charge (ρ_Q, $\mathbf{j}_Q$) is not independent; these four functions satisfy the continuity equation at every point in space and are mutually dependent. In the stationary case, the charge conservation law takes the form $\operatorname{div}\mathbf{j}_Q = 0$.

Note 1: A similar continuity equation holds for other conservation laws. A detailed discussion of the differential form of the conservation laws for energy, momentum, and angular momentum for the electromagnetic field is presented in Section 2.4.

Note 2: A distinction is made between extensive or additive quantities, increasing with the number of particles (charge, heat, energy, momentum...), and intensive quantities, independent of particle amount (electric field, magnetic field, temperature...). Flux can be introduced for both types. For additive quantities, flux equals the product of density and flow velocity, e.g., charge flux $\mathbf{j}_Q \equiv \rho_Q \mathbf{u}$. For vector intensive quantities, flux is the surface integral of the second kind, e.g., magnetic field flux $\psi_\mathrm{B} \equiv \iint \mathbf{B}\cdot\mathrm{d}\mathbf{S}$.

2.1.2 Lorentz Equation of Motion

Observation of charged particle motion revealed that the electric field accelerates particles, whereas the magnetic field exerts a force perpendicular to both itself and the particle velocity. Results of numerous experiments can be summarized by a simple formula:

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = Q\mathbf{E} + Q\mathbf{v}\times\mathbf{B}\,, \tag{2.8}$$

which was previously encountered in theoretical mechanics. The temporal change in particle momentum consists of an electric part and a magnetic part. The magnetic part features a cross product, describing the fact that the force is perpendicular to both the velocity and the magnetic field, and that the magnetic field exerts no force on a stationary particle. The magnetic force causes particles to helix along the magnetic field lines. Particle trajectories can be predicted by means of this equation when the fields are known. The equation can also serve as the defining equation, since both fields **E** and **B**. can be determined (preferably separately) from the force acting on the particle.

Interest lies not only in the reaction of particles to given fields, but also in the reverse: What fields are generated by moving particles? A definitive answer to this question is provided by the system of Maxwell equations. However, prior to analyzing them, familiarity with two concepts is required: electric and magnetic dipole moments.

2.1.3 Electric and Magnetic Dipole

The concepts of the electric and magnetic dipole are essential for understanding the interaction of matter with the electromagnetic field. At this point, their simplest definition is introduced, which is later generalized for any system of charged particles.

Electric Dipole

The fact that atoms and molecules are electrically neutral does not mean that they do not interact electrically. They usually contain regions with both positive and negative charge. The simplest approximation of their behavior is replacement with an electric dipole: equal positive and negative charges separated by a certain distance.

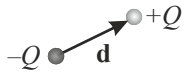


Fig. 2.2: Electric dipole.

Such an object is electrically neutral as a whole, yet it still responds to an electric field. The electric dipole moment is defined as the product of the charge magnitude and the vector connecting the negative charge to the positive charge:

► $$\mathbf{p}_\mathrm{E} \equiv Q\,\mathbf{d}\,; \qquad [\mathbf{p}_\mathrm{E}] = \mathrm{Cm}\,. \tag{2.9}$$

Consider matter consisting of a large number of such elementary electric dipoles (for instance, polar water molecules). It is useful to introduce the electric dipole moment density of all such dipoles:

► $$\mathbf{P} \equiv \lim_{\Delta V\to 0} \frac{\Delta\mathbf{p}_\mathrm{E}}{\Delta V}\,; \qquad [\mathbf{P}] = \frac{\mathrm{C}}{\mathrm{m}^2}\,, \tag{2.10}$$

where $\Delta\mathbf{p}_\mathrm{E}$ is the total dipole moment of all entities within a unit volume. The quantity **P** is called the *polarization* of the material, and it is evident that if the dipoles are oriented randomly, the resulting polarization is zero. However, when an electric field is applied, the dipoles tend to align in the direction of the field, and the resulting polariza-

tion becomes non-zero. The dimension of polarization is easily determined to be C/m^2, which dimensionally corresponds to a surface charge density.

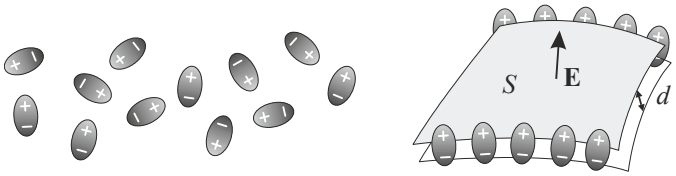


Fig. 2.3: System of electric dipoles under the influence of a field.

Indeed, under the influence of a field, a non-zero surface electric charge appears on the boundary of the dipole-filled region (such as a dielectric) due to the dipole alignment. Consider originally neutral atoms polarized by an applied electric field, meaning the positive charge shifts in the field direction and the negative charge against it, creating electric dipoles with a separation distance d between the two charges. The polarization (the density of dipole moments) in the layer adjacent to the surface (the figure on the right) is determined as follows. The magnitude of the electric dipole moment density is:

$$P = \frac{NQd}{\Delta V} = \frac{NQd}{Sd} = \frac{NQ}{S} \tag{2.11}$$

Thus, for a field perpendicular to the surface, the magnitude of the polarization vector equals the surface density of the bound charge σ_b (the subscript b stands for "*bound*") induced on the surface of the dielectric:

$$P = \sigma_b . \tag{2.12}$$

In this case, the polarization vector points in the direction of the electric field applied to the dielectric. If the electric field does not act perpendicularly to the surface of the dielectric but at an angle, the dipoles align in the direction of the field, and the boundary layer has a thickness of $d \cos \alpha$ is the deviation of the electric field, and thus of the dipoles, from the surface normal. This results in the relation $P \cos \alpha = \sigma_b$; hence, only the normal component of the polarization vector is effective, $P_n = \mathbf{P} \cdot \mathbf{n} = P \cos \alpha$:

► $$\mathbf{P} \cdot \mathbf{n} = \sigma_b . \tag{2.13}$$

Magnetic Dipole

Observation reveals that the most natural motion of charged particles in a magnetic field is circling. However, a circling particle generates a magnetic field. If the loop is small, a magnetic dipole field is created. Such a circling particle is described by a quantity called the magnetic dipole moment, which is defined as the product of the area and the electric current created by the orbiting particle:

► $$\mathbf{p}_M \equiv I\,\mathbf{S}\,; \qquad [\mathbf{p}_M] = \mathrm{A\,m^2} . \tag{2.14}$$

The magnetic moment points in the direction of the normal to the surface orbited by the charged particle. Later, the magnetic moment will be defined for a general system of moving charged particles. If a large number of such elementary magnetic moments are present, the magnetic moment density can be introduced by the relation:

► $$\mathbf{M} \equiv \lim_{\Delta V \to 0} \frac{\Delta \mathbf{p}_\mathrm{M}}{\Delta V}; \qquad [\mathbf{M}] = \frac{\mathrm{A}}{\mathrm{m}}. \tag{2.15}$$

This quantity is called magnetization, and its dimension is A/m. The same dimension belongs to the current per unit length (for example, flowing across a curve). This is not a coincidence. If no magnetic field acts on a system of magnetic dipoles, they can be oriented completely randomly, and the resulting magnetization is zero.

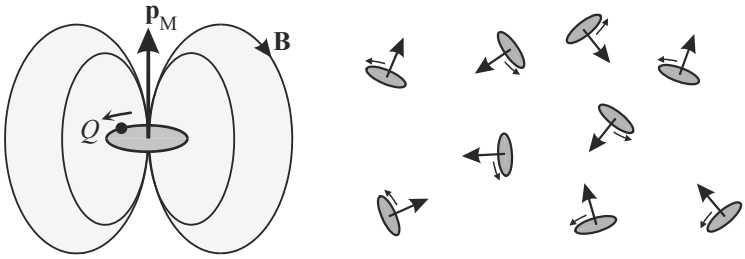


Fig. 2.4: Magnetic dipoles.

Consider a magnetically active material to which a magnetic field pointing along the surface is applied. Charged particles begin to execute a circular motion in the magnetic field, as shown in the figure. Inside the material, the electric current resulting from this circular motion cancels out; however, a non-zero surface current (magnetization current) remains on the surface. Individual circling particles are considered elementary dipoles with a magnetic moment *IS*. Near the surface of the magnetic material, a single cylinder of circling particles with volume *Sl* is selected, and the magnitude of magnetization in this surface cylinder is determined:

$$M \equiv \frac{N I S}{\Delta V} = \frac{N I S}{S l} = \frac{N I}{l}.$$

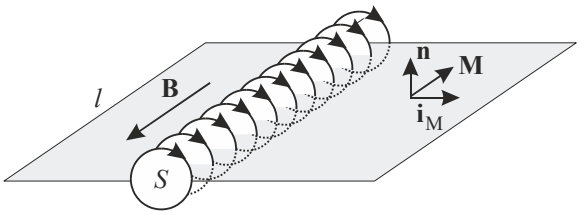


Fig. 2.5: Surface current caused by magnetic dipoles.

Thus, near the surface of the magnetic material, magnetization equals the total surface magnetization current per unit transverse length. Such a quantity is denoted as *i*:

$$M = i_\mathrm{M}. \tag{2.16}$$

In bulk bodies, the quantity of interest is the current flowing per unit area, whereas on the surface of bodies, the current flowing per unit transverse length is important. Considering the directions of the vectors (the magnetic dipole moment is perpendicular to both the flowing current and the surface normal), the following vector relation holds:

$$\mathbf{M} \times \mathbf{n} = \mathbf{i}_\mathrm{M}. \tag{2.17}$$

This relation holds even for a general direction of the magnetic field and material magnetization. If the magnetization points obliquely to the surface, only the tangential component, equaling $M \sin \alpha$ (α being the angle between magnetization and the surface normal), contributes to the current generation. Therefore, a cross product is present in relation (2.17), introducing the sine of both vectors. For freely circling particles, the direction of magnetization is opposite to the magnetic field (as occurs, for example, in plasma). This phenomenon is referred to as *diamagnetic* behavior.

2.1.4 Maxwell Equations

Everything is now prepared to describe individual equations from the Maxwell system published in 1873. The modern form was given to the equations by Oliver Heaviside. Maxwell wrote the equations in component notation and, moreover, utilized quaternions – a generalization of complex numbers into four dimensions. The equations provide a description of experimentally acquired experience, including the Coulomb law, Ampère law, and Faraday induction law.

In the passage dedicated to the charge conservation law, the flux of an extensive quantity was introduced, specifically the charge flux $\mathbf{j}_Q$. Similarly, mass flux, energy flux, heat flux, etc., can be defined. Flux is always defined as the product of the density of the respective quantity and its flow velocity. For vector intensive quantities (electric field, magnetic field, velocity field, etc.), the flux is defined as the surface integral:

$$\psi_K \equiv \iint_S \mathbf{K} \cdot \mathrm{d}\mathbf{S} \,. \tag{2.18}$$

If the respective field points in the direction of the surface normal, the field flux is maximal; if it points along the surface, the flux through the surface is zero. The respective surface may or may not be closed. Both magnetic field flux and electric field flux will be useful. The second important quantity is the line integral, in which the field is summed along a curve:

$$C_K \equiv \int_\gamma \mathbf{K} \cdot \mathrm{d}\mathbf{l} \,. \tag{2.19}$$

If the field is a force, the mechanical work performed by the force along the given path is obtained. In the case of an electric field, the voltage between the initial and end points of the curve results. The respective curve may or may not be closed. If it is closed, integral (2.19) is called the field circulation $\mathbf{K}$. Field fluxes and circulations are key concepts in the system of Maxwell equations that will be encountered later.

Gauss Law of Magnetostatic Fields

Experiments reveal that the magnetic field only passes through all points. It possesses no sources anywhere; magnetic monopoles from which it would emanate do not exist. They might have existed in the early phases of the universe, but during the inflationary phase, the universe expanded so much that only a few monopoles occur in the observable universe, offering no chance of detection. If any macroscopic volume is singled out in space, the flux of magnetic field lines through the enclosing surface is always zero (those entering the volume also leave it):

$$\oiint_{S=\partial V} \mathbf{B} \cdot \mathrm{d}\mathbf{S} = 0 \,. \tag{2.20}$$

Application of the Gauss theorem yields:

$$\iiint_V \operatorname{div}\mathbf{B}\,\mathrm{d}V = 0\,. \tag{2.21}$$

Since the relation holds for any arbitrary volume, the integrand must be zero (mathematically rigorous language would state "zero except for sets of a lower measure or dimension than the integration domain," but these subtleties will not be addressed or pointed out in subsequent cases):

► $$\operatorname{div}\mathbf{B} = 0\,. \tag{2.22}$$

This first Maxwell equation expresses the sourceless nature of the magnetic field.

Coulomb Law

Introductory physics courses demonstrate that a charge creates an electric field:

$$\mathbf{E} = \frac{Q}{4\pi\varepsilon_0 r^2}\frac{\mathbf{r}}{r}\,. \tag{2.23}$$

The field flux remains the same for all enclosing surfaces (see introductory mathematics or physics courses); therefore, when calculating the electric field flux, it suffices to construct a spherical surface with radius r around the charge:

$$\oiint_S \mathbf{E}\cdot\mathrm{d}\mathbf{S} = \oiint_S E\,\mathrm{d}S = E\oiint_S \mathrm{d}S = E4\pi r^2 = \frac{Q}{4\pi\varepsilon_0 r^2}4\pi r^2 = \frac{Q}{\varepsilon_0}\,. \tag{2.24}$$

This construction can be generalized to any charge, i.e., the electric field flux through a closed surface must be proportional to the free charge Q plus the bound charge Q_b located within the volume enclosed by this surface:

$$\oiint_S \mathbf{E}\cdot\mathrm{d}\mathbf{S} = \frac{Q+Q_\mathrm{b}}{\varepsilon_0}\,. \tag{2.25}$$

In a material medium, such as a dielectric, polarization occurs. Inside the material, the dipole charges cancel out, but near its boundary, they do not. According to (2.13), a surface charge $\mathbf{P}\cdot\mathbf{n}$ is created on the enclosing surface, and $-\mathbf{P}\cdot\mathbf{n}$ appears below the surface. The total bound charge in the considered region is:

$$Q_\mathrm{b} = -\oiint_S \mathbf{P}\cdot\mathbf{n}\cdot\mathrm{d}S = -\oiint_S \mathbf{P}\cdot\mathrm{d}\mathbf{S}\,. \tag{2.26}$$

Substitution into (2.25) yields:

$$\oiint_S \varepsilon_0\mathbf{E}\cdot\mathrm{d}\mathbf{S} = Q - \oiint_S \mathbf{P}\cdot\mathrm{d}\mathbf{S}\,. \tag{2.27}$$

Moving the integral from the right-hand side to the left yields:

$$\oiint_S (\varepsilon_0\mathbf{E}+\mathbf{P})\cdot\mathrm{d}\mathbf{S} = Q\,, \tag{2.28}$$

$$\oiint_S \mathbf{D}\cdot\mathrm{d}\mathbf{S} = Q\,; \qquad \mathbf{D} \equiv \varepsilon_0\mathbf{E}+\mathbf{P}(\mathbf{E})\,. \tag{2.29}$$

The newly introduced vector **D** is called the electric displacement, incorporating the material response to the electric field (described by the polarization vector **P**, representing the density of the induced electric dipoles). The polarization vector is a function of the applied field **E**. The charge on the right-hand side is the free charge enclosed by the integration surface. Relation (2.29) is the integral form of another Maxwell equation. To obtain the differential form, the Gauss theorem is applied on the left-hand side (the integral will be over the volume enclosed by the integration surface), and the right-hand side is written as a volume integral of the electric charge density:

$$\iiint_V \operatorname{div} \mathbf{D}\, \mathrm{d}V = \iiint_V \rho_Q\, \mathrm{d}V. \tag{2.30}$$

This relation must hold for any volume; therefore, the integrands must be identical, i.e.:

► $$\operatorname{div} \mathbf{D} = \rho_Q\,. \tag{2.31}$$

This differential form of the *Gauss law of electrostatic fields* (the generalized Coulomb law) states that the electric displacement **D** emanates from positive charges and sinks into regions of negative charges. The free charge density appears on the right-hand side.

Faraday Law of Induction

During extensive experiments with electricity and magnetism, English physicist Michael Faraday discovered a noteworthy phenomenon. If a time-varying magnetic flux passed through a conducting loop, an electric voltage was generated, subsequently causing an electric current to flow. Crucially, the magnetic flux must be time-varying.

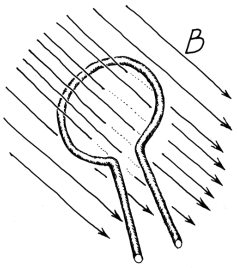


Fig. 2.6: Faraday law of induction.

This is achieved by rotating the loop, rotating the magnet, or by moving the magnet and the loop closer together or further apart. At the microscopic level, the electric current is naturally generated by the Lorentz force acting on the charged particles. The Faraday induction law can be written very simply:

$$U = -\frac{\mathrm{d}\psi_B}{\mathrm{d}t}\,. \tag{2.32}$$

The electric voltage equals the temporal change in the magnetic flux through the loop. The minus sign symbolizes that the generated voltage and the current generated by it act against the change that caused it (Lenz rule). Denoting the loop area by the symbol S and the boundary curve by the symbol $\gamma = \partial S$, the Faraday induction law can easily be rewritten into the form:

$$\oint_{\gamma=\partial S} \mathbf{E}\cdot\mathrm{d}\mathbf{l} = -\frac{\mathrm{d}}{\mathrm{d}t}\iint_S \mathbf{B}\cdot\mathrm{d}\mathbf{S}\,. \tag{2.33}$$

If the time derivative is moved to the right-hand side into the integrand argument, a partial derivative must be used, as the magnetic field is a function of both time and space, and a choice among multiple variables is available during differentiation:

$$\oint_{\gamma=\partial S} \mathbf{E}\cdot\mathrm{d}\mathbf{l} = -\iint_S \frac{\partial\mathbf{B}}{\partial t}\cdot\mathrm{d}\mathbf{S}\,. \tag{2.34}$$

The Faraday induction law in form (2.34) is the integral form of another Maxwell equation. Again, interest lies in the differential form of this equation. Therefore, the integral on the left-hand side is converted into a surface integral using the Stokes theorem:

$$\iint_S \operatorname{curl}\mathbf{E}\cdot\mathrm{d}\mathbf{S} = -\iint_S \frac{\partial\mathbf{B}}{\partial t}\cdot\mathrm{d}\mathbf{S}\,. \tag{2.35}$$

Both integrals must be equal for any chosen surface; therefore, the integrands must be identical, leading to the differential form of another Maxwell equation – the Faraday induction law:

► $$\operatorname{curl}\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}\,. \tag{2.36}$$

Ampere Law

A conductor carrying an electric current generates a magnetic field around itself. The magnitude of this field is directly proportional to the flowing current and inversely proportional to the distance from the conductor:

$$B \sim \frac{I}{r}\,. \tag{2.37}$$

In the SI system of units, this proportionality is expressed as:

$$B = \frac{\mu_0 I}{2\pi r}\,, \tag{2.38}$$

which can be understood as the defining relation for the permeability of vacuum μ_0. Integration of the field along any closed field line at a distance R from the conductor (along a circle forming the boundary of the surface S, where $\gamma = \partial S$; the curve symbol must not be confused with differential conductivity) yields:

$$\oint_{\gamma=\partial S} \mathbf{B}\cdot\mathrm{d}\mathbf{l} = \oint_{\gamma=\partial S} B\,\mathrm{d}l = B\oint_{\gamma=\partial S}\mathrm{d}l = \frac{\mu_0 I}{2\pi R}\,2\pi R = \mu_0 I\,.$$

This result holds for any surface and its boundary curve, not just for a field line. Moving the curve further from the conductor increases its length proportionally, while the magnetic field decreases inversely. These opposing dependencies compensate for each other, leaving the integral unchanged. Deforming the curve does not affect the outcome, and the result remains identical. The final Ampere law is therefore:

$$\oint_{\gamma=\partial S} \mathbf{B}\cdot\mathrm{d}\mathbf{l} = \mu_0 I_C\,. \tag{2.39}$$

The *conduction current* I_C is responsible for generating the magnetic field **B**. The line integral of the magnetic field is proportional to the conduction current passing through the surface bounded by the curve. The coefficient of proportionality is the vacuum permeability. This law applies only to stationary situations in non-magnetic media.

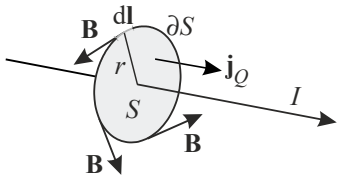


Fig. 2.7: Line integral of magnetic field.

In a general medium, particles along the boundary curve circulate, creating secondary currents known as *magnetization currents*. In the resulting law, it is necessary to write:

$$\oint_{\gamma=\partial S} \mathbf{B}\cdot\mathrm{d}\mathbf{l} = \mu_0(I_C + I_\mathrm{M})\,. \tag{2.40}$$

From relations (2.16) and (2.17), the magnetization current per unit transverse length is numerically equal to the magnitude of magnetization. Regarding its direction, it points perpendicular to the magnetization (which, in simple media, is parallel to the magnetic field). In this case, the magnetization current is given by the integration:

$$I_\mathrm{M} = \oint_{\gamma=\partial S} i_Q\,\mathrm{d}l = \oint_{\gamma=\partial S} M\,\mathrm{d}l = \oint_{\gamma=\partial S} \mathbf{M}\cdot\mathrm{d}\mathbf{l}\,. \tag{2.41}$$

For circulating particles, the magnetization **M** points opposite to the applied or generated field **B**, in whose direction the boundary curve γ points; therefore, the dot product in the last term is negative. The vector directions are apparent from the figure.

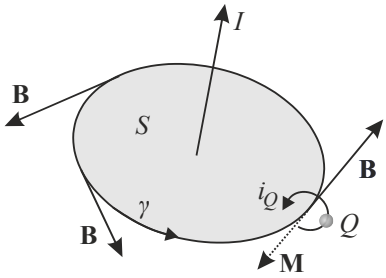


Fig. 2.8: Vector directions; magnetization is oriented opposite to magnetic field.

Now the magnetization current (2.41) is substituted into the Ampere law (2.40):

$$\frac{1}{\mu_0}\oint_{\gamma=\partial S} \mathbf{B}\cdot\mathrm{d}\mathbf{l} = I_\mathrm{C} + \oint_{\gamma=\partial S} \mathbf{M}\cdot\mathrm{d}\mathbf{l}\,.$$

Next, both line integrals are combined into one by moving the line integral of magnetization from the right-hand side to the left-hand side of the equation:

$$\oint_{\gamma=\partial S}\left(\frac{\mathbf{B}}{\mu_0}-\mathbf{M}\right)\cdot\mathrm{d}\mathbf{l}=I_C\,.$$

The Ampere law supplemented by the material response therefore takes the form:

$$\oint_{\gamma=\partial S}\mathbf{H}\cdot\mathrm{d}\mathbf{l}=I_\mathrm{C}\,;\qquad \mathbf{H}\equiv\frac{\mathbf{B}}{\mu_0}-\mathbf{M}(\mathbf{B})\,. \tag{2.42}$$

The second relation defines the magnetic field intensity, which incorporates both the original field **B** and the material response **M**. The current on the right-hand side is the conduction current flowing through the surface enclosed by the bounding curve γ. The grouping of individual terms is historically different from that of electric fields (2.29). The equation is now transformed into its differential form. On the left-hand side, the Stokes theorem is applied, and on the right-hand side, the current is expressed as the flux of free charge through the surface bounded by the chosen curve:

$$\iint_S \mathrm{curl}\,\mathbf{H}\cdot\mathrm{d}\mathbf{S}=\iint_S\mathbf{j}_Q\cdot\mathrm{d}\mathbf{S}\,. \tag{2.43}$$

Since these relations must hold for any surface, the integrands must be identical:

$$\mathrm{curl}\,\mathbf{H}=\mathbf{j}_Q\,. \tag{2.44}$$

On the right-hand side is the conduction current density passing through the surface. The Ampere law written in this manner applies only to steady-state electric currents. This can easily be demonstrated by taking the divergence of the last equation. The left-hand side yields zero (since the divergence of a curl is always zero), meaning that:

$$\mathrm{div}\,\mathbf{j}_Q=0\,,$$

which is the stationary form of the charge conservation law (2.7). For non-stationary processes, this law would be violated. Therefore, James C. Maxwell added another term to the equation (the so-called Maxwell displacement current) to ensure validity:

► $$\mathrm{curl}\,\mathbf{H}=\mathbf{j}_Q+\frac{\partial\mathbf{D}}{\partial t}\,. \tag{2.45}$$

This can easily be verified. Applying the divergence yields:

$$0=\mathrm{div}\,\mathbf{j}_Q+\mathrm{div}\left(\frac{\partial\mathbf{D}}{\partial t}\right)=\mathrm{div}\,\mathbf{j}_Q+\frac{\partial}{\partial t}\mathrm{div}\,\mathbf{D}\,.$$

Substituting div **D** = ρ_Q from the Maxwell equation (2.31) yields:

$$0=\mathrm{div}\,\mathbf{j}_Q+\frac{\partial\rho_Q}{\partial t}\,,$$

which is precisely the charge conservation law. Equation (2.45) states that a curl of the magnetic field can arise either near a current-carrying conductor or due to a time-varying electric field. Such a situation occurs, for example, around a parallel-plate ca-

pacitor (assuming an empty gap between the plates). If an alternating current flows through a conductor, it generates a magnetic field around it (the direction of the curls alternates in accordance with the instantaneous direction of the electric current). A positive charge is delivered to the capacitor plates during certain phases, and a negative charge during others. This establishes a time-varying electric field between the plates, acting as the source of the magnetic field around them. A distant observer cannot tell from the magnetic field alone that the circuit is interrupted by the capacitor plates.

2.1.5 System Completeness

The entire set of Maxwell equations represents a mathematical description of experimental results (magnetostatics, Coulomb law, Ampere law, Faraday law of electromagnetic induction). It constitutes a generalized experience gained over decades of experiments with electric and magnetic phenomena. James Clerk Maxwell wrote his system in components and, as previously mentioned, used the quaternion formalism, making the Maxwell system in its original notation difficult to read for a modern reader. For current purposes, interest lies in the differential form of the equations. All the obtained equations are compiled into a single whole:

► $$\operatorname{div}\mathbf{B} = 0, \tag{2.46}$$

► $$\operatorname{div}\mathbf{D} = \rho_Q, \tag{2.47}$$

► $$\operatorname{curl}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{2.48}$$

► $$\operatorname{curl}\mathbf{H} = \mathbf{j}_Q + \frac{\partial \mathbf{D}}{\partial t}. \tag{2.49}$$

Ordering the equations holds little significance; justification can always be found for any pairing. The first two equations are tests for field sources, while the second two test for field curls. Furthermore, the latter contain time derivatives, governing temporal evolution. However, grouping can differ: the first and third contain no field sources, whereas the second and fourth contain them (charge density and flux). Alternatively, it can be argued that the first and fourth belong together as magnetic field equations, while the second and third form the electric field equations.

Regardless of grouping, the input consists of four quantities ρ_Q and $\mathbf{j}_Q$ describing charge distribution and motion, while the output comprises electric and magnetic fields **E**, **D**, **H**, **B**. An inconsistency appears: 12 unknowns (four vectors) must be determined from only 8 equations (two scalar and two vector). A successful solution requires supplementary information regarding material response to applied fields, i.e., constitutive relations between both electric and both magnetic vectors:

► $$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}(\mathbf{E}); \tag{2.50}$$

► $$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M}(\mathbf{B}). \tag{2.51}$$

The polarization vector **P** and the magnetization vector **M** describe how the material responds to the applied fields, i.e., whether and in what quantity electric and magnetic

dipoles form. These *constitutive relations* can be obtained either experimentally or through theoretical calculation of internal processes. Note that the grouping of terms in both relations differs due to historical reasons. Both relations are usually highly nonlinear; polarization and magnetization are typically complex functions of the field intensities **E** and **H**. The material can even "remember history," meaning the relations can be affected by past states of the system. Consider, for example, the hysteresis loop of magnetically active materials. Only for very weak fields can both relations be Taylor-expanded to obtain a linear dependence between vectors **D** and **E**, or **B** and **H**.

The constitutive relations introduce six additional equations into the Maxwell system, increasing the total number of equations to 14. Since 12 unknowns are present, two equations in the system must be redundant. It can now be shown that the first two equations for the divergences of the fields are actually only initial conditions for the second two equations, which determine the temporal evolution of the fields. The divergence operation is applied to the third equation of system (2.48):

$$0 = -\operatorname{div}\frac{\partial \mathbf{B}}{\partial t} \quad \Rightarrow \quad \frac{\partial}{\partial t}\operatorname{div}\mathbf{B} = 0 \quad \Rightarrow \quad \operatorname{div}\mathbf{B} = f(\mathbf{r}).$$

Evidently, the Faraday induction law implies that div **B** cannot be a function of time, but only a function of spatial variables. The value of div **B** is therefore identical to its initial value. The equation div **B** states that this function was zero initially, representing an initial condition for equation (2.48). A similar situation occurs with the second equation. The divergence operation is applied to the Ampere law (2.49):

$$0 = \operatorname{div}\mathbf{j}_Q + \operatorname{div}\frac{\partial \mathbf{D}}{\partial t}.$$

The first term on the right-hand side is expressed via the charge conservation law (2.7):

$$0 = -\frac{\partial \rho_Q}{\partial t} + \operatorname{div}\frac{\partial \mathbf{D}}{\partial t} \quad \Rightarrow \quad \frac{\partial}{\partial t}\left(-\rho_Q + \operatorname{div}\mathbf{D}\right) = 0 \quad \Rightarrow \quad \operatorname{div}\mathbf{D} - \rho_Q = g(\mathbf{r}).$$

The combination div **D** $-\rho_Q$ is solely a spatial function and does not depend on time. Equation (2.47) merely states that this function is zero, meaning it was also zero initially. Equation (2.47) thus has the character of an initial condition for the Ampere law. Consequently, the result consists of six equations for the temporal evolution of fields (2.48), (2.49) and six constitutive relations, yielding 12 equations for 12 unknowns.

2.1.6 Constitutive Relations

The previous section demonstrated that the Maxwell equations must be supplemented by constitutive relations describing the material response to applied fields:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}(\mathbf{E}); \tag{2.52}$$

$$\mathbf{H} = \frac{\mathbf{B}}{\mu_0} - \mathbf{M}(\mathbf{B}). \tag{2.53}$$

Fields **E** and **B** are supplemented by material responses **P**(**E**) and **M**(**B**), ensuring that only conduction currents and free charges act as source terms in the Maxwell equations.

New fields incorporating the material response are denoted **D** and **H**. The fields **E** and **H** are called intensities, while the fields **D** and **B** are called displacements or inductions.

The material responses **P**(**E**) and **M**(**B**) can be calculated theoretically in exceptional cases, as shown later using the example of an electromagnetic wave passing through a plasma. Under the influence of the wave, electrons begin to oscillate, generating a volume density of the electric dipole moment (a non-zero material polarization). In most cases, the material response to applied fields is determined experimentally. The dependencies **P**(**E**) and **M**(**B**) are generally highly non-linear, and in certain cases, they do not depend solely on the currently applied fields, but also on the previous history of the material. This category of phenomena includes, for example, various residual fields resulting from past field exposure. In such cases, polarization and magnetization depend not only on the instantaneous field values, but also on their temporal changes, i.e.:

$$\mathbf{P} = \mathbf{P}(\mathbf{E}, \dot{\mathbf{E}}, \ddot{\mathbf{E}}, \dddot{\mathbf{E}}, \ldots) , \tag{2.54}$$

$$\mathbf{M} = \mathbf{M}(\mathbf{B}, \dot{\mathbf{B}}, \ddot{\mathbf{B}}, \dddot{\mathbf{B}}, \ldots) . \tag{2.55}$$

For simplicity, assuming only a linear material response throughout the entire history and a one-dimensional case:

$$M = \int_{-\infty}^{t} K(t-t')B(t')\,\mathrm{d}t' . \tag{2.56}$$

The function K determines how far into the past the material remembers the field exposure. The substitution $\tau = t - t'$ is performed, where the parameter τ is the time elapsed from the past to the present:

$$M = \int_{0}^{\infty} K(\tau)B(t-\tau)\,\mathrm{d}\tau . \tag{2.57}$$

Now the Taylor expansion is performed:

$$B(t-\tau) = \sum_{k=0}^{\infty} \frac{B^{(k)}(t)}{k!}\left(-\tau\right)^{k} = B(t) - \tau\frac{\mathrm{d}B}{\mathrm{d}t} + \frac{\tau^2}{2}\frac{\mathrm{d}^2B}{\mathrm{d}t^2} \pm \cdots \tag{2.58}$$

After substitution into integration (2.57), the magnetization takes the form:

$$M(t) = c_1 B(t) + c_2 \dot{B}(t) + c_3 \ddot{B}(t) + \cdots \tag{2.59}$$

If the material remembers history, magnetization is a function of both the field and its time derivatives, even in the linear case. The coefficients c_k represent the results of integrations over τ, reflecting the specific material properties and memory attenuation. Consider further a simple material that does not remember history, i.e., **P** = **P**(**E**). For weak fields, the system response can be considered linear, allowing a first-order Taylor expansion of the polarization vector (the summation convention is used):

$$P_k \doteq \frac{\partial P_k}{\partial E_l} E_l \equiv \varepsilon_0 \kappa_{kl}\ E_l \ ; \tag{2.60}$$

► $$\kappa_{kl} \equiv \frac{1}{\varepsilon_0}\frac{\partial P_k}{\partial E_l} \tag{2.61}$$

The matrix κ_{kl} is called the *electric susceptibility tensor*. The vacuum permittivity in the definition ensures the dimensionless nature of susceptibility. In vector notation:

$$\mathbf{P} = \varepsilon_0 \, \ddot{\boldsymbol{\kappa}} \cdot \mathbf{E} \, . \tag{2.62}$$

In the limit of weak fields, the relation $\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}$ yields:

$$D_k = \varepsilon_{kl} \, E_l \, ; \qquad \varepsilon_{kl} \equiv \varepsilon_0 (\delta_{kl} + \kappa_{kl}) \, , \tag{2.63}$$

► $$\mathbf{D} = \ddot{\boldsymbol{\varepsilon}} \cdot \mathbf{E} \, ; \qquad \ddot{\boldsymbol{\varepsilon}} \equiv \varepsilon_0 \left(\ddot{\mathbf{1}} + \ddot{\boldsymbol{\kappa}} \right) . \tag{2.64}$$

The matrix ε_{kl} is called the *permittivity tensor*. The fields **D** and **E** do not necessarily point in the same direction. In statistical physics, it is proven that the permittivity and susceptibility tensors must be symmetric. In a homogeneous isotropic medium, the susceptibility matrix is even diagonal, with all diagonal elements being identical. In the limit of weak fields, a simple relation holds in this case:

$$\mathbf{D} = \varepsilon \mathbf{E} \, . \tag{2.65}$$

The situation in a magnetic field is analogous. The defining relation for the magnetic field intensity (2.53) is rewritten into the form:

$$\mathbf{B} = \mu_0 \left[\mathbf{H} + \mathbf{M}(\mathbf{H}) \right] . \tag{2.66}$$

Magnetization is written as a function of the field **H** (ensuring that only one field appears on the right-hand side). This is naturally possible, as both fields are convertible via the defining relation (2.53). For weak fields, the system response can be considered linear, allowing a first-order Taylor expansion of the magnetization vector:

$$M_k \doteq \frac{\partial M_k}{\partial H_l} H_l = \chi_{kl} \, H_l \, ; \tag{2.67}$$

► $$\chi_{kl} \equiv \frac{\partial M_k}{\partial H_l} \, . \tag{2.68}$$

The coefficient matrix χ_{kl} is called the m*agnetic susceptibility tensor*. In vector notation:

$$\mathbf{M} = \ddot{\boldsymbol{\chi}} \cdot \mathbf{H} \, . \tag{2.69}$$

In the limit of weak fields, the relation between induction and intensity yields:

$$B_k = \mu_{kl} \, H_l \, ; \qquad \mu_{kl} \equiv \mu_0 (\delta_{kl} + \chi_{kl}) \, . \tag{2.70}$$

► $$\mathbf{B} = \ddot{\boldsymbol{\mu}} \cdot \mathbf{H} \, ; \qquad \ddot{\boldsymbol{\mu}} \equiv \mu_0 \left(\ddot{\mathbf{1}} + \ddot{\boldsymbol{\chi}} \right) . \tag{2.71}$$

The matrix μ_{kl} is called the permeability tensor. The fields **B** and **H** do not necessarily point in the same direction. The susceptibility and permeability tensors must be symmetric. In a homogeneous isotropic medium, the susceptibility matrix is diagonal with identical elements. In the limit of weak fields, an even simpler relation holds:

$$\mathbf{B} = \mu \mathbf{H} \, . \tag{2.72}$$

The electric and magnetic susceptibility tensors serve as indicators of a phase transition, exhibiting characteristic peaks in the transition region depending on a specific parameter. The results of the linear approximation are summarized in a concise table:

Electric Field	Magnetic Field
$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P}(\mathbf{E})$	$\mathbf{B} = \mu_0 \left[\mathbf{H} + \mathbf{M}(\mathbf{H})\right]$
$\mathbf{P} = \varepsilon_0 \, \vec{\boldsymbol{\kappa}} \cdot \mathbf{E}$	$\mathbf{M} = \vec{\boldsymbol{\chi}} \cdot \mathbf{H}$
$\kappa_{kl} \equiv \frac{1}{\varepsilon_0} \frac{\partial P_k}{\partial E_l}$	$\chi_{kl} \equiv \frac{\partial M_k}{\partial H_l}$
$\mathbf{D} = \vec{\boldsymbol{\varepsilon}} \cdot \mathbf{E}$	$\mathbf{B} = \vec{\boldsymbol{\mu}} \cdot \mathbf{H}$
$\vec{\boldsymbol{\varepsilon}} \equiv \varepsilon_0 \left(\vec{\mathbf{1}} + \vec{\boldsymbol{\kappa}}\right)$	$\vec{\boldsymbol{\mu}} \equiv \mu_0 \left(\vec{\mathbf{1}} + \vec{\boldsymbol{\chi}}\right)$

In strong fields, such expansions fail, and general relations (2.52) and (2.53) must be used.

2.1.7 Boundary Conditions

Electric and magnetic fields frequently pass through boundaries where their direction changes. This directional change follows rules derived from the integral form of the Maxwell equations. According to the following figure, the equations containing divergences are integrated near the boundary between two media, I and II, over a small cylinder with volume V and boundary S. To determine the boundary conditions, the cylinder height h must approach zero in the limit. The shape of the cylinder base is arbitrary.

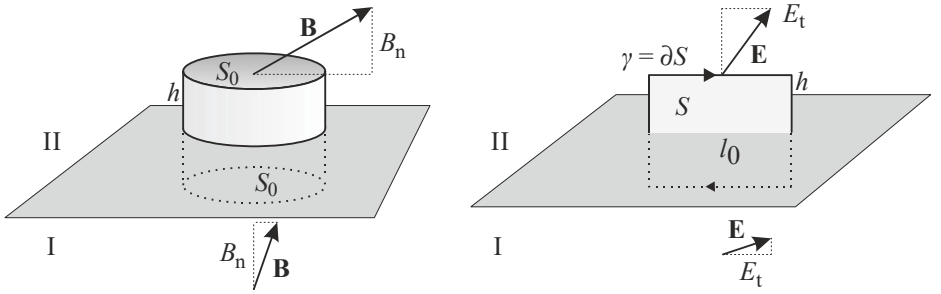


Fig. 2.9: Boundary conditions between two media.

First, consider the Gauss law of magnetostatic fields:

$$\operatorname{div}\mathbf{B} = 0 \quad \Rightarrow \quad \iiint_V \operatorname{div}\mathbf{B}\, \mathrm{d}V = 0 \quad \Rightarrow \quad \oiint_{S=\partial V} (\mathbf{B}\cdot\mathbf{n})\mathrm{d}S = 0\,.$$

Integration is performed over the surface enclosing the cylinder. As the height decreases, only the integration over the top and bottom bases remains. The outward normal points in opposite directions on these bases:

$$\mathbf{B}_{\mathrm{II}} \cdot \mathbf{n}_{\mathrm{II}} S_0 + \mathbf{B}_{\mathrm{I}} \cdot \mathbf{n}_{\mathrm{I}} S_0 = 0$$

Assuming the normal $\mathbf{n}$ to the boundary points upward, then $\mathbf{n}_{\mathrm{II}} = \mathbf{n}$ and $\mathbf{n}_{\mathrm{I}} = -\mathbf{n}$:

►
$$B_{n\mathrm{II}} - B_{n\mathrm{I}} = 0\,, \tag{2.73}$$

meaning the normal component of the magnetic field is always conserved at the boundary (see figure on left). Similar reasoning applies to the second divergence equation:

$$\operatorname{div}\mathbf{D} = \rho_Q \quad \Rightarrow \quad \iiint\limits_V \operatorname{div}\mathbf{D}\,\mathrm{d}V = \iiint\limits_V \rho_Q\,\mathrm{d}V \quad \Rightarrow \quad \oiint\limits_{S=\partial V} (\mathbf{D}\cdot\mathbf{n})\mathrm{d}S = \iiint\limits_V \rho_Q\,\mathrm{d}V.$$

The left-hand side is treated analogously, decreasing cylinder height until the difference between the normal components of electric displacement is obtained. The right-hand side situation differs. If only volume charge is present inside the cylinder, integration yields zero (as the volume approaches zero), and the normal components of electric displacement are continuous – just as for the magnetic field. A different result is obtained if a surface charge is localized at the boundary:

$$\left(D_{n\mathrm{II}} - D_{n\mathrm{I}}\right)S_0 = \lim_{h\to 0}\rho_Q V = \lim_{h\to 0}\frac{\Delta Q}{S_0 h}S_0 h = \Delta Q\,.$$

Moving the area term from the left-hand side to the right-hand side demonstrates that:

►
$$D_{n\mathrm{II}} - D_{n\mathrm{I}} = \sigma\,. \tag{2.74}$$

If the boundary contains a surface charge, the normal component of the electric displacement changes discontinuously. If no charge is present at the boundary, it is continuous. Next, the Faraday induction law is analyzed. Integration is performed over a rectangle as shown in the figure on the right:

$$\operatorname{curl}\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t} \quad \Rightarrow \quad \iint\limits_S (\operatorname{curl}\mathbf{E})\cdot\mathrm{d}\mathbf{S} = -\iint\limits_S \frac{\partial\mathbf{B}}{\partial t}\cdot\mathrm{d}\mathbf{S} \quad \Rightarrow \quad \oint\limits_{\gamma=\partial S} \mathbf{E}\cdot\mathrm{d}\mathbf{l} = -\iint\limits_S \frac{\partial\mathbf{B}}{\partial t}\cdot\mathrm{d}\mathbf{S}\,.$$

As the rectangle height approaches zero in the limit, the integral on the right-hand side approaches zero (the integration area vanishes). On the left-hand side, only the segments parallel to the boundary remain from the bounding curve after the limiting transition:

$$\mathbf{E}_{\mathrm{II}}\cdot\mathbf{t}_{\mathrm{II}}\,l_0 + \mathbf{E}_{\mathrm{I}}\cdot\mathbf{t}_{\mathrm{I}}l_0 = 0\,,$$

where $\mathbf{t}$ is the tangent vector to the boundary pointing in the direction of the integration curve. Since $\mathbf{t}_{\mathrm{II}} = \mathbf{t}$ and $\mathbf{t}_{\mathrm{I}} = -\mathbf{t}$, the following relation is obtained:

►
$$E_{t\mathrm{II}} - E_{t\mathrm{I}} = 0\,, \tag{2.75}$$

meaning the tangential component of the electric field intensity is always continuous at the boundary (see the right part of the figure). The last of the Maxwell equations remains – the Ampere law. Again, integration is performed over the rectangle in the right part of the figure:

$$\operatorname{curl}\mathbf{H} = \mathbf{j}_Q + \frac{\partial\mathbf{D}}{\partial t} \quad \Rightarrow \quad \iint\limits_S (\operatorname{curl}\mathbf{H})\cdot\mathrm{d}\mathbf{S} = \iint\limits_S \mathbf{j}_Q\cdot\mathrm{d}\mathbf{S} + \iint\limits_S \frac{\partial\mathbf{D}}{\partial t}\cdot\mathrm{d}\mathbf{S} \quad \Rightarrow$$

$$\oint\limits_{\gamma=\partial S} \mathbf{H}\cdot\mathrm{d}\mathbf{l} = \iint\limits_S \mathbf{j}_Q\cdot\mathrm{d}\mathbf{S} + \iint\limits_S \frac{\partial\mathbf{D}}{\partial t}\cdot\mathrm{d}\mathbf{S}.$$

In the limiting transition, the integral on the left-hand side approaches the difference between the tangential components multiplied by the length of the curve along the boundary. Both integrals on the right-hand side vanish in the case of volume currents (since the integration area approaches zero in the limit). In such a case, the tangential components of the magnetic field intensity are continuous, analogous to the electric field intensity. However, if an electric current flows within the boundary surface (a surface current), the situation differs:

$$\left(H_{t\mathrm{II}} - H_{t\mathrm{I}}\right) l_0 = \lim_{h\to 0} j_Q S = \lim_{h\to 0} \frac{\Delta I}{l_0 h} l_0 h = \Delta I \,.$$

Moving the curve length from the left-hand side to the right-hand side yields:

► $$H_{t\mathrm{II}} - H_{t\mathrm{I}} = i \,. \tag{2.76}$$

If the boundary contains flowing currents, the tangential component of the magnetic field intensity changes discontinuously. The magnitude of the discontinuity equals the current flowing through the boundary per unit length. If no currents flow within the boundary, the tangential component of the magnetic field intensity is continuous. In conclusion, all boundary conditions are summarized in a single relation:

► $$\begin{aligned} B_{n\mathrm{II}} - B_{n\mathrm{I}} &= 0, \\ D_{n\mathrm{II}} - D_{n\mathrm{I}} &= \sigma, \\ E_{t\mathrm{II}} - E_{t\mathrm{I}} &= 0, \\ H_{t\mathrm{II}} - H_{t\mathrm{I}} &= i. \end{aligned} \tag{2.77}$$

Without surface currents and charges, the normal components of the displacements **B**, **D** and the tangential components of the intensities **E**, **H** are conserved. In the presence of surface charges and currents at the boundary, the electric displacement and the magnetic field intensity (the quantities incorporating the material response to applied fields) change discontinuously.

◖ **Example 2.1**: **Field in a Cavity**

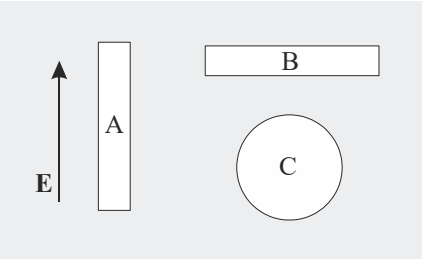


Fig. 2.10: Field in a cavity.

An electric field affects a dielectric in which three cavities are cut. The first cavity (A) is oriented predominantly in the direction of the field. The tangential component of the electric field, which is continuous under all circumstances, permeates into the cavity; therefore, the field inside cavity A is identical to that in the surrounding dielectric:

$$\mathbf{E}_{\mathrm{A}} = \mathbf{E} \,. \tag{2.78}$$

Cavity B is oriented predominantly perpendicular to the field, meaning the normal component of the electric field permeates into it. However, this component is discontinuous. Conversely, the normal components of the electric displacement $\mathbf{D} = \varepsilon_0\mathbf{E} + \mathbf{P}$ are continuous (if a non-zero surface charge density were present on the cavity surface, the electric displacement would also change discontinuously). Polarization in the dielectric is non-zero, whereas it is zero in the cavity, yielding:

$$\mathbf{E}_{\mathrm{B}} = \mathbf{E} + \frac{\mathbf{P}}{\varepsilon_0}. \tag{2.79}$$

The most complex case is a spherical cavity, where the field must be decomposed into normal and tangential parts at every point on the surface, and all contributions must then be summed. If this effort is undertaken, the result is:

$$\mathbf{E}_{\mathrm{C}} = \mathbf{E} + \frac{\mathbf{P}}{3\varepsilon_0}. \tag{2.80}$$

2.2 Electrostatics

In time-varying fields, electric and magnetic phenomena are coupled. Time-varying magnetic fields generate electric fields, and vice versa. In this chapter, fields are assumed to be time-independent, and the medium is considered non-conducting (a dielectric) with no electric current flowing through it. In such a case, electric phenomena within the Maxwell system are described solely by the equations:

$$\operatorname{div} \mathbf{D} = \rho_Q \,, \tag{2.81}$$

$$\operatorname{curl} \mathbf{E} = 0 \,. \tag{2.82}$$

Evidently, electric phenomena become independent of magnetic ones. Furthermore, in this chapter, the material in which the phenomena occur (the dielectric) is assumed to possess no memory, meaning the material response depends only on the currently applied field. In addition, a homogeneous and isotropic material is assumed, yielding the simplest possible constitutive relation:

$$\mathbf{D} = \varepsilon \mathbf{E} \,. \tag{2.83}$$

2.2.1 Poisson Equation and Its Solution

Conservative Nature of Electric Fields

Equation (2.82) states that the electric field creates no curls in the stationary case. Its curl is zero. After integration over an arbitrary surface, the Stokes theorem yields the following relation for the bounding curve:

$$\oint_{\gamma=\partial S} \mathbf{E} \cdot \mathrm{d}\mathbf{l} = 0 \,. \tag{2.84}$$

This merely expresses the irrotational nature of the field. A non-zero field circulation value can only be possessed by a vortex field. Irrotational fields are called *conservative*; thus the electric field is a conservative. If equation (2.84) is multiplied by charge, the integral of the force acting on a particle appears on the left-hand side:

$$\oint_{\gamma=\partial S} \mathbf{F} \cdot \mathrm{d}\mathbf{l} = 0 \,. \tag{2.85}$$

The line integral of force represents the performed work. For a closed curve, this work is always zero in a conservative field. The force of a conservative field can be written as the negative gradient of potential energy. Analogously, an electric field satisfying equation (2.82) can always be written as:

► $$\mathbf{E} = -\nabla \phi \,. \tag{2.86}$$

The quantity ϕ is called the electric field potential. The minus sign ensures that particles move toward the minimum of potential energy, rather than the maximum. Equation (2.82) is automatically satisfied since the curl of a gradient is always zero. The function

ϕ always exists; equation (2.86) is a simple partial differential equation always possessing a solution, implying that a single scalar function suffices to describe the electrostatic field. From this function, the respective electric field can always be generated via relation (2.86). The scalar potential is not uniquely determined. Multiple potentials can exist for a single field. For example, two potentials differing by a constant yield the same electric field. However, the freedom in choosing a potential for a given field is much broader and will be addressed in detail in the non-stationary case.

Poisson Equation

Equation (2.81) states that electric charges are sources of the electric field. Let electric displacement be converted to electric field intensity, and the intensity to field potential:

$$\operatorname{div}\mathbf{D} = \rho_Q \quad \Rightarrow \quad \operatorname{div}\varepsilon\mathbf{E} = \rho_Q \quad \Rightarrow$$

$$\operatorname{div}\varepsilon\left(-\nabla\phi\right) = \rho_Q \quad \Rightarrow \quad \operatorname{div}(\nabla\phi) = -\frac{\rho_Q}{\varepsilon_0}.$$

Recognizing that the divergence of a gradient is the Laplace operator yields the fundamental equation of electrostatics (*Poisson equation*):

► $$\Delta\phi = -\frac{\rho_Q}{\varepsilon}. \tag{2.87}$$

This derivation naturally holds only for the stationary case, a linear relation between the electric displacement and intensity, and a homogeneous isotropic medium in which the permittivity is constant. In the first step, the solution to the Poisson equation for the potential is found for a given charge distribution. In the second step, the potential is converted to the electric field using relation (2.86).

Green Function of Laplace Operator

For a point charge located at the origin of the coordinate system, the electric field potential is:

► $$\phi(\mathbf{r}) = \frac{Q}{4\pi\varepsilon r}. \tag{2.88}$$

From this scalar function, all three components of the electric field near the point charge are easily generated. During differentiation, the following well-known relation is useful:

$$\frac{\partial r}{\partial x_k} = \frac{x_k}{r}. \tag{2.89}$$

This relation is easily derived by recognizing that $r = (x^2 + y^2 + z^2)^{1/2}$. It can now be shown that potential (2.88) is suitable for a point charge:

$$E_k = -\frac{\partial\phi}{\partial x_k} = -\frac{\partial}{\partial x_k}\left(\frac{Q}{4\pi\varepsilon r}\right) = -\frac{Q}{4\pi\varepsilon}\frac{\partial}{\partial x_k}r^{-1} = -\frac{Q}{4\pi\varepsilon}\frac{\mathrm{d}}{\mathrm{d}r}\left(r^{-1}\right)\frac{\partial r}{\partial x_k} =$$

$$= -\frac{Q}{4\pi\varepsilon}\left(-r^{-2}\right)\frac{x_k}{r} = \frac{Q}{4\pi\varepsilon r^2}\frac{x_k}{r},$$

which is the correct relation for the electric field intensity, expressible in vector form as:

$$\mathbf{E} = \frac{Q}{4\pi\varepsilon r^2}\mathbf{e}_\mathbf{r} \,. \tag{2.90}$$

Note that the potential of a point charge is a function of $1/r$, whereas the field intensity possesses a magnitude proportional to $1/r^2$. If the particle is not located at the origin but at position $\mathbf{r}'$ (a prime always denotes the positions of field sources, while a symbol without a prime denotes the position of the observer), the potential at position $\mathbf{r}$, where the observer/detector is located, has the value:

$$\phi(\mathbf{r}) = \frac{Q}{4\pi\varepsilon\left|\mathbf{r}-\mathbf{r}'\right|} \,. \tag{2.91}$$

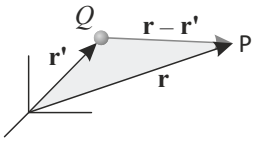


Fig. 2.11: Positions of charge and observer.

As before, the denominator features the distance between the charge and the observer/detector. The potential of a point charge, whose charge density is given by the Dirac delta distribution $\rho_Q = Q\,\delta(\mathbf{r}-\mathbf{r}')$, must satisfy the Poisson equation, i.e.:

$$\Delta\left(\frac{Q}{4\pi\varepsilon\left|\mathbf{r}-\mathbf{r}'\right|}\right) = -\frac{Q\,\delta(\mathbf{r}-\mathbf{r}')}{\varepsilon}$$

A simple rearrangement isolates the Dirac delta distribution on the right-hand side:

$$\Delta\left(-\frac{1}{4\pi\left|\mathbf{r}-\mathbf{r}'\right|}\right) = \delta(\mathbf{r}-\mathbf{r}') \,.$$

However, the solution to the Poisson equation with a Dirac impulse on the right-hand side is the Green function of the respective operator, in this case, the Laplace operator in a three-dimensional unbounded space:

► $$G(\mathbf{r},\mathbf{r}') = -\frac{1}{4\pi\left|\mathbf{r}-\mathbf{r}'\right|} \,. \tag{2.92}$$

The Green function is useful when constructing the solution to the electrostatic Poisson equation for a completely general charge distribution in space.

General Solution to Poisson Equation

The general solution to the Poisson equation $\Delta\phi = -\rho_Q/\varepsilon$ is the convolution of the Green function with the right-hand side, i.e.:

$$\phi = G * \left(-\frac{\rho_Q}{\varepsilon}\right) = \int_V G(\mathbf{r},\mathbf{r}')\left(-\frac{\rho_Q(\mathbf{r}')}{\varepsilon}\right)\mathrm{d}^3\mathbf{r}' \,, \quad \text{tj.}$$

►
$$\phi(\mathbf{r}) = \int_V \frac{\rho_Q(\mathbf{r}')}{4\pi\varepsilon\left|\mathbf{r}-\mathbf{r}'\right|}\mathrm{d}^3\mathbf{r}'. \tag{2.93}$$

This represents the solution to the Poisson equation for a general charge distribution. Volume integrals are henceforth written with a single integration symbol; its multiplicity is apparent from the differential exponent. Vector **r** denotes the observer or detector, while the vector **r′**, over which the integration is performed, sweeps through individual charges. This result can also be obtained via simple physical reasoning. Consider a small volume element cut from continuously distributed charges, as shown in the figure:

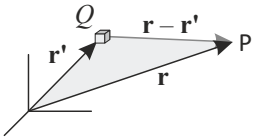


Fig. 2.12: Positions of charge element and observer (detector).

According to relation (2.91), this element contributes to the potential at the observer location as follows:

$$\Delta\phi = \frac{\Delta Q}{4\pi\varepsilon\left|\mathbf{r}-\mathbf{r}'\right|} = \frac{\rho_Q(\mathbf{r}')\Delta V}{4\pi\varepsilon\left|\mathbf{r}-\mathbf{r}'\right|} \tag{2.94}$$

The charge element is now treated as infinitesimal, and the total potential is determined by integration over all such elements:

$$\phi(\mathbf{r}) = \int_V \frac{\rho_Q(\mathbf{r}')}{4\pi\varepsilon\left|\mathbf{r}-\mathbf{r}'\right|}\mathrm{d}^3\mathbf{r}'.$$

The result is identical to that obtained via the Green function. This relation holds for the stationary case and is useful in electrostatics. If charges were moving, the signal would propagate at a finite speed, making the relation more complex. This is derived later.

2.2.2 Interesting Examples

In certain specific cases, solutions to the Poisson equation can be obtained using various techniques. Several of these methods are demonstrated below.

Network of Charged Conductors

Consider a network of uniformly distributed charged conductors, as shown in Fig. 2.13. The generated electric field constitutes a sum of individual Fourier components:

$$\phi(x,y) = \sum_{n=0}^{\infty}\phi_n = \sum_{n=0}^{\infty} A_n(y)\cos\left(\frac{2\pi nx}{a}\right) \tag{2.95}$$

Substituting a single Fourier component into the Laplace equation (the Poisson equation in a vacuum, i.e., without a right-hand side) yields an ordinary differential equation for the potential amplitudes $A_n(y)$:

$$\frac{\mathrm{d}^2A_n}{\mathrm{d}y^2} - \left(\frac{2\pi n}{a}\right)^2 A_n = 0 \tag{2.96}$$

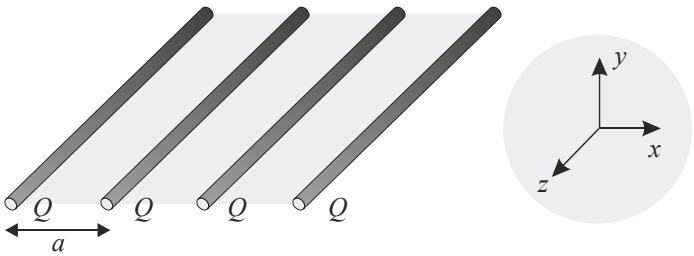


Fig. 2.13: Network of charged conductors in a plane.

The solution to equation (2.96) is straightforward:

$$A_n(y) = \begin{cases} c_0 y\,; & n = 0, \\ c_n \exp\left[-2\pi n y/a\right]; & n = 1, 2, 3\ldots \end{cases} \tag{2.97}$$

All harmonic solutions for $n > 1$ are already strongly attenuated at a distance of a few a. The fundamental mode with $n = 1$ decays the slowest. For $y > a$, the field behaves with a high degree of accuracy like a homogeneous field represented by the $n = 0$ term. The conductor network provides shielding as effective as a metal plane. At a large distance, only a single component is significant:

$$\begin{aligned} \phi_0 &= c_0 y = -Ey\,, \\ \mathbf{E}_0 &= -\nabla\phi_0 = (0, E, 0). \end{aligned} \tag{2.98}$$

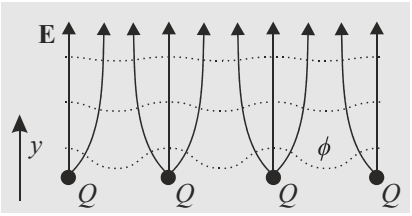


Fig. 2.14: Potential and intensity of a system of charged conductors.

Since the electric field intensity is the gradient of the potential, the electric field (and thus the field lines) always points perpendicular to the equipotentials.

Two-Dimensional Sources and Complex Variable

This example focuses on a 2D problem. Planar charged electrodes are distributed in a plane, and the objective is to find the electric field generated by them. In finding the fields, a complex function of a complex variable can be utilized successfully:

$$\begin{aligned} f(z)&: \quad \mathbb{C} \to \mathbb{C}\,; \\ f(z) &= U(x, y) + \mathrm{i}V(x, y). \end{aligned} \tag{2.99}$$

The functions U and V represent the real and imaginary parts of the complex function. Its derivative with respect to $z = x + \mathrm{i}y$ must not depend on the path; therefore, the derivative with respect to x must yield the same result as the one with respect to iy:

$$\frac{\partial}{\partial x}(U+\mathrm{i}V)=\frac{\partial}{\partial \mathrm{i}\,y}(U+\mathrm{i}V)\,. \tag{2.100}$$

Separating the real and imaginary parts yields the Cauchy–Riemann (CR) conditions (see the "*Selected Topics I*," Section I.3.2.5:

$$\frac{\partial U}{\partial x}=\frac{\partial V}{\partial y}\,;\qquad \frac{\partial U}{\partial y}=-\frac{\partial V}{\partial x}\,. \tag{2.101}$$

Further differentiation of the Cauchy–Riemann conditions reveals that both the real and imaginary parts of the complex function satisfy the Laplace equation:

$$\Delta U=0\,;\qquad \Delta V=0\,. \tag{2.102}$$

Thus, both functions U and V can serve as the electric field potential. If, for instance, the real part is utilized, the curves

$$U(x,y)=0 \tag{2.103}$$

define the equipotentials. In such a case, the curves defined by the second function:

$$V(x,y)=0 \tag{2.104}$$

are perpendicular to the previous set (following directly from the CR conditions) and therefore represent the electric field lines. In a two-dimensional case, a single complex function of a complex variable thus enables the description of both the equipotentials and the electric field lines around the electrodes.

Colloidal Particles in Electrolyte and Debye Length

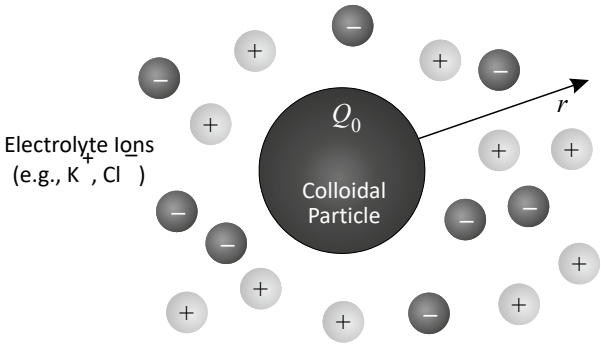


Fig. 2.15: Colloidal particle in an electrolyte.

Monitoring the potential distribution near a colloidal particle reveals the influence of other neighboring charged particles – ions. If the plasma is not far from thermodynamic equilibrium, particles of opposite polarity shift toward the selected source, providing shielding. This results in an exponential decay of the source field with a characteristic distance λ_D, at which the potential and field decrease to 1/e of the value given by the Coulomb law. This distance is called the *Debye shielding distance*, named after the Dutch physicist and chemist Peter Debye. The electric potential $\phi(r)$ around the source is determined from the Poisson equation:

$$\Delta\phi = -\frac{\rho_Q}{\varepsilon_0}$$

For simplicity, the positive and negative ions are assumed to possess identical charges, with their number density given by the Boltzmann statistical distribution:

$$\Delta\phi = -\frac{Q_+ n_+ + Q_- n_-}{\varepsilon_0} = -\frac{+Qn_+ - Qn_-}{\varepsilon_0} \qquad \Rightarrow$$

$$\Delta\phi = -\frac{1}{\varepsilon_0}\left[+Qn_0 \exp\left(-\frac{Q\phi(\mathbf{x})}{k_\mathrm{B}T}\right) - Qn_0 \exp\left(+\frac{Q\phi(\mathbf{x})}{k_\mathrm{B}T}\right)\right].$$

If the system is not far from thermodynamic equilibrium, a first-order Taylor expansion of the exponentials is performed. The zeroth-order terms cancel out (the solution is neutral), and the remainder yields:

$$\Delta\phi = \alpha\phi\,; \qquad \alpha = \frac{2Q^2 n_0}{\varepsilon_0 k_\mathrm{B} T}\,.$$

The equation is solved in spherical coordinates with the origin at the colloidal particle:

$$\frac{1}{r}\frac{\mathrm{d}^2}{\mathrm{d}r^2}\left[r\phi(r)\right] = \alpha\phi \qquad \Rightarrow \qquad \frac{\mathrm{d}^2\psi}{\mathrm{d}r^2} = \alpha\psi\,; \qquad \psi(r) \equiv r\phi(r) \qquad \Rightarrow$$

$$\psi(r) = C_1\,\mathrm{e}^{\sqrt{\alpha}\,r} + C_2\,\mathrm{e}^{-\sqrt{\alpha}\,r} \qquad \Rightarrow \qquad \phi(r) = \frac{C_1}{r}\mathrm{e}^{\sqrt{\alpha}\,r} + \frac{C_2}{r}\mathrm{e}^{-\sqrt{\alpha}\,r}\,.$$

Since the potential of a point source cannot diverge at an infinite distance, $C_1 = 0$. The constant C_2 is determined such that the potential in the limit of small distance from the source transitions into the classical Coulomb potential of a source with charge Q_0:

► $$\phi(r) = \frac{Q_0}{4\pi\varepsilon_0 r}\,\mathrm{e}^{-\frac{r}{\lambda_\mathrm{D}}}\,; \qquad \lambda_\mathrm{D} \equiv \sqrt{\frac{\varepsilon_0 k_\mathrm{B} T}{2Q^2 n_0}}\,. \tag{2.105}$$

Ions shield the colloidal particle from the surroundings and other particles at the characteristic distance λ_D (the Debye length). Adding salt to the solution increases the ion number density and decreases the Debye length. Better shielded colloidal particles aggregate more easily, forming larger structures; consequently, the colloid can precipitate.

Field of a Tip

Consider a metallic (conducting) element with two ends possessing different curvatures.

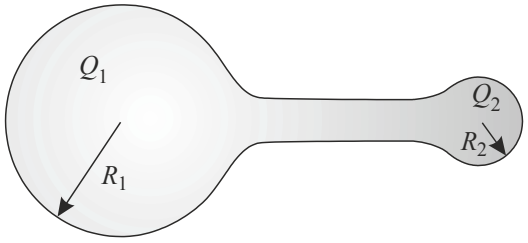


Fig. 2.16: Field of a tip.

The potentials on the surface of the end parts are approximately:

$$\phi_1 = \frac{Q_1}{4\pi\varepsilon_0 R_1}; \qquad \phi_2 = \frac{Q_2}{4\pi\varepsilon_0 R_2} \tag{2.106}$$

Since this is a conductor, everything is conductively connected, and a single potential exists across the entire surface (otherwise, charges would shift until the potentials of different parts of the conductor equalized):

$$\phi_1 = \phi_2 \qquad \Rightarrow \qquad \frac{Q_1}{R_1} = \frac{Q_2}{R_2} \tag{2.107}$$

Inside a conductor, repelling charges of identical polarity are pushed to the surface. The electric field intensity near the surface is proportional to the surface charge density:

$$E \sim \sigma \qquad \Rightarrow \qquad \frac{E_1}{E_2} = \frac{Q_1 / 4\pi R_1^2}{Q_2 / 4\pi R_2^2} = \frac{Q_1 R_2^2}{Q_2 R_1^2}$$

Substituting the charge ratio from (2.107) yields:

$$\frac{E_1}{E_2} = \frac{R_2}{R_1} \qquad \Rightarrow \qquad E \sim \frac{1}{R} \tag{2.108}$$

The electric field intensity near the conductor surface is inversely proportional to the radius of curvature of the region. Near sharp tips, the field is therefore extremely large, giving rise to various discharges (corona discharge, St. Elmo's fire, etc.).

2.2.3 Electric Field Energy

The electric field intensity and the electric potential are related by the formula:

$$\mathbf{E} = -\nabla\phi$$

A particle with charge Q placed in an electric field experiences a force::

$$\mathbf{F} = Q\mathbf{E} = -\nabla(Q\phi)$$

Since the electrostatic field is conservative, i.e., the force is the negative gradient of the potential energy, the potential energy must satisfy the following relation:

$$W_\mathrm{p} = Q\phi\,.$$

Multiplying the field by the charge yields the force, while multiplying the potential by the charge yields the particle energy:

► $$\begin{aligned} \mathbf{F} &= Q\,\mathbf{E}\,; \\ W_\mathrm{p} &= Q\,\phi. \end{aligned} \tag{2.109}$$

The energy of particle a in the field of particle b is now determined as:

$$U_{ab} = Q_a\phi_b = \frac{Q_a Q_b}{4\pi\varepsilon_0 \left|\mathbf{r}_a - \mathbf{r}_b\right|} \tag{2.110}$$

For a system of particles, such expressions are summed over all possible pairs of particles a and b. However, the energy of each pair is thereby counted twice (for example, as W_{37} and W_{73}). Only half of the resulting sum must therefore be used, since all terms appear twice. Furthermore, cases where particles would form a pair with themselves must be excluded by requiring $a \neq b$:

$$U = \frac{1}{2} \sum_{\substack{a,b \\ a \neq b}} \frac{Q_a Q_b}{4\pi\varepsilon_0 \left| \mathbf{r}_a - \mathbf{r}_b \right|} \tag{2.111}$$

The sum is now split:

$$U = \frac{1}{2} \sum_a Q_a \sum_{b \neq a} \frac{Q_b}{4\pi\varepsilon_0 \left| \mathbf{r}_a - \mathbf{r}_b \right|} . \tag{2.112}$$

The inner sum represents the total potential of all particles at the position $\mathbf{r}_a$, which is the total potential generated by all particles at the position of particle a:

$$U = \frac{1}{2} \sum_a Q_a \phi(\mathbf{r}_a) . \tag{2.113}$$

For a continuous charge distribution, the expression becomes the integral:

$$U = \frac{1}{2} \int_V \rho_Q(\mathbf{r}) \phi(\mathbf{r}) \mathrm{d}^3\mathbf{r} . \tag{2.114}$$

Denoting the integration variable as $\mathbf{r}'$ is unnecessary since only a single vector figures in the relation.. The total energy of all charged particles has been calculated. However, these particles interact through electric fields, which can also be used to express the energy. Expressions of the type (2.111) and (2.114) clearly relate the energy to individual particles. If the relation is consistently rewritten in terms of fields, the energy can also be deposited outside the region of non-zero charge, since the fields are non-zero even in a vacuum. The field thus acquires the character of a real entity that is a carrier (and literally a transmitter) of energy. To rewrite the expression in terms of fields, the charge density is expressed from the Maxwell equation div $\mathbf{D} = \rho_Q$:

$$U = \frac{1}{2} \int_V \phi \operatorname{div} \mathbf{D} \, \mathrm{d}^3\mathbf{r} . \tag{2.115}$$

Integration by parts in three dimensions is now utilized (see Section I.3.6.3 in the first volume of "*Selected Topics*"):

$$U = \frac{1}{2} \int_V \phi \, (\partial_k D_k) \, \mathrm{d}^3\mathbf{r} = -\frac{1}{2} \int_V (\partial_k \phi) D_k \, \mathrm{d}^3\mathbf{r} + \oint_{S=\partial V} \phi D_k n_k \, \mathrm{d}S .$$

The first integral is modified using the electric field ($E_k = -\partial_k \phi$), while the second is taken over the boundary. The potential decreases with distance from the sources as $1/r$, and the electric displacement field as $1/r^2$; the integrand thus decreases as $1/r^3$. If the integration surface is taken sufficiently far from the sources, ideally at infinity, the contribution of the last integral is zero, and the total energy is obtained as:

$$U = \frac{1}{2} \int_V \mathbf{E} \cdot \mathbf{D} \, \mathrm{d}^3\mathbf{r}. \tag{2.116}$$

The integrand represents the energy density of the electromagnetic field (energy per unit volume, with the unit J/m^3), which is half of the scalar product of both electric vectors:

► $$u = \mathbf{E} \cdot \mathbf{D} / 2 \,. \tag{2.117}$$

The same relation is derived in introductory physics courses for a capacitor. The field energy density in a capacitor is likewise $\mathbf{E}\cdot\mathbf{D}/2$.

2.2.4 Multipole Expansion

Localized field sources are assumed to form a cluster of particles in a region where the origin of the coordinate system is placed:

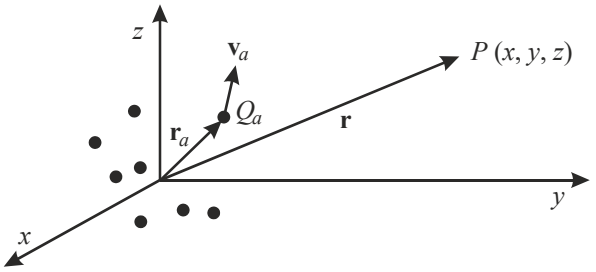


Fig. 2.17: Localized particles and distant observer (detector).

Each charge has a position vector $\mathbf{r}_a$, velocity $\mathbf{v}_a$, mass m_a, and charge Q_a. The observer is located at a large distance $\mathbf{r}$ from the field source, The charge density is given by:

$$\rho_Q(\mathbf{r}') = \sum_a Q_a \delta(\mathbf{r}' - \mathbf{r}_a) \,, \tag{2.118}$$

In this case, the electric potential has a simple expression:

$$\phi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0} \sum_a \frac{Q_a}{|\mathbf{r} - \mathbf{r}_a|} \,, \tag{2.119}$$

Since the field is observed at a distant location $\mathbf{r}$ where $r \gg r_a$, a Taylor expansion can be performed for the argument $\mathbf{r}$ and the increment $\mathbf{r}_a$:

$$\frac{1}{|\mathbf{r} - \mathbf{r}_a|} = \frac{1}{r} - x_k^{(a)} \frac{\partial}{\partial x_k}\left(\frac{1}{r}\right) + \frac{1}{2!} x_k^{(a)} x_l^{(a)} \frac{\partial^2}{\partial x_k \partial x_l}\left(\frac{1}{r}\right) \pm \cdots \tag{2.120}$$

The derivatives are now evaluated using the relation $\partial r/\partial x_k = x_k/r$:

$$\frac{1}{|\mathbf{r} - \mathbf{r}_a|} = \frac{1}{r} + x_k^{(a)} \frac{x_k}{r^3} + \frac{1}{2!} x_k^{(a)} x_l^{(a)} \frac{3x_k x_l - r^2 \delta_{kl}}{r^5} \pm \cdots \tag{2.121}$$

Substituting the derived expansion into relation (2.119) for the potential and factoring out the quantities independent of the summation yields:

$$\phi = \frac{1}{4\pi\varepsilon_0 r} \sum_a Q_a + \frac{x_k}{4\pi\varepsilon_0 r^3} \sum_a Q_a x_k^{(a)} + \frac{3x_k x_l - r^2 \delta_{kl}}{8\pi\varepsilon_0 r^5} \sum_a Q_a x_k^{(a)} x_l^{(a)} + \cdots \tag{2.122}$$

In the last written of the scalar potential expansion, the expression inside the summation can be modified into a form analogous to the expression preceding the summation:

$$\frac{3x_k x_l - r^2\delta_{kl}}{8\pi\varepsilon_0 r^5}\sum_a Q_a x_k^{(a)} x_l^{(a)} = \frac{3x_k x_l - r^2\delta_{kl}}{24\pi\varepsilon_0 r^5}\sum_a Q_a\left(3x_k^{(a)} x_l^{(a)}\right) = \\ = \frac{3x_k x_l - r^2\delta_{kl}}{24\pi\varepsilon_0 r^5}\sum_a Q_a\left(3x_k^{(a)} x_l^{(a)} - r_a^2\delta_{kl}\right). \tag{2.123}$$

Subtracting the term $r_a^2\delta_{kl}$ renders the tensor in parentheses traceless, which leaves the result unchanged because the following identity holds:

$$\left(3x_k x_l - r^2\delta_{kl}\right) r_a^2\delta_{kl} = \left(3r^2 - 3r^2\right) r_a^2 = 0\,. \tag{2.124}$$

Overall, the first expansion terms required for further analysis take the form:

$$\begin{aligned} &\phi^{(0)} = \frac{Q}{4\pi\varepsilon_0 r}; && Q \equiv \sum_a Q_a\,, \\ \blacktriangleright\quad &\phi^{(1)} = \frac{\mathbf{p}_\mathrm{E}\cdot\mathbf{r}}{4\pi\varepsilon_0 r^3}; && \mathbf{p}_\mathrm{E} \equiv \sum_a Q_a\mathbf{r}_a, \\ &\phi^{(2)} = \frac{3x_k x_l - r^2\delta_{kl}}{24\pi\varepsilon_0 r^5} Q_{kl}\,; && Q_{kl} \equiv \sum_a Q_a\left(3x_k^{(a)} x_l^{(a)} - r_a^2\delta_{kl}\right). \end{aligned} \tag{2.125}$$

These contributions are called the monopole (0), dipole (1), and quadrupole (2) terms. For an electrically neutral system, the total charge Q is zero, vanishing the monopole term. The monopole contribution decreases with distance as 1/, the dipole as $1/r^2$, and the quadrupole as $1/r^3$. The corresponding forces (derivatives) decrease as $1/r^2$, $1/r^3$, and $1/r^4$ for the monopole, dipole, and quadrupole, respectively. The force between two interacting dipoles thus decreases with the third power of the distance. For two particles with equal but opposite charges, the dipole moment yields the original definition (2.9):

$$\mathbf{p}_\mathrm{E} \equiv \sum_a Q_a\mathbf{r}_a = Q_+\mathbf{r}_+ + Q_-\mathbf{r}_- = Q(\mathbf{r}_+ - \mathbf{r}_-) = Q\mathbf{d}\,. \tag{2.126}$$

The electric quadrupole contribution can be written invariantly as:

$$\blacktriangleright\quad \phi^{(2)} = \frac{3\mathbf{r}\otimes\mathbf{r} - r^2\ddot{\mathbf{1}}}{24\pi\varepsilon_0 r^5} : \ddot{\mathbf{Q}}\,; \qquad \ddot{\mathbf{Q}} \equiv \sum_a Q_a\left(3\mathbf{r}_a\otimes\mathbf{r}_a - r_a^2\mathbf{1}\right). \tag{2.127}$$

Monopole

Differentiating the potentials (2.125) via $E_k = -\partial_k\phi$ determines the components of the electric field. For the monopole term, the result is the expected Coulomb law:

$$\mathbf{E}^{(0)} = \frac{Q}{4\pi\varepsilon_0 r^2}\frac{\mathbf{r}}{r}. \tag{2.128}$$

The field decreases with the square of the distance from the charge system, and the field lines have a radial direction. The field is non-zero only if the total charge Q is non-zero.

Dipole

The definition of the dipole moment is important in numerical simulations. Tracking particle motions provides position vectors at each instant, which enables determining both the total dipole moment $\mathbf{p}_\mathrm{E}$ and its density, i.e., the polarization $\mathbf{P}$. This is useful, for example, when calculating the permittivity tensor, as well as in other situations. The electric field generated by a dipole moment is determined as:

$$E_k = -\partial_k \phi^{(1)} = -\frac{p_{\mathrm{E}l}}{4\pi\varepsilon_0}\frac{\partial}{\partial x_k}\left(\frac{x_l}{r^3}\right) = \frac{3p_{\mathrm{E}l}x_l x_k - p_{\mathrm{E}k}r^2}{4\pi\varepsilon_0 r^5}. \tag{2.129}$$

The relation $\partial_k r = x_k/r$ is utilized. The covariant expression for the dipole field is:

$$\mathbf{E} = \frac{3(\mathbf{p}_\mathrm{E}\cdot\mathbf{n})\mathbf{n} - \mathbf{p}_\mathrm{E}}{4\pi\varepsilon_0 r^3}; \qquad \mathbf{n} \equiv \frac{\mathbf{r}}{r}. \tag{2.130}$$

The field decreases as $1/r^3$ and exhibits a relatively complex directional dependence.

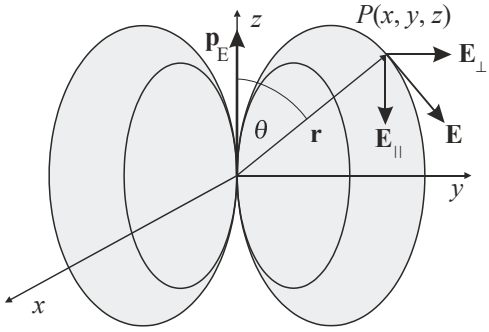


Fig. 2.18: Electric dipole field.

If the dipole axis coincides with the z-axis, the individual components take the values:

$$\mathbf{E} = \frac{p_\mathrm{E}}{4\pi\varepsilon_0 r^3}\left(3n_z n_x, 3n_z n_y, 3n_z^2 - 1\right);$$

$$E_\perp = \frac{p_\mathrm{E}}{4\pi\varepsilon_0 r^3} 3\cos\theta\sin\theta; \tag{2.131}$$

$$E_\parallel = \frac{p_\mathrm{E}}{4\pi\varepsilon_0 r^3}(3\cos^2\theta - 1).$$

Quadrupole

The quadrupole moment expresses the deviation of the charge distribution from spherical symmetry. The tensor is symmetric and traceless, satisfying:

$$Q_{kl} = Q_{lk}, \tag{2.132}$$

$$\mathrm{Tr}(\ddot{\mathbf{Q}}) = Q_{11} + Q_{22} + Q_{33} = 0. \tag{2.133}$$

In a general system, a symmetric matrix possesses 6 independent components; the trace condition reduces this number to five. For real symmetric matrices, the eigenvalues are real and the eigenvectors form an orthogonal system. If these eigenvectors are chosen as the axes of the coordinate system, the quadrupole moment tensor possesses only diagonal components. In such a coordinate system, due to the vanishing trace, only two components of the quadrupole moment tensor remain independent. For a diagonal quadrupole moment tensor, $Q_{kl}\delta_{kl} = 0$ holds because this is the trace of the tensor, and only the first term remains for the potential $\phi^{(2)}$ in expression (2.125):

$$\phi^{(2)} = \frac{x_k x_l}{8\pi\varepsilon_0 r^5} Q_{kl} = \frac{1}{2}\frac{Q_{kl} n_k n_l}{4\pi\varepsilon_0 r^3} \tag{2.134}$$

In covariant notation, the expression is given by:

$$\phi^{(2)} = \frac{1}{2}\ \frac{(\ddot{\mathbf{Q}}\cdot\mathbf{n})\cdot\mathbf{n}}{4\pi\varepsilon_0 r^3}; \qquad \mathbf{n} \equiv \frac{\mathbf{r}}{r} \tag{2.135}$$

For an axially symmetric charge distribution (around the z-axis), the first two diagonal components are identical, meaning the quadrupole moment tensor possesses a single independent component since the trace must vanish:

$$Q_{kl} = \begin{pmatrix} -Q/2 & 0 & 0 \\ 0 & -Q/2 & 0 \\ 0 & 0 & Q \end{pmatrix}. \tag{2.136}$$

Note that the quantity Q in this expression does not denote charge. The directional vector in the potential is a unit vector (the sum of the component squares equals unity). This fact can be used for the potential calculation in axial symmetry around the z-axis:

$$\begin{aligned}(\ddot{\mathbf{Q}}\cdot\mathbf{n})\cdot\mathbf{n} = Q_{kl} n_k n_l = Q_{11} n_1^2 + Q_{22} n_2^2 + Q_{33} n_3^2 = \\ = -\frac{Q}{2}\left(n_1^2 + n_2^2\right) + Q n_3^2 = \\ = -\frac{Q}{2}\left(1 - n_3^2\right) + Q n_3^2 = \\ = \frac{Q}{2}\left(3n_3^2 - 1\right) = \frac{Q}{2}\left(3\cos^2\theta - 1\right).\end{aligned}$$

Assuming axial symmetry around the z-axis, the quadrupole part of the potential in the coordinate system formed by the eigenvectors can be written as:

$$\phi^{(2)} = \frac{Q}{16\pi\varepsilon_0 r^3}(3\cos^2\theta - 1), \tag{2.137}$$

which represents a simple and readily applicable relation. The angle θ is the deviation of the observer position vector from the axis of symmetry (z). The corresponding electric field obtained by differentiation exhibits a complex structure and is utilized with an additional magnetic field to confine particles in an electromagnetic trap.

2.2.5 Electric Dipole in External Field

In a sense, the opposite problem is now considered. Thus far, the field generated by a particle system localized in space has been examined, with the coordinate origin placed somewhere within the system, e.g., at its center of mass. The response of such a system to an external electric field is now investigated. If the system is neutral, the zeroth term of the multipole expansion vanishes, and the system interacts with the external field via the dipole or quadrupole terms. Consider a system of charges localized near the coordinate origin, where the position vectors of the particles are $\mathbf{r}_a$. The particles are immersed in an external field with potential ϕ. The energy of the system in the external field is obtained by performing a Taylor expansion to the first order around the origin:

$$U = \sum_a Q_a \phi(\mathbf{r}_a) = \sum_a Q_a \left[\phi(0) + (\mathbf{r}_a \cdot \nabla)\phi + \cdots \right] \tag{2.138}$$

Writing the partial derivatives of the potential as the components of the electric field ($E_k = -\partial_k \phi$) and factoring out the quantities independent of the summation yields:

$$U = \left(\sum_a Q_a \right) \phi(0) - \left(\sum_a Q_a \mathbf{r}_a \right) \cdot \mathbf{E} + \cdots$$

The first term vanishes for a neutral system, and the second term provides the dipole contribution to the particle energy in the external field:

► $$U^{(1)} = -\mathbf{p}_\mathrm{E} \cdot \mathbf{E} \,. \tag{2.139}$$

An electric dipole has the lowest energy when aligned with the electric field and the highest energy when oriented oppositely. At low temperatures elementary dipoles therefore tend to align in the direction of the field. The probability density (and the dipole number density, which is proportional to it) is given by the Boltzmann distribution:

$$n = n_0 \exp\left[-\frac{U^{(1)}}{k_\mathrm{B} T} \right] = n_0 \exp\left[+\frac{p_\mathrm{E} E \cos\theta}{k_\mathrm{B} T} \right], \tag{2.140}$$

where θ is the angle between the field and the dipole, and k_B is the Boltzmann constant. Dipoles align along the field more effectively as the field becomes stronger, the temperature lower, and the dipole moment larger.

* * *

The torque acting on the system of charged particles is also of interest:

$$\mathbf{M} = \sum_a \mathbf{r}_a \times \mathbf{F} = \sum_a \mathbf{r}_a \times Q_a \mathbf{E} = \left(\sum_a \mathbf{r}_a Q_a \right) \times \mathbf{E} \,.$$

► $$\mathbf{M} = \mathbf{p}_\mathrm{E} \times \mathbf{E} \,. \tag{2.141}$$

A non-zero torque acts on the dipoles even in a homogeneous field.

2.3 Magnetostatics

As previously demonstrated in electrostatics, in the stationary case where fields are time-independent, the equations for the electric and magnetic fields decouple. For the magnetic field, two equations remain from the Maxwell system:

$$\operatorname{div}\mathbf{B} = 0\,, \tag{2.142}$$

$$\operatorname{curl}\mathbf{H} = \mathbf{j}_Q\,. \tag{2.143}$$

In this chapter, analogously to electrostatics, the medium is assumed to have no memory; its response depends only on the currently applied field and not on its time derivatives. The fields are assumed to be sufficiently weak to justify a linear response (it thus suffices to perform a Taylor expansion of the magnetization only to the first order). Furthermore, a homogeneous and isotropic medium is assumed, meaning that the simplest possible constitutive relations are sufficient for the calculations in this chapter:

$$\mathbf{B} = \mu\,\mathbf{H}\,. \tag{2.144}$$

2.3.1 Vector Potential

The magnetic field is not conservative, the force acting on particles cannot be written as the gradient of a potential energy. Such a function does not exist in a magnetic field. However, the equation div **B** = 0 allows the introduction of a vector potential. It can be easily shown that the identity div(curl **K**) = 0 holds for any field. If the field **B** can be written as the curl of another field **A**, equation (2.142) is automatically satisfied:

$$\mathbf{B} = \operatorname{curl}\mathbf{A} \quad \Rightarrow \quad \operatorname{div}\mathbf{B} = \operatorname{div}\operatorname{curl}\mathbf{A} = 0\,. \tag{2.145}$$

A fundamental question is whether such a vector field can be found. The answer is positive; the equation curl **A** = **B** is a partial differential equation for the field **A**, where the right-hand side is formed by the field **B**. Except for exotic sources such as a magnetic monopole, such an equation always has a solution, and in fact, infinitely many. The function **A** is called the *vector potential of the magnetic field*. Once the vector potential is determined, the magnetic field is easily found from the equation:

► $$\mathbf{B} = \operatorname{curl}\mathbf{A}\,. \tag{2.146}$$

However, the potential is not uniquely determined for a given field. If the gradient of any scalar function is added to the potential, the magnetic field remains unchanged. Such a gradient transformation is called a *gauge transformation*:

► $$\tilde{\mathbf{A}} = \mathbf{A} + \nabla f \quad \Rightarrow \quad \tilde{\mathbf{B}} = \mathbf{B}\,. \tag{2.147}$$

The proof of this statement is straightforward:

$$\tilde{\mathbf{B}} = \operatorname{curl}\tilde{\mathbf{A}} = \operatorname{curl}\left(\mathbf{A} + \nabla f\right) = \operatorname{curl}\mathbf{A} = \mathbf{B}\,.$$

The magnetic field indeed remains unchanged because the curl of a gradient is zero. A vast number of potentials lead to the same field. This freedom can be exploited to select a potential (a procedure called *gauge fixing*) for which the corresponding equations are as simple as possible. In electrostatics, the electric field can be described by a single function – the scalar potential. No such simplification holds in magnetostatics. Instead of one vector field (**B**), another vector field (**A**) is utilized, the equations for which are simpler. Overall, a set of four functions (the scalar and vector potentials) is sufficient to describe the electromagnetic field. This corresponds well with the fact that physical processes occur in a four-dimensional spacetime and are described by ordered quadruplets. It is demonstrated later that the quadruplet (ϕ, **A**) describing the electromagnetic field forms a four-vector that transforms via the Lorentz transformation. The fields in another coordinate system are determined from the transformed potentials.

◖ Example 2.2: Potential of Homogeneous Magnetic Field

The magnetic field is assumed to be homogeneous. The coordinate system is chosen such that the field points along the *z*-axis:

$$\mathbf{B} = (0, 0, B)\,. \tag{2.148}$$

An attempt is now made to find the vector potential directly from relation (2.146). The equation curl **A** = **B** is expanded into components:

$$\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = 0\,, \qquad \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} = 0\,, \qquad \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = B\,. \tag{2.149}$$

The third equation is naturally the most essential; to yield the correct right-hand side upon differentiation, the components A_x and A_y must be chosen appropriately. Verifying that suitable potentials include the following is left as an exercise:

$$\begin{aligned} &\mathbf{A} = \big(0,\, Bx,\, 0\big)\,; \\ &\mathbf{A} = \big(-By,\, 0,\, 0\big)\,; \\ &\mathbf{A} = \tfrac{1}{2}\big(-By,\, Bx,\, 0\big)\,; \\ &\mathbf{A} = \tfrac{1}{2}\,\mathbf{B}\times\mathbf{r}\,. \end{aligned} \tag{2.150}$$

All these potentials are equally valid; yielding the magnetic field (0, 0, *B*). Potentials of simple fields are listed in the table at the end of Section 2.6.1. ◗

2.3.2 Poisson Equation and Its Solution

Introducing the vector potential automatically satisfies the first magnetostatic equation (2.142). Substituting the vector potential into the second equation (2.143) yields:

$$\operatorname{curl}\mathbf{H} = \mathbf{j}_Q \quad \Rightarrow \quad \operatorname{curl}\mathbf{B} = \mu\mathbf{j}_Q \quad \Rightarrow \quad \operatorname{curl}\operatorname{curl}\mathbf{A} = \mu\mathbf{j}_Q \quad \Rightarrow$$

$$\blacktriangleright \qquad \operatorname{grad}\operatorname{div}\mathbf{A} - \Delta\mathbf{A} = \mu\mathbf{j}_Q\,. \tag{2.151}$$

This relation represents a partial differential equation for the vector potential. The free current density figures here as the source term for the vector potential. Were it not for the first term on the left-hand side, this expression would be the Laplace-Poisson equa-

tion, just as in electrostatics. However, the first term complicates the equation. Infinitely many vector potentials are known to exist, differing from each other by the gradient of an arbitrary function. Choosing a potential such that the first term of equation (2.151) vanishes is therefore considered. Assuming the potential **A** is known, a new potential with a vanishing divergence is sought. Performing the transformation

$$\tilde{\mathbf{A}} = \mathbf{A} + \nabla f \tag{2.152}$$

and requiring the divergence of the new potential to be zero

$$\operatorname{div} \tilde{\mathbf{A}} = 0 \,. \tag{2.153}$$

leads to the following equation:

$$\operatorname{div}\left(\mathbf{A} + \nabla f\right) = 0 \qquad \Rightarrow \qquad \Delta f = -\operatorname{div} \mathbf{A} \,.$$

Since a known function figures on the right-hand side, the equation for f generating the transformation to the new potential reads:

$$\Delta f = F(\mathbf{r}). \tag{2.154}$$

Such an equation always has a solution; hence, the function f transforming the old potential into the new one exists. For the new potential, a simpler equation holds (the tildes are omitted in the following text):

► $$\Delta \mathbf{A} = -\mu \mathbf{j}_Q \,; \tag{2.155}$$

► $$\operatorname{div} \mathbf{A} = 0 \,. \tag{2.156}$$

The first relation is the Laplace-Poisson equation of magnetostatics for the vector potential, and the second is an additional condition (the *Coulomb gauge* condition, also termed as the *Lorenz gauge*) that the selected potential must satisfy. If any solution to equation (2.155) is found, verification must be performed to ensure it satisfies the gauge condition. Since the Green function of the Laplace operator (2.92) in a three-dimensional unbounded space is known, the general solution can be written immediately:

$$\mathbf{A} = G * \mathbf{g} \,, \tag{2.157}$$

where **g** is the right-hand side of the Laplace-Poisson equation. Substitution yields:

$$\mathbf{A}(\mathbf{r}) = \int_V G(\mathbf{r}, \mathbf{r}')\, \mathbf{g}(\mathbf{r}')\, \mathrm{d}^3\mathbf{r}' = \int_V \left(-\frac{1}{4\pi \left| \mathbf{r} - \mathbf{r}' \right|} \right) \left(-\mu \mathbf{j}_Q(\mathbf{r}') \right) \mathrm{d}^3\mathbf{r}' \qquad \Rightarrow$$

► $$\mathbf{A}(\mathbf{r}) = \frac{\mu}{4\pi} \int_V \frac{\mathbf{j}_Q(\mathbf{r}')}{\left| \mathbf{r} - \mathbf{r}' \right|} \mathrm{d}^3\mathbf{r}' . \tag{2.158}$$

The solution is completely analogous to the scalar potential (2.93) for the electric field. However, the source of the magnetic field is not the charge density, but the electric charge flux. The primed variables again denote the source positions, while the unprimed variables denote the observer position. The denominator contains the distance between the field source and the observer. Direct calculation shows that the potential chosen in this manner satisfies the gauge condition $\operatorname{div} \mathbf{A} = 0$ provided that $\operatorname{div} \mathbf{j}_Q = 0$, which represents the charge conservation law in the stationary case.

2.3.3 Biot-Savart Law

Consider now a situation where the magnetic field is excited not by freely moving particles, but by a conductor of general shape (described by a curve γ) carrying an electric current. In this case, the charge flux is given by:

$$\mathbf{j}_Q(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}' = I\,\mathrm{d}\mathbf{l}', \tag{2.159}$$

where d $\mathbf{l}'$ is a curve element along the conductor. The vector potential is therefore:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu}{4\pi}\int_\gamma \frac{I\,\mathrm{d}\mathbf{l}'}{|\mathbf{r}-\mathbf{r}'|}. \tag{2.160}$$

The magnetic field is calculated as the curl of this vector potential:

► $$\mathbf{B} = \operatorname{curl}\mathbf{A} = \frac{\mu}{4\pi}\int_\gamma \frac{I\,\mathrm{d}\mathbf{l}'\times(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3}. \tag{2.161}$$

The calculation is straightforward: $B_k = \varepsilon_{klm}\partial_l A_m$ ... The derivative is moved inside the integral, and the denominator is differentiated as a composite function, utilizing the familiar relation $\partial r/\partial x_k = x_k/r$. Equation (2.161) is frequently used in electrical engineering to calculate magnetic fields generated by current-carrying conductors. This relation was first derived by French physicist and astronomer Jean-Baptiste Biot and French physicist and mathematician Félix Savart in 1820. Notably, the law in form (2.161) holds only for the stationary case, i.e., for constant electric currents.

2.3.4 Multipole Expansion

Analogously to electrostatics, a localized charge system is assumed in magnetostatics. The origin of the coordinate system can be placed, for example, at the center of mass of the system. Each charge possesses a position vector $\mathbf{r}_a$, a velocity vector $\mathbf{v}_a$, a mass m_a, and a charge Q_a. The observer is located at a large distance $\mathbf{r}$ from the field source.

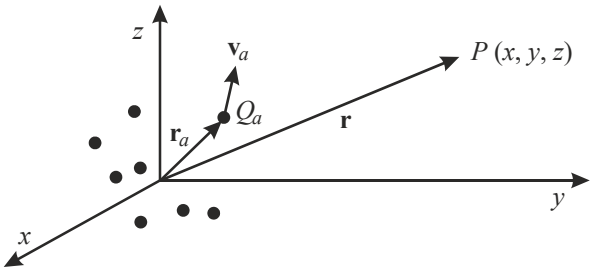


Fig. 2.19: Localized particles and distant observer (detector).

The charge flux represents a superposition over the positions of individual particles:

$$\mathbf{j}_Q(\mathbf{r}') = \sum_a Q_\alpha \mathbf{v}_a \delta(\mathbf{r}'-\mathbf{r}_a). \tag{2.162}$$

For the potential (2.158), a simple expression is obtained in this case:

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi} \sum_a \frac{Q_a \mathbf{v}_a}{|\mathbf{r} - \mathbf{r}_a|} . \tag{2.163}$$

Since the field is observed far from the source ($r >> r_a$), a Taylor expansion is performed for the argument $\mathbf{r}$ and the increment $\mathbf{r}_a$, identically to electrostatics:

$$\frac{1}{|\mathbf{r} - \mathbf{r}_a|} = \frac{1}{r} - x_k^{(a)} \frac{\partial}{\partial x_k} \left(\frac{1}{r} \right) \pm \cdots \tag{2.164}$$

The derivatives are now evaluated using the relation $\partial r / \partial x_k = x_k / r$:

$$\frac{1}{|\mathbf{r} - \mathbf{r}_a|} = \frac{1}{r} + x_k^{(a)} \frac{x_k}{r^3} \pm \cdots \tag{2.165}$$

Substituting the derived expansion into the relation for the potential and factoring out the quantities independent of the summation yields:

$$\mathbf{A} = \frac{\mu_0}{4\pi r} \sum_a Q_a \mathbf{v}_a + \frac{\mu_0}{4\pi r^3} x_k \sum_a Q_a \mathbf{v}_a x_k^{(a)} + \cdots \tag{2.166}$$

The first term after the equals sign is zero (the sum of the contributions of all electric currents must vanish in the stationary case for an isolated charge system). In the next term, the resulting combination can be decomposed into symmetric and antisymmetric parts with respect to the velocity and position of the charged particle ($\mathbf{r}_a$, $\mathbf{v}_a$):

$$\begin{aligned} &\sum_a Q_a \mathbf{v}_a (\mathbf{r} \cdot \mathbf{r}_a) = \\ &= \sum_a Q_a \left(\frac{1}{2} \left[\mathbf{v}_a (\mathbf{r} \cdot \mathbf{r}_a) + \mathbf{r}_a (\mathbf{r} \cdot \mathbf{v}_a) \right] + \frac{1}{2} \left[\mathbf{v}_a (\mathbf{r} \cdot \mathbf{r}_a) - \mathbf{r}_a (\mathbf{r} \cdot \mathbf{v}_a) \right] \right) = \\ &= \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{2} \sum_a Q_a \mathbf{r}_a (\mathbf{r} \cdot \mathbf{r}_a) + \frac{1}{2} \sum_a Q_a \left(\mathbf{r}_a \times \mathbf{v}_a \right) \times \mathbf{r} . \end{aligned} \tag{2.167}$$

Due to the assumption of stationarity, the symmetric part vanishes (as shown later, this no longer holds for radiation), and the antisymmetric part yields a vector triple product:

$$\mathbf{A} = 0 + \frac{\mu_0}{4\pi r^3} \left(\sum_a \frac{1}{2} Q_a (\mathbf{r}_a \times \mathbf{v}_a) \right) \times \mathbf{r} + \cdots . \tag{2.168}$$

The first non-zero expansion term takes the form:

► $$\mathbf{A}^{(1)} = \frac{\mu_0}{4\pi r^3} \mathbf{p}_\mathrm{M} \times \mathbf{r} ; \qquad \mathbf{p}_\mathrm{M} \equiv \sum_a \frac{1}{2} Q_a (\mathbf{r}_a \times \mathbf{v}_a) \tag{2.169}$$

The quantity $\mathbf{p}_\mathrm{M}$ is called the magnetic dipole moment. It represents a generalization of the simple relation $\mathbf{p}_\mathrm{M} = I\,\mathbf{S}$ for a general localized system of charged particles. This formula is suitable for numerical simulations, where the magnetic dipole moment of the system can be calculated immediately from the known positions and velocities of the particles, and from it, the magnetization, i.e., the response of the system to an applied

magnetic field. It can be shown that for a single charged particle orbiting in a circle, the new definition of the magnetic moment leads to the previously used relation $\mathbf{p}_\mathrm{M} = I\,\mathbf{S}$. First, the magnitude of the magnetic moment is determined:

$$p_\mathrm{M} = \left| \sum_a \frac{1}{2} Q_a (\mathbf{r}_a \times \mathbf{v}_a) \right| = \frac{1}{2} QRv = \frac{1}{2} QR \frac{2\pi R}{T} = \frac{Q}{T} \pi R^2 = IS\,.$$

The symbol T denotes the orbital period of the charged particle, and R is the radius of the circle. Considering the area as a vector $\mathbf{S} = S\,\mathbf{n}$ yields the original relation.

2.3.5 Magnetic Dipole

The magnetic field of a dipole is determined as the curl of the potential (2.169) at a constant magnetic moment:

$$\mathbf{B} = \operatorname{curl} \mathbf{A}^{(1)} = \operatorname{curl} \frac{\mu_0}{4\pi r^3} \mathbf{p}_\mathrm{M} \times \mathbf{r}\,,$$

$$B_k = \frac{\mu_0}{4\pi} \varepsilon_{klm} \partial_l \varepsilon_{mop} p_o \frac{x_p}{r^3}.$$

Differentiating the ratio x_p/r^3 in the standard manner, transforming the Levi-Civita symbols into Kronecker deltas, and performing a straightforward calculation yields:

$$\mathbf{B} = \frac{\mu_0}{4\pi r^3} \left[3(\mathbf{p}_\mathrm{M} \cdot \mathbf{n})\mathbf{n} - \mathbf{p}_\mathrm{M} \right]; \qquad \mathbf{n} \equiv \frac{\mathbf{r}}{r}. \tag{2.170}$$

The shape of the magnetic field lines of a magnetic dipole is identical to the shape of the electric field lines of an electric dipole (2.130). Both fields decrease with the third power of the distance from the dipole and are mostly described in a coordinate system with the z-axis aligned with the dipole axis.

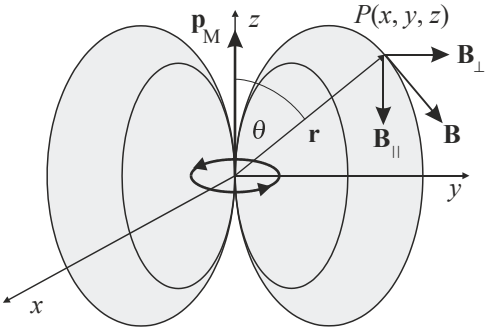


Fig. 2.20: Magnetic dipole field.

If the dipole axis coincides with the z-axis, the individual components take the values:

$$\mathbf{B} = \frac{\mu_0 p_\mathrm{M}}{4\pi r^3} \left(3n_z n_x, 3n_z n_y, 3n_z^2 - 1 \right); \tag{2.171}$$

$$B_{\perp} = \frac{\mu_0 p_{\mathrm{M}}}{4\pi r^3} 3\cos\theta\sin\theta\,;$$
$$B_{\parallel} = \frac{\mu_0 p_{\mathrm{M}}}{4\pi r^3}(3\cos^2\theta - 1)\,. \tag{2.172}$$

Relations analogous to those for the electric dipole can be derived for the magnetic dipole through similar procedures. The individual results are presented without derivation and are summarized together with the previous relations into a single table:

Quantity	Electric Dipole	Magnetic Dipole
Dipole Moment of the System	$\mathbf{p}_{\mathrm{E}} \equiv \sum_{k=1}^{N} Q_k \mathbf{r}_k$	$\mathbf{p}_{\mathrm{M}} \equiv \sum_{k=1}^{N} \frac{1}{2} Q_k \mathbf{r}_k \times \mathbf{v}_k$
Simplest Realization	Two opposite charges: $\mathbf{p}_{\mathrm{E}} = Q\,\mathbf{d}$	Orbiting charge: $\mathbf{p}_{\mathrm{M}} = I\,\mathbf{S}$
Energy in External Field	$U = -\,\mathbf{p}_{\mathrm{E}} \cdot \mathbf{E}$	$U = -\,\mathbf{p}_{\mathrm{M}} \cdot \mathbf{B}$
Torque Acting on Dipole	$\mathbf{M} = \mathbf{p}_{\mathrm{E}} \times \mathbf{E}$	$\mathbf{M} = \mathbf{p}_{\mathrm{M}} \times \mathbf{B}$
Potential of Dipole	$\phi = \frac{\mathbf{p}_{\mathrm{E}} \cdot \mathbf{r}}{4\pi\varepsilon_0 r^3}$	$\mathbf{A} = \frac{\mu_0}{4\pi} \frac{\mathbf{p}_{\mathrm{M}} \times \mathbf{r}}{r^3}$
Field of Dipole $\mathbf{n} = \mathbf{r}/r$	$\mathbf{E} = \frac{3(\mathbf{p}_{\mathrm{E}} \cdot \mathbf{n})\mathbf{n} - \mathbf{p}_{\mathrm{E}}}{4\pi\varepsilon_0 r^3}$	$\mathbf{B} = \frac{\mu_0}{4\pi r^3}\left[3(\mathbf{p}_{\mathrm{M}} \cdot \mathbf{n})\mathbf{n} - \mathbf{p}_{\mathrm{M}}\right]$
Field decomposition	$E_{\perp} = \frac{3p_{\mathrm{E}}}{4\pi\varepsilon_0 r^3}\cos\theta\sin\theta\,,$ $E_{\parallel} = \frac{p_{\mathrm{E}}}{4\pi\varepsilon_0 r^3}\left(3\cos^2\theta - 1\right).$	$B_{\perp} = \frac{3\mu_0 p_{\mathrm{M}}}{4\pi r^3}\cos\theta\sin\theta\,,$ $E_{\parallel} = \frac{\mu_0 p_{\mathrm{M}}}{4\pi r^3}\left(3\cos^2\theta - 1\right).$

Magnetic Dipole in Astronomy

In astronomy, the magnetic dipole moment of individual objects is presented in somewhat non-standard units. Since the dipole field decreases with the third power of the distance from the source, astronomers express the magnetic dipole moment as the product of the horizontal field component at the equator and the third power of the radius of the body. Such a quantity possesses the dimension Tm^3 and is proportional to the magnetic moment defined herein. Among the eight planets of the Solar System, only Venus and Mars lack a magnetic moment. The magnetic moments of the planets are:

Body	Magnetic Moment (Tm^3)	Body	Magnetic Moment (Tm^3)
Mercury	5×10^{12}	Jupiter	160×10^{18}
Venus	0	Saturn	4.6×10^{18}
Earth	8×10^{15}	Uranus	0.39×10^{18}
Mars	0	Neptune	0.21×10^{18}

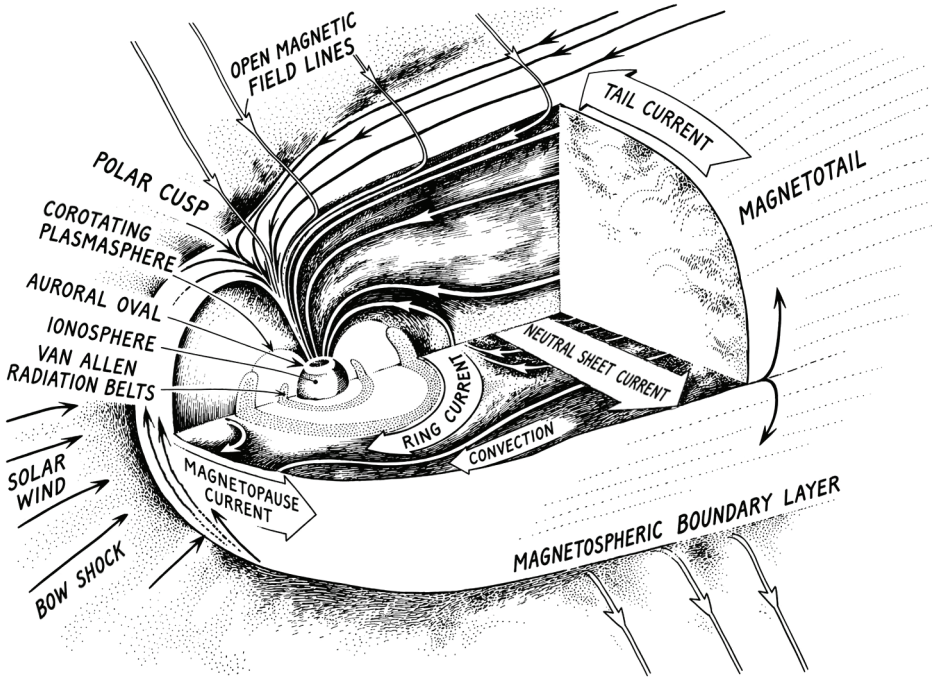


Fig. 2.21: Earth magnetosphere characteristically deformed by solar wind.

2.4 Conservation Laws

The electromagnetic field, introduced by English physicist Michael Faraday as a medium for transmitting interaction, is not merely a mathematical tool for describing electric and magnetic phenomena, but a real entity capable of transferring energy, momentum, and angular momentum from one location to another. The conservation laws for these quantities are encoded within the Maxwell equations, and this chapter provides a systematic overview of their properties. For convenient reference during the derivation, all Maxwell equations are rewritten below:

$$\operatorname{div}\mathbf{B} = 0\,, \tag{2.173}$$

$$\operatorname{div}\mathbf{D} = \rho_Q\,, \tag{2.174}$$

$$\operatorname{curl}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}\,, \tag{2.175}$$

$$\operatorname{curl}\mathbf{H} = \mathbf{j}_Q + \frac{\partial \mathbf{D}}{\partial t}\,. \tag{2.176}$$

2.4.1 Conservation of Energy

Conservation laws can be formulated locally, i.e., to hold at each point of spacetime. This formulation was previously demonstrated for charge, resulting in the continuity equation. This procedure can be performed for any extensive quantity, yielding a continuity equation of the form:

$$\frac{\partial \rho_A}{\partial t} + \operatorname{div}\mathbf{j}_A = 0\,. \tag{2.177}$$

The quantity ρ_A represents the volume density of A, with the unit [A]/m^3, such as mass density (kg/m^3), energy density (J/m^3), and momentum density (kg m s/m^3). The quantity $\mathbf{j}_A$ is the flux of A, signifying the amount flowing through a unit area per unit time along the direction of the flux, i.e., the velocity field. The unit of flux is [A]/(m^{2}s). For instance, the energy flux considered in this chapter has the dimension J·m^{-2}s^{-1}.

The electric field energy density is given by expression (2.117) as half of the scalar product of both electric vectors, namely $\mathbf{E}\cdot\mathbf{D}/2$. An example is the energy density in a capacitor. Analogously, the energy density in the magnetic field is given by the relation $\mathbf{H}\cdot\mathbf{B}/2$, with a solenoid serving as an example. The time variation of the energy density is calculated from the Maxwell equations (2.173) to (2.176) and modified into the form of a conservation law. To simplify the calculations, linear constitutive relations between both the electric and magnetic vectors are considered ($\mathbf{D} = \varepsilon\,\mathbf{E}$, $\mathbf{B} = \mu\,\mathbf{H}$):

$$\frac{\partial}{\partial t}\left(\frac{\mathbf{E}\cdot\mathbf{D}}{2} + \frac{\mathbf{H}\cdot\mathbf{B}}{2}\right) = \mathbf{E}\cdot\frac{\partial \mathbf{D}}{\partial t} + \mathbf{H}\cdot\frac{\partial \mathbf{B}}{\partial t} =$$

$$= \mathbf{E}\cdot\left(\operatorname{curl}\mathbf{H} - \mathbf{j}_Q\right) + \mathbf{H}\cdot\left(-\operatorname{curl}\mathbf{E}\right) =$$

$$= \mathbf{E}\cdot\operatorname{curl}\mathbf{H} - \mathbf{H}\cdot\operatorname{curl}\mathbf{E} - \mathbf{j}_Q\cdot\mathbf{E} =$$
$$= -\operatorname{div}\left(\mathbf{E}\times\mathbf{H}\right) - \mathbf{j}_Q\cdot\mathbf{E}\,.$$

The vector identity I.3.220 from the first volume of "*Selected Topics*" is utilized in the derivation. The resulting energy conservation law for the field takes the form of a continuity equation, the terms of which are readily identified:

► $$\frac{\partial u}{\partial t} + \nabla\cdot\mathbf{S} = -\mathscr{P}\,; \tag{2.178}$$

► $$u \equiv \rho_U \equiv \frac{\mathbf{E}\cdot\mathbf{D}}{2} + \frac{\mathbf{H}\cdot\mathbf{B}}{2}, \qquad \mathbf{S} \equiv \mathbf{j}_U = \mathbf{E}\times\mathbf{H}, \qquad \mathscr{P} \equiv \mathbf{j}_Q\cdot\mathbf{E}. \tag{2.179}$$

The time derivative contains the total field energy density u, consisting of the electric and magnetic components. The spatial derivative (within the divergence) represents the energy flux, which equals the vector product of both field intensities, $\mathbf{E}\times\mathbf{H}$. This flux is called the Poynting vector, named after English physicist John Henry Poynting, and is typically denoted by the symbol **S**. The energy always flows perpendicularly to both the electric field intensity and the magnetic field intensity. The right-hand side is non-zero, indicating that the electromagnetic field energy is not conserved on its own. This occurs because the electromagnetic field can transfer its energy to particles. For instance, in a microwave oven containing a meal, heating by the electromagnetic field takes place. The term $\mathbf{j}_Q\cdot\mathbf{E}$ on the right-hand side represents the Joule power density (W/m^3):

$$\mathbf{j}_Q\cdot\mathbf{E} = \rho_Q\,\mathbf{u}\cdot\mathbf{E} = Qn\,\mathbf{u}\cdot\mathbf{E} = n\,\mathbf{u}\cdot\mathbf{F} =$$
$$= n\,\mathbf{F}\cdot\mathbf{u} = n\frac{\mathbf{F}\cdot\mathrm{d}\mathbf{l}}{\mathrm{d}t} = n\,P = \mathscr{P}.$$

The Joule power describes the field energy transfer to particles. For the field, this represents a loss, explaining the minus sign. Only the energy conservation law for the entire system (field and particles) has zero on the right-hand side. The time derivative would then include both the field energy density and the particle energy density, while the divergence would be followed by both the field energy flux and the particle energy flux.

2.4.2 Conservation of Momentum

For scalar quantities, such as energy or charge, the corresponding fluxes are vectors. For vector quantities, however, the situation differs; the fluxes are second-rank tensors. To understand the underlying reasoning, the conservation of momentum of an ordinary conducting fluid is examined first, followed by the conservation of momentum of the electromagnetic field.

Conducting Fluid

For an object of mass m, the Newton motion equation is given by:

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \mathbf{F}\,. \tag{2.180}$$

Instead of a single particle a mass element of a flowing fluid experiencing a force d**F** is investigated. The velocity **v** of a single particle is replaced by the velocity field $\mathbf{u}(t, \mathbf{r})$:

$$\mathrm{d}m\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \mathrm{d}\mathbf{F}\,. \tag{2.181}$$

Dividing both sides of the equation by the volume element dV, yields the densities:

$$\rho\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \mathbf{f}\,. \tag{2.182}$$

The symbol ρ represents the mass density of the flowing fluid, and the symbol **f** represents the applied force density. Recognizing that the velocity field is a function of time and space, i.e., $\mathbf{u} = \mathbf{u}(t, \mathbf{r})$, yields:

$$\rho\frac{\partial\mathbf{u}}{\partial t} + \rho(\mathbf{u}\cdot\nabla)\mathbf{u} = \mathbf{f}\,. \tag{2.183}$$

The first term on the left-hand side expresses the explicit variation of the velocity field, while the second represents the variation caused by the flow. The force density on the right-hand side varies according to the described processes, arising from pressure, viscosity, or the Lorentz force. In this introductory example, viscous processes are omitted. The pressure energy is $\mathrm{d}W_\mathrm{p} = F\,\mathrm{d}x = pS\,\mathrm{d}x = p\,\mathrm{d}V$; pressure is therefore an energy density. The force density is given by the negative gradient of the potential energy density:

$$\mathbf{f}_\mathrm{p} = -\nabla p \tag{2.184}$$

To determine the Lorentz force density, a force element acting on a fluid element is:

$$\mathrm{d}\mathbf{F}_\mathrm{L} = \mathrm{d}Q\,\mathbf{E} + \mathrm{d}Q\,\mathbf{u}\times\mathbf{B}\,. \tag{2.185}$$

Dividing the equation by the volume element, i.e., transitioning to densities, yields:

► $$\mathbf{f}_\mathrm{L} = \rho_Q\mathbf{E} + \rho_Q\mathbf{u}\times\mathbf{B} = \rho_Q\mathbf{E} + \mathbf{j}_Q\times\mathbf{B}\,. \tag{2.186}$$

Both force densities are added to the right-hand side of the motion equation (2.183); the Lorentz force density is not expanded for now, as it represents pre-determined forces caused by external fields:

► $$\rho\frac{\partial\mathbf{u}}{\partial t} + \rho(\mathbf{u}\cdot\nabla)\mathbf{u} = -\nabla p + \mathbf{f}_\mathrm{L}\,. \tag{2.187}$$

This motion equation and the continuity equation for mass

► $$\frac{\partial\rho}{\partial t} + \mathrm{div}\,(\rho\mathbf{u}) = 0 \tag{2.188}$$

are sufficient to derive the conservation of momentum for a conducting fluid. Recall that ρ is the mass density and ρ_Q is the charge density. The mechanical momentum density of the fluid is given by:

$$\mathbf{g} \equiv \rho\,\mathbf{u}\,. \tag{2.189}$$

The time variation of one component of the momentum density is determined as:

$$\frac{\partial}{\partial t}g_k = \frac{\partial}{\partial t}(\rho u_k) = \frac{\partial\rho}{\partial t}u_k + \rho\frac{\partial u_k}{\partial t}\,.$$

Substituting for the time variation of the density from the continuity equation (2.188) and for the time variation of the velocity from the motion equation (2.187) yields:

$$\frac{\partial}{\partial t}(\rho u_k) = -\partial_l(\rho u_l)\, u_k - \rho(u_l \partial_l)\, u_k - \partial_k p + f_{\mathrm{L}k}.$$

Moving all terms containing derivatives to the left-hand side and rearranging yields:

$$\frac{\partial}{\partial t}(\rho u_k) + \frac{\partial(\rho u_l)}{\partial x_l} u_k + (\rho u_l)\frac{\partial u_k}{\partial x_l} + \frac{\partial p}{\partial x_k} = f_{\mathrm{L}k} \quad \Rightarrow$$

$$\frac{\partial}{\partial t}(\rho u_k) + \frac{\partial}{\partial x_l}\left(\rho u_l u_k\right) + \frac{\partial p}{\partial x_k} = f_{\mathrm{L}k} \quad \Rightarrow$$

$$\frac{\partial}{\partial t}(\rho u_k) + \frac{\partial}{\partial x_l}\left(p\delta_{kl} + \rho u_k u_l\right) = f_{\mathrm{L}k}\,.$$

The momentum conservation law has been obtained. The term in parentheses following the time derivative is the momentum density. The term following the spatial derivatives is the momentum flux, also called the pressure tensor. Since momentum is a vector quantity, its flux forms a second-rank tensor. The symmetry of the pressure tensor ensures the conservation of angular momentum in the flowing fluid. The pressure tensor consists of two parts: a scalar part, which represents the normal pressure acting equally in all directions, and a tensor part related to the fluid flow. The right-hand side is nonzero; the momentum of the conducting fluid is not conserved because it is acted upon by the Lorentz force density, which transfers momentum to it. The momentum conservation law with the expanded Lorentz force can be written in component notation as:

► $$\frac{\partial}{\partial t}(\rho u_k) + \frac{\partial}{\partial x_l}\left(T^{\mathrm{P}}_{kl}\right) = \rho_Q E_k + \varepsilon_{klm} j_{Ql} B_m\,; \qquad T^{\mathrm{P}}_{kl} \equiv p\delta_{kl} + \rho u_k u_l \qquad (2.190)$$

or in invariant form as:

► $$\frac{\partial}{\partial t}(\rho \mathbf{u}) + \vec{\nabla}\cdot\vec{\mathbf{T}}^{\mathrm{P}} = \rho_Q \mathbf{E} + \mathbf{j}_Q \times \mathbf{B}\,; \qquad \vec{\mathbf{T}}^{\mathrm{P}} \equiv p\vec{\mathbf{1}} + \rho \mathbf{u}\otimes\mathbf{u}\,. \qquad (2.191)$$

Electromagnetic Field

The electromagnetic field can be understood as a system of photons with zero rest mass and energy:

$$\mathcal{E} = pc\,. \qquad (2.192)$$

From relation (2.192), the relation between the energy density ρ_U and the field momentum density $\boldsymbol{\gamma}$ is straightforwardly derived:

$$\rho_U = \gamma c\,. \qquad (2.193)$$

For the momentum density $\boldsymbol{\gamma}$ as a vector, the expression yields:

$$\boldsymbol{\gamma} = \frac{\rho_U}{c}\frac{\mathbf{c}}{c}\,. \qquad (2.194)$$

The numerators contain the product of the energy density and the propagation velocity, i.e., the energy flux, which is given by the Poynting vector; therefore:

$$\boldsymbol{\gamma} = (\mathbf{E}\times\mathbf{H})/c^2\,. \qquad (2.195)$$

In the chapter on electromagnetic waves, the relation between the speed of light, permittivity, and permeability is derived ($c^2 = 1/\varepsilon\mu$), by which the expression is rewritten:

$$\boldsymbol{\gamma} = \mathbf{D}\times\mathbf{B}\,. \tag{2.196}$$

Notably, the vector product of the field intensities represents the energy flux, whereas the vector product of field displacements represents the momentum density. The objective to construct the electromagnetic field conservation law by finding the time derivative of relation (2.196), utilizing the Maxwell equations (2.173) to (2.176):

$$\frac{\partial\boldsymbol{\gamma}}{\partial t} = \frac{\partial}{\partial t}\left(\mathbf{D}\times\mathbf{B}\right) \;=\; \frac{\partial\mathbf{D}}{\partial t}\times\mathbf{B} + \mathbf{D}\times\frac{\partial\mathbf{B}}{\partial t} \;=\; \left(\operatorname{curl}\mathbf{H} - \mathbf{j}_{\mathrm{Q}}\right)\times\mathbf{B} \;+\; \mathbf{D}\times\left(-\operatorname{curl}\mathbf{E}\right) \;=\; \cdots$$

Standard modifications for the k-th component of the time variation of the momentum density follow, in which the terms with spatial derivatives are converted into the form of a divergence. This can be performed, for example, by rewriting the vector products using the Levi-Civita tensor. The result of elementary modifications is:

► $$\frac{\partial}{\partial t}\left(\mathbf{D}\times\mathbf{B}\right) + \vec{\nabla}\cdot\ddot{\mathbf{T}}^{\,\mathrm{EM}} = -\rho_Q\mathbf{E} - \mathbf{j}_Q\times\mathbf{B}\,; \tag{2.197}$$

► $$\ddot{\mathbf{T}}^{\,\mathrm{EM}} \equiv \frac{\mathbf{E}\cdot\mathbf{D}}{2}\ddot{\mathbf{1}} - \mathbf{E}\otimes\mathbf{D} + \frac{\mathbf{H}\cdot\mathbf{B}}{2}\ddot{\mathbf{1}} - \mathbf{H}\otimes\mathbf{B}, \tag{2.198}$$

► $$T_{kl}^{\mathrm{EM}} \equiv \frac{\mathbf{E}\cdot\mathbf{D}}{2}\delta_{kl} - E_k D_l + \frac{\mathbf{H}\cdot\mathbf{B}}{2}\delta_{kl} - H_k B_l\,. \tag{2.199}$$

In the conservation law (2.197), the electromagnetic field momentum density follows the time derivative. Following the spatial derivative (divergence) is the field momentum flux tensor $\mathbf{T}^{\mathrm{EM}}$, which is called the *Maxwell stress tensor* or the *pressure tensor*. It consists of an electric and a magnetic part, each possessing two terms – scalar (diagonal) and tensor. This is analogous to the particle pressure tensor, where the scalar pressure $p\delta_{kl}$ and the dynamic pressure $\rho u_k u_l$ are present. The momentum of the electromagnetic field is seen not to be conserved. This is due to the transfer of field momentum to particles. Only the total sum of the momentum of all particles and the field takes the form of a conservation law. It is obtained by adding equations (2.191) and (2.197):

► $$\frac{\partial}{\partial t}\left(\rho\,\mathbf{u} + \mathbf{D}\times\mathbf{B}\right) + \vec{\nabla}\cdot\left(\ddot{\mathbf{T}}^{\mathrm{P}} + \ddot{\mathbf{T}}^{\mathrm{EM}}\right) = 0\,. \tag{2.200}$$

The derived equation is the conservation of momentum for a conducting fluid or a system of charged particles and the electromagnetic field. The first term in the time derivative represents the matter momentum density $\mathbf{g} \equiv \rho\mathbf{u}$, which is simultaneously the mass flux from the continuity equation. The second term $\boldsymbol{\gamma} \equiv \mathbf{D}\times\mathbf{B}$ is the electromagnetic field momentum density. The spatial derivatives contain the particle momentum flux tensor and the electromagnetic field momentum flux tensor.

A time variation of momentum is a force, and a time variation of a momentum density is the Lorentz force density, which appears on the right-hand sides of the conservation laws (positive for particles and negative for fields). Only the sum of both equations yields zero on the right-hand side. The total result is rewritten below:

$$\frac{\partial \vec{\boldsymbol{\pi}}}{\partial t} + \vec{\nabla} \cdot \ddot{\mathbf{T}} = 0 ;$$

► $$\vec{\boldsymbol{\pi}} = \mathbf{g} + \boldsymbol{\gamma} = \rho \mathbf{u} + \mathbf{D} \times \mathbf{B} , \tag{2.201}$$

$$\ddot{\mathbf{T}} = \ddot{\mathbf{T}}^{\mathrm{P}} + \ddot{\mathbf{T}}^{\mathrm{EM}} = p \ddot{\mathbf{1}} + \rho \mathbf{u} \otimes \mathbf{u} + \frac{\mathbf{E} \cdot \mathbf{D}}{2} \ddot{\mathbf{1}} - \mathbf{E} \otimes \mathbf{D} + \frac{\mathbf{H} \cdot \mathbf{B}}{2} \ddot{\mathbf{1}} - \mathbf{H} \otimes \mathbf{B} .$$

In this total balance, the total particle and field momentum density $\boldsymbol{\pi}$ and the total particle and field momentum flux tensor $\mathbf{T}$ are introduced.

2.4.3 Field Pressure

The pressure that the electromagnetic field is capable of exerting on its surroundings is examined in this section.

Static Pressure

In a fluid, the scalar part of the momentum flux plays the role of ordinary pressure. The same holds for the field momentum flux (the Maxwell stress tensor). It contains two scalar parts, one magnetic and one electric, which can be identified with the magnetic and electric pressures:

► $$p_{\mathrm{M}} = \frac{1}{2} \mathbf{H} \cdot \mathbf{B} = \frac{\mathbf{B}^2}{2\mu} , \tag{2.202}$$

► $$p_{\mathrm{E}} = \frac{1}{2} \mathbf{E} \cdot \mathbf{D} = \frac{\mathbf{E}^2}{2\varepsilon} . \tag{2.203}$$

Notably, both pressures equal the corresponding field energy density. Similarly, in mechanics, mgh is the gravitational energy, and the density ρgh represents hydrostatic pressure. In sunspots, the local magnetic field reaches 0.3 tesla, corresponding to a pressure $B^2/2\mu_0$ of approximately 400,000 Pa. Around current-carrying plasma filaments, the generated magnetic field creates a pressure gradient that compresses the filament until balanced by the particle pressure gradient. The resulting equilibrium is analogous to spherical structures in gravitation, where objects are compressed by gravity. In a plasma, such configurations are usually unstable; stability is achieved only for currents flowing along a helical path, see the "*Selected Topics III*". Similarly, the electric field can exert enormous pressures responsible for matter transport and other phenomena.

Radiation Pressure

Electromagnetic radiation consists of regularly varying electric and magnetic fields. In statistical physics, radiation is derived to exert a pressure equal to one-third of the average energy density, see (1.166). This relation is derived here by simply averaging the electromagnetic field momentum tensor, the individual terms of which represent scalar and tensor pressures. Integrating the electromagnetic field momentum conservation law in a vacuum over an arbitrarily chosen volume yields (using the Gauss theorem):

$$\frac{\partial \boldsymbol{\gamma}}{\partial t} + \boldsymbol{\nabla} \cdot \mathbf{T}^{\mathrm{EM}} = 0 \qquad \Rightarrow \qquad \frac{\mathrm{d}}{\mathrm{d}t} \iiint_V \boldsymbol{\gamma} \, \mathrm{d}V + \oiint_{S=\partial V} \mathbf{T}^{\mathrm{EM}} \cdot \mathrm{d}\mathbf{S} = 0 .$$

Since the first term represents the time variation of momentum, i.e., the acting force, the field flux tensor must signify the force acting per unit area (this follows directly from its identification with pressure, which is also a force per unit area). Denoting the directional vector as **n**, the electromagnetic radiation pressure is:

$$p = \overline{T_{kl}^{\mathrm{EM}}} \langle \, n_k n_l \, \rangle \, . \tag{2.204}$$

The bar over the symbol denotes averaging over all possible values, and the angle brackets denote averaging over all possible directions. The averaging over angles is determined first:

$$\langle \, n_k n_l \, \rangle = \frac{1}{4\pi} \int n_k n_l \, \mathrm{d}\Omega = A \, \delta_{kl} \, . \tag{2.205}$$

The mean value of directional vectors over all directions must be independent of the choice of the coordinate system; it is therefore a diagonal symmetric tensor proportional to the Kronecker delta. The constant A is determined from the trace of the left- and right-hand sides:

$$\langle \, n_k n_k \, \rangle = A \, \delta_{kk} \qquad \Rightarrow \qquad \langle \, 1 \, \rangle = 3A \qquad \Rightarrow \qquad A = 1/3 \, .$$

Therefore:

$$\langle n_k n_l \rangle = \frac{1}{3} \delta_{kl} \, . \tag{2.206}$$

For the electromagnetic radiation pressure, the following expression is valid:

$$p = \overline{T_{kl}^{\mathrm{EM}}} \frac{1}{3} \delta_{kl} = \frac{1}{3} \overline{T_{kk}^{\mathrm{EM}}} =$$

$$= \frac{1}{3} \overline{\left(\frac{1}{2} \mathbf{E} \cdot \mathbf{D} \, \delta_{kk} - E_k D_k + \frac{1}{2} \mathbf{H} \cdot \mathbf{B} \, \delta_{kk} - H_k B_k \right)} \qquad \Rightarrow$$

$$p = \frac{1}{3} \overline{\left(\frac{3}{2} \mathbf{E} \cdot \mathbf{D} - \mathbf{E} \cdot \mathbf{D} + \frac{3}{2} \mathbf{H} \cdot \mathbf{B} - \mathbf{H} \cdot \mathbf{B} \right)} =$$

$$= \frac{1}{3} \overline{\left(\frac{1}{2} \mathbf{E} \cdot \mathbf{D} + \frac{1}{2} \mathbf{H} \cdot \mathbf{B} \right)} \qquad \Rightarrow$$

► $$p = \frac{1}{3} \bar{u} \, . \tag{2.207}$$

The electromagnetic radiation pressure equals one-third of its mean energy density. This pressure is capable of propelling solar sails, blowing away a comet tail, or creating a bubble around a newborn star in its stellar nebula..

2.4.4 Conservation of Angular Momentum

Angular momentum is standardly defined by the following expression:

$$\mathbf{b} \equiv \mathbf{r} \times \mathbf{p} \,. \tag{2.208}$$

In fact, this quantity represents a pseudovector rather than a vector, which is an antisymmetric matrix formed by the components:

$$M_{kl} = x_k p_l - x_l p_k \,.$$

Such a matrix possesses only three independent terms, which are interpreted as the components of the vector product:

$$\begin{aligned} b_1 &= M_{23} = x_2 p_3 - x_3 p_2 \,; \\ b_2 &= M_{31} = x_3 p_1 - x_1 p_3 \,; \\ b_3 &= M_{12} = x_1 p_2 - x_2 p_1 \,. \end{aligned} \tag{2.210}$$

The matrix M_{kl} can therefore be identified with the angular momentum. The angular momentum density of the electromagnetic field is now introduced by the relation:

► $$\mathcal{M}_{kl} = x_k \pi_l - x_l \pi_k \,, \tag{2.211}$$

meaning that the momenta p_k are replaced by their densities π_k. This represents the total momentum density of the field and particles, $\boldsymbol{\pi} = \mathbf{g} + \boldsymbol{\gamma}$. Analogously, $\mathbf{T} = \mathbf{T}^{\mathrm{P}} + \mathbf{T}^{\mathrm{EM}}$ is the total momentum flux tensor of the particles and fields. The time variation of the angular momentum density is now investigated. Expressing the time variations of the momentum density from the momentum conservation law (2.200) yields:

$$\frac{\partial}{\partial t}\mathcal{M}_{kl} = x_k \frac{\partial \pi_l}{\partial t} - x_l \frac{\partial \pi_k}{\partial t} = x_k \left(-\partial_n T_{nl}\right) - x_l \left(-\partial_n T_{nk}\right).$$

The derivatives are now rearranged according to the product rule $fg' = (fg)' - f'g$:

$$\frac{\partial}{\partial t}\mathcal{M}_{kl} = -\partial_n \left(x_k T_{nl}\right) + T_{nl}\delta_{nk} + \partial_n \left(x_l T_{nk}\right) - T_{nk}\delta_{nl} \quad \Rightarrow$$

$$\frac{\partial}{\partial t}\mathcal{M}_{kl} = -\partial_n \left(x_k T_{ln} - x_l T_{nk}\right) + T_{kl} - T_{lk} \,.$$

As previously noted, the symmetry of the momentum flux tensor is related to the conservation of angular momentum. For a symmetric tensor, the last two terms cancel, yielding the conservation of angular momentum for the electromagnetic field and the particles interacting with it:

$$\frac{\partial}{\partial t}\mathcal{M}_{kl} + \partial_n \mathcal{N}_{kln} = 0 \,;$$

► $$\begin{aligned} \mathcal{M}_{kl} &\equiv x_k \pi_l - x_l \pi_k \,, \\ \mathcal{N}_{kln} &\equiv x_k T_{ln} - x_l T_{kn} \,. \end{aligned} \tag{2.212}$$

The angular momentum flux $\mathcal{N}_{kln}$ is a third-rank tensor. If the momentum flux tensor T_{kl} were not symmetric, the conservation of angular momentum would not hold, and the fluid or field could spontaneously begin to rotate, which contradicts experimental observation.

2.4.5 Field as a Real Entity

Consider two charged particles, such as two electrons, separated by a distance of one meter. Moving the first particle causes this action to affect the second particle after a certain time interval, leading to its motion as well. However, this effect is not instantaneous. The motion of the first electron generates an electromagnetic field, transferring a certain portion of energy, momentum, and angular momentum to it. These three attributes are carried by the field at the speed of light to the second particle. The field is thus demonstrated to be physical reality rather than a mere mathematical construction facilitating calculations. The field is a carrier of energy, momentum, and angular momentum.



Fig. 2.22: Transmission of energy, momentum, and angular momentum by the field.



2.5 Electromagnetic Waves

2.5.1 Basic Concepts

Let a quantity whose values vary in time and space be denoted by $\psi(t, \mathbf{x})$. Whether this quantity is scalar or structured, such as a vector, is immaterial; examples include pressure, medium density, temperature, velocity, electric or magnetic fields, and ocean wave height. Several concepts utilized in wave theory are introduced below.

Wave Function

The wave function $\psi(t, \mathbf{x})$ describes a wave in time and space. Setting t = const yields a temporal snapshot of the wave, analogous to viewing a photograph of a wavy ocean surface. Setting $\mathbf{x}$ = const yields the time evolution of the monitored quantity at a single specific location. Waves are generally described by a complex wave function, as the use of complex numbers significantly simplifies calculations. However, only the real part of the wave function typically possesses physical significance. As with any complex function, the wave function can be written using two real functions, the amplitude A and the phase φ:

► $$\psi(t, \mathbf{x}) = A(t, \mathbf{x})\, \mathrm{e}^{\mathrm{i}\,\varphi(t, \mathbf{x})} . \tag{2.213}$$

Wavefront

A surface connecting points with a constant value of the phase φ of the wave function is called a wavefront. On a wavefront, the wave is in the same phase (for example, a wavefront connecting points where the pressure is 75 % of the maximum value).

Angular Frequency

Angular frequency is the time derivative of phase:

► $$\omega \equiv -\frac{\partial \varphi}{\partial t} . \tag{2.214}$$

The minus sign in the definition is not essential; it merely ensures that a plane wave propagates in the direction of the wave vector. This sign is also useful later for the correct relativistic formulation of relations. The angular frequency can vary both with time and from place to place. If the angular frequency is constant, its magnitude can be written in terms of the period T as $\omega = 2\pi/T$.

Wave Vector

The wave vector is defined as the spatial gradient of the wave phase.

► $$\mathbf{k} \equiv \frac{\partial \varphi}{\partial \mathbf{x}} = \nabla \varphi . \tag{2.215}$$

The wave vector, being a gradient, points perpendicularly to the wavefront, i.e., in the direction of wave propagation. Both its magnitude and direction can vary with time and from place to place. If the wave vector is constant, its magnitude can be written in terms of the wavelength λ as $k = 2\pi/\lambda$.

Dispersion Relation

At each point, a wave is described by four numbers (ω, $\mathbf{k}$), forming a relativistic four-vector, as demonstrated later. These dependent numbers satisfy a relation derived from the specific wave equations. Generally, the dependence takes the form

$$\phi(\omega, \mathbf{k}) = 0 \tag{2.216}$$

and is called a *dispersion relation*. In certain cases, the angular frequency can be explicitly calculated from the dispersion relation as a function of the wave vector, or the wave vector as a function of the angular frequency:

$$\omega = \omega(\mathbf{k}), \tag{2.217}$$

$$\mathbf{k} = \mathbf{k}(\omega). \tag{2.218}$$

Where explicit calculation is impossible, the implicit function theorem can be applied to determine the dispersion relation in the form $\omega(\mathbf{k})$ at least locally.

Plane (Monochromatic) Wave

This type of wave represents the simplest form, possessing a constant amplitude and a phase that is a linear function of time and space:

$$\begin{gathered} A(t, \mathbf{x}) = A; \\ \varphi(t, \mathbf{x}) = c_0 t + c_1 x + c_2 y + c_3 z = -\omega t + k_x x + k_y y + k_z z = \mathbf{k} \cdot \mathbf{x} - \omega t . \end{gathered} \tag{2.219}$$

The significance of the coefficients c_k follows from the definitions of the angular frequency and the wave vector. The term monochromatic signifies that a single frequency or color (chromos) is represented. The plane wave thus takes the form:

$$\psi(t, \mathbf{x}) = A\, \mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{x} - \omega t]} . \tag{2.220}$$

It is immediately apparent that surfaces of constant phase $\varphi(t, \mathbf{x}) = \text{const}$ represent the equations of moving planes:

$$\begin{gathered} \varphi(t, \mathbf{x}) = \text{const} \quad \Rightarrow \\ k_x x + k_y y + k_z z - \omega t = \text{const} \quad \Rightarrow \\ ax + by + cz + d(t) = 0 . \end{gathered}$$

The movement of the plane is perpendicular to itself, as oblique movements can be replaced by a perpendicular one with a velocity equal to the velocity projection onto the normal. The direction of movement is determined as the gradient of the plane equation:

$$\varphi(t, \mathbf{x}) = k_x x + k_y y + k_z z - \omega t \quad \Rightarrow \quad \nabla\varphi = \mathbf{k} .$$

The wave vector therefore points in the direction of wave propagation.

Phase Velocity

Phase velocity is the speed at which a surface of constant phase moves. Aligning the x-axis with the direction of plane propagation yields $\mathbf{k} = (k, 0, 0)$:

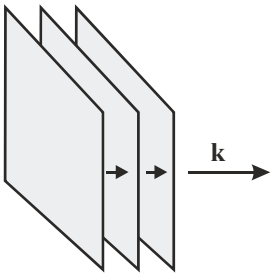


Fig. 2.23: Plane wavefront.

Differentiating the equation of a surface of constant phase yields the propagation velocity of the surface (the phase velocity):

$$kx - \omega t = \text{const} \qquad \Rightarrow \qquad k\,\mathrm{d}x - \omega\,\mathrm{d}t = 0 \qquad \Rightarrow \qquad v_\mathrm{f} = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\omega}{k}.$$

For a general choice of the coordinate system, the following expressions hold:

► $$v_\mathrm{f} = \frac{\omega}{k}; \qquad \mathbf{v}_\mathrm{f} = \frac{\omega}{k}\mathbf{e}_\mathbf{k}. \tag{2.221}$$

The first expression determines only the magnitude of the phase velocity, while the second indicates that the phase velocity vector points in the direction of the wave vector. Phase velocity is related solely to the movement of a location with identical wave phase; it is unrelated to the actual macroscopic transport of matter (ripples propagating on a water surface possess a different velocity than the water itself). The phase velocity can be, and in many cases is, superluminal. The form of the dispersion relation determines the value of the phase velocity for different frequencies. The phenomenon where waves of different frequencies propagate at different velocities is called *dispersion*.

General Wave

Plane waves are straightforward to manipulate, and a wave of a more general shape can be constructed from them:

$$\psi(t, \mathbf{x}) = \int a(\omega, \mathbf{k})\, \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{x} - \omega t)} \mathrm{d}^3\mathbf{k}\,. \tag{2.222}$$

This expression represents the Fourier transform $\psi(t, \mathbf{x}) \leftrightarrow a(\omega, \mathbf{k})$. The wave amplitudes are the Fourier transform of the wave function. Integration is performed only over the components of the wave vector. Since the angular frequency depends on the wave vector via the dispersion relation (2.217), it is not integrated over. Formally, the integration can be written four-dimensionally using the Dirac distribution:

$$\psi(t, \mathbf{x}) = \int a(\omega, \mathbf{k})\, \mathrm{e}^{\mathrm{i}(\mathbf{k}\cdot\mathbf{x} - \omega t)} \delta[\omega - \omega(\mathbf{k})]\, \mathrm{d}\omega \mathrm{d}^3\mathbf{k}\,. \tag{2.223}$$

Group Velocity

The velocity of a wave packet – a cluster of waves with similar frequencies and wave vectors – is now investigated. For simplicity, a packet propagating along the x-axis is considered (defining the coordinate system choice):

$$\psi(t,x) = \int_{k_0-\Delta k}^{k_0+\Delta k} a(\omega,k)\,\mathrm{e}^{\mathrm{i}(kx-\omega t)}\,\mathrm{d}k\,. \tag{2.224}$$

The wave amplitude is non-zero only in the interval ($k_0 - \Delta k$, $k_0 + \Delta k$) and is replaced by a constant amplitude:

$$\psi(t,x) \approx a(\omega_0,k_0) \int_{k_0-\Delta k}^{k_0+\Delta k} \mathrm{e}^{\mathrm{i}(kx-\omega t)}\,\mathrm{d}k\,.$$

In the next step, the central wave is factored out of the integral:

$$\psi(t,x) \approx a(\omega_0,k_0)\,\mathrm{e}^{\mathrm{i}(k_0x-\omega_0 t)} \int_{k_0-\Delta k}^{k_0+\Delta k} \mathrm{e}^{\mathrm{i}[(k-k_0)x-(\omega-\omega_0)t]}\,\mathrm{d}k\,.$$

Crucially, $\omega = \omega(k)$ holds, and integration is implicitly performed over ω as well. Further modifications are straightforward:

$$\psi(t,x) \approx a(\omega_0,k_0)\,\mathrm{e}^{\mathrm{i}(k_0x-\omega_0 t)} \int_{k_0-\Delta k}^{k_0+\Delta k} \exp\left[\mathrm{i}(k-k_0)\left(x-\frac{\omega(k)-\omega_0}{k-k_0}t\right)\right]\mathrm{d}k\,.$$

The fraction in the exponent can be replaced by a derivative (for $\Delta k \to 0$):

$$\frac{\omega(k)-\omega_0}{k-k_0} \approx \left.\frac{\partial\omega}{\partial k}\right|_{k_0} = v_\mathrm{g}(k_0)\,.$$

The quantity v_g merely serves as notation for the partial derivative defined above. The wave packet now takes the form:

$$\psi(t,x) \approx a(\omega_0,k_0)\,\mathrm{e}^{\mathrm{i}(k_0x-\omega_0 t)} \int_{k_0-\Delta k}^{k_0+\Delta k} \exp\left[\mathrm{i}(k-k_0)\left(x-v_\mathrm{g}\,t\right)\right]\mathrm{d}k\,.$$

Integration over the wave vector evidently yields a function of $x - v_\mathrm{g}t$:

$$\psi(t,x) \approx a(\omega_0,k_0)\,\mathrm{e}^{\mathrm{i}(k_0x-\omega_0 t)}\; F(x-v_\mathrm{g}t) = A(x-v_\mathrm{g}t)\,\mathrm{e}^{\mathrm{i}(k_0x-\omega_0 t)}\,.$$

The wave packet possesses an envelope propagating at the group velocity v_g. The inner packet is filled with a carrier wave $\exp[\mathrm{i}(k_0x - \omega t)]$. For simplicity, the calculation was performed in one dimension. For a generally oriented wave vector, the expression is:

► $$\mathbf{v}_\mathrm{g} = \frac{\partial\omega}{\partial\mathbf{k}} = \left(\frac{\partial\omega}{\partial k_x}, \frac{\partial\omega}{\partial k_y}, \frac{\partial\omega}{\partial k_z}\right). \tag{2.225}$$

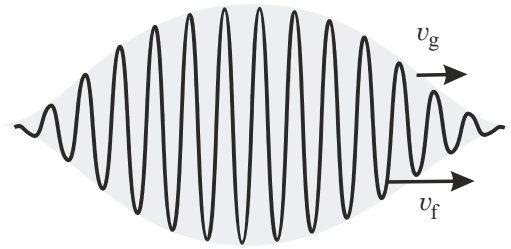


Fig. 2.24: Wave packet.

The propagation velocity of the wave packet as a whole is called the *group velocity*, representing the speed of information and energy transfer. It must be subluminal. Utilizing the Louis de Broglie matter-wave relations and the Hamilton canonical equations $dq_k/dt = \partial H/\partial p_k$, this velocity is straightforwardly shown to be the mechanical velocity of a particle quantum-mechanically associated with the wave packet:

$$\mathbf{v}_g = \frac{\partial \omega}{\partial \mathbf{k}} = \frac{\partial \hbar\omega}{\partial \hbar\mathbf{k}} = \frac{\partial E}{\partial \mathbf{p}} = \frac{\partial H}{\partial \mathbf{p}} = \mathbf{v}_{\text{mech}}.$$

Graphical Interpretation of Phase and Group Velocity

The graphical interpretation is illustrated in the figure. The phase velocity is given by the tangent of the angle between the horizontal axis and the line connecting a dispersion curve point to the origin, whereas the group velocity equals the tangent slope:

$$v_f = c\,\mathrm{tg}\,\gamma_1\,; \tag{2.226}$$

$$v_g = c\,\mathrm{tg}\,\gamma_2\,. \tag{2.227}$$

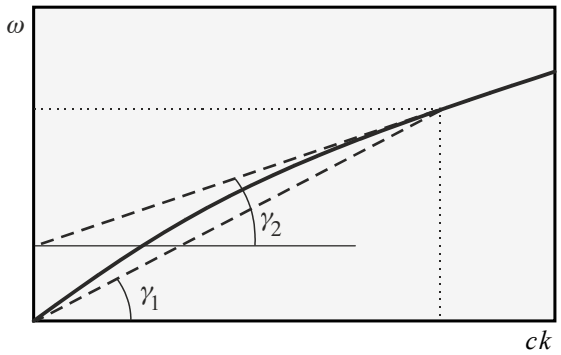


Fig. 2.25: Graphical interpretation of phase and group velocity.

Calculation Scheme

Arranging a theoretical model into a system of equations constitutes a major advantage. The basic calculation procedure is presented below:

1. Some variables can be eliminated from the system to reduce the total number of variables. The ideal outcome is to obtain a single equation for a single unknown. If the model describes wave propagation, the resulting relation is a form of wave equation. This procedure can be interrupted at any point to proceed to the next step.

2. If the system is non-linear, linearization is performed. The solution is written as a known baseline solution with an added small first-order perturbation. Powers higher than the first order are neglected. Since the Maxwell system of equations is linear, this step can be omitted.

3. The general solution is composed of plane waves according to relation (2.222). Due to the linearity of the system, each plane wave can be examined separately. The complete solution can be reconstructed from these partial waves at any later stage. The plane waves are exceptionally straightforward to manipulate. Differentiating the plane wave with respect to the temporal and spatial variables yields:

$$\frac{\partial}{\partial t}\psi = \frac{\partial}{\partial t} A\mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{x}-\omega t]} = -\mathrm{i}\omega A\mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{x}-\omega t]} = -\mathrm{i}\omega\psi\,,$$

$$\frac{\partial}{\partial x_l}\psi = \frac{\partial}{\partial x_l} A\mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{x}-\omega t]} = +\mathrm{i}k_l A\mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{x}-\omega t]} = +\mathrm{i}k_l\psi\,.$$

The partial derivatives for a plane wave are seen to transform into algebraic expressions. Any combination of partial derivatives can be replaced by algebraic expressions resulting from both relations. These are arranged in a well-organized table:

Expression	Example
$\frac{\partial}{\partial t} \to -\mathrm{i}\omega$	$\frac{\partial f}{\partial t} = -\mathrm{i}\omega f$
$\frac{\partial}{\partial x_l} \to +\mathrm{i}k_l$	$\frac{\partial f}{\partial x_l} = +\mathrm{i}k_l f$
$\boldsymbol{\nabla} \to +\mathrm{i}\mathbf{k}$	$\boldsymbol{\nabla} f = +\mathrm{i}\mathbf{k} f$
$\mathrm{div} \to \mathrm{i}\mathbf{k}\cdot$	$\mathrm{div}\,\mathbf{V} = \mathrm{i}\mathbf{k}\cdot\mathbf{V}$
$\mathrm{curl} \to \mathrm{i}\mathbf{k}\times$	$\mathrm{curl}\,\mathbf{V} = \mathrm{i}\mathbf{k}\times\mathbf{V}$
$\boldsymbol{\nabla}^2 \to -k^2$	$\boldsymbol{\nabla}^2 f = -k^2 f$

(2.228)

According to these rules, the initial system is converted into an algebraic system of equations, which is more straightforward to manipulate. This step is equivalent to performing a Fourier transform.

4. Since a non-zero solution is sought, the determinant of the system must vanish. The resulting system is assumed to have no right-hand side, which is typically the case. From this condition, the relation between ω and $\mathbf{k}$, i.e., the dispersion relation, is obtained. Reducing the number of system variables by elimination, thereby lowering the order of the calculated determinant, is often advantageous. Reducing the number of variables can be performed either before applying the Fourier transform rules or afterward within the algebraic system.

5. If the dispersion relation is complex, analyzing the stability or instability of the found solution is appropriate. A complex angular frequency or wave vector implies that exponential, non-oscillating terms appear in the expression $\exp[\mathrm{i}(\mathbf{k}\cdot\mathbf{x} - \omega t)]$. These terms can lead to attenuation or exponential growth of the solution.

6. Calculating the angular frequency from the dispersion relation is attempted. The phase and group velocities of the waves are determined from (2.221) and (2.225).
7. Returning to the original system, relations between quantities, mutual vector directions, and whether the wave is transverse or longitudinal are investigated.

2.5.2 Waves in Vacuum

As a simple example of the described procedure, electromagnetic waves in vacuum are analyzed, where $\rho_Q = 0$, $\mathbf{j}_Q = 0$, and the constitutive relations reduce to $\mathbf{D} = \varepsilon_0\mathbf{E}$, $\mathbf{B} = \mu_0\mathbf{H}$. In system (2.46) to (2.49), vectors **E** and **B** are retained:

$$\begin{aligned} \operatorname{div}\mathbf{E} &= 0, \\ \operatorname{div}\mathbf{B} &= 0, \\ \operatorname{curl}\mathbf{B} &= \varepsilon_0\mu_0\,\partial\mathbf{E}/\partial t, \\ \operatorname{curl}\mathbf{E} &= -\partial\mathbf{B}/\partial t. \end{aligned} \tag{2.229}$$

Two solution procedures are demonstrated. In the first, variables are eliminated before the Fourier transform; in the second, elimination is performed after the transform.

Procedure 1

The magnetic field is eliminated from the Maxwell equations to obtain an equation for the electric field. The curl operation is applied to the fourth equation, and the third equation is substituted for curl **B** on the right side:

$$\operatorname{curl}\operatorname{curl}\mathbf{E} = -\frac{\partial \operatorname{curl}\mathbf{B}}{\partial t} \quad \Rightarrow \quad \operatorname{curl}\operatorname{curl}\mathbf{E} = -\varepsilon_0\mu_0\frac{\partial^2\mathbf{E}}{\partial t^2} \quad \Rightarrow$$

$$\left(\operatorname{grad}\operatorname{div} - \boldsymbol{\nabla}^2\right)\mathbf{E} = -\varepsilon_0\mu_0\frac{\partial^2\mathbf{E}}{\partial t^2}.$$

Since div **E** = 0, the resulting equation is obtained:

$$\boldsymbol{\nabla}^2\mathbf{E} - \varepsilon_0\mu_0\frac{\partial^2\mathbf{E}}{\partial t^2} = 0. \tag{2.230}$$

This is the well-known wave equation for the electric field. Analogously, the same equation for the magnetic field is obtained by eliminating the electric field. Now, the Fourier transform is performed according to the rules presented on the previous page:

$$\left(-k^2 + \varepsilon_0\mu_0\omega^2\right)\mathbf{E} = 0.$$

The partial differential equation is converted into an algebraic equation without a right-hand side. A non-zero solution exists only if:

$$-k^2 + \varepsilon_0\mu_0\omega^2 = 0 \quad \Rightarrow$$

► $$\omega(k) = \sqrt{\frac{1}{\varepsilon_0 \mu_0}}\, k\,. \tag{2.231}$$

The dispersion relation is derived from the condition for a non-zero solution. The phase velocity of propagation (the speed of light) is given by:

► $$c \equiv v_{\mathrm{f}} = \frac{\omega}{k} = \sqrt{\frac{1}{\varepsilon_0 \mu_0}}\,. \tag{2.232}$$

The obtained dispersion relation of the form $\omega = ck$ is the simplest possible (linear), the phase and group velocities are identical, and the wave exhibits no dispersion, i.e., the velocity is independent of the wavelength or the wave vector.

Procedure 2

It is assumed that the elimination of the equation for the electric or magnetic field from the system of Maxwell equations was not performed initially. Therefore, the Fourier transform is applied directly to the original system of equations (2.229):

$$\begin{aligned} \mathbf{k}\cdot\mathbf{E} &= 0\,, \\ \mathbf{k}\cdot\mathbf{B} &= 0\,, \\ \mathbf{k}\times\mathbf{B} &= -\omega\varepsilon_0\mu_0\mathbf{E}\,, \\ \mathbf{k}\times\mathbf{E} &= +\omega\,\mathbf{B}\,. \end{aligned} \tag{2.233}$$

The elimination of variables can be performed at this stage. Magnetic field **B** from the last equation is substituted into the second to last equation:

$$\frac{1}{\omega}\mathbf{k}\times(\mathbf{k}\times\mathbf{E}) = -\omega\varepsilon_0\mu_0\mathbf{E} \quad \Rightarrow$$
$$\mathbf{k}\,(\mathbf{k}\cdot\mathbf{E}) - k^2\mathbf{E} = -\omega^2\,\varepsilon_0\,\mu_0\mathbf{E}\,.$$

The first term is zero according to the first equation in (2.233), and the equation for the electric field is therefore:

$$\left(k^2 - \varepsilon_0\mu_0\,\omega^2\right)\mathbf{E} = 0\,.$$

The condition for a non-zero electric field is again the dispersion relation:

$$k^2 - \varepsilon_0\mu_0\omega^2 \;=\; 0 \qquad \Rightarrow \qquad \omega(k) = \sqrt{\frac{1}{\varepsilon_0\mu_0}}\;k \;=\; ck\,.$$

From the original system (2.233), it is easily determined that vectors **E**, **B**, and **k** are mutually perpendicular; therefore, the electromagnetic wave is transverse. Vectors **E** and **B** are not independent in an electromagnetic wave; the equation curl $\mathbf{E} = -\partial\mathbf{B}/\partial t$ yields $kE = \omega B$ after the Fourier transform, and thus:

► $$\frac{E}{B} = \frac{\omega}{k} = v_f \tag{2.234}$$

This relation has a deeper significance. The ratio of the electric field to the magnetic field always represents a characteristic velocity in a given system. For instance, in

plasma, it corresponds to the drift velocity of particles; for an electromagnetic wave, it corresponds to the phase velocity of propagation. In the case of vacuum, the phase velocity equals c. In a material medium, the refractive index is introduced

► $$n \equiv \frac{c}{v_\mathrm{f}}, \tag{2.235}$$

and the ratio of both fields in an electromagnetic wave is equal to $E/B = v_\mathrm{f} = c/n$.

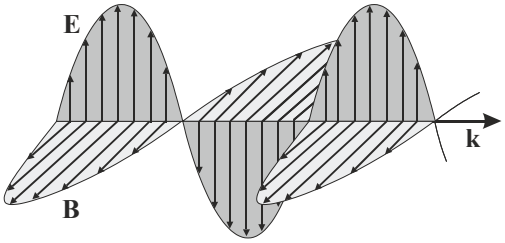


Fig. 2.26: Conditions in a plane electromagnetic wave.

Polarization

The polarization of an electromagnetic wave is defined as the plane of oscillation of the electric field. This polarization can be *planar* (the darker plane in the figure), where the electric field vector oscillates within a fixed plane. However, the plane of oscillation can also rotate (and with it, the plane of oscillation of the magnetic field). Polarization can be, for example, *circular* (the endpoint of the electric field traces a circle) or *chaotic* (the electromagnetic wave consists of many emission acts of various polarizations).

Note 1: The calculation stage for the Fourier transform is not critical, as both procedures lead to the same result. If partial differential equations are difficult to manipulate, it is advantageous to perform the transform as early as possible. If the elimination of variables fails both before and after the transform, the condition for a non-zero solution is the vanishing of the determinant of the entire system.

Note 2: The speed of light depends solely on the permittivity and permeability of vacuum and is identical in all coordinate systems. This stems from the Maxwell equations having different transformation properties from classical mechanics, where the velocities add up. The discrepancy between classical mechanics and electrodynamics led to a series of experiments which demonstrated that the behavior of nature corresponds to Maxwell electrodynamics. Therefore, classical mechanics had to be modified to be consistent with electrodynamics. This modification was performed by Albert Einstein in 1905 and is called the special theory of relativity. It leads to the currently well-known length contraction and time dilation.

Note 3: It follows from the Maxwell equations that electromagnetic waves can propagate even in vacuum, far from the sources from which they originated.

Note 4: Maxwell demonstrated from his equations that light is an electromagnetic wave and that electromagnetic waves of shorter and longer wavelengths than light must also exist. Such waves were discovered later; for example, radio waves were discovered by Heinrich Hertz in 1888.

2.5.3 Waves in Anisotropic Media

Material properties can introduce significant anisotropy into the propagation of electromagnetic waves. A similar effect can be caused by an external electric field or, more frequently, a magnetic field. This section briefly addresses anisotropic materials. For simplicity, a linear response and an exclusively electrically anisotropic medium (such as ordinary salt) are assumed, where $D_k = \varepsilon_{kl} E_l$:

$$\begin{pmatrix} D_1 \\ D_2 \\ D_3 \end{pmatrix} = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{pmatrix} \cdot \begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix}. \tag{2.236}$$

Vector **D** no longer points in the direction of vector **E**; the permittivity is not a single scalar but a set of nine components forming a second-rank permittivity tensor. Fundamental physical considerations, which lie beyond the scope of this course, imply that the permittivity tensor must be symmetric, i.e., $\varepsilon_{kl} = \varepsilon_{lk}$. Symmetric matrices possess real eigenvalues and three mutually perpendicular eigenvectors, serving as a natural coordinate system. In a crystalline substance, for instance, these correspond to axes along which wave propagation is substantially simpler than in other directions. If the medium is described in such a coordinate system, the permittivity tensor contains only diagonal components, resulting in three permittivity values differing along the three axes:

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 & 0 & 0 \\ 0 & \varepsilon_2 & 0 \\ 0 & 0 & \varepsilon_3 \end{pmatrix} \tag{2.237}$$

Varying permittivity in different directions implies distinct propagation velocities of the electromagnetic wave and different refractive indices:

$$n_k \equiv \frac{c}{v_k} = \frac{1/\sqrt{\varepsilon_0 \mu_0}}{1/\sqrt{\varepsilon_k \mu_0}} = \sqrt{\frac{\varepsilon_k}{\varepsilon_0}} \tag{2.238}$$

If all three permittivities, and consequently also the corresponding refractive indices, are distinct $n_1 \neq n_2 \neq n_3$, the medium is referred to as a *biaxial crystal*. In many crystals, however, two of the three refractive indices are identical, for example:

$$n_1 = n_2 \neq n_3 \,; \tag{2.239}$$

In this case, the crystals are referred to as *uniaxial*, where both identical indices are called the *ordinary* refractive index and the third, distinct index is called the *extraordinary* refractive index:

► $$\begin{aligned} n_\mathrm{o} &= n_1 = n_2 \,, \\ n_\mathrm{e} &= n_3 \end{aligned} \tag{2.240}$$

Examples of refractive indices for selected uniaxial crystals are presented in the table:

Crystal	n_o	n_e
Quartz	1.544	1.553
Rutile	2.616	2.903
Iceland Spar	1.658	1.486
Tourmaline	1.669	1.638
Beryl	1.598	1.590

Novel properties are encountered in electrically anisotropic media. Two wavefronts appear simultaneously in crystals, corresponding to the propagation of the ordinary and extraordinary waves. This results in two wave vectors and a double image of objects viewed through a birefringent crystal. However, an axis always exists along which both wave vectors coincide and the wavefronts touch. This axis is called the *optical axis*.

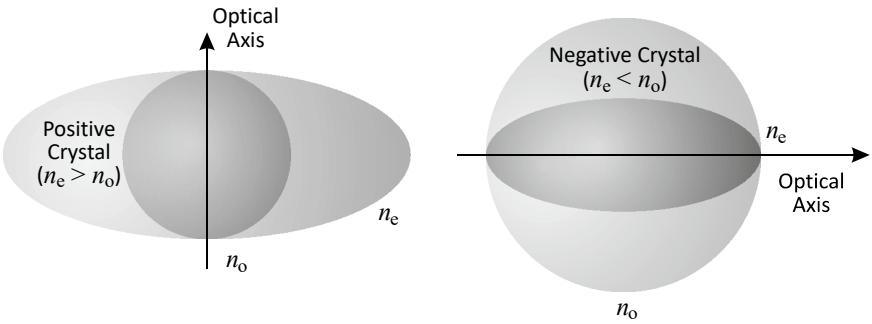


Fig. 2.27: Two most common types of birefringent crystals.

The directions of individual vectors are obtained from the Maxwell equations, where space charges and flowing currents are neglected, both electric and both magnetic vectors are retained, and the Fourier transform is performed:

$$\operatorname{div}\mathbf{B}=0\,;\quad \operatorname{div}\mathbf{D}=0\,;\quad \operatorname{curl}\mathbf{E}=-\frac{\partial\mathbf{B}}{\partial t};\quad \operatorname{curl}\mathbf{H}=\frac{\partial\mathbf{D}}{\partial t}\quad\Rightarrow,$$

$$\mathbf{k}\cdot\mathbf{B}=0\,;\quad \mathbf{k}\cdot\mathbf{D}=0\,;\quad \mathbf{k}\times\mathbf{E}=\omega\,\mathbf{B}\,;\quad \mathbf{k}\times\mathbf{H}=-\omega\mathbf{D}\,. \tag{2.241}$$

Wave vector **k** is perpendicular to vectors **D** and **B**, meaning it points in the direction of the momentum density **D**×**B** and thus the phase velocity of the wave. Conversely, the group velocity points in the direction of energy flux **E**×**H**. The ray is defined as the direction of energy propagation, which is no longer perpendicular to the wavefronts.

- Phase velocity and wave vector align with the momentum density **D**×**B**;
- Group velocity and ray point in the direction of the energy flux **E**×**H**;
- Optical axis is a direction characterized by a single wave vector.
- Crystals with $n_1 \neq n_2 \neq n_3$ also exist.

When a ray enters a birefringent crystal, the original wave splits into two components: the ordinary wave and the extraordinary wave. The ordinary wave refracts according to the refractive index n_o, while the extraordinary wave refracts according to the refractive index n_e. Both new waves are polarized, and their polarizations (planes of oscillation of the electric field) are mutually perpendicular.

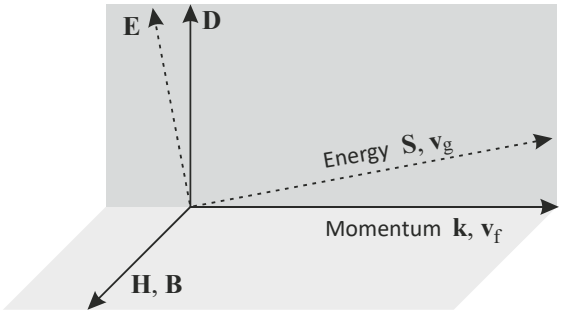


Fig. 2.28: Directions of fundamental vectors in a birefringent medium.

Birefringent media can be utilized in various devices. A *quarter-wave plate* is cut such that a normally incident ray enters perpendicularly to the optical axis. Both resulting waves propagate in the same direction, but at different velocities. The thickness of the plate is selected to induce a phase shift of 90° between both polarizations. At the output, both waves recombine. The oscillations of the electric fields of the waves are mutually perpendicular and phase-shifted by 90°. A circularly polarized electromagnetic signal is thereby generated. Another interesting device is a *polarizing prism* (Nicol prism), inside of which a cutting plane is bonded with Canada balsam. An unpolarized ray enters an anisotropic crystal of Iceland spar and splits into two polarizations. The refractive index of Canada balsam is such that total internal reflection occurs for the ordinary ray, causing it to exit the crystal. The second wave passes through the Canada balsam. Both polarizations are separated, and the prism thus serves as a source of polarized light.

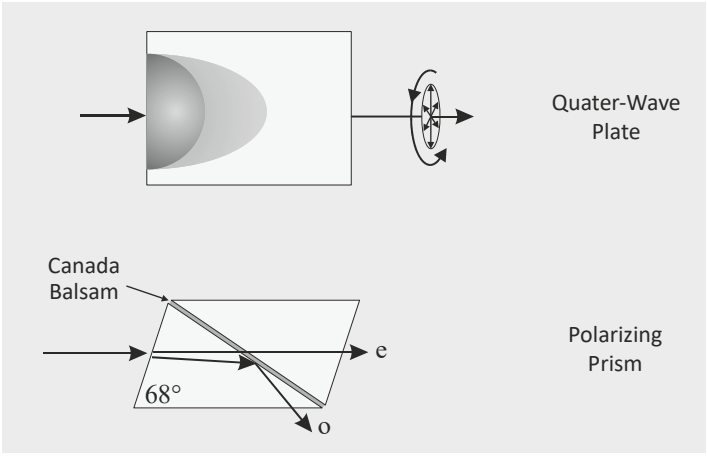


Fig. 2.29: Quarter-wave plate and polarizing prism.

2.5.4 Waves in Conducting Media

The wave equation for an electromagnetic wave in a metal is derived. The current density is substituted as $\mathbf{j}_Q = \sigma\mathbf{E}$, where σ represents the differential conductivity:

$$\operatorname{div}\mathbf{D} = \rho_Q\,, \quad \operatorname{div}\mathbf{B} = 0\,, \quad \operatorname{curl}\mathbf{H} = \sigma\mathbf{E} + \frac{\partial\mathbf{D}}{\partial t}\,, \quad \operatorname{curl}\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}\,. \tag{2.242}$$

The divergence operation is applied to the third equation, and the first equation is substituted for div **D**:

$$\frac{\partial\rho_Q}{\partial t} + \frac{\sigma}{\varepsilon}\rho_Q = 0 \qquad \Rightarrow \qquad \rho_Q \approx \rho_0 \exp\left[-\frac{\sigma}{\varepsilon}t\right]. \tag{2.243}$$

The spatial charge density in a conductor decays exponentially and is therefore neglected. The initial set of Maxwell equations for waves in a conductor is given by:

$$\operatorname{div}\mathbf{E} = 0\,, \quad \operatorname{div}\mathbf{B} = 0\,, \quad \operatorname{curl}\mathbf{B} = \mu\sigma\mathbf{E} + \varepsilon\mu\frac{\partial\mathbf{E}}{\partial t}\,, \quad \operatorname{curl}\mathbf{E} = -\frac{\partial\mathbf{B}}{\partial t}\,. \tag{2.244}$$

The electric field is eliminated by applying the curl operation to the third equation:

$$\operatorname{curl}\operatorname{curl}\mathbf{B} = \sigma\mu\operatorname{curl}\mathbf{E} + \varepsilon\mu\frac{\partial\operatorname{curl}\mathbf{E}}{\partial t}\,,$$

$$\operatorname{grad}\operatorname{div}\mathbf{B} - \nabla^2\mathbf{B} = -\sigma\mu\frac{\partial\mathbf{B}}{\partial t} - \varepsilon\mu\frac{\partial^2\mathbf{B}}{\partial t^2}\,,$$

$$\nabla^2\mathbf{B} - \sigma\mu\frac{\partial\mathbf{B}}{\partial t} - \varepsilon\mu\frac{\partial^2\mathbf{B}}{\partial t^2} = 0\,.$$

Analogously, the equation for the electric field is obtained. In a conductor, electromagnetic waves satisfy the *telegrapher equation*:

► $$\left(\nabla^2 - \sigma\mu\frac{\partial}{\partial t} - \varepsilon\mu\frac{\partial^2}{\partial t^2}\right)\begin{pmatrix}\mathbf{E}\\ \mathbf{B}\end{pmatrix} = 0\,.$$

Substituting a plane wave yields the dispersion relation:

$$\omega^2 = c^2k^2 - \mathrm{i}\,c^2\sigma\mu\omega\,. \tag{2.245}$$

If conductivity is zero ($\sigma = 0$), this formula reduces to the relation for a non-conducting medium. In a conductor, the dispersion relation is complex, which implies attenuation.

Spatial Attenuation

Spatial attenuation (the solution in terms of k is analyzed first:

$$c^2k^2 = \omega^2 + \mathrm{i}\,c^2\sigma\mu\omega \approx \mathrm{i}\,c^2\sigma\mu\omega\,.$$

Due to the high conductivity of metals, the first term on the right-hand side is neglected. The square root of this expression is evaluated utilizing $\mathrm{i}^{1/2} = (1+\mathrm{i})/2^{1/2}$:

$$k = k_1 + \mathrm{i}\, k_2 \,; \qquad k_1 = k_2 = \sqrt{\frac{\sigma\mu\omega}{2}} \,. \tag{2.246}$$

The real and imaginary parts of the wave vector are equal in magnitude, which is characteristic of metals. Spatially, the wave exhibits the following behavior:

$$\boldsymbol{\psi} = \mathbf{A} \exp\left[\mathrm{i}\sqrt{\sigma\mu\omega/2}\; x - \sqrt{\sigma\mu\omega/2}\; x \right] \tag{2.247}$$

The wave is exponentially attenuated. The characteristic distance over which the wave is attenuated by a factor e is readily determined:

► $$\delta = \sqrt{\frac{2}{\sigma\mu\omega}} \,. \tag{1.248}$$

This distance, representing the penetration depth of the wave into a conducting medium, is called the *skin depth*.

Temporal Attenuation

Temporal attenuation (the solution in terms of ω) is analyzed next. The dispersion relation is a quadratic equation for the frequency ω with the solution:

$$\omega_{1,2} = -\tfrac{\mathrm{i}}{2} c^2 \sigma\mu \pm \tfrac{1}{2}\sqrt{-c^4\sigma^2\mu^2 + 4c^2k^2} \,.$$

The conductivity term is dominant in the discriminant; hence, the last term is neglected. In this case, the single non-zero solution remains:

$$\omega \cong -\mathrm{i}\, c^2 \sigma\mu \,. \tag{2.249}$$

The solution for the frequency is purely imaginary and exhibits an attenuating behavior

$$\boldsymbol{\psi} = \mathbf{A}\, \mathrm{e}^{-c^2\sigma\mu t} \tag{2.250}$$

with a characteristic attenuation time over which the wave is attenuated by a factor of e:

► $$\tau = 1/(c^2\sigma\mu) \,. \tag{2.251}$$

Adhering to the sign convention (+ space, − time) in plane wave expressions of the type $\exp[\mathrm{i}\,(\mathbf{k}\cdot\mathbf{x} - \omega t)]$, an electromagnetic wave penetrating a conducting medium yields attenuation both in time and in space, which is consistent with physical expectations.

2.5.5 Light at Interface of Two Media

Law of Reflection and Refraction

The laws of reflection and refraction were first derived by the Arabic scholar Ibn Sahl in 984. In the Western world, the law of refraction was rediscovered by the Dutch astronomer and mathematician Willebrord Snellius in 1621. His predecessors described refraction of light using experimentally compiled tables. The law of refraction was also investigated by the French philosopher, physicist, and mathematician René Descartes, utilizing an analogy of balls penetrating a thin barrier, losing energy and changing di-

rection. Unfortunately, this procedure yielded a result where, although the sines of the angles of incidence and refraction appear correctly, the ratio of the velocities of light in the two media was inverted. The correct formula was also derived by the French mathematician and lawyer Pierre de Fermat from the principle of least time.

Numerous methods derive the law of refraction and reflection, ranging from the Huygens and Fermat principles to direct calculation from interface conditions following from Maxwell equations. The derivation is performed according to the following figure:

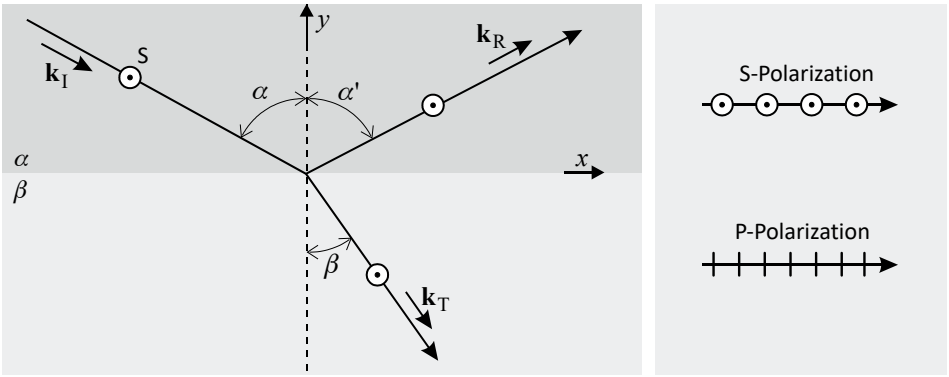


Fig. 2.30: Refraction and reflection at the interface of two media.

The interface plane is $y = 0$. The plane of incidence is $z = 0$. If the electric vector points perpendicularly to the plane of incidence, it is referred to as *S-polarization*, if it oscillates within this plane, it is *P-polarization*. Any other polarization can be constructed via superposition. The wave vectors of the *incident*, *reflected* and *transmitted* rays are denoted as $\mathbf{k}_I$, $\mathbf{k}_R$, and $\mathbf{k}_T$, respectively. The plane of oscillation of the electric field is indicated by red dots, where the electric field points toward the observer; this represents S-polarization, for which the calculation is performed. In this mode, the electric field possesses only a tangential component to the interface, which must be continuous. The angles are measured from the normal to the interface, where the angle of incidence is denoted as α, the angle of reflection as α' and the angle of refraction as β. The wave vectors possess the following components:

$$\begin{aligned}\mathbf{k}_I &= (k_I \sin\alpha, -k_I \cos\alpha, 0)\,;\\ \mathbf{k}_R &= (k_R \sin\alpha', k_R \cos\alpha', 0)\,;\\ \mathbf{k}_T &= (k_T \sin\beta, -k_T \cos\beta, 0)\,.\end{aligned} \tag{2.252}$$

The tangential field components are continuous, i.e.:

$$\mathbf{E}_{0I}\, e^{i[\mathbf{k}_I\cdot\mathbf{r}-\omega_I t]} + \mathbf{E}_{0R}\, e^{i[\mathbf{k}_R\cdot\mathbf{r}-\omega_R t]} = \mathbf{E}_{0T}\, e^{i[\mathbf{k}_T\cdot\mathbf{r}-\omega_T t]}. \tag{2.253}$$

The wave vectors and the interface condition ($y = 0$) are substituted into this relation:

$$E_{0I}\, e^{i[k_I x\sin\alpha-\omega_I t]} + E_{0R}\, e^{i[k_R x\sin\alpha'-\omega_R t]} = E_{0T}\, e^{i[k_T x\sin\beta-\omega_T t]}. \tag{2.254}$$

This interface condition must hold at any time t and at any position x, which is satisfied only if:

$$\omega_\mathrm{I} = \omega_\mathrm{R} = \omega_\mathrm{T}, \tag{2.255}$$

$$k_\mathrm{I} \sin\alpha = k_\mathrm{R} \sin\alpha' = k_\mathrm{T} \sin\beta. \tag{2.256}$$

The frequency of electromagnetic radiation does not change during reflection or refraction. From the second condition, it is straightforward to derive:

$$\frac{\omega_\mathrm{I}}{v_\mathrm{I}} \sin\alpha = \frac{\omega_\mathrm{R}}{v_\mathrm{R}} \sin\alpha' = \frac{\omega_\mathrm{T}}{v_\mathrm{T}} \sin\beta \quad \Rightarrow$$
$$\frac{\sin\alpha}{v_\alpha} = \frac{\sin\alpha'}{v_\alpha} = \frac{\sin\beta}{v_\beta}. \tag{2.257}$$

From the latter relation, the laws of reflection and refraction follow immediately:

► $$\alpha = \alpha', \tag{2.258}$$

► $$\frac{\sin\alpha}{\sin\beta} = \frac{v_\alpha}{v_\beta} = \frac{n_\beta}{n_\alpha} \tag{2.259}$$

The refractive index is defined as the ratio of the speed of light in vacuum to the speed of light in the given medium, thus:

$$n \equiv \frac{c}{v_\mathrm{f}} \tag{2.260}$$

Fresnel Relations

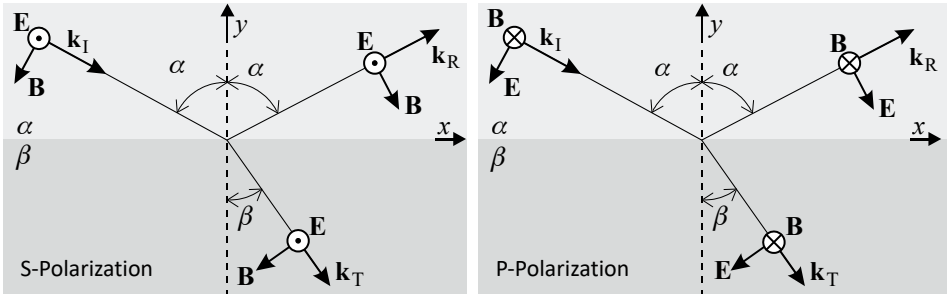


Fig. 2.31: Derivation of Fresnel relations.

A detailed analysis of light at an interface was performed in 1823 by the French physicist Augustin-Jean Fresnel, who advocated the wave nature of light over the Newtonian corpuscular theory. The reflection and transmission coefficients for both fundamental polarizations are derived here. Vector directions for the S-polarization and P-polarization are depicted in the figure. Fresnel investigated the field components parallel to the interface. The electric field **E** and magnetic field intensity **H** have continuous tangential components, whereas the tangential component of the magnetic field **B** exhibits a discontinuity. Assuming identical permeabilities, which holds for most transparent media, the tangential component of **B** is also continuous (representing the sum of the incident

and reflected wave components in medium 1). This simplifies the analysis of the continuity conditions for both fields. For the S-polarization, the continuity of the tangential components yields the following equations:

$$\begin{aligned} E_\mathrm{I} + E_\mathrm{R} &= E_\mathrm{T}, \\ -B_\mathrm{I}\cos\alpha + B_\mathrm{R}\cos\alpha &= -B_\mathrm{T}\cos\beta. \end{aligned} \tag{2.261}$$

From relation (2.234), the ratio of the electric field to the magnetic field is equal to the propagation velocity, $E/B = v_\mathrm{f} = c/n$. This is utilized to express the magnetic field:

$$\begin{aligned} E_\mathrm{I} + E_\mathrm{R} &= E_\mathrm{T}, \\ -E_\mathrm{I} n_\alpha \cos\alpha + E_\mathrm{R} n_\alpha \cos\alpha &= -E_\mathrm{T} n_\beta \cos\beta. \end{aligned} \tag{2.262}$$

The reflection amplitude $E_\mathrm{R}/E_\mathrm{I}$ and transmission amplitude $E_\mathrm{T}/E_\mathrm{I}$ are readily calculated from these equations. The ratio of the refractive indices is expressed using the Snell law of refraction (2.259). For S-polarization, the expressions are:

► $$r_\mathrm{S} \equiv \frac{E_\mathrm{R}}{E_\mathrm{I}} = -\frac{\sin(\alpha-\beta)}{\sin(\alpha+\beta)}, \tag{2.263}$$

► $$t_\mathrm{S} \equiv \frac{E_\mathrm{T}}{E_\mathrm{I}} = \frac{2\sin\beta\cos\alpha}{\sin(\alpha+\beta)} \tag{2.264}$$

Analogously, corresponding relations are determined for P-polarization:

► $$r_\mathrm{P} \equiv \frac{E_\mathrm{R}}{E_\mathrm{I}} = -\frac{\tan(\alpha-\beta)}{\tan(\alpha+\beta)}, \tag{2.265}$$

► $$t_\mathrm{P} \equiv \frac{E_\mathrm{T}}{E_\mathrm{I}} = \frac{2\sin\beta\cos\alpha}{\sin(\alpha+\beta)\cos(\alpha-\beta)} \tag{2.266}$$

These relations for the reflection and transmission amplitudes contain substantial information regarding the behavior of individual wave polarizations. For instance, if a wave is incident on an optically denser medium (with a higher refractive index), it reflects in opposition of phase – both reflection amplitudes are negative because $\alpha > \beta$ (refraction toward the normal). A wave in a rubber hose with one fixed end behaves analogously. A propagating wave reflects from the fixed end in opposition of phase.

Occasionally, the energy balance must be determined. The energy flux (the Poynting vector) is proportional to the square of the fields. For reflectivity, directional considerations are unnecessary because the incident and reflected rays exhibit mirror symmetry with respect to the normal to the interface; the reflection coefficient is simply the square of the reflection amplitude:

$$R \equiv r^2. \tag{2.267}$$

The *reflection amplitude* is the ratio of the field amplitudes, whereas the *reflectivity* is the ratio of the energy fluxes. For *transmissivity*, the situation differs. The ray is incident at a different angle than that of refraction. Therefore, angular dependence would have to be incorporated into the transmissivity; the incident energy depends on the angle of incidence, and the energy transmitted into the second medium depends on the angle

of refraction. However, transmissivity is readily determined from the law of conservation of energy, where the sum of reflectivity and transmissivity must equal unity, hence:

$$T = 1 - R\,. \tag{2.268}$$

Simple relations for the reflectivities and transmissivities of both polarizations are:

$$R_S = \frac{\sin^2(\alpha - \beta)}{\sin^2(\alpha + \beta)}, \tag{2.269}$$

$$R_P = \frac{\tan^2(\alpha - \beta)}{\tan^2(\alpha + \beta)}, \tag{2.270}$$

$$T_S = 1 - R_S, \tag{2.271}$$

$$T_P = 1 - R_P\,. \tag{2.272}$$

The reflectivities for both polarizations at an air-glass interface are plotted in the figure. The angle of refraction in relations (2.269) and (2.270) is given by the law of refraction.

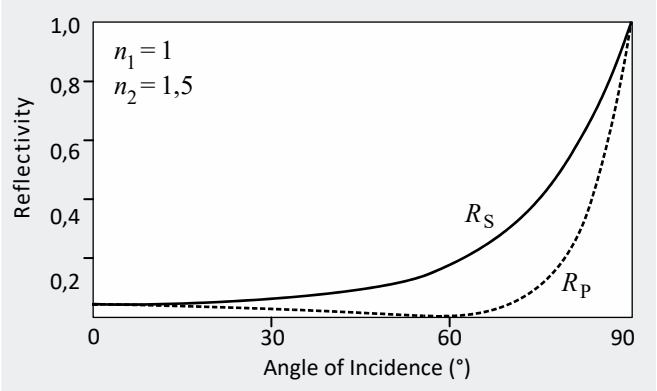


Fig. 2.32: Reflectivity for an air-glass interface.

Brewster Law

Each polarization exhibits distinct behavior upon reflection. The value for a zero angle of incidence must be determined from an expression of the 0/0 type. For this purpose, the sum formula for sines is utilized, and the sine of the angle of refraction is expressed from the law of refraction. The reflected light always contains the S-polarization, but for a specific angle of incidence, the P-polarization vanishes completely (the dashed curve reaches zero). If the incident light possessed a mixed polarization, the reflected light becomes completely polarized. Experimentally, this angle was first determined in 1815 by the Scottish inventor Sir David Brewster, who also invented the kaleidoscope and improved the optical microscope. Using biaxial crystals, the angle of maximum polarization was measured between 1813 and 1814, establishing that polarization is maximized when the tangent of the angle of incidence equals the refractive index. This relation is readily derived from the Fresnel relations. From relation (2.270), it is clear that

the reflectivity of the P-polarization is zero if the sum of the angle of incidence and the angle of refraction is exactly 90° (the tangent in the denominator diverges):

$$\alpha + \beta = 90° \,. \tag{2.273}$$

From the law of refraction, it follows that:

$$\frac{\sin\alpha}{\sin(90° - \alpha)} = \frac{n_\beta}{n_\alpha}. \tag{2.274}$$

The Brewster law for the angle of incidence at which the P-polarization of the incident light vanishes completely is then readily derived:

► $$\tan\alpha_B = n\,; \qquad n \equiv \frac{n_\beta}{n_\alpha}\,. \tag{2.275}$$

Light reflected from the Moon or a water surface is always partially polarized. If incident at the Brewster angle, the reflected light contains only the S-polarization.

Total Internal Reflection

Another interesting phenomenon is *total internal reflection*. This occurs when a ray is incident on an optically less dense medium (for instance, from water or glass into air). If the angle of incidence is increased, the ray refracts further away from the normal until it eventually propagates along the interface. An *evanescent wave* is generated at the interface, which decays exponentially with increasing distance from the interface (the wave vector possesses a complex component). Investigations of evanescent waves revealed that photons moving along the interface possess an atypical spin component in the direction of motion. At even larger angles, refraction no longer occurs. The light or electromagnetic wave reflects entirely from the interface and remains within the optically denser medium. This phenomenon is utilized, for example, in optical fibers.

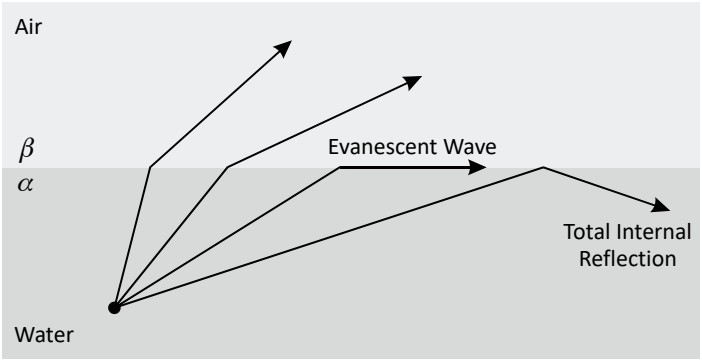


Fig. 2.33: Total internal reflection at the interface of two media.

For the critical angle, the angle of refraction equals 90°, and the Snell law yields:

► $$\sin\alpha_M = n\,; \qquad n \equiv \frac{n_\beta}{n_\alpha}. \tag{2.276}$$

The conditions for the S and P polarizations during total internal reflection can be analyzed in detail using the Fresnel relations, but this lies beyond the scope of this introductory course.

2.5.6 Raytracing

A highly interesting problem is the tracking of a ray propagating through a dispersive medium with a position-dependent refractive index (within the geometric optics approximation). Refined methods exist to determine the ray direction at a given point and its displacement to the subsequent position based on the knowledge of the permeability tensor and the permittivity tensor. However, such a formulation assumes knowledge of both tensors. Ray tracing can also be performed simply from the knowledge of the dispersion relation (the dependence of the angular frequency on the wave vector) for electromagnetic wave propagation. The analysis begins with the Hamilton equations for a quantum of electromagnetic radiation:

$$\dot{\mathbf{x}} = \frac{\partial H}{\partial \mathbf{p}}, \qquad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{x}}. \tag{2.277}$$

The relations for wave-particle duality are now utilized

$$H = \hbar\omega; \qquad \mathbf{p} = \hbar\mathbf{k} \tag{2.278}$$

and both relations are substituted into the Hamilton equations:

$$\begin{aligned} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} &= \frac{\partial \omega}{\partial \mathbf{k}}, \\ \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} &= -\frac{\partial \omega}{\partial \mathbf{x}}. \end{aligned} \tag{2.279}$$

►

The derived equations can be applied to ray tracking. However, several fundamental facts must be considered: 1) time figures here merely as a parameter of the ray trajectory; 2) when performing a numerical simulation, the vector **k**, representing the tangent to the tracked ray, is always obtained at a given point **x**. The displacement to the next position is performed along this tangent, and the procedure is repeated; 3) the derived equations for the *ray tracing* require explicit knowledge of the dispersion relation $\omega = \omega(\mathbf{x}, \mathbf{k})$, meaning that the angular frequency must be expressible from the dispersion relation, which can occasionally be problematic; 4) the spatial dependence on position **x** enters the dispersion relation through the known dependence of the number density or the external magnetic (electric) field on position in a plasma. Most frequently, the dispersion relation is known in an implicit form, i.e., unsolved with respect to variables ω or **k**:

$$\phi(\mathbf{x}, \omega, \mathbf{k}) = 0. \tag{2.280}$$

The differential of this relation is found (the frequency ω is fixed, determined by the frequency of the launched ray):

$$\frac{\partial \phi}{\partial \mathbf{x}} \cdot \mathrm{d}\mathbf{x} + \frac{\partial \phi}{\partial \mathbf{k}} \cdot \mathrm{d}\mathbf{k} = 0. \tag{2.281}$$

It is assumed that the ray trajectory is given parametrically, i.e.:

$$\begin{aligned} \mathbf{x} &= \mathbf{x}(\tau)\,; \\ \mathbf{k} &= \mathbf{k}(\tau)\,. \end{aligned} \tag{2.282}$$

From relation (2.281), it follows immediately that:

$$\frac{\partial \phi}{\partial \mathbf{x}} \cdot \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau} + \frac{\partial \phi}{\partial \mathbf{k}} \cdot \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}\tau} = 0\,. \tag{2.283}$$

The parameter τ can be arbitrary. This flexibility is utilized to choose τ such that:

► $$\begin{aligned} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau} &= \frac{\partial \phi}{\partial \mathbf{k}}\,; \\ \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}\tau} &= -\frac{\partial \phi}{\partial \mathbf{x}}\,. \end{aligned} \tag{2.284}$$

Through this choice, equation (2.283) is satisfied, and new equations for ray tracking are obtained, where the role of the Hamiltonian is assumed by the function ϕ, representing any implicit expression of the dispersion relation. Any finite-difference scheme can be applied to the system of Hamilton equations (2.284), allowing the propagation of a ray through a plasma to be tracked conveniently.



2.6 Electromagnetic Potentials

In the stationary case, a set of four potentials of the electromagnetic field can be defined from the source-free equations. Similarly, time-varying fields are considered by generalizing the definition of the potentials. Prior to reading the following text, familiarity with the technique of covariant and contravariant indices described in Chapter I.3.3.6 of the first volume of "*Selected Topics*" is recommended.

2.6.1 Definition of Potentials

The analysis begins with the source-free part of the Maxwell equations:

$$\operatorname{div}\mathbf{B} = 0; \tag{2.285}$$

$$\operatorname{curl}\mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \tag{2.286}$$

From the first equation, as in the stationary case, the existence of a function $\mathbf{A}(t, \mathbf{x})$ implies that:

► $$\mathbf{B} = \operatorname{curl}\mathbf{A}. \tag{2.287}$$

The field **A** is called the vector potential; through its introduction, Maxwell equation (2.285) is automatically satisfied since div curl **A** = 0. Now, the second equation (2.286) is addressed, substituting for **B** from the definition of the vector potential (2.287):

$$\operatorname{curl}\mathbf{E} = -\frac{\partial \operatorname{curl}\mathbf{A}}{\partial t} \quad \Rightarrow \quad \operatorname{curl}\left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}\right) = 0.$$

Twice continuously differentiable functions are assumed, allowing the order of the second derivatives to be interchanged. If the curl of a field is to be zero, i.e., curl **K** = 0, this equation is automatically satisfied by choosing **K** = grad *f*, since the curl of a gradient is always zero. Therefore, it can be written:

$$\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} = -\nabla\phi.$$

The minus sign is chosen, as in the stationary case, so that the forces point toward the minimum of the potential energy. The relation for the electric field thus has the form:

► $$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}. \tag{2.288}$$

This generalizes stationary relation (2.86). For time-varying fields, the electric field is also generated by time derivatives of potential **A**. The potentials are not uniquely determined, allowing considerable arbitrariness; a large number of potentials lead to identical electric and magnetic fields. This freedom is investigated in the section *Gauge Freedom of Potentials*. In Chapter 3.1, addressing special relativity, the potentials are shown to form a relativistic four-vector:

► $$A^{\mu} \equiv \begin{pmatrix} \phi/c \\ \mathbf{A} \end{pmatrix}; \qquad A_{\mu} \equiv \begin{pmatrix} -\phi/c \\ \mathbf{A} \end{pmatrix}; \tag{2.289}$$

Details regarding the introduction of four-vectors can be found in Section I.3.3.6 of the "*Selected Topics I*" and in Chapter 3.1. Here, it is only recalled that upper indices transform via the Lorentz matrix, whereas lower indices transform via the inverse Lorentz matrix. The four-gradient $\partial/\partial x^{\mu}$ contains the event components in the denominator; therefore, it does not transform via the Lorentz matrix but via its inverse, meaning the resulting index of the operation must be lower, i.e., $\partial_{\mu} = \partial/\partial x^{\mu}$ (see I.3.3.6), thus:

► $$\partial_{\mu} \equiv \begin{pmatrix} \partial/\partial ct \\ \boldsymbol{\nabla} \end{pmatrix}; \qquad \partial^{\mu} \equiv \begin{pmatrix} -\partial/\partial ct \\ \boldsymbol{\nabla} \end{pmatrix}. \tag{2.290}$$

Table of Potentials

The following table shows the potentials that, using relations (2.288) and (2.287), lead to the fields listed in the right column.

Source	Potential	Field
Homogeneous **E**	$\phi = -\mathbf{E}_0 \cdot \mathbf{r}$	$\mathbf{E} = \mathbf{E}_0$
Homogeneous **B**	$\mathbf{A} = \frac{1}{2}\,\mathbf{B} \times \mathbf{r}$	$\mathbf{B} = \mathbf{B}_0$
Point Charge	$\phi = \dfrac{Q}{4\pi\varepsilon_0 r}$	$\mathbf{E} = \dfrac{Q}{4\pi\varepsilon_0 r^2}\dfrac{\mathbf{r}}{r}$
Magnetic Monopole	$\mathbf{A}$ cannot be found on the entire $\mathcal{R}^3$	$\mathbf{B} = \dfrac{Q_{\mathrm{M}}}{r^2}\dfrac{\mathbf{r}}{r}$
Electric Dipole	$\phi = \dfrac{\mathbf{p}_{\mathrm{E}} \cdot \mathbf{r}}{4\pi\varepsilon_0 r^3}$	$\mathbf{E} = \dfrac{3(\mathbf{p}_{\mathrm{E}} \cdot \mathbf{n})\mathbf{n} - \mathbf{p}_{\mathrm{E}}}{4\pi\varepsilon_0 r^3}; \quad \mathbf{n} \equiv \dfrac{\mathbf{r}}{r}$
Magnetic Dipole	$\mathbf{A} = \dfrac{\mu_0}{4\pi}\dfrac{\mathbf{p}_{\mathrm{M}} \times \mathbf{r}}{r^3}$	$\mathbf{B} = \dfrac{\mu_0}{4\pi}\dfrac{3(\mathbf{p}_{\mathrm{M}} \cdot \mathbf{n})\mathbf{n} - \mathbf{p}_{\mathrm{M}}}{r^3}; \quad \mathbf{n} \equiv \dfrac{\mathbf{r}}{r}$

2.6.2 Potentials and Electromagnetic Field Tensor

The electromagnetic field is expressed via derivatives of potentials; relations (2.287) and (2.288) can be written using the electromagnetic field tensor (see also I.1.6.3):

► $$F^{\mu\nu} \equiv \partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{pmatrix}. \tag{2.291}$$

This is a second-rank antisymmetric tensor possessing six independent components (the three components of the electric field and the three components of the magnetic field). The field components can be easily read from the respective positions of the tensor.

Gauge Freedom of Potentials

The potentials are not uniquely determined; two different potentials can correspond to the same electromagnetic field. If new, transformed potentials are introduced via a gradient-based *gauge transformation*

► $$\tilde{A}^{\mu} \equiv A^{\mu} + \partial^{\mu} f \,, \tag{2.292}$$

where f is an arbitrary, twice continuously differentiable function, the fields remain unchanged:

$$\tilde{F}^{\mu\nu} = \partial^{\mu}\left(A^{\nu} + \partial^{\nu} f\right) - \partial^{\nu}\left(A^{\mu} + \partial^{\mu} f\right) = \partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu} = F^{\mu\nu} \,. \tag{2.293}$$

The gauge transformation (2.292) of the potentials can also be written separately as:

► $$\begin{aligned} \tilde{\phi} &= \phi - \frac{\partial f}{\partial t}, \\ \tilde{\mathbf{A}} &= \mathbf{A} + \nabla f \,. \end{aligned} \tag{2.294}$$

This transformation does not alter the field values, and the function f can be arbitrary. This freedom in the potentials can be utilized to construct the simplest possible variant of the Maxwell equations in terms of potentials. Maxwell equations (2.285) and (2.286) were used to introduce the electromagnetic field potentials.

2.6.3 Maxwell Equations in Potentials

The remaining two equations with source terms can be easily rewritten using the electromagnetic field tensor in the form:

► $$F^{\mu\nu}{}_{,\nu} = \mu_0 J^{\mu} \,, \tag{2.295}$$

where the four-vector J^{μ} represents the sources of the electric and magnetic fields, and indices following a comma denote the corresponding derivatives, e.g., $A_{\alpha,\beta}$ means $\partial_{\beta} A_{\alpha}$. This form of the Maxwell equations is manifestly relativistic. The Maxwell equations in the form (2.295) are rewritten in terms of the potentials as:

$$\begin{aligned} F^{\mu\nu}{}_{,\nu} &= \mu_0 J^{\mu} \,; \\ \partial_{\nu}\left(\partial^{\mu} A^{\nu} - \partial^{\nu} A^{\mu}\right) &= \mu_0 J^{\mu} \,; \\ \partial^{\mu}\partial_{\nu} A^{\nu} - \partial_{\nu}\partial^{\nu} A^{\mu} &= \mu_0 J^{\mu} \,. \end{aligned} \tag{2.296}$$

The second term on the left-hand side is the d'Alembert wave operator applied to the four-potential of the field, and the right-hand side represents the field sources. The only "aesthetic flaw" of the equations written in terms of potentials is the first term. Here, the considerable arbitrariness in the potentials given by the gauge transformation is utilized. It is assumed that the quantity $\partial_{\nu} A^{\nu}$ is equal to some function of time and space, $F(t, \mathbf{x})$:

$$\partial_\nu A^\nu = F(t,\mathbf{x})\,.$$

Another four-potential is chosen via a gauge transformation (2.292), requiring that:

$$\partial_\nu \tilde{A}^\nu = 0 \quad \Rightarrow \quad \partial_\nu A^\nu + \partial_\nu \partial^\nu f = 0 \quad \Rightarrow \quad F(t,\mathbf{x}) + \Box f = 0\,.$$

Such a gradient-based gauge transformation always exists. It is sufficient to choose the function f that generates the transformation such that it satisfies the equation:

$$\Box f = -F(t,\mathbf{x})\,. \tag{2.297}$$

In the new potentials, the first term in equation (2.296) is zero, and the Maxwell equations take the simple form:

$$\Box \tilde{A}^\mu = -\mu_0 J^\mu\,; \tag{2.298}$$

$$\partial_\mu \tilde{A}^\mu = 0\,. \tag{2.299}$$

These are the wave equations for the four-potential A^μ, where the source terms are the components of the four-vector J^μ. The equations are supplemented by the Lorenz gauge condition (2.299). In the subsequent text, the tildes are naturally omitted:

► $$\begin{gathered} \Box A^\mu = -\mu_0 J^\mu\,; \\ \partial_\mu A^\mu = 0\,. \end{gathered} \tag{2.300}$$

These equations can be written separately for the temporal and spatial components:

► $$\begin{gathered} \Box \phi = -\frac{\rho_Q}{\varepsilon_0}\,; \\ \Box \mathbf{A} = -\mu_0 \mathbf{j}_Q\,; \\ \frac{1}{c^2}\frac{\partial \phi}{\partial t} + \operatorname{div} \mathbf{A} = 0\,. \end{gathered} \tag{2.301}$$

In the case of stationary phenomena, the d'Alembert operator reduces to the Laplace operator, and the equations yield the Poisson equations of electrostatics and magnetostatics, including the Lorenz gauge condition.

Lagrangian Function of Electromagnetic Field

In theoretical mechanics, the central quantity is the Lagrangian function. For field problems, this corresponds to the Lagrangian density. The Lagrangian density leading to the Maxwell equations with source terms is given by (see Section I.1.6.3 in the first volume of "*Selected Topics*"):

► $$\mathscr{L} = J_\mu A^\mu - \frac{1}{4\mu_0} F^{\mu\nu} F_{\mu\nu}\,. \tag{2.302}$$

The first term describes the interaction of particles with the field, while the second describes the field itself. Without the first term, the Maxwell equations in vacuum would be obtained, i.e. without the source terms.

In classical electrodynamics, either fields or potentials can be used, as both variants are equivalent. In quantum theory, where the process probability depends on the action integral – and thus on the potentials and their derivatives in the Lagrangian density (2.302) – the potentials are primary. A particle interacts with the field even in a region where the field itself is zero but the potentials are non-zero. This was first demonstrated by the well-known Aharonov-Bohm experiment (see Section I.2.6.4 in the first volume of the "*Selected Topics*").

2.6.4 Potentials in Free Space

It has been demonstrated that the arbitrariness of the potentials can be utilized to satisfy the Lorenz gauge condition. Remarkably, even the requirement to satisfy this condition does not uniquely determine the potentials. From equation (2.297), it is evident that the function f is not unique, and any solution to the wave equation can be added to it:

$$\Box f_0 = 0\,. \tag{2.303}$$

Therefore, a further gradient-based gauge transformation is possible:

$$\tilde{\tilde{A}}^{\mu} = \tilde{A}^{\mu} + \partial^{\mu} f_0,$$

which can be used, for instance, to vanish the scalar potential in free space devoid of charges and currents. For the potential ϕ, the transformation takes the form:

$$\tilde{\tilde{\phi}} = \tilde{\phi} - \frac{\partial f_0}{\partial t}. \tag{2.304}$$

Evidently, the new scalar potential vanishes if the original potential can be expressed as:

$$\tilde{\phi} = \frac{\partial f_0}{\partial t}. \tag{2.305}$$

Such a potential satisfies the wave equation with a zero right-hand side:

$$\Box\tilde{\phi} = \Box\frac{\partial f_0}{\partial t} = \frac{\partial}{\partial t}\Box f_0 = \frac{\partial}{\partial t}0 = 0\,, \tag{2.306}$$

which represents the equation for the original potential in the absence of a source term (empty space). Such a transformation always exists, and any electromagnetic field in free space can be described solely by the vector potential (omitting the tildes):

$$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}, \tag{2.307}$$

$$\mathbf{B} = \operatorname{curl}\mathbf{A}\,. \tag{2.308}$$

2.6.5 Retarded and Advanced Potentials

In electrostatics and magnetostatics, the solutions derived from the Poisson equations for the potentials are given by:

$$\phi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho_Q(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}; \tag{2.309}$$

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{j}_Q(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}. \tag{2.310}$$

In electrodynamics, the Laplace operator is replaced by the d'Alembert operator, yielding the equations:

$$\Box\phi = -\frac{\rho_Q}{\varepsilon_0}; \qquad \Box\mathbf{A} = -\mu_0\mathbf{j}_Q; \qquad \frac{1}{c^2}\frac{\partial\phi}{\partial t} + \operatorname{div}\mathbf{A} = 0. \tag{2.311}$$

These equations account for the finite propagation time of information and possess the following solutions:

► $$\phi(t,\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho_Q(t-\Delta t,\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}; \tag{2.312}$$

► $$\mathbf{A}(t,\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{j}_Q(t-\Delta t,\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}. \tag{2.313}$$

The potentials are evaluated at the retarded time, which is the time of the observer minus the propagation time of the signal from the source:

► $$\Delta t = |\mathbf{r}-\mathbf{r}'|/c. \tag{2.314}$$

The obtained potentials are termed *retarded potentials*. Their values are calculated from the charge distribution within the past light cone of the observer. However, this is not the unique solution to the wave equations (2.311). The wave equation itself remains invariant under the substitution $c \to -c$; therefore, according to (2.314), the solutions also satisfy it:

$$\phi_+(t,\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho_Q(t+\Delta t,\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}, \tag{2.315}$$

$$\mathbf{A}_+(t,\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{j}_Q(t+\Delta t,\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}. \tag{2.316}$$

Such a solution is unphysical, appearing to arrive from the future, meaning it depends on the charge distribution within the future light cone of the observer and violates the

causality principle (where the cause must precede the effect in all coordinate systems). This solution is termed the *advanced potential*. Both solutions (the physical and the unphysical) can be written collectively using a single formula:

► $$\phi_{\pm}(t,\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho_Q(t \pm \Delta t, \mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{\left|\mathbf{r}-\mathbf{r}'\right|}, \tag{2.317}$$

► $$\mathbf{A}_{\pm}(t,\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{\mathbf{j}_Q(t \pm \Delta t, \mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{\left|\mathbf{r}-\mathbf{r}'\right|}. \tag{2.318}$$

Analogously, the solution to the Maxwell equations for the entire four-potential can be written as:

► $$A^{\alpha}_{\pm}(t,\mathbf{r}) = \frac{\mu_0}{4\pi}\int \frac{J^{\alpha}(t \pm \Delta t, \mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{\left|\mathbf{r}-\mathbf{r}'\right|}. \tag{2.319}$$

Advanced solutions possess significance in quantum theory and are also formally applied in classical electrodynamics when describing radiation. Direct differentiation demonstrates that the obtained potentials satisfy the Lorenz gauge condition. The potentials are calculated for an observer at position $\mathbf{r}$. The charges are localized at positions $\mathbf{r}'$, over which the integration is performed. Sufficiently far from the sources, a classical multipole expansion can be performed to determine the radiation fields generated by the moving particles. If the reaction of a particle to its own field is of interest, the reverse procedure must be adopted to determine the field generated in the immediate vicinity of the tracked particle.

2.7 Radiation of Charged Particles

The multipole expansion of the scalar and vector potentials has previously been analyzed for the stationary case, introducing definitions of dipole moments, the electric quadrupole, and other quantities. Currently, the potentials are treated as components of a single four-vector, and the stationary approach has been abandoned for electromagnetic waves. Therefore, a multipole expansion must be performed for the non-stationary case to understand the fundamental mechanisms of radiation from charged particles.

2.7.1 Radiative Multipole Expansion

Time-varying sources of the potentials are now considered:

$$A^{\alpha} = \frac{\mu_0}{4\pi}\int \frac{J^{\alpha}(t',\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}^3\mathbf{r}'\,; \tag{2.320}$$

$$t' \equiv t - \Delta t = t - \frac{|\mathbf{r}-\mathbf{r}'(t')|}{c}. \tag{2.321}$$

The analysis is more complex in this scenario. First, the retarded time is not defined explicitly but is specified by an implicit formula (2.321). Second, a Taylor expansion must be performed in both the denominator and the numerator, followed by the selection of the radiative terms. These terms decrease with distance as $1/r$, where r represents the distance of the observer from the origin of the coordinate frame localized in the source region. Only fields capable of carrying energy to an arbitrary distance from the source are considered; therefore, the integration of the Poynting vector $\mathbf{E}\times\mathbf{H}$, representing the energy flux, over the solid angle $r^2\mathrm{d}\Omega$ must yield a finite contribution at any distance. Conversely, this fact simplifies the problem, as it is sufficient to select only terms that decrease as $1/r$ from the expansions of the potentials and fields.

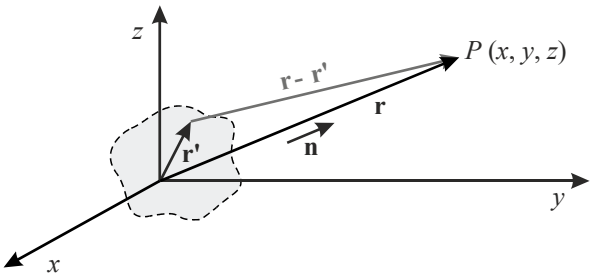


Fig. 2.34: Localized radiation source and distant observer.

The denominator expansion was performed in the stationary case, see (2.120); for radiation theory, it is sufficient to retain only the first term, which decays as $1/r$:

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} \approx \frac{1}{r}\,. \tag{2.322}$$

In the argument of the four-current, the retarded time must be expanded for an observer located at a large distance from the source:

$$t' = t - \frac{|\mathbf{r} - \mathbf{r}'|}{c} = t - \frac{r}{c} - \frac{\partial}{\partial x_k}\left(\frac{r}{c}\right)\left(-x_k'\right) + \cdots = t - \frac{r}{c} + \frac{1}{c}\frac{x_k}{r}x_k' + \cdots = t - \frac{r}{c} + \frac{\mathbf{n}\cdot\mathbf{r}'}{c} + \cdots$$

The first term represents the observer time, the second is the retarded time with respect to the origin (common to all sources), and the last represents the subtle nuances of retardation for individual elements of the field source. This can be written as:

$$t' \approx \tau + \frac{\mathbf{n}\cdot\mathbf{r}'}{c}; \tag{2.323}$$

$$\tau \equiv t - r/c\,. \tag{2.324}$$

The symbol τ denotes the common retarded time with respect to the origin, and $\mathbf{n} = \mathbf{r}/r$ represents the normal vector to the distant integration surface (the unit vector connecting the origin of the system to the observer). Naturally, the implicit dependence remains, and vector $\mathbf{r}'$ is again expressed in terms of the retarded time. Considering expansions (2.322) and (2.323), relation (2.320) for the four-potential takes the form:

$$A^\alpha = \frac{\mu_0}{4\pi r}\int J^\alpha(\tau + \mathbf{n}\cdot\mathbf{r}'/c, \mathbf{r}')\,\mathrm{d}^3\mathbf{r}'\,. \tag{2.325}$$

Now, the four-current is expanded in terms of time τ and the increment $\mathbf{n}\cdot\mathbf{r}'$:

$$A^\alpha(\tau) = \frac{\mu_0}{4\pi r}\int J^\alpha(\tau,\mathbf{r}')\,\mathrm{d}^3\mathbf{r}' + \frac{\mu_0}{4\pi r}\int \frac{\partial J^\alpha}{\partial\tau}\left(\frac{\mathbf{n}\cdot\mathbf{r}'}{c}\right)\mathrm{d}^3\mathbf{r}' + \cdots. \tag{2.326}$$

The scalar potential can be chosen as zero in free space, and the radiation fields can be calculated from the relations:

$$\mathbf{E}_{\mathrm{rad}} = -\frac{\partial \mathbf{A}_{\mathrm{rad}}}{\partial t}, \tag{2.327}$$

$$\mathbf{B}_{\mathrm{rad}} = \mathrm{curl}\,\mathbf{A}_{\mathrm{rad}}\,. \tag{2.328}$$

Consequently, only the spatial part of relation (2.326) is analyzed further, and the derivative with respect to the common retarded time τ is factored out before integration:

$$\mathbf{A}(\tau) = \frac{\mu_0}{4\pi r}\int \mathbf{j}_Q(\tau,\mathbf{r}')\,\mathrm{d}^3\mathbf{r}' + \frac{\mu_0}{4\pi r}\frac{\partial}{\partial\tau}\int \mathbf{j}_Q(\tau,\mathbf{r}')\left(\frac{\mathbf{n}\cdot\mathbf{r}'}{c}\right)\mathrm{d}^3\mathbf{r}' + \cdots. \tag{2.329}$$

The first term will be shown to describe electric dipole radiation, whereas the second term describes magnetic dipole and electric quadrupole radiation. The integration is now performed for a system of charged particles localized near the origin, where:

$$\mathbf{j}_Q(\tau,\mathbf{r}') = \sum_a Q_a \mathbf{v}_a(\tau)\,\delta(\mathbf{r}' - \mathbf{r}_a)\,. \tag{2.330}$$

This immediately yields:

$$\mathbf{A}(\tau) = \frac{\mu_0}{4\pi r}\sum_a (Q_a\mathbf{v}_a) + \frac{\mu_0}{4\pi rc}\frac{\partial}{\partial\tau}\sum_a Q_a\mathbf{v}_a(\mathbf{n}\cdot\mathbf{r}_a) + \cdots. \tag{2.331}$$

In the first term, the time derivative is factored out before the summation (all is in the common retarded time τ, and symmetrization (2.167) is performed in the second; however, the time derivative of the symmetric part is not zero, unlike in the stationary case:

$$\mathbf{A}(\tau) = \mathbf{A}^{(\mathrm{E1})} + \mathbf{A}^{(\mathrm{M1})} + \mathbf{A}^{(\mathrm{E2})} + \cdots ; \tag{2.332}$$

$$\mathbf{A}^{(\mathrm{E1})} = \frac{\mu_0}{4\pi r}\frac{\partial}{\partial \tau}\sum_a Q_a \mathbf{r}_a \,, \tag{2.333}$$

$$\mathbf{A}^{(\mathrm{M1})} = \frac{\mu_0}{4\pi rc}\frac{\partial}{\partial \tau}\sum_a \frac{1}{2} Q_a \left(\mathbf{r}_a \times \mathbf{v}_a\right)\times \mathbf{n} \,, \tag{2.334}$$

$$\mathbf{A}^{(\mathrm{E2})} = \frac{\mu_0}{8\pi rc}\frac{\partial^2}{\partial \tau^2}\sum_a Q_a \mathbf{r}_a (\mathbf{n}\cdot\mathbf{r}_a) \,. \tag{2.335}$$

All three radiative potentials can be rewritten using quantities from the stationary expansion. The first involves the electric dipole moment, the second the magnetic dipole moment, and only the third potential must still be modified. From previous results – see (2.292) – it is known that the potential can be altered by the gradient of any function:

$$\mathbf{A} \to \mathbf{A} + \nabla f(\mathbf{r}) \,. \tag{2.336}$$

Choosing a function solely of distance, the transformation to the new potential is:

$$\mathbf{A} \to \mathbf{A} + f'(r)\frac{\mathbf{r}}{r} \tag{2.337}$$

and potential $\mathbf{A}^{(\mathrm{E2})}$ is transformed into a form involving the electric quadrupole moment:

$$\begin{aligned} \mathbf{A}^{(\mathrm{E2})} &= \frac{\mu_0}{8\pi rc}\frac{\partial^2}{\partial \tau^2}\frac{1}{3}\sum_a 3 Q_a \mathbf{r}_a (\mathbf{n}\cdot\mathbf{r}_a) \;\to \\ &\to \frac{\mu_0}{24\pi rc}\frac{\partial^2}{\partial \tau^2}\sum_a Q_a \left[3\mathbf{r}_a (\mathbf{n}\cdot\mathbf{r}_a) - r_a^2 \mathbf{n}\right]. \end{aligned} \tag{2.338}$$

The resulting relations for the radiative potentials are thus (all is expressed in the common retarded time τ, and the time derivatives are also with respect to τ):

► $$\mathbf{A}^{(\mathrm{E1})} = \frac{\mu_0}{4\pi r}\dot{\mathbf{p}}_{\mathrm{E}} \,, \tag{2.339}$$

► $$\mathbf{A}^{(\mathrm{M1})} = \frac{\mu_0}{4\pi rc}\dot{\mathbf{p}}_{\mathrm{M}} \times \mathbf{n} \,, \tag{2.340}$$

► $$\mathbf{A}^{(\mathrm{E2})} = \frac{\mu_0}{24\pi rc}\ddot{\mathbf{Q}}\cdot\vec{\mathbf{n}} \,. \tag{2.341}$$

The last expression represents the contraction of tensor $\ddot{\mathbf{Q}}$ and vector $\vec{\mathbf{n}}$.

2.7.2 Electric Dipole Radiation

The electric and magnetic fields for the potential in the form (2.339) are now determined. It must be noted that the common retarded time is also a function of r (the fixed position of the observer), and therefore:

$$\tau = t - r/c \qquad \Rightarrow \tag{2.342}$$

$$\frac{\partial \tau}{\partial t} = \frac{\partial t}{\partial \tau} = 1; \quad \frac{\partial \tau}{\partial x_k} = -\frac{1}{c}\frac{x_k}{r}. \tag{2.343}$$

The electric and magnetic fields are now calculated. During differentiation, only radiative terms with a radial dependence of $1/r$ are selected, while higher powers are neglected. The radiation fields constructed in this manner are denoted by script symbols:

$$\mathcal{E} = -\frac{\partial \mathbf{A}^{(\mathrm{E1})}}{\partial t} = -\frac{\partial \mathbf{A}^{(\mathrm{E1})}}{\partial \tau}\frac{\partial \tau}{\partial t} = -\frac{\mu_0}{4\pi r}\ddot{\mathbf{p}}_{\mathrm{E}};$$

$$\mathcal{B}_k = \left(\operatorname{curl}\mathbf{A}^{(\mathrm{E1})}\right)_k = \frac{\mu_0}{4\pi}\varepsilon_{klm}\frac{\partial}{x_l}\left(\frac{\dot{p}_m}{r}\right) = \frac{\mu_0}{4\pi}\varepsilon_{klm}\frac{\frac{\partial \dot{p}_m}{\partial \tau}\frac{\partial \tau}{\partial x_l}r - \dot{p}_m\frac{\partial r}{\partial x_l}}{r^2}.$$

The second term behaves as $1/r^2$, is non-radiative, and is omitted:

$$\mathcal{B}_k = \frac{\mu_0}{4\pi r}\varepsilon_{klm}\ddot{p}_m\left(-\frac{1}{c}\frac{x_l}{r}\right) = -\frac{\mu_0}{4\pi cr}\left(\mathbf{n}\times\ddot{\mathbf{p}}_{\mathrm{E}}\right)_k.$$

The result is thus obtained:

► $$\mathcal{E}^{(\mathrm{E1})} = -\frac{\mu_0}{4\pi r}\ddot{\mathbf{p}}_{\mathrm{E}}; \tag{2.344}$$

► $$\mathcal{B}^{(\mathrm{E1})} = -\frac{\mu_0}{4\pi cr}\left(\mathbf{n}\times\ddot{\mathbf{p}}_{\mathrm{E}}\right) = \frac{1}{c}\left(\mathbf{n}\times\mathcal{E}^{(\mathrm{E1})}\right). \tag{2.345}$$

Sufficiently far from the source, the electromagnetic field resembles a plane wavefront. Near the dipole, the field is much more complex; the derived solution is valid only in the radiation zone. When calculating any radiation field, it is sufficient to compute only the electric field in this zone and determine the magnetic field from the plane wavefront relation (which is faster than utilizing the curl of the vector potential):

► $$\mathcal{B} = \frac{1}{c}\left(\mathbf{n}\times\mathcal{E}\right). \tag{2.346}$$

The energy flux (the Poynting vector) is now determined for the radiation fields. Only terms proportional to $1/r^2$ are retained:

$$\mathcal{S} = \mathcal{E}\times\mathcal{H} = \frac{1}{\mu_0 c}\mathcal{E}\times\left(\mathbf{n}\times\mathcal{E}\right) = \frac{1}{\mu_0 c}\left[\mathcal{E}^2\mathbf{n} - \left(\mathcal{E}\cdot\mathbf{n}\right)\mathcal{E}\right] \tag{2.347}$$

The quantity of interest is the projection of the Poynting vector into the direction from the charges to the observer (the radial energy flux, intensity), i.e.:

$$\mathcal{S}_n = \boldsymbol{\mathcal{S}} \cdot \mathbf{n} = \frac{1}{\mu_0 c}\left[\mathcal{E}^2 - (\boldsymbol{\mathcal{E}} \cdot \mathbf{n})^2\right] = \frac{1}{\mu_0 c}\left[\mathcal{E}^2 - \mathcal{E}^2 \cos^2\theta\right] = \frac{\mathcal{E}^2}{\mu_0 c}\sin^2\theta\,. \tag{2.348}$$

The electric field from (2.344) is substituted into the radial energy flux, and permeability is converted to permittivity using the relation $c^2 = 1/\varepsilon_0\mu_0$. The radial flux is obtained

$$\mathcal{S}_n = \frac{\ddot{p}_{\mathrm{E}}^2}{16\pi^2\varepsilon_0 c^3 r^2}\sin^2\theta\,; \quad \theta = \sphericalangle(\boldsymbol{\mathcal{E}}, \mathbf{n}) = \sphericalangle(\ddot{\mathbf{p}}_{\mathrm{E}}, \mathbf{n})\,, \quad \left[\mathcal{S}_n\right] = \frac{\mathrm{W}}{\mathrm{m}^2}\,, \tag{2.349}$$

where θ is the angle between the distant electric field (the second time derivative of $\mathbf{p}_{\mathrm{E}}$) and the direction to the observer. Consequently, the particle radiates most intensely in the direction perpendicular to the second time derivative of the electric dipole moment (for a single particle, this corresponds to the direction perpendicular to the acceleration).

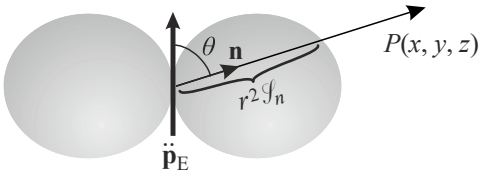


Fig. 2.35: Electric dipole radiation.

The power element radiated into a solid angle is given by:

$$\mathrm{d}\mathcal{P} = \boldsymbol{\mathcal{S}}\,\mathrm{d}\mathbf{S} = (\boldsymbol{\mathcal{S}} \cdot \mathbf{n})\,\mathrm{d}S = \mathcal{S}_n\,\mathrm{d}S = \mathcal{S}_n\, r^2\,\mathrm{d}\Omega\,. \tag{2.350}$$

Substituting from (2.349) yields the *Larmor differential formula*:

► $$\mathrm{d}\mathcal{P} = \frac{\ddot{p}_{\mathrm{E}}^2 \sin^2\theta}{16\pi^2\varepsilon_0 c^3}\mathrm{d}\Omega\,. \tag{2.351}$$

The total radiated power is obtained via integration:

$$\mathcal{P} = \int \frac{\ddot{p}_{\mathrm{E}}^2 \sin^2\theta}{16\pi^2\varepsilon_0 c^3}\mathrm{d}\Omega =$$

$$= \frac{\ddot{p}_{\mathrm{E}}^2}{16\pi^2\varepsilon_0 c^3}\int \sin^2\theta \sin\theta\,\mathrm{d}\varphi\,\mathrm{d}\theta =$$

$$= \frac{\ddot{p}_{\mathrm{E}}^2}{16\pi^2\varepsilon_0 c^3}\,2\pi\int_0^{\pi}(1-\cos^2\theta)\sin\theta\,\mathrm{d}\theta =$$

$$= \frac{\ddot{p}_{\mathrm{E}}^2}{8\pi\varepsilon_0 c^3}\int_1^{-1}(1-\xi^2)(-\mathrm{d}\xi) = \frac{\ddot{p}_{\mathrm{E}}^2}{8\pi\varepsilon_0 c^3}\left[\xi - \frac{\xi^3}{3}\right]_{-1}^{+1} = \frac{\ddot{p}_{\mathrm{E}}^2}{6\pi\varepsilon_0 c^3} \quad \Rightarrow$$

► $$\mathcal{P} = \frac{\ddot{p}_{\mathrm{E}}^2}{6\pi\varepsilon_0 c^3}\,; \qquad \left[\mathcal{P}\right] = \mathrm{W}. \tag{2.352}$$

Several interesting consequences follow from the *Larmor formula* (2.352):

1) If particles move with acceleration, they radiate because:

$$\ddot{\mathbf{p}}_{\mathrm{E}} = \frac{\partial^2}{\partial \tau^2} \sum_a Q_a \mathbf{r}_a = \sum_a Q_a \ddot{\mathbf{r}}_a \ . \tag{2.353}$$

2) An isolated system of identical charges ($Q_a = q$) localized in a single region (around the origin of the coordinate system) does not radiate, since the second time derivative of the electric dipole moment is proportional to the sum of all internal forces, which is zero in an isolated system:

$$\ddot{\mathbf{p}}_{\mathrm{E}} = \frac{\partial^2}{\partial \tau^2} \sum_a Q_a \mathbf{r}_a = \frac{q}{m} \sum_a m \ddot{\mathbf{r}}_a = \frac{q}{m} \sum_a \mathbf{F}_a = 0 \, . \tag{2.354}$$

3) The simplest realization of electric dipole radiation is the so-called *Hertz dipole*. This is an open linear element carrying an alternating current (fed, for example, inductively from an alternating current circuit). Argument 2) does not apply here, as the system is not isolated. Electric charges travel back and forth in this element, leading to a simple expression for the electric dipole moment:

$$\mathbf{p}_{\mathrm{E}} = p_0 \cos(\omega t) \ \mathbf{e}_3 \, . \tag{2.355}$$

Considering that the mean value of the squared cosine is ½, relation (2.352) yields the mean power:

$$\langle \mathcal{P} \rangle = \frac{p_0^2 \omega^4}{12 \pi \varepsilon_0 c^3} \, . \tag{2.356}$$

A slight increase in frequency brings a significant increase in the radiated power.

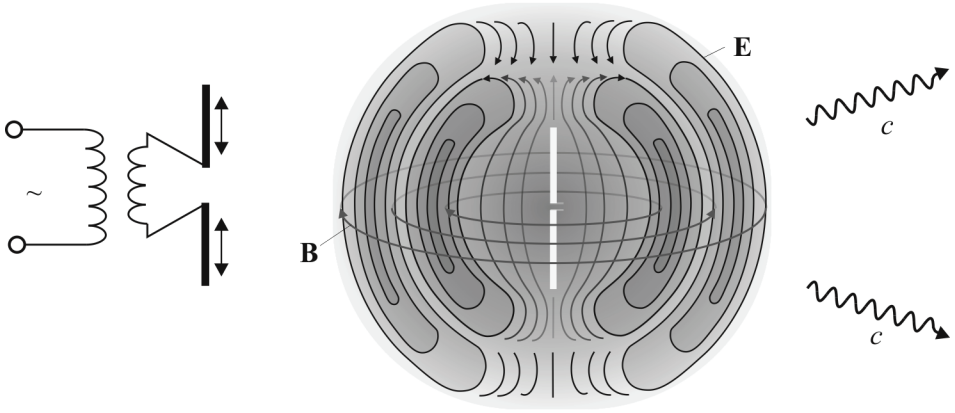


Fig. 2.36: Hertz dipole. Left: simple realization. Right: near fields in the vicinity of a Hertz dipole. Two wavelengths are displayed, corresponding to two reversals of the dipole. The relations derived here are valid only at a large distance from the dipole. Source: Averse/Wikipedia.

2.7.3 Thomson Scattering

The radiation of a free electron interacting with a low-energy electromagnetic wave (the scattering of an electron by electromagnetic radiation) is analyzed. The electric field component exerts a force $\mathbf{F} = -e\mathbf{E}$ on the electron, imparting an acceleration:

$$\ddot{\mathbf{r}}_\mathrm{e} = -\frac{e}{m_\mathrm{e}}\mathbf{E}\,. \tag{2.357}$$

The effect of the magnetic field is neglected because the analysis is restricted to the low-energy limit, where $v \ll c$. This interaction induces a non-zero electric dipole moment, whose second time derivative is given by:

$$\ddot{\mathbf{p}}_\mathrm{E} = -e\ddot{\mathbf{r}}_\mathrm{e} = \frac{e^2}{m_\mathrm{e}}\mathbf{E}\,. \tag{2.358}$$

The free electron oscillates in unison with the incident wave, acquiring a non-zero periodic acceleration and subsequently radiating. The power radiated by the electron is determined from the Larmor formula (2.351), substituting the second time derivative of the electric dipole moment from relation (2.358):

$$\mathrm{d}\mathcal{P}_\mathrm{out} = \frac{\ddot{p}_\mathrm{E}^2 \sin^2\theta}{16\pi^2\varepsilon_0 c^3}\mathrm{d}\Omega = \frac{e^4E^2}{16\pi^2\varepsilon_0 c^3 m_\mathrm{e}^2}\sin^2\theta\,\mathrm{d}\Omega\,. \tag{2.359}$$

The scattering probability is determined by the ratio of the outgoing energy to the incoming energy. If the outgoing energy is represented by the Larmor power (2.359) and the incoming energy by the Poynting vector of the incident wave (utilizing $E/B = c$):

$$\mathscr{S}_\mathrm{in} = EH = \frac{EB}{\mu_0} = \frac{E^2}{c\mu_0}\,. \tag{2.360}$$

The ratio is given by:

$$\frac{\mathrm{d}\mathcal{P}_\mathrm{out}}{\mathscr{S}_\mathrm{in}} = \frac{\mathrm{d}W/\mathrm{d}t}{\mathrm{d}W/(\mathrm{d}t\,\mathrm{d}S)} = \mathrm{d}S = \mathrm{d}\sigma \tag{2.361}$$

and equals an element of the area over which the electron is capable of receiving the incident radiation, which defines the differential cross-section. For the scattering cross-section of an electron by an electromagnetic wave, the following expression is obtained:

$$\mathrm{d}\sigma = \frac{\mathrm{d}\mathcal{P}_\mathrm{out}}{\mathscr{S}_\mathrm{in}} = \frac{\dfrac{e^4E^2}{16\pi^2\varepsilon_0 c^3 m_\mathrm{e}^2}\sin^2\theta\,\mathrm{d}\Omega}{\dfrac{E^2}{c\mu_0}} = \frac{e^4 c\mu_0}{16\pi^2\varepsilon_0 c^3 m_\mathrm{e}^2}\sin^2\theta\,\mathrm{d}\Omega\,. \tag{2.362}$$

Converting permeability to permittivity yields the final differential cross-section:

► $$\mathrm{d}\sigma = \frac{e^4}{16\pi^2\varepsilon_0^2 c^4 m_\mathrm{e}^2}\sin^2\theta\,\mathrm{d}\Omega\,. \tag{2.363}$$

The total cross-section is obtained via integration over the solid angle:

► $$\sigma_{\text{tot}} = \frac{e^4}{6\pi\varepsilon_0^2 c^4 m_e^2}. \tag{2.364}$$

The total cross-section for electromagnetic radiation scattering by an electron is given by a universal constant. For historical reasons, this constant is conventionally expressed using the classical electron radius, defined as the dimension at which the electrostatic field generates the rest mass of the electron:

$$\frac{e^2}{4\pi\varepsilon a_e} = m_e c^2 \quad \Rightarrow \quad a_e = \frac{e^2}{4\pi\varepsilon m_e c^2}; \quad a_e \simeq 2.8\times10^{-15}\ \text{m}. \tag{2.365}$$

In this notation, the cross-section is a small multiple of the electron cross-section represented by the classical electron radius:

► $$\sigma_{\text{tot}} = \frac{8}{3}\pi a_e^2; \qquad \sigma_{\text{tot}} \simeq 6.7\times10^{-29}\ \text{m}^2. \tag{2.366}$$

Note 1: The derivation performed here is non-relativistic and holds exclusively for low radiation energies. It yields identical frequencies for the incident wave and the wave generated by the electron. Furthermore, the cross-section does not depend on the frequency. For high energies, the conservation of energy and momentum given by the relativistic four-momentum must be considered. The photon alters its direction of motion, which changes its momentum and consequently its energy (and thus frequency), linked to momentum via the magnitude of the four-momentum vector. The impact of a high-energy photon results in photon-electron scattering that alters the frequency. This phenomenon is called the Compton effect, and Thomson scattering represents its low-energy limit.

Note 2: The incident wave is scattered simultaneously in various directions according to formula (2.363). If a specific scattered wave is observed in the direction θ, the plane of polarization remains unchanged. The electric field naturally remains perpendicular to the new propagation direction. The scattering angle is denoted as χ in the figure.

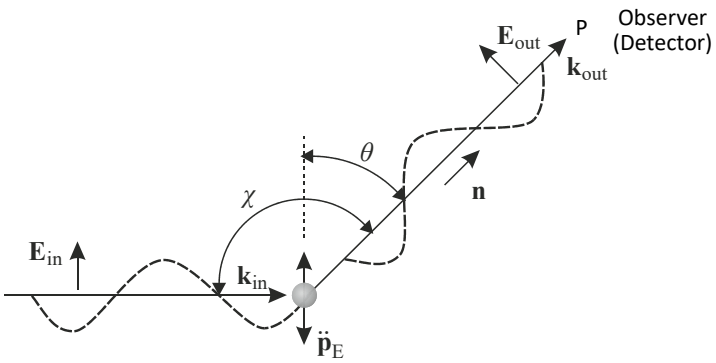


Fig. 2.37: Thomson scattering – incident and scattered waves.

2.7.4 Rayleigh Scattering

Thomson scattering applies to free electrons. However, electrons are frequently bound – either in atomic shells, molecules, or within small dust particles. Such an electron responds to the electric field of an incident wave, exhibiting a specific natural frequency ω_0 of natural oscillations within the host system. It is also characterized by a natural damping factor δ, caused by the radiation reaction of the electron. The equation of motion takes a slightly more complex form than (2.357):

$$\ddot{\mathbf{r}}_\mathrm{e} + 2\delta\dot{\mathbf{r}}_\mathrm{e} + \omega_0^2\mathbf{r}_\mathrm{e} = -\frac{e}{m_\mathrm{e}}\mathbf{E}_0\,\mathrm{e}^{-\mathrm{i}\omega t}\,. \tag{2.367}$$

This represents forced oscillations, a well-known problem in classical mechanics. The driving electric field **E** possesses a frequency ω. Written in an exponential form that is straightforward to differentiate, the actual solution for $\mathbf{r}(t)$ is the real part of the of the sum of the homogeneous and particular solutions. The homogeneous solution represents damped oscillations at the natural frequency; thus, from a long-term perspective, it is a transient phenomenon and is neglected. The particular solution has the form:

$$\mathbf{r}_\mathrm{e}(t) = A\,\mathbf{E}_0\,\mathrm{e}^{-\mathrm{i}\omega t}\;; \qquad A = \frac{-e/m}{(\omega_0^2 - \omega^2 - 2\mathrm{i}\,\delta\omega)}. \tag{2.368}$$

The amplitude is generally a complex number, which can be expressed in polar form:

$$A = |A|\mathrm{e}^{\mathrm{i}\varphi}\;; \qquad |A| = \frac{e/m}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\delta^2\omega^2}}. \tag{2.369}$$

The actual oscillations of the electron are the real part of the obtained solution, i.e.:

$$\mathbf{r}_\mathrm{e}(t) = |A|\,\mathbf{E}_0\cos(\omega t - \varphi) = \frac{e\mathbf{E}_0/m}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\delta^2\omega^2}}\cos(\omega t - \varphi). \tag{2.370}$$

The second derivative of the electric dipole moment $\mathbf{p}_\mathrm{e} = -e\mathbf{r}_\mathrm{e}$ is given by:

$$\ddot{\mathbf{p}}_\mathrm{e}(t) = \frac{e^2\omega^2\,\mathbf{E}_0/m}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\delta^2\omega^2}}\cos(\omega t - \varphi). \tag{2.371}$$

Subsequently, the procedure is identical to the derivation of the Thomson cross-section. The only difference is the averaging of the squared cosine over a period, which yields:

$$\left\langle \cos^2(\omega t - \varphi) \right\rangle_T = \tfrac{1}{2}\,. \tag{2.372}$$

For the radiated power, the following expression is obtained:

$$\mathrm{d}\mathcal{P}_\mathrm{out} = \frac{\ddot{p}_\mathrm{E}^2\sin^2\theta}{16\pi^2\varepsilon_0 c^3}\mathrm{d}\Omega = \frac{e^4\omega^4\,E_0^2}{32\pi^2\varepsilon_0 c^3 m_\mathrm{e}^2\left((\omega_0^2 - \omega^2)^2 + 4\delta^2\omega^2\right)}\sin^2\theta\,\mathrm{d}\Omega \tag{2.373}$$

The incident power is also averaged over a period:

$$\mathcal{S}_{\text{in}} = \langle EH \rangle = \langle EB/\mu_0 \rangle = \frac{E_0^2}{2c\mu_0}. \tag{2.374}$$

The differential cross-section is the ratio of these two quantities, see (2.361), i.e.:

► $$\mathrm{d}\sigma = \frac{e^4\omega^4}{16\pi^2\varepsilon_0^2 c^4 m_\mathrm{e}^2 \left((\omega_0^2-\omega^2)^2+4\delta^2\omega^2\right)} \sin^2\theta \,\mathrm{d}\Omega \tag{2.375}$$

In the case of Thomson scattering, averaging was not performed because the generated and incident waves were in phase, causing the time-dependent terms to cancel out. Integrating over the spatial dependence yields the total cross-section:

► $$\sigma_{\text{tot}} = \frac{e^4\omega^4}{6\pi\varepsilon_0^2 c^4 m_\mathrm{e}^2 \left((\omega_0^2-\omega^2)^2+4\delta^2\omega^2\right)}. \tag{2.376}$$

It is noted that for a free electron, i.e., $\omega_0 = 0$ and $\delta = 0$, both formulas reduce to relations (2.363) and (2.364) for Thomson scattering. For low frequencies $\omega << \omega_0$ (which holds, for example, for electrons bound in atoms and molecules and for the frequency of sunlight incident from the Sun), the expression reduces to:

► $$\sigma_{\text{tot}} = \frac{e^4}{6\pi\varepsilon_0^2 c^4 m_\mathrm{e}^2}\left(\frac{\omega}{\omega_0}\right)^4. \tag{2.377}$$

This represents the well-known Rayleigh formula for the scattering of sunlight in the atmosphere. The cross-section decreases with the fourth power of the wavelength. Blue light is scattered much more efficiently than red light, which explains the blue color of the sky. The scattering mechanism is named after the English physicist John William Strutt, who was raised to the peerage for his research and became Lord Rayleigh.

2.7.5 Bremsstrahlung of Electrons on Ions

If an electron enters the field of a heavy Z -times ionized ion, an elastic collision occurs, altering the direction of the electron motion. During the collision, a centripetal acceleration (directed toward the ion) acts on the electron, resulting in electron radiation. If the collision is described in a coordinate system attached to the heavy ion, the magnitude of the electron acceleration is

$$\ddot{r}_\mathrm{e} = \frac{F}{m_\mathrm{e}} = \frac{Ze^2}{4\pi\varepsilon_0 r_\mathrm{e}^2 m_\mathrm{e}}, \tag{2.378}$$

yielding the magnitude of the second time derivative of the electric dipole moment:

$$\ddot{p}_\mathrm{E} = e\ddot{r}_\mathrm{e} = \frac{Ze^3}{4\pi\varepsilon_0 r_\mathrm{e}^2 m_\mathrm{e}}. \tag{2.379}$$

The Larmor formula (2.352) shows that radiation depends solely on the magnitude of electron acceleration, regardless of whether it is tangential or centripetal. For a Coulomb collision, acceleration is determined exclusively by the electron position relative to the scattering ion. Thus, radiation intensity or power depends only on the distance from the

nucleus; the closer the electron approaches the ion, the more intensely it radiates. The instantaneous radiated power is readily determined:

$$\mathcal{P}(t) = \frac{\ddot{p}_{\mathrm{E}}^2}{6\pi\varepsilon_0 c^3} = \frac{Z^2 e^6}{96\pi^3 \varepsilon_0^3 c^3 m_{\mathrm{e}}^2 \left[r_{\mathrm{e}}(t)\right]^4}. \tag{2.380}$$

This formula can be used for further analysis. The Fourier spectrum of the radiated power can be calculated from the electron trajectory. Another possibility is to consider a homogeneous sample of electrons with varying impact parameters and average the radiated power for this sample of electrons distributed symmetrically around the ion:

$$\left\langle \mathcal{P} \right\rangle = \frac{Z^2 e^6}{96\pi^3 \varepsilon_0^3 c^3 m_{\mathrm{e}}^2} \int_{r_{\min}}^{\infty} \frac{n_{\mathrm{e}} 4\pi r_{\mathrm{e}}^2 \, \mathrm{d}r_{\mathrm{e}}}{r_{\mathrm{e}}^4} = \frac{4\pi Z^2 e^6 n_{\mathrm{e}}}{96\pi^3 \varepsilon_0^3 c^3 m_{\mathrm{e}}^2} \frac{1}{r_{\min}}. \tag{2.381}$$

However, this introduces additional problems: the lower cutoff of the minimum distance cannot be zero, and the electron exhibits quantum behavior when interacting with the atomic shell. If the de Broglie wavelength calculated from the most probable (thermal) velocity is chosen as the lower bound of the impact parameter

$$r_{\min} \approx \frac{2\pi\hbar}{\sqrt{2 m_{\mathrm{e}} k_{\mathrm{B}} T_{\mathrm{e}}}}, \tag{2.382}$$

the averaged power of a collection of electrons scattering off a single ion is obtained as:

$$\left\langle \mathcal{P} \right\rangle \approx \frac{Z^2 e^6 n_{\mathrm{e}}}{48\pi^3 \varepsilon_0^3 c^3 m_{\mathrm{e}} \hbar} \left(\frac{2 k_{\mathrm{B}} T_{\mathrm{e}}}{m_{\mathrm{e}}} \right)^{1/2}. \tag{2.383}$$

If the scattering occurs on ions with a number density n_{i}, the power radiated per unit volume (power density) equals:

► $$\mathscr{P} = \left\langle \mathcal{P} \right\rangle n_{\mathrm{i}} \approx \frac{Z^2 e^6 n_{\mathrm{e}} n_{\mathrm{i}}}{48\pi^3 \varepsilon_0^3 c^3 m_{\mathrm{e}} \hbar} \left(\frac{2 k_{\mathrm{B}} T_{\mathrm{e}}}{m_{\mathrm{e}}} \right)^{1/2}; \qquad \left[\mathscr{P}\right] = \frac{\mathrm{W}}{\mathrm{m}^3}. \tag{2.384}$$

Note: The derivation performed here is non-quantum. In a quantum calculation, a factor termed the Gaunt factor (named after John Arthur Gaunt) appears in the final relation; this factor is weakly frequency-dependent and its value is approximately unity. In a first approximation, it can therefore be neglected.

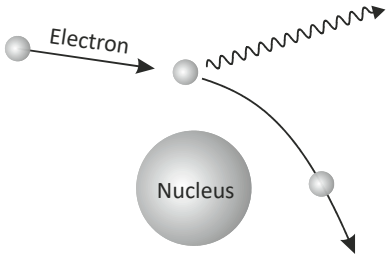


Fig. 2.38: Bremsstrahlung of electrons on ions.

2.7.6 Cyclotron Radiation

A charged particle in a magnetic field undergoes gyration, propagating along a helix with a non-zero centripetal acceleration determined by the Lorentz force. Since cyclotron radiation is important for both electrons and ions, the calculation is performed for a general charge Q. The magnitude of the acceleration of the charged particle is

$$\ddot{r}_0 = \frac{F}{m} = \frac{Q v_\perp B}{m}, \tag{2.385}$$

yielding the magnitude of the second time derivative of the electric dipole moment:

$$\ddot{p}_\mathrm{E} = Q \ddot{r}_0 = \frac{Q^2 v_\perp B}{m}. \tag{2.386}$$

From the Larmor formula (2.352), the instantaneous power radiated by the particle is:

► $$\mathcal{P} = \frac{\ddot{p}_\mathrm{E}^2}{6\pi\varepsilon_0 c^3} = \frac{Q^4 v_\perp^2 B^2}{6\pi\varepsilon_0 c^3 m^2}. \tag{2.387}$$

Due to the quadratic dependence on the particle mass, cyclotron radiation is more intense for electrons than for ions. Specifically for an electron, using the classical electron radius (2.365), the power radiated by a single particle can be rewritten as:

$$\mathcal{P} = \frac{8}{3}\pi a_\mathrm{e}^2 \varepsilon_0 c v_\perp^2 B^2. \tag{2.388}$$

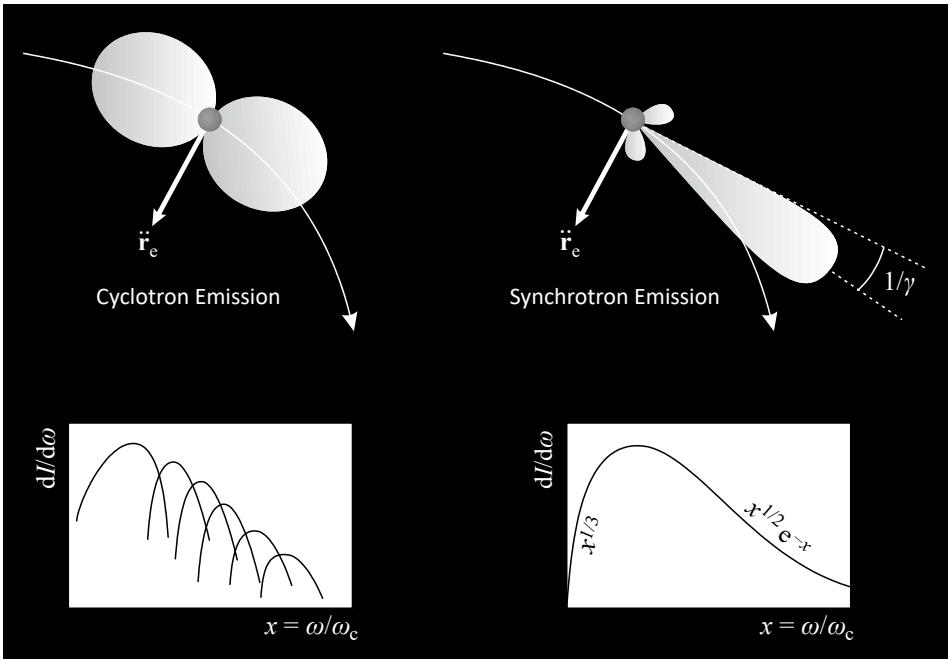


Fig. 2.39: Cyclotron and synchrotron emission.

◖ Example 2.3: Radiation Cooling

The energy balance for the electron component of a plasma is formulated as:

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{3}{2}n_\mathrm{e}k_\mathrm{B}T_\mathrm{e}\right)=\mathcal{P}\,n_\mathrm{e}\,. \tag{2.389}$$

The left-hand side represents the decay of the electron energy density, and the right-hand side represents the radiated power density, where both perpendicular velocity components are replaced by the most probable (thermal) velocities:

$$-\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{3}{2}n_\mathrm{e}k_\mathrm{B}T_\mathrm{e}\right)=\frac{8}{3}\pi a_\mathrm{e}^2\varepsilon_0 c\frac{2k_\mathrm{B}T_\mathrm{e}}{m_\mathrm{e}}B^2\,n_\mathrm{e}\,. \tag{2.390}$$

The derived differential equation for the temperature leads to an exponential decay of the temperature of the electron plasma caused by the cyclotron radiation of electrons:

$$\begin{aligned} T_\mathrm{e} &= T_0\exp[-\alpha t]\,; \\ \alpha &\equiv \frac{32}{9}\frac{\pi a_\mathrm{e}^2\varepsilon_0 c}{m_\mathrm{e}}B^2\,. \end{aligned} \tag{2.391}$$

◗

Note 1: The derivation performed here for cyclotron radiation is non-relativistic. Cyclotron radiation occurs primarily at the cyclotron frequency and its higher harmonics. At high electron energies, however, this calculation is invalid. The cyclotron frequency depends on the particle velocity through the relativistic mass, and it must also be transformed into the laboratory frame, where the relativistic Doppler effect applies. Similarly, the radiated power must be transformed. The velocity dependence of the cyclotron frequency causes the cyclotron emission peaks of a large collection of particles to broaden at individual harmonics, and at high energies, the spectrum becomes continuous. In this scenario, the phenomenon is termed synchrotron radiation. The angular distribution of the radiation is distorted by powers of the Lorentz gamma factor into a characteristic forward lobe shape. The formula for synchrotron radiation can be obtained by transforming the relations for cyclotron radiation; however, a direct calculation from the Feynman formula is more elegant (see *Liénard-Wiechert potentials*, Chapter 2.7.8).

Note 2: Within the Solar System, the planet Jupiter represents a very strong source of cyclotron radiation. Jupiter possesses the strongest magnetic field among all planets, and its magnetosphere features a tail 5.5 astronomical units long, extending to the orbit of Saturn. Electrons gyrating around the magnetic field lines generate intense electromagnetic waves in a narrow band of radio frequencies, abbreviated as ECE (*Electron Cyclotron Emission*). For Jupiter, most waves originate from electrons rotating along the field lines that pass through the moon Io, whose volcanic activity leaves a characteristic plasma torus around the planet composed mainly of sulfur ions. Jupiter is utilized as a reference radio source for calibrating instruments on space probes (such as Planck).

2.7.7 Higher Multipole Radiation

Charged particles do not radiate only via dipole emission; magnetic dipole and electric quadrupole radiation are less intense but also significant. The procedure is identical to that for electric dipole radiation. At a large distance from the source, wavefronts are planar; therefore, it is sufficient to determine only the electric field from the relation

$$\boldsymbol{\mathcal{E}} = -\partial \mathbf{A}/\partial t \tag{2.392}$$

and the magnetic field is readily calculated from the relations:

$$\boldsymbol{\mathcal{B}} = \frac{1}{c}\left(\boldsymbol{\mathcal{E}} \times \mathbf{n}\right); \tag{2.393}$$

$$\boldsymbol{\mathcal{E}} = c\left(\boldsymbol{\mathcal{B}} \times \mathbf{n}\right). \tag{2.394}$$

From formulas (2.340) and (2.341), the electric and magnetic fields of magnetic dipole and electric quadrupole radiation are readily determined:

► $$\boldsymbol{\mathcal{E}}^{(\mathrm{M1})} = -\frac{\mu_0}{4\pi r c}\ddot{\mathbf{p}}_{\mathrm{M}} \times \mathbf{n}, \tag{2.395}$$

► $$\boldsymbol{\mathcal{B}}^{(\mathrm{M1})} = -\frac{\mu_0}{4\pi r c^2}\ddot{\mathbf{p}}_{\mathrm{M}}; \tag{2.396}$$

► $$\boldsymbol{\mathcal{E}}^{(\mathrm{E2})} = -\frac{\mu_0}{24\pi r c}\dddot{\mathbf{Q}} \cdot \vec{\mathbf{n}}, \tag{2.397}$$

► $$\boldsymbol{\mathcal{B}}^{(\mathrm{E2})} = -\frac{\mu_0}{24\pi r c^2}\left(\left[\dddot{\mathbf{Q}} \cdot \vec{\mathbf{n}}\right] \times \mathbf{n}\right); \tag{2.398}$$

Analogously to dipole radiation, the normal component of the Poynting vector is determined, and integration over the entire solid angle yields the total radiated power of the system of particles. In the resulting relations, the vacuum permeability is expressed using the permittivity via the relation $\mu_0 = 1/(c^2\varepsilon_0)$:

$$\mathcal{I}_n^{(\mathrm{M1})} = \frac{\ddot{p}_{\mathrm{M}}^2}{16\pi^2 \varepsilon_0 c^5 r^2}\sin^2\theta; \tag{2.399}$$

$$\mathcal{I}_n^{(\mathrm{E2})} = \frac{\left(\ddot{Q}_{kl} n_l\right)\left(\ddot{Q}_{km} n_m\right) - \left(\ddot{Q}_{kl} n_k n_l\right)\left(\ddot{Q}_{mo} n_m n_o\right)}{576\pi^2 \varepsilon_0 c^5 r^2}; \tag{2.400}$$

► $$\mathcal{P}^{(\mathrm{M1})} = \frac{\ddot{p}_{\mathrm{M}}^2}{6\pi\varepsilon_0 c^5}, \tag{2.401}$$

► $$\mathcal{P}^{(\mathrm{E2})} = \frac{\dddot{Q} : \dddot{Q}}{720\pi\varepsilon_0 c^5}. \tag{2.402}$$

The following figure illustrates electric quadrupole radiation in the immediate vicinity of the source, visualized using the LIC (*Line Integral Convolution*) method. A computer-generated noise texture is distorted in the direction of the electric field lines by means of convolution with the electric field.

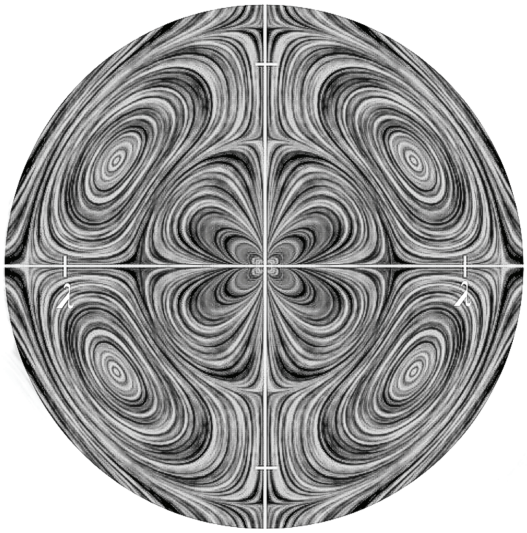


Fig. 2.40: Electric quadrupole radiation. Calculation performed at Duke University.

2.7.8 Liénard-Wiechert Potentials

If particles move at high velocities and are not localized near the origin of the coordinate system, expanding the potentials for a distant observer is not feasible. Furthermore, the corresponding expansion removes the relativistic behavior of the particles. In the general case, a different approach must be adopted. The analysis begins again with the retarded potentials:

$$A^{\alpha} = \frac{\mu_0}{4\pi}\int \frac{J^{\alpha}(t',\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}^3\mathbf{r}'\,; \qquad t' \equiv t - \frac{|\mathbf{r}-\mathbf{r}'(t')|}{c}\,. \tag{2.403}$$

For the scalar and vector potentials, this relation yields the expressions:

$$\phi = \frac{1}{4\pi\varepsilon_0}\int \frac{\rho_Q(t',\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}^3\mathbf{r}'\,, \tag{2.404}$$

$$\mathbf{A} = \frac{\mu_0}{4\pi}\int \frac{\mathbf{j}_Q(t',\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,\mathrm{d}^3\mathbf{r}'\,, \tag{2.405}$$

For a point particle, the source terms are given by:

$$\rho_Q = Q\,\delta(\mathbf{r}'-\mathbf{r}_0)\,, \tag{2.406}$$

$$\mathbf{j}_Q = Q\,\mathbf{v}(t')\,\delta(\mathbf{r}'-\mathbf{r}_0)\,. \tag{2.407}$$

In relations (2.404) and (2.405), the particle parameters must be evaluated at the retarded time; therefore, a temporal distribution is added for a single particle:

$$\phi = \frac{1}{4\pi\varepsilon_0}\int \frac{Q\,\delta(\mathbf{r}'-\mathbf{r}_0)\,\delta\left(t'-t+\left|\mathbf{r}-\mathbf{r}'\right|/c\right)}{\left|\mathbf{r}-\mathbf{r}'\right|}\,\mathrm{d}t'\,\mathrm{d}^3\mathbf{r}', \tag{2.408}$$

$$\mathbf{A} = \frac{\mu_0}{4\pi}\int \frac{Q\,\mathbf{v}(t')\delta(\mathbf{r}'-\mathbf{r}_0)\,\delta\left(t'-t+\left|\mathbf{r}-\mathbf{r}'\right|/c\right)}{\left|\mathbf{r}-\mathbf{r}'\right|}\,\mathrm{d}t'\,\mathrm{d}^3\mathbf{r}'. \tag{2.409}$$

The temporal distribution is non-zero only for

► $$t' = t - \frac{\left|\mathbf{r}-\mathbf{r}'(t')\right|}{c}, \tag{2.410}$$

which corresponds precisely to the retarded time, as required. First, the integration over the spatial distribution is performed:

$$\phi = \frac{Q}{4\pi\varepsilon_0}\int \frac{\delta\left(t'-t+\left|\mathbf{r}-\mathbf{r}_0\right|/c\right)}{\left|\mathbf{r}-\mathbf{r}_0\right|}\,\mathrm{d}t', \tag{2.411}$$

$$\mathbf{A} = \frac{\mu_0}{4\pi}Q\int \frac{\mathbf{v}(t')\,\delta\left(t'-t+\left|\mathbf{r}-\mathbf{r}_0\right|/c\right)}{\left|\mathbf{r}-\mathbf{r}_0\right|}\,\mathrm{d}t'. \tag{2.412}$$

Integration over time is more complex, involving a function distribution for which the relation holds:

► $$\delta\big(F(\xi)\big) = \frac{\delta(\xi-\xi_0)}{F'(\xi_0)}; \qquad F(\xi_0) = 0\,. \tag{2.413}$$

This relation is now utilized. First, the derivative of function F is found:

$$\frac{\mathrm{d}F}{\mathrm{d}t'} = \frac{\mathrm{d}}{\mathrm{d}t'}\big(t'-t+\left|\mathbf{r}-\mathbf{r}_0(t')\right|/c\big) = \frac{\mathrm{d}}{\mathrm{d}t'}\left(t'-t+\frac{1}{c}\sqrt{\big(\mathbf{r}-\mathbf{r}_0(t')\big)^2}\right) \quad \Rightarrow$$

$$\frac{\mathrm{d}F}{\mathrm{d}t'} = 1+\frac{1}{c}\left(-\frac{\mathrm{d}\mathbf{r}_0}{\mathrm{d}t'}\right)2\big(\mathbf{r}-\mathbf{r}_0(t')\big)\frac{1}{2}\frac{1}{\sqrt{\big(\mathbf{r}-\mathbf{r}_0(t')\big)^2}} = 1-\frac{\mathbf{v}\cdot\mathbf{R}}{c\,R} = 1-\boldsymbol{\beta}\cdot\mathbf{n}\,. \tag{2.414}$$

For a non-localized particle, the normal to the surface at the observer position must be chosen as the direction from the particle to the observer, i.e., $\mathbf{n} = (\mathbf{r}-\mathbf{r}_0)/|\,\mathbf{r}-\mathbf{r}_0\,|$; furthermore, the notation $\boldsymbol{\beta} = \mathbf{v}/c$. is introduced.

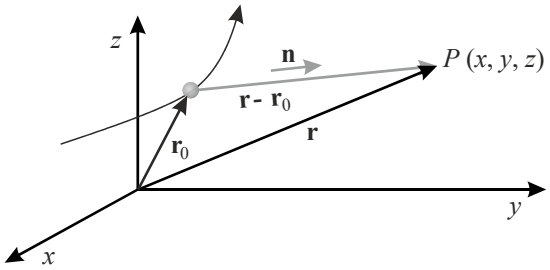


Fig. 2.41: Radiation of a non-localized particle.

After performing the integration over the temporal distribution, the resulting potentials (the *Liénard-Wiechert potentials*) are readily obtained:

$$\phi = \frac{Q}{4\pi\varepsilon_0 R}\frac{1}{\left(1-\boldsymbol{\beta}\cdot\mathbf{n}\right)}, \tag{2.415}$$

$$\mathbf{A} = \frac{Q\boldsymbol{\beta}}{4\pi\varepsilon_0 cR}\frac{1}{\left(1-\boldsymbol{\beta}\cdot\mathbf{n}\right)}, \tag{2.416}$$

where the following notation is introduced:

$$\mathbf{n} \equiv \frac{\mathbf{R}}{R} = \frac{\mathbf{r}-\mathbf{r}_0(t')}{\left|\mathbf{r}-\mathbf{r}_0(t')\right|}, \tag{2.417}$$

$$\boldsymbol{\beta} \equiv \frac{\mathbf{v}(t')}{c}. \tag{2.418}$$

All arguments related to the particle are taken at the retarded time, i.e., $\mathbf{r}_0(t')$ and $\mathbf{v}(t')$. The Liénard-Wiechert potentials are relativistic, and no expansions or simplifications have been made. The electromagnetic fields are now determined via direct calculation. For this purpose, the individual required derivatives must be precalculated:

$$\frac{\partial f}{\partial t} = \frac{\partial f}{\partial t'}\frac{\partial t'}{\partial t} \tag{2.419}$$

$$\frac{\partial f}{\partial x_k} = \frac{\partial f}{\partial t'}\frac{\partial t'}{\partial x_k} \tag{2.420}$$

For instance, $\partial t'/\partial t$ is determined:

$$\frac{\partial t'}{\partial t}\frac{\partial t}{\partial t'} = 1; \qquad \Rightarrow \qquad \frac{\partial t'}{\partial t}\frac{\partial}{\partial t'}\left[t' + \frac{\left|\mathbf{r}-\mathbf{r}_0(t')\right|}{c}\right] = 1; \qquad \Rightarrow \qquad \frac{\partial t'}{\partial t}\left[1 - \frac{\mathbf{v}\cdot\mathbf{R}}{cR}\right] = 1.$$

From this, the following expressions are readily obtained:

$$\frac{\partial t'}{\partial t} = \frac{1}{1-\boldsymbol{\beta}\cdot\mathbf{n}}; \qquad \frac{\partial t}{\partial t'} = 1-\boldsymbol{\beta}\cdot\mathbf{n}. \tag{2.421}$$

Analogously, it is found:

$$\frac{\partial t'}{\partial x_k} = -\frac{n_k}{c\left(1-\boldsymbol{\beta}\cdot\mathbf{n}\right)}. \tag{2.422}$$

The calculation of electric fields from the Liénard-Wiechert potentials is straightforward but somewhat lengthy. The electric field is obtained from the relation $\mathbf{E} = -\nabla\phi - \partial\mathbf{A}/\partial t$, and the magnetic field from the relation $\mathbf{B} = \text{curl}\ \mathbf{A}$. This yields the Feynman formula:

$$\mathbf{E} = \frac{Q}{4\pi\varepsilon_0 R^2}\frac{\left(1-\beta^2\right)\left(\mathbf{n}-\boldsymbol{\beta}\right)}{\left(1-\boldsymbol{\beta}\cdot\mathbf{n}\right)^3} + \frac{Q}{4\pi\varepsilon_0 cR}\frac{\mathbf{n}\times\left[\left(\mathbf{n}-\boldsymbol{\beta}\right)\times\dot{\boldsymbol{\beta}}\right]}{\left(1-\boldsymbol{\beta}\cdot\mathbf{n}\right)^3}, \tag{2.423}$$

► $$\mathbf{B} = \frac{\mathbf{n}\times\mathbf{E}}{c}. \tag{2.424}$$

The first part of the field represents the generalized Coulomb field, while the second part represents the general radiation field, in which no expansions have been made, and is fully relativistic. From the formulas determined in this manner, the Poynting vector and, by its integration, the radiated power can be calculated further. Such a procedure can be used, for example, to obtain the exact relation for synchrotron radiation.

2.7.9 Relativistic Bremsstrahlung and Synchrotron Radiation

If the field in the immediate vicinity of the particle is not of interest and only the radiation field at a larger distance is considered, the first (Coulombic) part in the Feynman formula can be neglected, retaining only the second (radiative) part:

$$\boldsymbol{\mathcal{E}} = \frac{Q}{4\pi\varepsilon_0 cR}\frac{\mathbf{n}\times\left[(\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}\right]}{(1-\boldsymbol{\beta}\cdot\mathbf{n})^3} \tag{2.425}$$

The Poynting vector is now determined:

$$\boldsymbol{\mathcal{S}} = \boldsymbol{\mathcal{E}}\times\boldsymbol{\mathcal{H}} = \frac{1}{\mu_0 c}\boldsymbol{\mathcal{E}}\times(\mathbf{n}\times\boldsymbol{\mathcal{E}}) = \frac{1}{\mu_0 c}\left[\mathcal{E}^2\mathbf{n} - (\boldsymbol{\mathcal{E}}\cdot\mathbf{n})\boldsymbol{\mathcal{E}}\right]. \tag{2.426}$$

The second term in the brackets is clearly zero since the electric field is perpendicular to **n** according to (2.425). The normal component of the energy flux is therefore:

$$\mathcal{S}_\mathrm{n} = \boldsymbol{\mathcal{S}}\cdot\mathbf{n} = \frac{\mathcal{E}^2}{\mu_0 c} \quad \Rightarrow \tag{2.427}$$

$$\mathcal{S}_\mathrm{n} = \frac{1}{\mu_0 c}\left(\frac{Q}{4\pi\varepsilon_0 cR}\right)^2 \frac{\left|\mathbf{n}\times\left[(\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}\right]\right|^2}{(1-\boldsymbol{\beta}\cdot\mathbf{n})^6} \tag{2.428}$$

The entire right-hand side is a function of the retarded time; thus, the energy radiated into a solid angle equals:

$$\mathrm{d}\mathcal{W} = \mathcal{S}_n\,\mathrm{d}t\,R^2\,\mathrm{d}\Omega = \mathcal{S}_n \mathrm{d}t'\frac{\partial t}{\partial t'}R^2\,\mathrm{d}\Omega\,. \tag{2.429}$$

The radiated power is given by:

$$\mathrm{d}\mathcal{P} = \frac{\mathrm{d}\mathcal{W}}{\mathrm{d}t'} = \mathcal{S}_n\frac{\partial t}{\partial t'}R^2\,\mathrm{d}\Omega\,. \tag{2.430}$$

Substituting for the Poynting vector (2.428) and the time derivative (2.421) yields:

► $$\mathrm{d}\mathcal{P} = \frac{Q^2}{16\pi^2\varepsilon_0 c}\frac{\left|\mathbf{n}\times\left[(\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}\right]\right|^2}{(1-\boldsymbol{\beta}\cdot\mathbf{n})^5}\mathrm{d}\Omega\,. \tag{2.431}$$

Three limiting cases of this general relativistic radiation formula are presented here:

Low-Velocity Limit

First, it should be verified that in the low-velocity limit, the new formula reduces to the original Larmor relation. Assuming $\beta \ll 1$, i.e., $1 - \boldsymbol{\beta}\cdot\mathbf{n} \approx 1$ and $\mathbf{n} - \boldsymbol{\beta} \approx \mathbf{n}$:

$$\begin{aligned} \mathrm{d}\mathcal{P} &= \frac{Q^2}{16\pi^2\varepsilon_0 c}\left|\mathbf{n}\times(\mathbf{n}\times\dot{\boldsymbol{\beta}})\right|^2 \mathrm{d}\Omega = \\ &= \frac{Q^2}{16\pi^2\varepsilon_0 c}\left|(\mathbf{n}\cdot\dot{\boldsymbol{\beta}})\,\mathbf{n} - \dot{\boldsymbol{\beta}}\right|^2 \mathrm{d}\Omega = \\ &= \frac{Q^2}{16\pi^2\varepsilon_0 c}\left|\dot{\boldsymbol{\beta}}_\perp\right|^2 \mathrm{d}\Omega = \\ &= \frac{Q^2\dot{v}^2\sin^2\theta}{16\pi^2\varepsilon_0 c^3}\,\mathrm{d}\Omega . \end{aligned}$$

Rewriting this relation using the electric dipole moment $\mathbf{p}_\mathrm{e} = Q\mathbf{r}$ yields:

► $$\mathrm{d}\mathcal{P} = \frac{\ddot{p}_\mathrm{e}^{\,2}\sin^2\theta}{16\pi^2\varepsilon_0 c^3}\,\mathrm{d}\Omega , \tag{2.432}$$

which is the Larmor formula (2.351) for slow particles. It is recalled that θ is the angle between the particle acceleration and the direction from the particle to the observer, i.e.:

$$\theta \equiv \sphericalangle(\dot{\boldsymbol{\beta}}, \mathbf{n}) . \tag{2.433}$$

Relativistic Bremsstrahlung

This case concerns a particle accelerated or decelerated along its motion, i.e., $\boldsymbol{\beta} \parallel \mathrm{d}\boldsymbol{\beta}/\mathrm{d}t$. A typical example is relativistic bremsstrahlung. The general formula simplifies to:

$$\mathrm{d}\mathcal{P}_\parallel = \frac{Q^2}{16\pi^2\varepsilon_0 c}\frac{\left|\mathbf{n}\times(\mathbf{n}\times\dot{\boldsymbol{\beta}})\right|^2}{\left(1-\boldsymbol{\beta}\cdot\mathbf{n}\right)^5}\mathrm{d}\Omega . \tag{2.434}$$

For a particle accelerated in the direction of its motion, the following applies:

$$\theta \equiv \sphericalangle(\dot{\boldsymbol{\beta}}, \mathbf{n}) = \sphericalangle(\boldsymbol{\beta}, \mathbf{n}) . \tag{2.435}$$

The angular dependence of the radiated power is therefore given by:

► $$\mathrm{d}\mathcal{P}_\parallel = \frac{Q^2}{16\pi^2\varepsilon_0 c}\frac{\dot{\beta}^2\sin^2\theta}{\left(1-\beta\cos\theta\right)^5}\mathrm{d}\Omega . \tag{2.436}$$

Integrating over all angular dependencies yields the total radiated power:

► $$\mathcal{P}_\parallel = \frac{Q^2}{6\pi^2\varepsilon_0 c}\gamma^6\dot{v}^2 . \tag{2.437}$$

Synchrotron Radiation

This scenario addresses a particle whose acceleration is perpendicular to its velocity, $\mathrm{d}\boldsymbol{\beta}/\mathrm{d}t \perp \boldsymbol{\beta}$. This occurs, for instance, during circular motion, where synchrotron radiation is generated. The calculation is more complex than in the previous case. An instantaneous coordinate frame is chosen where the z-axis points in the direction of the moving particle and the x-axis in the direction of the particle acceleration (toward the center of the circle in circular motion):

$$\mathbf{e}_z \equiv \frac{\boldsymbol{\beta}}{\beta}; \qquad \mathbf{e}_x \equiv \frac{\dot{\boldsymbol{\beta}}}{\dot{\beta}}; \qquad \boldsymbol{\beta}\cdot\dot{\boldsymbol{\beta}} = 0. \tag{2.438}$$

Standard spherical coordinates are introduced, where ϑ represents the deviation from the z-axis and φ is the polar angle in the (x, y) plane. In this frame:

$$\boldsymbol{\beta}\cdot\mathbf{n} = \beta\cos\vartheta; \qquad \dot{\boldsymbol{\beta}}\cdot\mathbf{n} = \dot{\beta}\sin\vartheta\cos\varphi., \tag{2.439}$$

After converting all vector triple products into scalar products, the general formula (2.431) for acceleration perpendicular to velocity reduces to the relation:

► $$\mathrm{d}\mathcal{P}_\perp = \frac{Q^2}{16\pi^2\varepsilon_0 c}\frac{(1-\beta\cos\vartheta)^2 + (\sin\vartheta\cos\varphi)^2(\beta^2-1)}{(1-\beta\cos\vartheta)^5}\dot{\beta}^2\,\mathrm{d}\Omega\,. \tag{2.440}$$

This is the general formula for synchrotron radiation. Integration over the angles yields:

► $$\mathcal{P}_\perp = \frac{Q^2}{6\pi^2\varepsilon_0 c^3}\gamma^4\dot{v}^2\,. \tag{2.441}$$

If the analysis were not restricted to cases where the acceleration is parallel or perpendicular to the velocity direction and a general direction of velocity and acceleration were retained, the total radiated power would be a simple sum of the two limiting cases (2.437) and (2.441). The integration of the cross-terms yields a zero contribution.

The following figure illustrates the radiation pattern of synchrotron radiation calculated in Wolfram Mathematica according to formula (2.440) for values of $\beta = 0.004$ (left) and $\beta = 0.4$ (right). At high velocities, the radiation strongly dominates in the direction of the particle motion. The minimum amount of radiation (the depression in the figure) is emitted along the line connecting the particle to the center of the circular trajectory.

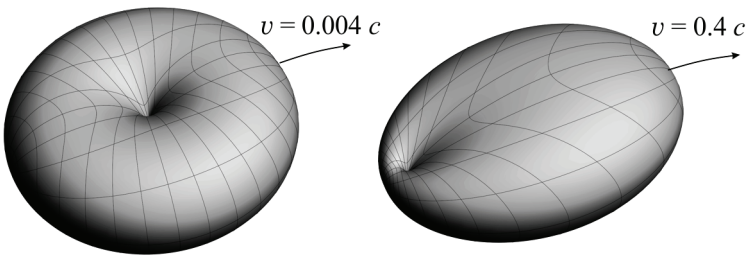


Fig. 2.42: Cyclotron and synchrotron radiation.

The synchrotron radiation power is proportional to γ^4. The energy radiated per unit time is inversely proportional to the fourth power of the particle mass ($E = \gamma m_0 c^2$). Therefore, for identical changes in velocity and energy, electrons radiate much more intensely than protons or ions. In certain devices, synchrotron radiation represents an unwanted parasitic effect, whereas in others, it is utilized for scientific experiments.

An example of unwanted radiation is the largest particle accelerator in the world, the LHC (*Large Hadron Collider*) near Geneva, Switzerland, which features a circumference of 27 kilometers and circulating protons possessing energies of several teraelectronvolts. The accelerator does not utilize a circular tunnel, as might appear at first glance. Instead, there are 8 straight sections with a length of approximately 528 meters. These sections are localized near the detectors. The trajectory of the circulating protons within these sections is not accompanied by synchrotron emission. These represent the only sections where the protons exhibit no electromagnetic signatures. This fact is crucial for the entry of protons into the detector, where collisions occur and a vast number of particles are generated; synchrotron photons would merely represent an impediment here. Furthermore, the straight sections are utilized to focus the proton beam by means of quadrupole magnets. Previously, the LEP II electron-positron collider operated in the same tunnel, where the problem of synchrotron radiation was even more significant due to the lower mass of those particles.

Conversely, other facilities are directly constructed as sources of synchrotron radiation. Generally, these involve accelerated electrons accumulated in a storage ring, where they generate synchrotron radiation during their motion. This category includes, for example, the American synchrotron source NSLS-II (*National Synchrotron Light Source*) at the Brookhaven National Laboratory. The ring circumference is 792 meters, and the electron energy is 3 GeV. In Europe, the operation of the ESRF (*European Synchrotron Radiation Facility*) source, with a circumference of 844 meters and an electron energy of 6 GeV, was concluded in 2019. The device was replaced by the newly constructed fourth-generation source ESRF-EBS (*Extremely Brilliant Source*).

Another source of synchrotron radiation is represented by linear electron accelerators, where electrons are initially accelerated via a system of resonant cavities. Subsequently, high-energy electron bunches enter an *undulator* – a specific magnetic structure with a periodically alternating magnetic field orientation – causing the electrons to propagate along a wavy trajectory where they emit synchrotron radiation. At large electron oscillation amplitudes, the device is termed a *wiggler*. The emitted beams generate an extremely intense pulse of coherent and monochromatic radiation. These devices are called *free-electron lasers* (FELs). Extremely large facilities of this type are designated by the abbreviation XFEL. The largest in the world is the European XFEL, commissioned in 2017. This laser has a length of 3.4 kilometers and is located at the DESY German Research Centre in Hamburg. In second place is the American LCLS (*Linac Coherent Light Source*) at the SLAC National Accelerator Laboratory in California.

The advantage of all these facilities is their tunability, typically ranging from the ultraviolet to the hard X-ray region; for instance, the European XFEL features a range from 0.05 to 4.7 nanometers. Synchrotron radiation sources are utilized for microscopy (where the resolution corresponds roughly to the radiation wavelength) or for the generation and investigation of extreme states of matter.

2.8 Field in Cavity

An electromagnetic field is a medium capable of transferring energy, momentum, and angular momentum. Field energy is quantized, and the elementary quanta are individual photons. Overall, the electromagnetic field can be understood as a system of photons, and each photon as a harmonic oscillator. This concept is expanded here for photons trapped inside a cavity resonator. In this cavity, the field reflects off the internal walls and resides in the resonator for an extended period. Cavity resonators are utilized, for example, in the LHC particle accelerator to accelerate charged particles. These devices were perfected by French physicist Serge Haroche, who succeeded in manufacturing a cavity a few centimeters in size with a superconducting surface in which a single photon reflected for a full 130 ms. During this time, the photon traversed a distance equal to the circumference of the Earth. Serge Haroche was the first to perform non-destructive measurements of the photon number in such cavities. A single photon in the cavity could be repeatedly detected several hundred times. For this work, the Nobel Prize in Physics was awarded for the year 2012. A field enclosed in a resonant cavity is sometimes colloquially referred to as a "field in a box." The cavity field forms standing waves with nodes at the walls. Therefore, only certain modes of the electromagnetic field are allowed in the cavity. These modes are analyzed in this chapter.

2.8.1 Cavity Potentials

The cavity field is source-free; neither charges nor electric currents are present. The Maxwell equations for the potentials along with the Lorenz gauge condition possess a simple form:

$$\Box \mathbf{A} = 0; \tag{2.442}$$

$$\Box \phi = 0; \tag{2.443}$$

$$\frac{1}{c^2}\frac{\partial \phi}{\partial t} + \operatorname{div} \mathbf{A} = 0. \tag{2.444}$$

It was shown in the preceding text that in free space without currents and charges, the scalar potential can even be chosen as zero due to gauge freedom. The equations for the cavity field then take the form:

► $$\Box \mathbf{A} = 0; \tag{2.445}$$

► $$\operatorname{div} \mathbf{A} = 0. \tag{2.446}$$

The electromagnetic fields are given by the relations:

► $$\mathbf{E} = -\frac{\partial \mathbf{A}}{\partial t}; \tag{2.447}$$

► $$\mathbf{B} = \operatorname{curl} \mathbf{A}. \tag{2.448}$$

2.8.2 Natural Field Oscillations

An electromagnetic field is assumed to be located within a rectangular cavity with edges L_1, L_2, and L_3. The coordinate axes are aligned with the edges. A standing wave with nodes at the walls is assumed to form in the cavity. For the desired potential, it is sufficient to require that it has the same value on opposite walls (the so-called periodic boundary condition). The potential is expressed as a superposition of Fourier modes:

$$\mathbf{A}_\mathbf{k}(t,\mathbf{r}) = \mathbf{a}_\mathbf{k}(t)\,\mathrm{e}^{\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}}\,. \tag{2.449}$$

Each mode must satisfy equation (2.445), yielding for the amplitudes $\mathbf{a}_k$:

$$-k^2\mathbf{a}_\mathbf{k} - \frac{1}{c^2}\ddot{\mathbf{a}}_\mathbf{k} = 0 \tag{2.450}$$

Since the cavity field must satisfy the dispersion relation

$$\omega_\mathbf{k} = c\,k\,, \tag{2.451}$$

the equation for the amplitude can be rewritten into the form:

$$\ddot{\mathbf{a}}_\mathbf{k} + \omega_\mathbf{k}^2\mathbf{a}_\mathbf{k} = 0\,. \tag{2.452}$$

This equation has two solutions $\exp[\pm\mathrm{i}\,\omega_\mathbf{k}t]$, the general solution being their linear combination. If the wave vector is required to point in the direction of the phase velocity, and the total phase forms a relativistic scalar $\mathbf{k}\cdot\mathbf{x} - \omega_\mathbf{k}\,t$, the solution with the minus sign must be chosen (the plus solution is also utilized later), yielding:

► $$\mathbf{a}_\mathbf{k}(t) = \mathbf{c}_\mathbf{k}\,\mathrm{e}^{-\mathrm{i}\,\omega_\mathbf{k}\,t} \tag{2.453}$$

The total solution for the vector potential is therefore an ordinary plane wave

► $$\mathbf{A}_\mathbf{k}(t,\mathbf{r}) = \mathbf{a}_\mathbf{k}(t)\,\mathrm{e}^{\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}} = \mathbf{c}_\mathbf{k}\,\mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{r}-\omega_\mathbf{k}\,t]}\,, \tag{2.454}$$

which is a consequence of the simple rectangular shape of the cavity. The Lorenz gauge condition must be taken into account, which implies a constraint on the direction of the constant vectors $\mathbf{c}_\mathbf{k}$:

$$\mathbf{k}\cdot\mathbf{a}_\mathbf{k} = 0 \quad\Rightarrow\quad \mathbf{k}\cdot\mathbf{c}_\mathbf{k} = 0\,. \tag{2.455}$$

The walls of the cavity do not allow the generation of waves of arbitrary wavelengths; therefore, the wave vector $\mathbf{k}$ is not continuous but takes on only certain discrete values. The potential must be identical for $x = 0$ and $x = L_1$, similarly $y = 0$ and $y = L_2$, and the same applies to the third axis for $z = 0$ and $z = L_3$. For instance, along the x-axis:

$$\mathrm{e}^{k_1 0 + k_2 y + k_3 z} = \mathrm{e}^{k_1 L_1 + k_2 y + k_3 z} \quad\Rightarrow\quad 1 = \mathrm{e}^{k_1 L_1} \quad\Rightarrow\quad k_1 L_1 = n_1 2\pi\,. \tag{2.456}$$

An analogous procedure applied to the other axes yields the wave vector values:

► $$\mathbf{k} = \left(\frac{2\pi}{L_1}n_1, \frac{2\pi}{L_2}n_2, \frac{2\pi}{L_3}n_3\right); \qquad n_l = 0, \pm 1, \pm 2\ldots \tag{2.457}$$

$$k = 2\pi\sqrt{\frac{n_1^2}{L_1^2}+\frac{n_2^2}{L_2^2}+\frac{n_3^2}{L_3^2}}\,. \tag{2.458}$$

Upon transitioning to normalized Fourier modes (with $1/\sqrt{L_k}$ for each dimension)

$$\varphi_{\mathbf{k}} \equiv \frac{1}{\sqrt{L_1 L_2 L_3}}\,\mathrm{e}^{\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}} = \frac{1}{\sqrt{V}}\,\mathrm{e}^{\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}}\,, \tag{2.459}$$

the resulting solution for the potential can be written as:

► $$\mathbf{A}(t,\mathbf{r}) = \frac{1}{\sqrt{V}}\sum_{\mathbf{k}}\left(\mathbf{a}_{\mathbf{k}}(t)\,\mathrm{e}^{\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}} + \mathbf{a}_{\mathbf{k}}^{*}(t)\,\mathrm{e}^{-\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}}\right). \tag{2.460}$$

The coefficient preceding the solution is the normalization constant of the basis functions. The complex conjugate solution is added to ensure a real result. Alternatively, the real part Re $\mathbf{A} = (\mathbf{A} + \mathbf{A}^*)/2$ could be utilized, and the factor of one-half absorbed into the amplitudes $\mathbf{a}_{\mathbf{k}}$ and $\mathbf{c}_{\mathbf{k}}$. In the complex conjugate solution, the second solution $\exp[+\mathrm{i}\omega t]$ manifests in the time-dependent part. The solution represents two identical waves traveling in opposite directions, creating a standing wave. The final task is to calculate the electric and magnetic fields. The obtained potential is a superposition of plane electromagnetic waves, where the time dependence is contained within the amplitudes $\mathbf{a}_{\mathbf{k}}$, $\mathbf{a}_{\mathbf{k}}^*$). For non-rectangular cavities, the time-dependent part would not have the form $\exp[\pm\mathrm{i}\,\omega t]$. For plane wavefronts, time and spatial derivatives lead to algebraic relations (2.228). The electric and magnetic fields are therefore immediately obtained:

► $$\mathbf{E} = -\frac{\partial\mathbf{A}}{\partial t} = \frac{\mathrm{i}}{\sqrt{V}}\sum_{\mathbf{k}}\omega_{\mathbf{k}}\left(\mathbf{a}_{\mathbf{k}}(t)\,\mathrm{e}^{\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}} - \mathbf{a}_{\mathbf{k}}^{*}(t)\,\mathrm{e}^{-\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}}\right); \tag{2.461}$$

► $$\mathbf{B} = \operatorname{curl}\mathbf{A} = \frac{\mathrm{i}}{\sqrt{V}}\sum_{\mathbf{k}}\mathbf{k}\times\left(\mathbf{a}_{\mathbf{k}}(t)\,\mathrm{e}^{\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}} - \mathbf{a}_{\mathbf{k}}^{*}(t)\,\mathrm{e}^{-\mathrm{i}\,\mathbf{k}\cdot\mathbf{r}}\right). \tag{2.462}$$

The amplitude of the **k**-th Fourier component is $\mathbf{a}_{\mathbf{k}}$, where the numbers **k** are discrete according to relation (2.457). Both fields are real superpositions of Fourier modes.

2.8.3 Energy and Momentum of Cavity Field

The calculation of the total energy and total momentum of all electromagnetic field modes enclosed in the cavity is performed by evaluating the integrals of energy density (2.179) and momentum density (2.196):

$$\mathcal{E} = \int_V\left(\frac{\mathbf{E}\cdot\mathbf{D}}{2}+\frac{\mathbf{H}\cdot\mathbf{B}}{2}\right)\mathrm{d}^3\mathbf{r} = \varepsilon_0\int_V\left(\frac{\mathbf{E}^2}{2}+c^2\,\frac{\mathbf{B}^2}{2}\right)\mathrm{d}^3\mathbf{r}\,, \tag{2.463}$$

$$\mathcal{G} = \int_V \boldsymbol{\gamma}\,\mathrm{d}^3\mathbf{r} = \int_V \mathbf{D}\times\mathbf{B}\,\mathrm{d}^3\mathbf{r} = \varepsilon_0\int_V \mathbf{E}\times\mathbf{B}\,\mathrm{d}^3\mathbf{r}\,. \tag{2.464}$$

The fields from (2.461) and (2.462) are substituted into both relations. During the simplification of individual terms, the following relations are applied sequentially:

(1) $$\frac{1}{V}\int_V \mathrm{e}^{\mathrm{i}(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}\,\mathrm{d}^3\mathbf{r} = \delta_{\mathbf{k},\mathbf{k}'}, \qquad \frac{1}{V}\int_V \mathrm{e}^{\mathrm{i}(\mathbf{k}+\mathbf{k}')\cdot\mathbf{r}}\,\mathrm{d}^3\mathbf{r} = \delta_{\mathbf{k},-\mathbf{k}'}, \tag{2.465}$$

(2) $$\mathbf{a}_{-\mathbf{k}} = \mathbf{a}_{\mathbf{k}}^*, \qquad \mathbf{a}_{-\mathbf{k}}^* = \mathbf{a}_{\mathbf{k}}, \tag{2.466}$$

(3) $$\mathbf{k}\cdot\mathbf{a}_{\mathbf{k}} = 0, \qquad \mathbf{k}\cdot\mathbf{a}_{\mathbf{k}}^* = 0, \tag{2.467}$$

(4) $$[\mathbf{a}\times\mathbf{b}]\cdot[\mathbf{c}\times\mathbf{d}] = (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) - (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c}). \tag{2.468}$$

The first relation follows immediately from the orthonormality of the basis functions, with the Kronecker deltas of individual discrete indices on the right-hand side. The second relation follows from expression (2.460) for the vector potential, in which the summation index $\mathbf{k}$ is replaced with $-\mathbf{k}$. The index substitution must not alter the expression; after substitution, the second part becomes the first and vice versa. Comparison with the original relation yields the second relation. The third relation follows from the Lorenz gauge condition (2.455) and its complex conjugate. The last relation is the known Cramer formula (see "*Selected Topics I*," Section I.3.3.4 Vector Identities). The calculation of the field energy in the resonant cavity given by relation (2.463) can finally be performed. First, the expressions $\int\mathbf{E}^2\,\mathrm{d}^3\mathbf{r}$ and $\int\mathbf{B}^2\,\mathrm{d}^3\mathbf{r}$ are pre-calculated by substituting the fields from (2.461) and (2.462):

$$\begin{aligned}
\int_V \mathbf{E}^2\mathrm{d}^3\mathbf{r} &= -\frac{1}{V}\int_V \sum_{\mathbf{k},\mathbf{k}'} \omega_{\mathbf{k}}\omega_{\mathbf{k}'}\left(\mathbf{a}_{\mathbf{k}}\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} - \mathbf{a}_{\mathbf{k}}^*\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\right)\cdot\left(\mathbf{a}_{\mathbf{k}'}\,\mathrm{e}^{\mathrm{i}\mathbf{k}'\cdot\mathbf{r}} - \mathbf{a}_{\mathbf{k}'}^*\,\mathrm{e}^{-\mathrm{i}\mathbf{k}'\cdot\mathbf{r}}\right)\mathrm{d}^3\mathbf{r} \overset{(1)}{=} \\
&= -\sum_{\mathbf{k}} \omega_{\mathbf{k}}^2\left(\mathbf{a}_{\mathbf{k}}\cdot\mathbf{a}_{-\mathbf{k}} - \mathbf{a}_{\mathbf{k}}^*\cdot\mathbf{a}_{\mathbf{k}} - \mathbf{a}_{\mathbf{k}}\cdot\mathbf{a}_{\mathbf{k}}^* + \mathbf{a}_{\mathbf{k}}^*\cdot\mathbf{a}_{-\mathbf{k}}^*\right) \overset{(2)}{=} \\
&= -\sum_{\mathbf{k}} \omega_{\mathbf{k}}^2\left(\mathbf{a}_{\mathbf{k}}\cdot\mathbf{a}_{\mathbf{k}}^* - \mathbf{a}_{\mathbf{k}}^*\cdot\mathbf{a}_{\mathbf{k}} - \mathbf{a}_{\mathbf{k}}\cdot\mathbf{a}_{\mathbf{k}}^* + \mathbf{a}_{\mathbf{k}}^*\cdot\mathbf{a}_{\mathbf{k}}\right) = 0.
\end{aligned}$$

The square of the magnetic field is determined through an analogous procedure:

$$\begin{aligned}
\int_V \mathbf{B}^2\mathrm{d}^3\mathbf{r} &= -\frac{1}{V}\int_V \sum_{\mathbf{k},\mathbf{k}'}\left(\mathbf{k}\times\mathbf{a}_{\mathbf{k}}\,\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} - \mathbf{k}\times\mathbf{a}_{\mathbf{k}}^*\,\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\right)\cdot\left(\mathbf{k}'\times\mathbf{a}_{\mathbf{k}'}\,\mathrm{e}^{\mathrm{i}\mathbf{k}'\cdot\mathbf{r}} - \mathbf{k}\times\mathbf{a}_{\mathbf{k}'}^*\,\mathrm{e}^{-\mathrm{i}\mathbf{k}'\cdot\mathbf{r}}\right)\mathrm{d}^3\mathbf{r} \overset{(1,2)}{=} \\
&= -\sum_{\mathbf{k}}\Big[(\mathbf{k}\times\mathbf{a}_{\mathbf{k}})\cdot(-\mathbf{k}\times\mathbf{a}_{\mathbf{k}}^*) - (\mathbf{k}\times\mathbf{a}_{\mathbf{k}}^*)\cdot(\mathbf{k}\times\mathbf{a}_{\mathbf{k}}) - (\mathbf{k}\times\mathbf{a}_{\mathbf{k}})\cdot(\mathbf{k}\times\mathbf{a}_{\mathbf{k}}^*) + (\mathbf{k}\times\mathbf{a}_{\mathbf{k}}^*)\cdot(-\mathbf{k}\times\mathbf{a}_{\mathbf{k}})\Big] = \\
&= 4\sum_{\mathbf{k}}(\mathbf{k}\times\mathbf{a}_{\mathbf{k}}^*)\cdot(\mathbf{k}\times\mathbf{a}_{\mathbf{k}}) \overset{(4)}{=} \\
&= 4\sum_{\mathbf{k}} k^2(\mathbf{a}_{\mathbf{k}}\cdot\mathbf{a}_{\mathbf{k}}^*) - (\mathbf{k}\cdot\mathbf{a}_{\mathbf{k}})(\mathbf{k}\cdot\mathbf{a}_{\mathbf{k}}^*) \overset{(3)}{=} \\
&= 4\sum_{\mathbf{k}} k^2(\mathbf{a}_{\mathbf{k}}\cdot\mathbf{a}_{\mathbf{k}}^*).
\end{aligned}$$

The resulting electromagnetic field energy after substitution into (2.463) is therefore:

$$\mathcal{E} = \sum_{\mathbf{k}} 2\varepsilon_0 c^2 k^2(\mathbf{a}_{\mathbf{k}}\cdot\mathbf{a}_{\mathbf{k}}^*).$$

The dispersion relation states that $ck = \omega_{\mathbf{k}}$ holds in free space, leading to the final relation for the energy. The relation for the momentum is determined analogously:

$$\mathcal{E} = \sum_{\mathbf{k}} \mathcal{E}_{\mathbf{k}} = \sum_{\mathbf{k}} 2\varepsilon_0 \omega_{\mathbf{k}}^2 (\mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}_{\mathbf{k}}^*), \tag{2.469}$$

$$\mathcal{G} = \sum_{\mathbf{k}} \mathcal{G}_{\mathbf{k}} = \sum_{\mathbf{k}} 2\varepsilon_0 \omega_{\mathbf{k}} (\mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}_{\mathbf{k}}^*) \mathbf{k}. \tag{2.470}$$

The cavity field energy and momentum can be expressed as the sum of the energies and momenta of independent linear harmonic oscillators with "allowed" discrete numbers $\mathbf{k} = (2\pi n_1/L_1, 2\pi n_2/L_2, 2\pi n_3/L_3)$, a description compatible with the corpuscular picture of the electromagnetic field. Between the momentum modes pointing along the wave vector and the energy modes, simple relations hold:

$$\mathcal{G}_{\mathbf{k}} = \frac{\mathcal{E}_{\mathbf{k}}}{\omega_{\mathbf{k}}} \mathbf{k} = \frac{\mathcal{E}_{\mathbf{k}}}{c} \frac{\mathbf{k}}{k} \tag{2.471}$$

$$\mathcal{G}_{\mathbf{k}}^2 = \frac{\mathcal{E}_{\mathbf{k}}^2}{c^2}, \tag{2.472}$$

which is the known relation between photon momentum and energy following from special relativity for particles with zero rest mass (see Chapter 3.1).

2.8.4 Hamilton Equations

The Fourier amplitudes $\mathbf{a}_{\mathbf{k}}$, $\mathbf{a}_{\mathbf{k}}^*$ characterize wave propagation in the resonant cavity in the directions $\mathbf{k}$, $-\mathbf{k}$, as they represent coefficients of the waves $\exp[\pm \mathrm{i}\, \mathbf{k}\cdot\mathbf{r}]$. These complex numbers are used to define new real amplitudes $\mathbf{Q}_{\mathbf{k}}$ and $\mathbf{P}_{\mathbf{k}}$ via the relations:

$$\mathbf{Q}_{\mathbf{k}} \equiv \sqrt{\varepsilon_0} \left(\mathbf{a}_{\mathbf{k}}^* + \mathbf{a}_{\mathbf{k}} \right), \tag{2.473}$$

$$\mathbf{P}_{\mathbf{k}} \equiv \mathrm{i}\omega_{\mathbf{k}} \sqrt{\varepsilon_0} \left(\mathbf{a}_{\mathbf{k}}^* - \mathbf{a}_{\mathbf{k}} \right). \tag{2.474}$$

Both functions are real vector functions. The first is proportional to the real part of $\mathbf{a}_{\mathbf{k}}$, the second to the imaginary part. It is shown that these rescaled amplitudes play the role of generalized coordinates and generalized momenta of the oscillators describing the cavity electromagnetic field. By adding and subtracting the defining equations (with the coefficients moved to the left), the original amplitudes can be reconstructed:

$$\mathbf{a}_{\mathbf{k}} = \frac{1}{2\sqrt{\varepsilon_0}} \left(\mathbf{Q}_{\mathbf{k}} + \frac{\mathrm{i}}{\omega_{\mathbf{k}}} \mathbf{P}_{\mathbf{k}} \right), \tag{2.475}$$

$$\mathbf{a}_{\mathbf{k}}^* = \frac{1}{2\sqrt{\varepsilon_0}} \left(\mathbf{Q}_{\mathbf{k}} - \frac{\mathrm{i}}{\omega_{\mathbf{k}}} \mathbf{P}_{\mathbf{k}} \right). \tag{2.476}$$

The new variables $\mathbf{Q}_{\mathbf{k}}$, $\mathbf{P}_{\mathbf{k}}$ currently possess no meaning other than being real amplitudes constructed from the original complex amplitudes $\mathbf{a}_{\mathbf{k}}$, $\mathbf{a}^*_{\mathbf{k}}$. The original amplitudes incorporate the time factors $\exp[\pm\omega_{\mathbf{k}} t]$ and, according to (2.453) and (2.455), satisfy:

$$\dot{\mathbf{a}}_{\mathbf{k}} = -\omega_{\mathbf{k}} \mathbf{a}_{\mathbf{k}}\,, \tag{2.477}$$

$$\dot{\mathbf{a}}^{*}_{\mathbf{k}} = +\omega_{\mathbf{k}} \mathbf{a}_{\mathbf{k}}\,. \tag{2.478}$$

$$\mathbf{k} \cdot \mathbf{a}_{\mathbf{k}} = 0, \tag{2.479}$$

$$\mathbf{k} \cdot \mathbf{a}^{*}_{\mathbf{k}} = 0. \tag{2.480}$$

These relations imply that the newly introduced amplitudes satisfy the relations:

$$\dot{\mathbf{Q}}_{\mathbf{k}} = \mathbf{P}_{\mathbf{k}}\,, \tag{2.481}$$

$$\ddot{\mathbf{Q}}_{\mathbf{k}} + \omega^2_{\mathbf{k}} \mathbf{Q}_{\mathbf{k}} = 0\,. \tag{2.482}$$

$$\mathbf{k} \cdot \mathbf{Q}_{\mathbf{k}} = 0, \tag{2.483}$$

$$\mathbf{k} \cdot \mathbf{P}_{\mathbf{k}} = 0, \tag{2.484}$$

The amplitude $\mathbf{P}_{\mathbf{k}}$ plays the role of a generalized velocity/momentum for the amplitude $\mathbf{Q}_{\mathbf{k}}$. The second relation follows directly from the wave equation $\Box\mathbf{A} = 0$ for the original amplitudes $\mathbf{a}_{\mathbf{k}}$. The third and fourth relations imply that both new amplitudes are perpendicular to the propagation direction. The total energy of the electromagnetic field in the resonant cavity is now determined using the new amplitudes $\mathbf{Q}_{\mathbf{k}}$, $\mathbf{P}_{\mathbf{k}}$:

► $$\mathcal{E} = \mathcal{H} = \sum_{\mathbf{k}} 2\varepsilon_0 \omega^2_{\mathbf{k}} \,(\mathbf{a}_{\mathbf{k}} \cdot \mathbf{a}^{*}_{\mathbf{k}}) = \sum_{\mathbf{k}} \frac{1}{2}\left(\mathbf{P}^2_{\mathbf{k}} + \omega^2_{\mathbf{k}} \mathbf{Q}^2_{\mathbf{k}}\right). \tag{2.485}$$

The energy expressed via the new amplitudes plays the role of the Hamiltonian for the cavity electromagnetic field. It is quadratic in both variables, as is the case for a harmonic oscillator. It is shown that the new amplitudes are canonically conjugate to each other, meaning they satisfy the Hamilton equations. For this purpose, the Hamiltonian (2.485) is differentiated with respect to both new amplitudes:

$$\frac{\partial \mathcal{H}}{\partial \mathbf{P}_{\mathbf{k}}} = \mathbf{P}_{\mathbf{k}} = \dot{\mathbf{Q}}_{\mathbf{k}}\,;$$

$$\frac{\partial \mathcal{H}}{\partial \mathbf{Q}_{\mathbf{k}}} = \omega^2_{\mathbf{k}} \mathbf{Q}_{\mathbf{k}} = -\ddot{\mathbf{Q}}_{\mathbf{k}} = -\dot{\mathbf{P}}_{\mathbf{k}}\,.$$

Ultimately, the canonical Hamilton equations are obtained:

► $$\dot{\mathbf{Q}}_{\mathbf{k}} = \frac{\partial \mathcal{H}}{\partial \mathbf{P}_{\mathbf{k}}}; \tag{2.486}$$

► $$\dot{\mathbf{P}}_{\mathbf{k}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{Q}_{\mathbf{k}}}. \tag{2.487}$$

The new amplitudes satisfy the Hamilton equations and are therefore canonically conjugate. The cavity electromagnetic field can be described as a system of harmonic oscillators with generalized coordinates $\mathbf{Q}_{\mathbf{k}}$ and generalized momenta $\mathbf{P}_{\mathbf{k}}$.

2.8.5 Thermal Radiation

An electromagnetic field within a cavity in thermodynamic equilibrium with walls at temperature T exhibits a characteristic dependence of intensity/energy density on frequency. Whether intensity or energy density is considered is immaterial. Intensity represents energy flux, which is the energy density multiplied by the propagation velocity – the speed of light in this case. Both quantities differ only by a factor of c. For low frequencies, this dependence was derived by English physicist John Rayleigh and English mathematician and astronomer James Jeans; for high frequencies, the relation was derived by German physicist Wilhelm Wien. The correct relation over the entire frequency range was successfully derived by German physicist Max Planck in 1900 upon introducing the then-artificial assumption of electromagnetic radiation energy quantization.

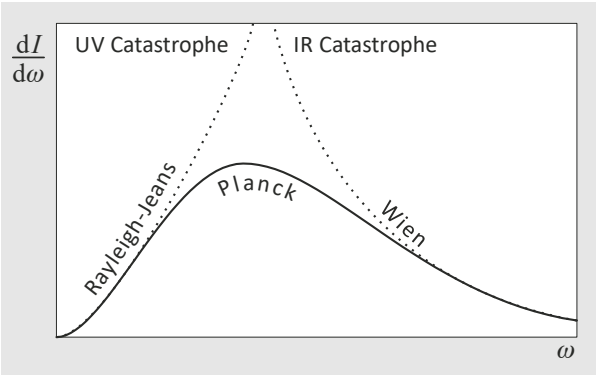


Fig. 2.43: Blackbody radiation.

Cavity Mode Count

The allowed wave vectors **k** from relation (2.457) are taken as a starting point. The number of natural oscillations in the interval d**k** is:

$$\mathrm{d}\Gamma = 2\,\mathrm{d}n_1\,\mathrm{d}n_2\,\mathrm{d}n_3 = 2\frac{L_1L_2L_3}{(2\pi)^3}\mathrm{d}k_1\,\mathrm{d}k_2\,\mathrm{d}k_3 =$$

$$= 2\frac{V}{(2\pi)^3}4\pi k^2\,\mathrm{d}k = 2\frac{V}{(2\pi)^3}4\pi\left(\frac{\omega}{c}\right)^2\frac{\mathrm{d}\omega}{c}.$$

The coefficient 2 preceding the expression accounts for two independent polarizations of electromagnetic radiation. The wave vector differentials are expressed in spherical coordinates, where 4π represents the full solid angle. The differential of the wave vector magnitude is expressed from the dispersion relation $\omega = ck$. The resulting relation for the number of states in the frequency interval dω is:

► $$\mathrm{d}\Gamma_\omega = \frac{V}{\pi^2c^3}\omega^2\,\mathrm{d}\omega. \qquad (2.488)$$

Rayleigh and Jeans Calculation

The radiation intensity is given by the energy flux, which equals the energy density multiplied by the propagation velocity:

$$\mathrm{d}I_\omega = uc\ \mathrm{d}\Gamma_\omega \tag{2.489}$$

If each electromagnetic radiation mode is assumed to correspond to the thermal energy $k_\mathrm{B}T$, i.e., $u = k_\mathrm{B}T/V$, the Rayleigh-Jeans law is obtained:

► $$\mathrm{d}I_\omega = \frac{k_\mathrm{B}T}{\pi^2 c^2}\omega^2\ \mathrm{d}\omega. \tag{2.490}$$

This law holds well for low frequencies and fails completely for high frequencies, where electromagnetic radiation energy quantization becomes essential.

Planck Calculation

Energy is assumed to be quantized, and oscillator states can only exist with energies:

$$\varepsilon_n = n\,\hbar\omega, \qquad n = 0, 1, 2\ldots \tag{2.491}$$

Each oscillator can occur in different energy states with a quantum number n. The probability that a random oscillator possesses exactly the energy ε_n is denoted by w_n:

$$w_n = C\exp\left[-\frac{\varepsilon_n}{k_\mathrm{B}T}\right] = C\exp\left[-n\ \beta\hbar\omega\right]; \qquad \beta \equiv \frac{1}{k_\mathrm{B}T}. \tag{2.492}$$

The normalization constant is calculated from the condition that the sum of all probabilities equals 1:

$$1 = C\sum_{n=0}^{\infty}\exp\left[-\beta\hbar\omega\right]^n = \frac{C}{1-\exp\left[-\beta\hbar\omega\right]} \qquad \Rightarrow$$

$$C = 1-\exp\left[-\beta\hbar\omega\right].$$

The convergence of the geometric series could be verified, but this is unnecessary since the quotient $q = \exp[-x] < 1$ always holds for $x > 0$. The probability is therefore:

► $$w_n = \left(1-\mathrm{e}^{-\beta\hbar\omega}\right)\mathrm{e}^{-\beta n\hbar\omega}; \qquad \beta \equiv \frac{1}{k_\mathrm{B}T}. \tag{2.493}$$

The mean energy of a single oscillator is now determined:

$$\overline{\varepsilon} = \sum_{n=0}^{\infty}\varepsilon_n w_n =$$

$$= \sum_{n=0}^{\infty} n\hbar\omega\left(1-\mathrm{e}^{-\beta\hbar\omega}\right)\mathrm{e}^{-\beta n\hbar\omega} =$$

$$= \hbar\omega\left(1-\mathrm{e}^{-\beta\hbar\omega}\right)\sum_{n=0}^{\infty} n\mathrm{e}^{-\beta n\hbar\omega}.$$

The required series is easily summed using a simple trick:

$$\sum_{n=0}^{\infty} n\mathrm{e}^{-\alpha n} = -\frac{\partial}{\partial\alpha}\sum_{n=0}^{\infty}\mathrm{e}^{-\alpha n} =$$

$$= -\frac{\partial}{\partial\alpha}\sum_{n=0}^{\infty}\left(\mathrm{e}^{-\alpha}\right)^n = -\frac{\partial}{\partial\alpha}\frac{1}{1-\mathrm{e}^{-\alpha}} = \frac{\mathrm{e}^{-\alpha}}{\left(1-\mathrm{e}^{-\alpha}\right)^2}.$$

This result is substituted into the relation for the mean energy. Dividing both the numerator and denominator by $\exp[-\beta\hbar\omega]$ yields the result:

► $$\bar{\varepsilon} = \frac{\hbar\omega}{\mathrm{e}^{\beta\hbar\omega}-1}. \tag{2.494}$$

The intensity is again calculated using relation (2.489), but the energy density is now replaced by the mean oscillator energy divided by the volume:

$$\mathrm{d}I_\omega = u\,c\,\mathrm{d}\Gamma_\omega = \frac{\bar{\varepsilon}}{V}c\,\mathrm{d}\Gamma_\omega = \frac{1}{V}\frac{\hbar\omega}{\mathrm{e}^{\beta\hbar\omega}-1}c\frac{V}{\pi^2c^3}\omega^2\,\mathrm{d}\omega.$$

After a rearrangement of the coefficients, the famous Planck radiation law is obtained:

► $$\mathrm{d}I_\omega = \frac{\hbar}{\pi^2c^2}\frac{\omega^3}{\mathrm{e}^{\beta\hbar\omega}-1}\mathrm{d}\omega. \tag{2.495}$$

In the low-frequency limit, a first-order expansion of the exponential in the denominator yields the Rayleigh-Jeans law. The Planck constant vanishes, as quantization is not significant for low frequencies. Conversely, for high frequencies, the exponential in the denominator is dominant, allowing the subtraction of unity to be neglected, which yields the Wien law for high frequencies:

► $$\mathrm{d}I_\omega \approx \begin{cases} \dfrac{1}{\pi^2c^2}k_\mathrm{B}T\,\omega^2\,\mathrm{d}\omega; & \beta\hbar\omega \ll 1, \\ \dfrac{\hbar}{\pi^2c^2}\omega^3\,\mathrm{e}^{-\beta\hbar\omega}\,\mathrm{d}\omega; & \beta\hbar\omega \gg 1. \end{cases} \tag{2.496}$$

* * *

The following figure illustrates a superconducting resonant cavity constructed by French physicist Serge Haroche and his team. It consists of two hemispherical copper mirrors whose surfaces are machined with an accuracy of a few nanometers, serving as a substrate for a thin layer of niobium, which is superconducting at temperatures below 9 kelvin. The microwave cavity has an open geometry to allow Rydberg atoms, serving as measuring devices, to fly through its center one after another. This setup is capable of non-destructively detecting the photon number in the cavity. A Rydberg atom is in a superposition of two energy states and behaves as a rotating dipole. In the presence of a photon, its rotation phase changes. At a temperature of 0.8 K, the cavity is empty for 95 % of the time according to the Planck radiation law, but a single photon occurs inside it for 5 % of the time. The Haroche microwave cavity is of such high quality that

each generated photon survives inside for approximately 130 milliseconds. This duration is long enough for a photon reflecting between the mirrors to traverse a total path of around 40,000 kilometers, corresponding to the circumference of the Earth. During this time, up to a thousand Rydberg atoms pass through the cavity, capable of repeatedly detecting individual photons several hundred times. Naturally, the microwave cavity does not need to contain only a single photon; several photons can be present (by raising the temperature) or a larger quantity of photons can be generated using an electromagnetic pulse. For the first non-destructive measurement of the photon number, Haroche was awarded the Nobel Prize in Physics for the year 2012.

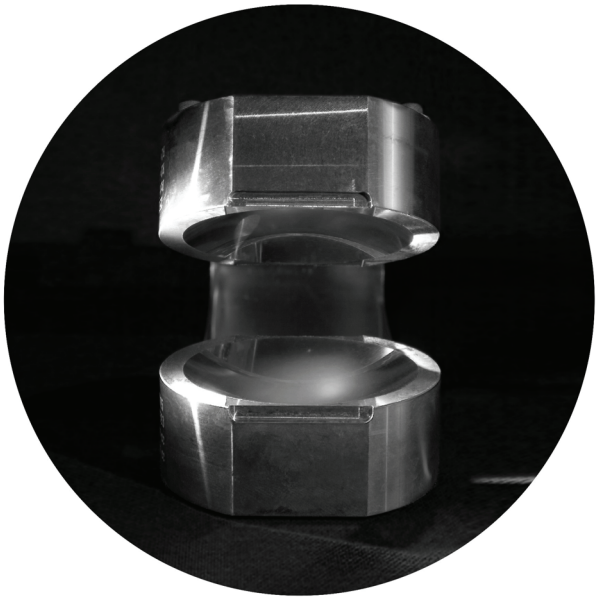
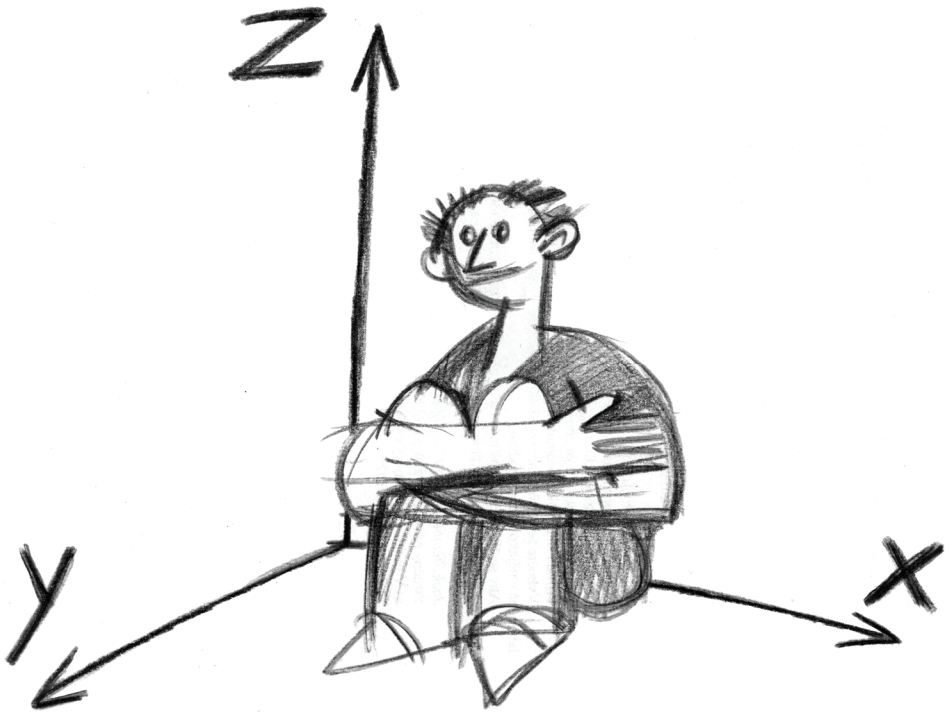


Fig. 2.44: Haroche resonant cavity.

3. Relativity



3.1 Special Relativity

Prior to reading the relativistic section of "*Selected Topics*," thorough familiarity with the concept of a metric and with the contravariant (upper) and covariant (lower) tensor components is required. Upper indices, such as A^{μ}, transform via the Lorentz matrix; lower indices, such as A_{μ}, transform via the inverse Lorentz matrix. The scalar product of two vectors is then expressed in the Einstein summation convention as $A_{\mu}B^{\mu}$ or $g_{\mu\nu}A^{\mu}B^{\nu}$. Indices written with Greek letters range over the values 0, 1, 2, 3. The index 0 corresponds to the time component, whereas indices 1, 2, 3 correspond to the spatial components of the described objects. Details are located in the mathematical section of "*Selected Topics I*," in Section 3.3.6 *Tensors and Metric*.

3.1.1 Lorentz Transformation

Electrodynamics and Classical Mechanics Discrepancy

The understanding of the relativity of motion emerged in three waves. The first dates back to the seventeenth century when Galileo Galilei and Isaac Newton established classical mechanics, wherein time and space form an immutable stage for motion. The second wave, from 1905, is termed special relativity; time and space remain absolute, but temporal or spatial intervals depend on the coordinate system of observation. The last is general relativity from 1916, where time and space become part of the action, co-created by all objects in the universe.

The Galileo principle of relativity introduces an inertial system, an ideal coordinate system in which the law of inertia holds true (the term inertial derives from the Latin *iners*, meaning sluggish or inactive); i.e., bodies remain at rest or move uniformly in a straight line unless acted upon by a force. Different inertial coordinate systems can only move at a constant velocity relative to each other; otherwise, both systems could not be simultaneously inertial. The transformation for the position vector of a particle is assumed to be (with one of the systems denoted by a tilde):

$$\tilde{\mathbf{r}} = \mathbf{r} - \mathbf{v}t \,, \tag{3.1}$$

where $\mathbf{v}$ is the mutual velocity of both systems and $\mathbf{r}$ is the particle position vector. Differentiation with respect to time yields the transformation of the particle velocity $\mathbf{u}$:

$$\tilde{\mathbf{u}} = \mathbf{u} - \mathbf{v} \,. \tag{3.2}$$

This represents the classical addition of velocities to which mechanics is well accustomed. Further differentiation with respect to time yields the relation between the accelerations, since $\mathbf{v}$ is constant:

$$\tilde{\mathbf{a}} = \mathbf{a} \,. \tag{3.3}$$

Identical accelerations, and thus identical forces, act in both inertial systems. Therefore, mechanical processes yield identical results in all inertial systems, and a preferred sys-

tem superior to others cannot be found. This constitutes the essence of the *Galileo principle of relativity*.

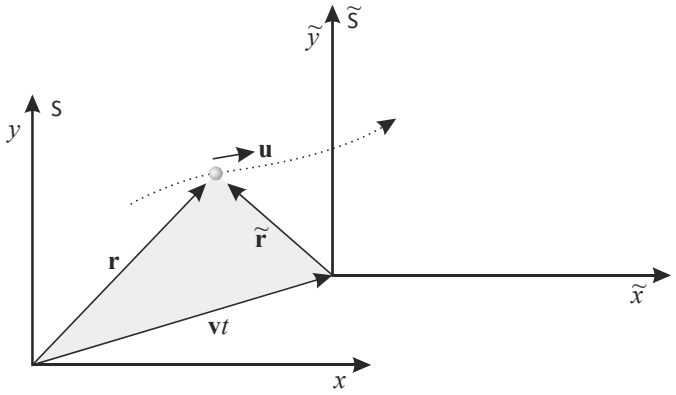


Fig. 3.1: Particle described from two different inertial systems.

In the 18th and 19th centuries, the laws of electricity and magnetism were gradually discovered, and correlations between both phenomena were established experimentally. In 1873, all achieved results were compiled by James Clerk Maxwell in the monumental two-volume work *A Treatise on Electricity and Magnetism* with the famous Maxwell equations, which imply that velocities should combine in a manner different from that assumed by Galileo Galilei, and that the speed of light should even be independent of the choice of coordinate system, possessing the identical value in all systems, as obtained in Section 2.5.2. This apparent contradiction between classical mechanics and Maxwell electrodynamics cannot be resolved analytically; an experimental determination was necessary to decide which theory is correct. This was first accomplished by Albert Abraham Michelson and Edward Morley in their famous 1887 experiment, in which they interferometrically measured the difference in the speed of light on the moving Earth in the direction of its motion around the Sun and in the perpendicular direction. The result validated Maxwell electrodynamics; the speed of light did not depend on the motion of the Earth. Since then, numerous other experiments have demonstrated the correctness of Maxwell electrodynamics.

Lorentz Transformation

The modification of classical mechanics into a form consistent with electrodynamics is termed *special relativity* and dates from 1905. The author is Albert Einstein, who extended the principle of relativity to electromagnetic processes and assumed that mechanical and electromagnetic experiments yield identical results in all inertial systems, meaning no preferred coordinate system can be found by any mechanical or electromagnetic experiment. This necessarily implies that the speed of light must be identical in all systems. For simplicity, two inertial coordinate systems are assumed to move relative to each other solely along the x-axis, and the Galileo transformation (2.7) is corrected using a correction factor $\gamma(v)$ that approaches unity for low velocities:

$$\tilde{x} = \gamma\left(x - vt\right), \tag{3.4}$$

The symmetry of both systems implies that the inverse transformation must possess the identical form but with the opposite direction of velocity:

$$x = \gamma\left(\tilde{x} + v\tilde{t}\right). \tag{3.5}$$

Consider the moment when both systems pass each other, and a flashlight is flashed at the origin of both systems. If light propagates at the identical velocity c in both systems, as required by the Maxwell equations, the following must hold:

$$x = ct; \qquad \tilde{x} = c\tilde{t}\,. \tag{3.6}$$

Therefore, from (3.4) and (3.5), the following is obtained:

$$\begin{aligned} c\tilde{t} &= \gamma\left(ct - vt\right); \\ ct &= \gamma\left(c\tilde{t} + v\tilde{t}\right). \end{aligned} \tag{3.7}$$

Multiplication of both equations yields the relation:

$$c^2 t\tilde{t} = \gamma^2\left(c - v\right)\left(c + v\right)t\tilde{t}\,, \tag{3.8}$$

in which both times cancel, allowing the calculation of the correction factor gamma:

► $$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}\,. \tag{3.9}$$

The new coordinate transformation will therefore be:

$$\tilde{x} = \frac{x - vt}{\sqrt{1 - v^2/c^2}}\,. \tag{3.10}$$

Velocity is distance divided by a time interval. If the speed of light is to be identical in both systems, both spatial coordinates and time must transform. To find this transformation, the system of equations (3.4) and (3.5) can be utilized, in which the coefficient γ is already known. Substituting $\tilde{x}$ from the first equation into the second equation yields, after a simple calculation:

$$\tilde{t} = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}\,. \tag{3.11}$$

The discovered transformation is called the Lorentz transformation, and for an entire event (both time and spatial coordinates), it can be written in the form:

► $$\begin{aligned} \tilde{t} &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}; \\ \tilde{x} &= \frac{x - vt}{\sqrt{1 - v^2/c^2}}; \\ \tilde{y} &= y\ ; \\ \tilde{z} &= z\,, \end{aligned} \tag{3.12}$$

A matrix notation of the Lorentz transformation is more elegant:

$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}_{\tilde{S}} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix}_{S}, \tag{3.13}$$

where the following notation is defined:

$$\beta = v/c\,; \tag{3.14}$$

$$\gamma = 1/\sqrt{1-v^2/c^2} \tag{3.15}$$

and the coordinates are $x_0 = ct$, $x_1 = x$, $x_2 = y$, and $x_3 = z$. The coefficient β signifies a dimensionless velocity, and the coefficient γ is called the Lorentz contraction factor (it will be seen that it expresses the relative length contraction in the direction of motion). The time coordinate x_0 possesses the identical dimension as the spatial coordinates. The Lorentz transformation is named after Dutch physicist Hendrik Antoon Lorentz, who investigated the transformation properties of the Maxwell equations. The transformation was applied to modify the equations of classical mechanics by Albert Einstein in 1905 when he formulated special relativity – a mechanics consistent with electrodynamics.

Different formulations of the Galileo principle of relativity:

- The course of mechanical processes does not depend on the choice of an inertial coordinate system.
- Mechanical experiments yield identical results in all inertial systems moving uniformly in a straight line relative to each other.
- Two inertial coordinate systems cannot be distinguished from each other by any mechanical experiment. A preferred inertial system does not exist.
- The states of motion and rest are relative; an absolute state of rest does not exist, and a unique inertial system at rest does not exist.

Different formulations of the Einstein principle of relativity:

- The course of mechanical and electromagnetic processes does not depend on the choice of an inertial coordinate system.
- Mechanical and electromagnetic experiments yield identical results in all inertial systems moving uniformly in a straight line relative to each other.
- Two inertial coordinate systems cannot be distinguished by any mechanical or electromagnetic experiment. A preferred inertial system does not exist.

New Velocity Addition

The velocity transformation is obtained by differentiating relation (3.12):

$$\tilde{u} = \frac{\mathrm{d}\tilde{x}}{\mathrm{d}\tilde{t}} = \frac{(\mathrm{d}x - v\,\mathrm{d}t)/\sqrt{....}}{(\mathrm{d}t - v\,\mathrm{d}x/c^2)/\sqrt{....}} = \frac{\mathrm{d}x - v\,\mathrm{d}t}{\mathrm{d}t - v\,\mathrm{d}x/c^2}.$$

Dividing both the numerator and the denominator by the differential dt yields:

$$\tilde{u} = \frac{u - v}{1 - uv/c^2}. \tag{3.16}$$

If the mutual motion of the systems is reversed ($v \to -v$), the following is obtained:

$$\tilde{u} = \frac{u + v}{1 + uv/c^2}. \tag{3.17}$$

Transformations (3.16) and (3.17) represent the new velocity addition. The addition in the numerator is identical to the Galileo addition, but the denominator contains a relativistic correction manifesting only at high velocities, which was undetectable from the experimental knowledge of Galileo or Newton. Setting of the velocities in relation (3.17), for instance u, to the speed of light yields:

$$\tilde{u} = \frac{c + v}{1 + cv/c^2} = \frac{c + v}{(c + v)/c} = c\,.$$

The speed of light is thus seen to remain independent of velocity addition in the new transformation, remaining identical in all inertial systems. For low velocities compared to the speed of light, the correction in the denominator is negligible, and velocities combine according to the Galileo transformation.

Time Dilation and Length Contraction

The requirement of a constant speed of light introduces new phenomena. The speed of light is distance divided by a time interval. If this ratio is to be identical in all coordinate systems, distances and time intervals must depend on the coordinate system of observation. An observer is assumed to be in the coordinate system S, while a clock and a rod aligned with the direction of motion are located in the moving system.

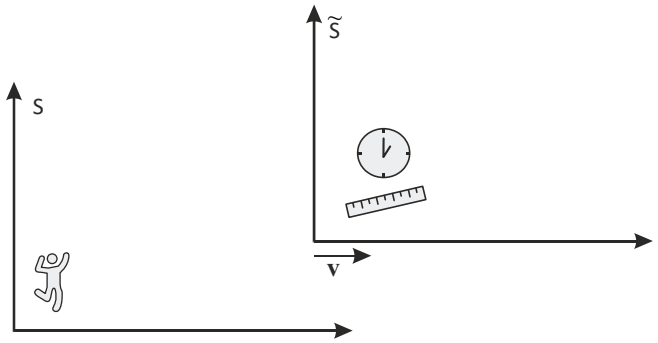


Fig. 3.2: Measurement of time and rod length.

The direct and inverse (differing by the sign of the velocity) Lorentz transformations for time and the x-coordinate are expressed as:

$$\begin{aligned} \tilde{t} &= \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}; \quad t = \frac{\tilde{t} + v\tilde{x}/c^2}{\sqrt{1 - v^2/c^2}}; \\ \tilde{x} &= \frac{x - vt}{\sqrt{1 - v^2/c^2}}; \quad x = \frac{\tilde{x} + v\tilde{t}}{\sqrt{1 - v^2/c^2}}, \end{aligned} \tag{3.18}$$

For finite time intervals and finite spatial distances, the following relations hold:

$$\Delta\tilde{t} = \frac{\Delta t - v\,\Delta x/c^2}{\sqrt{1-v^2/c^2}}; \quad \Delta t = \frac{\Delta\tilde{t} + v\,\Delta\tilde{x}/c^2}{\sqrt{1-v^2/c^2}};$$
$$\Delta\tilde{x} = \frac{\Delta x - v\,\Delta t}{\sqrt{1-v^2/c^2}}; \quad \Delta x = \frac{\Delta\tilde{x} + v\,\Delta\tilde{t}}{\sqrt{1-v^2/c^2}}, \tag{3.19}$$

These are linear dependencies and therefore exact relations. The clock does not move in its rest frame, meaning $\Delta\tilde{x} = 0$, $\Delta\tilde{t} = \Delta t_0$; hence, to transform the time interval, it is advantageous to use the second relation in (3.19), from which the *time dilation* follows:

$$\Delta t = \frac{\Delta t_0}{\sqrt{1-v^2/c^2}}. \tag{3.20}$$

The time interval is the shortest possible in its rest frame. In any other system, the time interval appears longer. Human lifespan thus lasts the shortest duration in the rest frame. The length of the rod is considered next. Its ends must be measured simultaneously in the coordinate system of the observer, i.e., $\Delta t = 0, \Delta\tilde{x} = \Delta x_0$. It is most advantageous to use the third relation in (3.19), from which the *length contraction* follows:

$$\Delta x = \sqrt{1-v^2/c^2}\ \Delta x_0\,. \tag{3.21}$$

The length of moving rods contracts in the direction of motion and is longest in the rest coordinate system. Both relations can be rewritten using the Lorentz contraction factor:

► $$\Delta t = \gamma\,\Delta t_0, \tag{3.22}$$

► $$\Delta x = \Delta x_0/\gamma\,. \tag{3.23}$$

The derivation performed above was purely formal, with both time dilation and length contraction obtained by merely differentiating the Lorentz transformation. Both key relations are derived once more below on the basis of a physical approach.

Physical Approach to Time Dilation

A change in the rate of time must not depend on the specific implementation of a clock. Therefore, a clock of the simplest design is devised, consisting of two parallel mirrors between which a trapped photon reflects back and forth. These reflections are considered the ticks of the clock, appearing different in the reference frame attached to this clock compared to the frame of its motion:

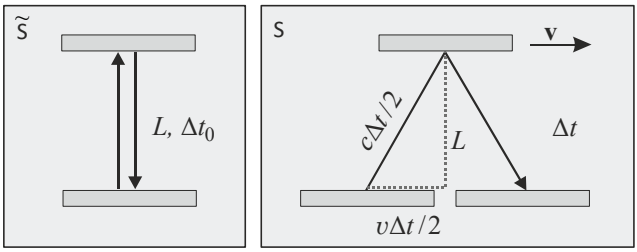


Fig. 3.3: Simple clock implementation – a photon between two mirrors.

The distance between the mirrors is denoted by L, one tick of the clock (the photon returning to the original position on the mirror) is denoted by Δt_0 in the frame attached to the clock, and by Δt in the observer frame. The Pythagorean theorem is now expressed for the triangle indicated on the right side of Fig. 3.3:

$$\frac{c^2\Delta t^2}{4} = \frac{v^2\Delta t^2}{4} + L^2 \tag{3.24}$$

From this relation, the time interval Δt in the observer frame is easily calculated:

$$\Delta t = \frac{2L/c}{\sqrt{1-v^2/c^2}} = \frac{\Delta t_0}{\sqrt{1-v^2/c^2}} \tag{3.25}$$

yielding again:

► $$\Delta t = \gamma\,\Delta t_0. \tag{3.26}$$

Time dilation can therefore be derived directly from the Lorentz transformation or indirectly from the behavior of photons in a simple photon clock. The result is identical.

Physical Approach to Length Contraction

Various high-energy particles, termed cosmic rays, arrive at the Earth from space. When a cosmic ray particle strikes the Earth atmosphere, it collides with an atmospheric atom or molecule, producing an extensive shower of secondary particles at an altitude of approximately 15 kilometers. These include muons (heavy electrons), discovered by American physicist Carl Anderson in 1936. The mean lifetime of muons is approximately two microseconds; if traveling at the highest possible velocity, i.e., the speed of light, they would traverse about 600 meters in 2 microseconds. Consequently, muons should not reach the surface of the Earth. Nevertheless, approximately 10,000 of them arrive per square meter every minute. How is this possible?

From the perspective of an observer on the Earth, muons possess a very high velocity upon creation, approximately 0.994 of the speed of light. For those muons traveling toward the observer on the surface, time dilation occurs, and their lifetime increases approximately tenfold ($\gamma \approx 10$), allowing at least some of them to reach the surface.

However, what is the explanation from the perspective of the coordinate system attached to the moving muon? From the muon perspective, its mean lifetime is a mere 2 microseconds. But the surface of the Earth approaches, distance contraction occurs, and the muon does not have to traverse the full 15 kilometers, but less.

From the perspective of the observer on the Earth, the arrival of muons at the surface is explained by time dilation; from the muon perspective, it is explained by distance contraction. Given that this is the identical phenomenon, the distance to the Earth must contract in the identical ratio as time dilation extends the muon lifetime, yielding:

► $$\Delta l = \Delta l_0/\gamma. \tag{3.27}$$

Length contraction can therefore be derived directly from the Lorentz transformation or indirectly from the equivalent nature of both relativistic phenomena..

Today, time dilation and length contraction are verified by numerous experiments; these phenomena play a primary role especially in large accelerators, which would not function without taking them into account. Length contraction occurs solely in the direction of the mutual motion of both systems, which in this notation corresponds to the x-axis direction. Moving objects are therefore observed to be spatially deformed.

Rapidity

The Lorentz transformation (3.13) possesses very simple properties. The inverse transformation is obtained by merely reversing the sign of the velocity:

$$\Lambda = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \Lambda^{-1} = \begin{pmatrix} \gamma & +\gamma\beta & 0 & 0 \\ +\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.28}$$

The determinant of both matrices equals unity; e.g., for the direct transformation:

$$\det \Lambda = \gamma^2 - \gamma^2\beta^2 = \gamma^2(1-\beta^2) = \frac{1}{1-v^2/c^2} \cdot (1-v^2/c^2) = 1. \tag{3.29}$$

Mathematically, this represents a rotational transformation. By substituting:

$$\begin{aligned} \cosh u &= \gamma; \\ \sinh u &= \gamma\beta \end{aligned} \tag{3.30}$$

the Lorentz matrix can be rewritten in a simple form:

$$\Lambda = \begin{pmatrix} +\cosh u & -\sinh u & 0 & 0 \\ -\sinh u & +\cosh u & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.31}$$

The determinant remains equal to unity, but the transformation matrix is now formally identical to a standard rotation matrix with ordinary sines and cosines. These are replaced by hyperbolic functions in the Lorentz transformation, implying that the rotation angle is imaginary, i.e., $\varphi = \mathrm{i}u$. The rotation clearly occurs in the (t, x) plane. The quantity u is termed *rapidity*, and its value is obtained by dividing equations (3.30):

$$u = \operatorname{atanh}\left(\frac{v}{c}\right). \tag{3.32}$$

For an infinitesimal Lorentz transformation, a Taylor expansion of (3.31) yields:

$$\Lambda_{\text{inf}} = \begin{pmatrix} 1 & -u & 0 & 0 \\ -u & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - u \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.33}$$

The infinitesimal Lorentz transformation can thus be decomposed into two matrix transformations; the first is formed by the identity matrix (the event remains unchanged) and the second matrix merely exchanges the temporal and the first spatial component of the event. The infinitesimal Lorentz transformation is thus one of the simplest mathematical transformations. A finite Lorentz transformation can be composed by multiple repetitions of transformation (3.33) with a small rapidity value.

General Transformation

The Lorentz transformation is typically introduced along the *x*-axis. The general transformation in all three axes simultaneously is considered here. Infinitesimal transformations along individual axes are first introduced analogously to relation (3.33):

► $$\Lambda_{\text{inf},x} \equiv \mathbb{I} - u_x \mathbb{N}_x\,; \quad \Lambda_{\text{inf},y} \equiv \mathbb{I} - u_y \mathbb{N}_y\,; \quad \Lambda_{\text{inf},z} \equiv \mathbb{I} - u_z \mathbb{N}_z, \tag{3.34}$$

where $\mathbb{I}$ is the identity matrix and $\mathbb{N}_k$ are the generators of the Lorentz transformation:

► $$\mathbb{N}_x \equiv \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad \mathbb{N}_y \equiv \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}; \quad \mathbb{N}_z \equiv \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \tag{3.35}$$

The rapidities u_k relate to the mutual velocity of both systems v_k via the relations:

$$u_k = \operatorname{atanh}\left(v_k / c\right), \tag{3.36}$$

The infinitesimal Lorentz transformations (3.34) are elegant and simple. The generators (3.35) only exchange the values of the temporal and one of the spatial components. Compared to the original notation (3.12), this represents a significant advancement. Consider a transformation with a finite rapidity u, for instance along the x-axis. Such a transformation can be composed as n repetitions of a transformation with rapidity u/n:

$$u = n\,\frac{u}{n}\,;. \tag{3.37}$$

Taking the limit $n \to \infty$ corresponds to multiple repetitions of the infinitesimal transformation (3.34):

$$\Lambda_x = \lim_{n\to\infty} \Lambda_{\text{inf}}^{\,n} = \lim_{n\to\infty} \left(\mathbb{I} - \frac{u_x}{n} \mathbb{N}_x \right)^n = \mathrm{e}^{-u_x \mathbb{N}_x} \tag{3.38}$$

This procedure is analogous to the plane rotations described in the "*Selected Topics I*", Section I.3.2.3. The exponential in the last expression is a function of an operator and can be determined, for example, from the spectral decomposition (see Section I.3.4.5). A general Lorentz transformation along all three axes simultaneously takes the form:

► $$\Lambda(\mathbf{u}) = \mathrm{e}^{-\mathbf{u}\cdot\mathbf{N}}\,; \quad u_k = \operatorname{atanh}(v_k/c)\,; \quad \mathbf{N} \equiv \left(\mathbb{N}_x, \mathbb{N}_y, \mathbb{N}_z\right). \tag{3.39}$$

The matrices of a finite Lorentz transformation do not commute, $\Lambda_x \Lambda_y \neq \Lambda_y \Lambda_x$. However, due to the introduction of rapidity, the infinitesimal generators in the argument of the exponential can be simply added, allowing the expression of a general Lorentz transformation (the so-called *Lorentz boost*) in the form (3.39)

3.1.2 Minkowski Metric

Prior to reading this section, the index raising and lowering technique, the concept of a metric, and covariant and contravariant indices are required. Details are located in the "*Selected Topics I*", Section I.3.3.6 Tensors and Metric. The most essential relations and concepts are summarized here, and the most important passages are reviewed.

Four-Vectors

A four-vector is a four-dimensional vector that can be decomposed in a basis $\{\mathbf{e}_\alpha\}$:

► $$\boldsymbol{\mathcal{A}} = \sum_{\alpha=0}^{3} A^\alpha \mathbf{e}_\alpha = A^\alpha \mathbf{e}_\alpha \,. \tag{3.40}$$

Greek indices range over the values 0 (time axis), 1, 2, 3 (spatial axes). The components of a four-vector are denoted by upper (contravariant) indices and transform via the Lorentz matrix. Lower (covariant) indices transform as basis elements. The transformation of lower indices must occur via the inverse Lorentz matrix; only then does combination (3.40) remain an object independent of the choice of basis. The Einstein summation convention remains applicable, but summation always occurs over one lower and one upper index. The result of such a sum is again independent of the choice of basis. The standard position of the transformation matrix indices is the first upper and the second lower. The transformation of four-vectors thus takes the form:

$$\tilde{A}^\alpha = \Lambda^\alpha{}_\beta A^\beta \,; \qquad \Lambda^\alpha{}_\beta \equiv \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.41}$$

The scalar product of two four-vectors can be written as:

$$\boldsymbol{\mathcal{A}} \cdot \boldsymbol{\mathcal{B}} = \left(A^\alpha \mathbf{e}_\alpha\right) \cdot \left(B^\beta \mathbf{e}_\beta\right) = A^\alpha B^\beta \mathbf{e}_\alpha \cdot \mathbf{e}_\beta \,. \tag{3.42}$$

To define the scalar product, it is sufficient to know those of the basis elements:

► $$\boldsymbol{\mathcal{A}} \cdot \boldsymbol{\mathcal{B}} = g_{\alpha\beta} A^\alpha B^\beta ; \tag{3.43}$$

► $$g_{\alpha\beta} \equiv \mathbf{e}_\alpha \cdot \mathbf{e}_\beta \tag{3.44}$$

The coefficients $g_{\alpha\beta}$ are termed the metric, as they determine the scalar product, and thereby the magnitudes of four-vectors and the angles between them, representing the metric properties. The inverse matrix to the metric is introduced via:

$$g^{\alpha\beta} \equiv \left(g_{\alpha\beta}\right)^{-1} \,; \qquad g^{\alpha\beta} g_{\beta\gamma} = \delta^\alpha{}_\gamma \,. \tag{3.45}$$

The identity matrix elements (Kronecker delta) are denoted by $\delta^\alpha{}_\gamma$. The metric can be utilized to lower and raise indices. The standard position is upper for four-vector components and lower for basis elements. Auxiliary dual objects are be introduced via:

$$\mathbf{e}^\alpha \equiv g^{\alpha\beta} \mathbf{e}_\beta \,; \qquad A_\alpha \equiv g_{\alpha\beta} A^\beta \,. \tag{3.46}$$

These represent formal linear combinations given by the metric rather than actual basis elements or actual vector components. An upper index implies transformation via the matrix used for vector components, and a lower index implies transformation via the matrix used for basis elements. With the aid of the metric, indices can be lowered or raised arbitrarily, provided the rule of summing over one upper and one lower index is maintained. Details and examples are located in Section I.3.3.6 of the first volume of "*Selected Topics*". The scalar product can now be expressed in several ways:

$$\mathbf{A}\cdot\mathbf{B} = g_{\alpha\beta}A^{\alpha}B^{\beta} = A^{\alpha}B_{\alpha}\,,$$

where the second index is lowered via the metric. The first index could also be lowered:

$$\mathbf{A}\cdot\mathbf{B} = g_{\alpha\beta}A^{\alpha}B^{\beta} = A_{\beta}B^{\beta} = A_{\alpha}B^{\alpha}\,.$$

yielding:

$$\mathbf{A}\cdot\mathbf{B} = g_{\alpha\beta}A^{\alpha}B^{\beta} = A^{\alpha}B_{\alpha} = A_{\alpha}B^{\alpha}\,. \tag{3.47}$$

The definition of the inverse metric (3.45) can again be understood as lowering or raising indices:

$$\left.\begin{aligned} g^{\alpha\beta}g_{\beta\gamma} &= \delta^{\alpha}{}_{\gamma}\,; \\ g^{\alpha\beta}g_{\beta\gamma} &= g^{\alpha}{}_{\gamma}\,; \end{aligned}\right\} \quad\Rightarrow\quad g^{\alpha}{}_{\gamma} = \delta^{\alpha}{}_{\gamma}\,. \tag{3.48}$$

The metric and the Kronecker delta are thus a single object. If both indices are lower, these are metric coefficients. If both indices are upper, this is the inverse matrix to the metric coefficients, and if the indices are mixed, this represents the Kronecker delta, the elements of the identity matrix. The metric is therefore nothing other than the identity matrix with appropriately shifted indices. Using tensor notation, it can be written as:

$$\mathbf{1} = \delta^{\alpha}{}_{\beta}\mathbf{e}_{\alpha}\otimes\mathbf{e}^{\beta} = g_{\alpha\beta}\mathbf{e}^{\alpha}\otimes\mathbf{e}^{\beta} = g^{\alpha\beta}\mathbf{e}_{\alpha}\otimes\mathbf{e}_{\beta}\,. \tag{3.49}$$

Minkowski Metric

An event (as well as the difference between nearby events) transforms according to the Lorentz matrix in special relativity:

$$\tilde{x}^{\alpha} = \Lambda^{\alpha}{}_{\beta}\,x^{\beta}\,;\quad \mathrm{d}\tilde{x}^{\alpha} = \Lambda^{\alpha}{}_{\beta}\,\mathrm{d}x^{\beta}\,;\quad \Lambda^{\alpha}{}_{\beta} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.50}$$

where $\Lambda^{\alpha}{}_{\beta}$ is the Lorentz transformation matrix. In special relativity, any set of four quantities that transforms via the Lorentz transformation (identically to an event or an infinitesimal difference of events) is understood as the components of a four-vector:

► $$\tilde{V}^{\alpha} = \Lambda^{\alpha}{}_{\beta}V^{\beta}\,. \tag{3.51}$$

The fundamental sets include the *event* (time and space coordinates of the phenomenon). From the Noether theorem, conservation laws of energy and momentum correspond to spacetime translations – these quantities form the *four-momentum* (energy and

momentum). Another set is the *wave four-vector* (angular frequency and wave vector), with individual components relating to the temporal and spatial change of the wave phase. In the section of this book dedicated to the electromagnetic field, the *four-potential* of the electromagnetic field (scalar and vector potential) and the *charge four-current* (the source terms of the Maxwell equations – charge density and charge flux) were introduced. Temporal and spatial derivatives can also be arranged into a set to form a *four-gradient*. In the SI system, all four components of four-vectors must possess the identical dimension. This can be achieved by multiplying or dividing the time component by the universal constant c (the speed of light in vacuum):

$$\begin{aligned} x^\mu &\equiv \begin{pmatrix} ct \\ \mathbf{x} \end{pmatrix}; \qquad P^\mu \equiv \begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix}; \qquad k^\mu \equiv \begin{pmatrix} \omega/c \\ \mathbf{k} \end{pmatrix}; \\ A^\mu &\equiv \begin{pmatrix} \phi/c \\ \mathbf{A} \end{pmatrix}; \qquad J^\mu \equiv \begin{pmatrix} \rho_Q c \\ \mathbf{j}_Q \end{pmatrix}; \qquad \partial_\mu \equiv \begin{pmatrix} \partial/\partial ct \\ \partial/\partial \mathbf{x} \end{pmatrix}. \end{aligned} \tag{3.52}$$

Note 1: Greek indices strictly denote only four-vector components (index 0 corresponds to the time component, indices 1, 2, 3 to the spatial components).

Note 2: The four-gradient possesses a lower index $\partial_\mu = \partial/\partial x^\mu$, as the vector components are in the denominator. If an index is written in the numerator, it must possess the opposite position, as the transformation matrix changes to the inverse.

From the condition of a constant speed of light in all inertial systems, successive operations yield:

$$\frac{\mathrm{d}l}{\mathrm{d}t} = c \quad \Rightarrow \quad \mathrm{d}l = c\mathrm{d}t \quad \Rightarrow$$

$$\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2} = c\,\mathrm{d}t \quad \Rightarrow$$

$$\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 = c^2\mathrm{d}t^2 \quad \Rightarrow$$

$$-c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 = 0 \quad \Rightarrow$$

$$-\mathrm{d}x_0^2 + \mathrm{d}x_1^2 + \mathrm{d}x_2^2 + \mathrm{d}x_3^2 = 0\,. \tag{3.53}$$

The last line represents the wave propagation equation in special relativity, from which the metric of special relativity, the so-called Minkowski metric, can be identified. It is denoted by $\eta_{\alpha\beta}$. It is diagonal and contains a minus sign in the time component. The same applies to the inverse matrix (the metric with upper indices). The metric with mixed indices is the identity matrix, meaning its elements are Kronecker delta symbols:

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}; \qquad \eta^{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}; \tag{3.54}$$

$$\eta^{\mu}{}_{\nu} = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}; \qquad \eta_{\mu}{}^{\nu} = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 \\ 0 & 0 & +1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix}. \tag{3.55}$$

The Minkowski metric is often written simply as $\eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$. Equation (3.53) can be expressed using the Minkowski metric as:

$$\eta_{\mu\nu}\, \mathrm{d}x^{\mu}\, \mathrm{d}x^{\nu} = 0\,. \tag{3.56}$$

With the aid of the metric, the covariant components of standard four-vectors and the contravariant component of the four-gradient are easily determined:

$$\begin{aligned} x_{\mu} &\equiv \begin{pmatrix} -ct \\ \mathbf{x} \end{pmatrix}; \qquad P_{\mu} \equiv \begin{pmatrix} -E/c \\ \mathbf{p} \end{pmatrix}; \qquad k_{\mu} \equiv \begin{pmatrix} -\omega/c \\ \mathbf{k} \end{pmatrix}; \\ A_{\mu} &\equiv \begin{pmatrix} -\phi/c \\ \mathbf{A} \end{pmatrix}; \qquad J_{\mu} \equiv \begin{pmatrix} -\rho_Q c \\ \mathbf{j}_Q \end{pmatrix}; \qquad \partial^{\mu} \equiv \begin{pmatrix} -\partial/\partial ct \\ \partial/\partial \mathbf{x} \end{pmatrix}. \end{aligned} \tag{3.57}$$

The scalar product of two four-vectors is defined as:

$$A_{\mu}B^{\mu} = A^{\mu}B_{\mu} = \eta_{\mu\nu}A^{\mu}B^{\nu} = -A^0B^0 + A^1B^1 + A^2B^2 + A^3B^3\,. \tag{3.58}$$

Typical scalar products leading to known expressions include:

$$k \cdot x = k^{\mu}x_{\mu} = k^0x_0 + k^1x_1 + k^2x_2 + k^3x_3 = -\omega t + \mathbf{k}\cdot\mathbf{x}\,,$$

$$\mathrm{d}s^2 \equiv \mathrm{d}x \cdot \mathrm{d}x = \mathrm{d}x_{\mu}\mathrm{d}x^{\mu} = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\,;$$

$$J \cdot A \equiv J_{\mu}A^{\mu} = -\rho_Q\phi + \mathbf{j}_Q \cdot \mathbf{A}\,;$$

$$\frac{\partial\rho}{\partial t} + \mathrm{div}\,\mathbf{j} = 0 \qquad \Leftrightarrow \qquad \partial_{\mu}j^{\mu} = 0\,;$$

$$\Box f = 0 \qquad \Leftrightarrow \qquad \partial_{\mu}\partial^{\mu}f = 0\,.$$

An abbreviated notation is frequently utilized, where the derivative is written after a comma. Indices before the comma are actual, while those after it represent derivatives:

$$\frac{\partial V^{\mu}}{\partial x^{\nu}} \equiv \partial_{\nu}V^{\mu} \equiv V^{\mu}{}_{,\nu}\,.$$

This represents the most concise notation for a derivative, from which the transformation properties are apparent at first glance. Further examples include:

$$\frac{\partial\varphi}{\partial x_{\mu}} \equiv \partial^{\mu}\varphi \equiv \varphi_{,}{}^{\mu}\,; \qquad \frac{\partial\varphi}{\partial x^{\mu}} \equiv \partial_{\mu}\varphi \equiv \varphi_{,\mu}\,;$$

$$\Box f \equiv \partial_{\mu}\partial^{\mu}f \equiv f_{,\mu}{}^{\mu}\,.$$

Interval and Causality

Relation (3.56) is a wave propagation equation for light, but its left-hand side also possesses independent significance. Let it be denoted as:

$$\mathrm{d}s^2 \equiv \eta_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu\,. \tag{3.59}$$

This combination is termed the *interval* in special relativity. By its very nature, the interval is a scalar, and thus an invariant independent of the choice of coordinate system. Writing the interval explicitly yields:

$$\mathrm{d}s^2 \equiv -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\,. \tag{3.60}$$

When observed from different coordinate systems, time intervals dt and spatial distances dx_k change due to dilation and contraction, but their combination (3.60) is identical in all coordinate systems. This is demonstrated in the following example:

◖ Example 3.1: Invariance of Interval

The interval is expressed in the tilde coordinate system, and individual differentials are sequentially transformed using the Lorentz transformation (3.12) or (3.13):

$$\mathrm{d}\tilde{s}^2 = -c^2\mathrm{d}\tilde{t}^2 + \mathrm{d}\tilde{x}^2 + \mathrm{d}\tilde{y}^2 + \mathrm{d}\tilde{z}^2 =$$

$$= -c^2\left[\gamma\mathrm{d}t - \gamma\beta\,\mathrm{d}x/c\right]^2 + \left[\gamma\mathrm{d}x - \gamma\beta\,\mathrm{d}ct\right]^2 + \mathrm{d}y^2 + \mathrm{d}z^2 =$$

$$= -c^2\gamma^2\mathrm{d}t^2 + 2c\gamma^2\beta\,\mathrm{d}t\mathrm{d}x - \gamma^2\beta^2\mathrm{d}x^2 + \gamma^2\mathrm{d}x^2 - 2\gamma^2\beta c\,\mathrm{d}t\,\mathrm{d}x + \gamma^2\beta^2c^2\mathrm{d}t^2 + \mathrm{d}y^2 + \mathrm{d}z^2 =$$

$$= -c^2\left[\gamma^2 - \gamma^2\beta^2\right]\mathrm{d}t^2 + \left[\gamma^2 - \gamma^2\beta^2\right]\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2\,.$$

However, the coefficients in the square brackets equal unity:

$$\gamma^2 - \gamma^2\beta^2 = \gamma^2(1-\beta^2) = (1-\beta^2)^{-1}(1-\beta^2) = 1\,,$$

and therefore:

$$\mathrm{d}\tilde{s}^2 = \mathrm{d}s^2. \qquad \text{◗}$$

In special relativity, everything changes; the interval alone is a certainty that yields the identical number in all coordinate systems. However, the interval can also be used within a single coordinate system to calculate the "spacetime separation" of two events, as shown below.

Events are represented by points in a spacetime diagram. For instance, in the following figure, event A possesses the coordinates (t_A, x_A, y_A, z_A). This can represent a grain of sand placed on a table. The sequence of events related to a single object is called the worldline of the object. If the grain of sand lies motionless on the table, it is associated with worldline *b*. Time elapses, but the spatial coordinates do not change. The interval between the initial event A and the final event B will be:

$$\Delta s^2 = -c^2(t_B - t_A)^2 + (x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2 = -c^2(t_B - t_A)^2 < 0\,.$$

If the interval between both events is calculated, the spatial part is zero (the grain does not move) and the temporal part is non-zero (time elapses). The result is therefore nega-

tive. If the wind blows, the grain moves, and its fate is described, for example, by worldline *c*. The interval between the initial and final states will be:

$$\Delta s^2 = -c^2(t_C - t_A)^2 + (x_C - x_A)^2 + (y_C - y_A)^2 + (z_C - z_A)^2$$

The result is again negative because the worldline is closer to the time axis, and the temporal part yields a negative contribution. If the interval between two events is negative, the first event can influence the second, since the object can move to the location of the second event at a subluminal velocity. Such a pair of events is said to be causally linked. Events B or C can be consequences of event A. Both are located within the so-called future light cone of event A.

Now consider that event A is related not to a grain of sand, but to a propagating photon. After a certain period, the photon reaches another location, corresponding to event D on the diagram. Its worldline *d* is inclined at an angle of 45° on the diagram, corresponding to the highest possible velocity. The negative temporal part and the positive spatial part contribute equally to the interval, and the interval between events A and D is zero, corresponding to the wave propagation equation (3.56). The initial and final events for light lie on the conical surface given by the relation $\Delta s^2 = 0$.

The figure also displays event E, which lies outside the future light cone of event A. To move to this position from the location corresponding to A, a superluminal velocity would be required. Such two events are not causally linked, and one cannot be the cause of the other. The interval between them is positive. Event E is not located within the future light cone of event A. Analogous considerations lead to the conclusion that event A could only be influenced by events lying within its past light cone.

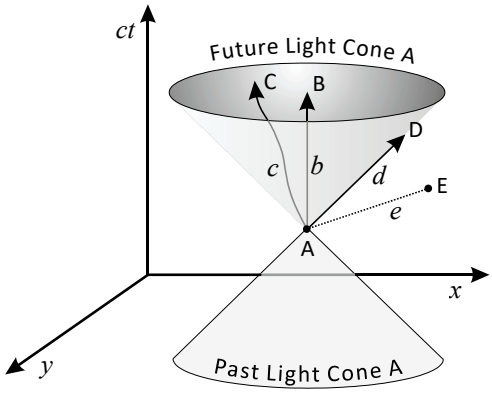


Fig. 3.4: Future light cone and past light cone.

Causal Relationship

Two events can be causally linked if $\Delta s^2 < 0$. Events are not in a causal relationship if $\Delta s^2 > 0$.

The interval is an invariant of the Lorentz transformation yielding the identical result in any coordinate system. This fact ensures that cause and effect are events whose order cannot be reversed by choosing a different coordinate system. Events that have already occurred cannot be influenced in any manner.

Proper Time

If two events are linked to a moving particle, the interval can be expressed both in the laboratory frame and in the frame attached to the particle ($\mathrm{d}\mathbf{x} = 0$, $t = \tau$):

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 = -c^2\mathrm{d}\tau^2 , \tag{3.61}$$

where τ is the proper time, i.e., the time elapsing for the particle itself. It holds that:

► $$\mathrm{d}s^2 = -c^2\mathrm{d}\tau^2 . \tag{3.62}$$

Thus, the proper time is also an invariant of the Lorentz transformation. The relation between the proper time τ and the coordinate time t is determined by expressing the interval in the rest frame of the particle and in a general coordinate system:

$$-c^2\mathrm{d}\tau^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \qquad \Rightarrow$$

$$c^2\mathrm{d}\tau^2 = \mathrm{d}t^2\left(c^2 - (\mathrm{d}x/\mathrm{d}t)^2 - (\mathrm{d}y/\mathrm{d}t)^2 - (\mathrm{d}z/\mathrm{d}t)^2\right) \qquad \Rightarrow$$

$$\mathrm{d}\tau = \sqrt{1 - \mathbf{v}^2/c^2}\;\mathrm{d}t .$$

The relation between both times is therefore:

► $$\mathrm{d}t = \gamma\mathrm{d}\tau ; \qquad \frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma . \tag{3.63}$$

This represents the time dilation formula. Time elapses fastest in the rest frame; the temporal interval between two proper events is the shortest possible. The proper time is utilized as a suitable parameter for the parametrization of the particle worldline. However, it cannot be used for light. For photons, proper time does not elapse ($\mathrm{d}s^2 = 0$), and another parameter must be utilized for parametrization, such as the traversed distance (the proper length of the trajectory).

Transformation of Physical Quantities

The Lorentz matrix can be used to transform the time and spatial coordinates of events from one inertial coordinate system to another. Analogously, any four-vector can be transformed:

► $$\tilde{A}^{\mu} = \Lambda^{\mu}{}_{\nu} A^{\nu} ; \qquad \text{or } \tilde{A} = \Lambda \cdot A , \tag{3.64}$$

which allows simple conversion of quantities between coordinate systems. In the case of a two-index quantity, such as the electromagnetic field tensor, the transformation of each index is performed separately:

► $$\tilde{F}^{\mu\nu} = \Lambda^{\mu}{}_{\alpha}\Lambda^{\nu}{}_{\beta} F^{\alpha\beta} ; \qquad \text{or } \tilde{F} = \Lambda \cdot F \cdot \Lambda^{\mathrm{T}} . \tag{3.65}$$

On the right is the corresponding transformation in matrix notation. A column four-vector is multiplied from the left by the Lorentz matrix. A two-index tensor is multiplied from the left by the Lorentz matrix and from the right by the transposed Lorentz matrix. The result is the four-vector or tensor in the new coordinate system. This allows a straightforward transformation of the electromagnetic field tensor, illustrated by three

examples: 1) calculation of the field of a moving charge, 2) calculation of the relativistic Doppler effect, and 3) transformation of electromagnetic fields between two frames.

Heaviside Field

A charged particle is assumed to fly past an observer at a constant velocity. The fields perceived by the observer are analyzed:

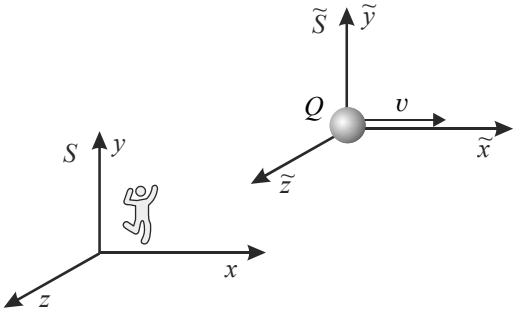


Fig. 3.5: Moving charge.

The transformation of the field four-potential (ϕ/c, **A**) is utilized. In the frame attached to the charge, the vector potential is zero and the scalar potential is given by the Coulomb law:

$$\tilde{\phi} = \frac{Q}{4\pi\varepsilon_0 \tilde{r}} . \tag{3.66}$$

An inverse Lorentz transformation into the observer frame S is performed:

$$\begin{pmatrix} \phi/c \\ A_x \\ A_y \\ A_z \end{pmatrix}_S = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} Q/(4\pi\varepsilon_0 c\tilde{r}) \\ 0 \\ 0 \\ 0 \end{pmatrix} . \tag{3.67}$$

Matrix multiplication yields the potentials in the observer frame:

$$\begin{aligned} \phi &= \frac{\gamma Q}{4\pi\varepsilon_0 \tilde{r}} ; \\ A_x &= \frac{\gamma\beta Q}{4\pi\varepsilon_0 c\tilde{r}} ; \\ A_y &= 0 ; \\ A_z &= 0 . \end{aligned} \tag{3.68}$$

where the following notation is defined:

$$\tilde{r} = \sqrt{\tilde{x}^2 + \tilde{y}^2 + \tilde{z}^2} = \sqrt{\gamma^2 (x - vt)^2 + y^2 + z^2} . \tag{3.69}$$

The magnetic field is non-zero, and the electric field is also modified. The new form of the fields is:

$$E_x = -\frac{\partial \phi}{\partial x} - \frac{\partial A_x}{\partial t} = \frac{\gamma Q\,(x - vt)}{4\pi\varepsilon_0 \left[\gamma^2 (x-vt)^2 + y^2 + z^2\right]^{3/2}},$$

$$E_y = -\frac{\partial \phi}{\partial y} - \frac{\partial A_y}{\partial t} = \frac{\gamma Q\, y}{4\pi\varepsilon_0 \left[\gamma^2 (x-vt)^2 + y^2 + z^2\right]^{3/2}}, \tag{3.70}$$

$$E_z = -\frac{\partial \phi}{\partial z} - \frac{\partial A_z}{\partial t} = \frac{\gamma Q\, z}{4\pi\varepsilon_0 \left[\gamma^2 (x-vt)^2 + y^2 + z^2\right]^{3/2}}.$$

The magnetic field is determined as the curl of the vector potential:

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = 0,$$

$$B_y = \frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} = -\frac{\gamma \beta Q\, z}{4\pi\varepsilon_0 c \left[\gamma^2 (x-vt)^2 + y^2 + z^2\right]^{3/2}}, \tag{3.71}$$

$$B_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} = +\frac{\gamma \beta Q\, y}{4\pi\varepsilon_0 c \left[\gamma^2 (x-vt)^2 + y^2 + z^2\right]^{3/2}}.$$

The perpendicular and parallel components of the electric field are essential, determined at the locations given in the figure by the observer positions $P_{\parallel}$ and $P_{\perp}$:

$$E_{\perp} = \sqrt{E_y^2 + E_z^2}\Big|_{x = vt} = \gamma \frac{Q}{4\pi\varepsilon_0 (y^2 + z^2)}, \tag{3.72}$$

$$E_{\parallel} = E_x\Big|_{\substack{y=0 \\ z=0}} = \frac{1}{\gamma^2} \frac{Q}{4\pi\varepsilon_0 (x - vt)^2}. \tag{3.73}$$

The electric field is stretched across the motion by a factor of γ and compressed along the motion by a factor of $1/\gamma^2$. The field moves along with the charge. The magnetic field forms circles around the axis of the particle motion and weakens with distance from the charge. For an infinite array of charges, the field around a wire would be obtained. The field is named after English physicist and mathematician Oliver Heaviside.

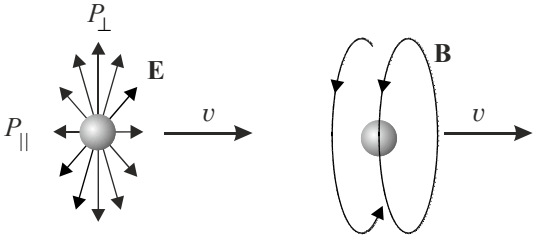


Fig. 3.6: Field of a moving charge.

Doppler Effect

The relativistic Doppler effect is now derived using the transformation of the wave four-vector ω/c, $\mathbf{k}$. The observer is in the frame S, while the source of the periodic signal is in the tilde frame. The wave propagates in all directions; the wave vector $\mathbf{k}$ pointing directly toward the observer is of interest. The angle between the direction of motion of the wave source and the direction toward the observer is denoted by α (a distinction must be made between α^0 in the source frame and α in the observer frame).

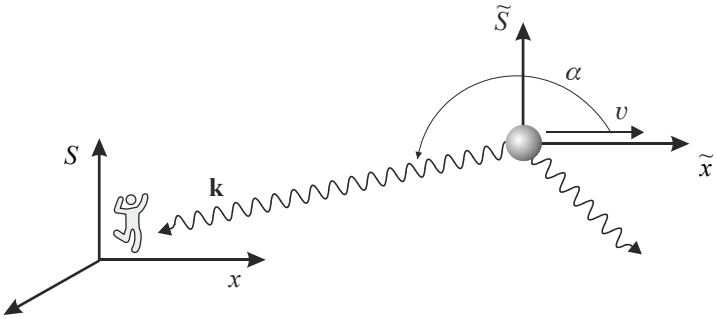


Fig. 3.7: Relativistic Doppler Effect.

In the frame $\tilde{S}$ attached to the radiation source, the solution is straightforward:

$$\begin{pmatrix} \omega/c \\ k_x \\ k_y \\ k_z \end{pmatrix}_{\tilde{S}} = \begin{pmatrix} \omega_0/c \\ k\cos\alpha_0 \\ k\sin\alpha_0 \\ 0 \end{pmatrix}_{\tilde{S}} = \begin{pmatrix} \omega_0/c \\ \omega_0/c\cdot\cos\alpha_0 \\ \omega_0/c\cdot\sin\alpha_0 \\ 0 \end{pmatrix}_{\tilde{S}} . \tag{3.74}$$

An inverse Lorentz transformation into the observer frame S is now performed:

$$\begin{pmatrix} \omega/c \\ \omega_0/c\cdot\cos\alpha \\ \omega_0/c\cdot\sin\alpha \\ 0 \end{pmatrix}_{S} = \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \omega_0/c \\ \omega_0/c\cdot\cos\alpha_0 \\ \omega_0/c\cdot\sin\alpha_0 \\ 0 \end{pmatrix}_{\tilde{S}} . \tag{3.75}$$

Since the frequency at the observer location is of interest, it is sufficient to find only the zeroth row of the matrix multiplication:

► $$\omega = \gamma\left(1+\frac{v}{c}\cos\alpha_0\right)\omega_0 . \tag{3.76}$$

This relation is known as the relativistic Doppler effect. In the low-velocity limit (neglecting quadratic and higher terms in v/c), $\gamma \to 1$ and $\omega = (1 + v/c \cos\alpha_0)\,\omega_0$. When the source moves away, $\alpha_0 = 180°$ and $\omega = (1 - v/c)\,\omega_0$; when the wave source approaches, $\alpha_0 = 0°$ and $\omega = (1 + v/c)\,\omega_0$, yielding the known non-relativistic relations. At higher velocities, these relations are modified by the factor γ. If the radiation source passes the observer ($\alpha_0 = \pm 90°$), $\omega = \gamma\,\omega_0$. A frequency change therefore occurs even when the source is neither moving away nor approaching. This phenomenon is termed the *trans-*

verse Doppler effect and represents a purely relativistic phenomenon with no counterpart in non-relativistic physics. It is caused by the change in the rate of time in the moving frame (time dilation). The spatial relations of the matrix transformation yield:

$$\omega\cos\alpha = \gamma\omega_0(\beta + \cos\alpha_0)\,, \tag{3.77}$$

$$\omega\sin\alpha = \omega_0\sin\alpha_0\,. \tag{3.78}$$

Dividing both equations yields the relation between both angles, which is independent of the frequencies and depends solely on the mutual velocity of the frames:

► $$\tan\alpha = \frac{\sin\alpha_0}{\gamma\beta + \gamma\cos\alpha_0} \tag{3.79}$$

This relation implies that the wavefront altered its direction a change depending solely on the mutual velocity of the frames v. The relativistic Doppler effect was derived by this procedure strictly for light ($\omega = ck$) rather than for a general wave of matter.

General Electromagnetic Field Transformation

The conversion of electromagnetic fields between two different, mutually moving inertial frames represents a critical task. The transformation can be performed via the four-potential according to relation (3.64). The advantage is a rapid conversion, while the disadvantage is the calculation of potentials from fields and subsequently fields from the new potentials. For this reason, the transformation of the electromagnetic field tensor (2.291) via (3.65) is more direct. Both methods lead to the transformation relations:

► $$\tilde{\mathbf{E}} = \gamma\left(\mathbf{E} + \mathbf{v}\times\mathbf{B}\right) - \frac{\gamma^2}{\gamma^2 - 1}\left(\frac{\mathbf{v}}{c}\cdot\mathbf{E}\right)\frac{\mathbf{v}}{c}\,, \tag{3.80}$$

► $$\tilde{\mathbf{B}} = \gamma\left(\mathbf{B} - \frac{\mathbf{v}\times\mathbf{E}}{c^2}\right) - \frac{\gamma^2}{\gamma^2 - 1}\left(\frac{\mathbf{v}}{c}\cdot\mathbf{B}\right)\frac{\mathbf{v}}{c}\,. \tag{3.81}$$

Fields alter even during slow motion, and the ratio of the electric and magnetic field components changes according to the choice of the coordinate system. In the non-relativistic case (where the mutual velocity of the coordinate systems is substantially lower than the speed of light), the field transformation possesses a simpler form:

$$\tilde{\mathbf{E}} = \mathbf{E} + \mathbf{v}\times\mathbf{B}\,, \tag{3.82}$$

$$\tilde{\mathbf{B}} = \mathbf{B}\,. \tag{3.83}$$

A typical application is the formulation of the Ohm law in a moving conductive fluid, for instance in a plasma, where the Ohm law holds in the form:

$$\mathbf{j}_Q = \sigma\tilde{\mathbf{E}}. \tag{3.84}$$

The electric field in the conductive fluid must be expressed using the fields measured in the laboratory frame:

► $$\mathbf{j}_Q = \sigma\left(\mathbf{E} + \mathbf{u}\times\mathbf{B}\right). \tag{3.85}$$

This relation is widely utilized, for example, in non-relativistic magnetohydrodynamics, in which plasma is described as a conductive fluid (see "*Selected Topics III*").

3.1.3 Relativistic Dynamics

Equation of Motion

To define the four-velocity correctly, the event cannot be differentiated with respect to the coordinate time. A four-velocity defined in such a way would not transform via the Lorentz matrix. The four-velocity must be introduced using the proper time, yielding:

► $$U^{\mu} \equiv \frac{\mathrm{d}x^{\mu}}{\mathrm{d}\tau}. \tag{3.86}$$

The proper time acts here as a parameter of the particle motion, i.e., $x^{\mu} = x^{\mu}(\tau)$. The proper time cannot be used as a parameter for light, as transitioning into a frame attached to light is impossible (in such a frame, light would be motionless, which is not permitted). The photon trajectory must be parametrized differently, for instance by the proper length λ of the traversed path. The relationship of the four-velocity to the commonly defined particle velocity is easily found:

► $$U^{\mu} \equiv \frac{\mathrm{d}x^{\mu}}{\mathrm{d}\tau} = \frac{\mathrm{d}x^{\mu}}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{\mathrm{d}x^{\mu}}{\mathrm{d}t}\gamma = \begin{pmatrix} \gamma c \\ \gamma \mathbf{v} \end{pmatrix}. \tag{3.87}$$

For low velocities, the time component is dominant. The four-momentum is introduced using the four-velocity by multiplying it by the rest mass of the particle, yielding:

► $$P^{\mu} \equiv m_0 \frac{\mathrm{d}x^{\mu}}{\mathrm{d}\tau} = \begin{pmatrix} \gamma m_0 c \\ \gamma m_0 \mathbf{v} \end{pmatrix}. \tag{3.88}$$

The transformation properties remain preserved; the four-momentum is a four-vector. If a "relativistic" mass is introduced via the relation

► $$m \equiv \gamma m_0\,, \tag{3.89}$$

the following simple relation for the four-momentum is obtained

► $$P^{\mu} = \begin{pmatrix} mc \\ m\mathbf{v} \end{pmatrix}. \tag{3.90}$$

However, the notation m is not the actual mass, but merely an abbreviation for γm_0. Einstein himself did not introduce the relativistic mass. Comparing the four-momentum (3.52) with expression (3.90) immediately yields the known relations:

► $$E = mc^2\,; \qquad \mathbf{p} = m\mathbf{v}\,; \qquad m \equiv \gamma m_0\,. \tag{3.91}$$

The relation for the energy is the famous Einstein relation, which expresses the equivalence between mass and energy. The significance thereof is addressed below. The second relation replicates the classical definition of momentum. In both relations, it must be kept in mind that this is not the actual mass; the symbol m denotes the product of the Lorentz contraction factor and the rest mass of the particle or object.

The four-acceleration is understood as a four-vector obtained by differentiating the four-velocity with respect to the proper time, yielding:

► $$a^{\mu} \equiv \frac{\mathrm{d}U^{\mu}}{\mathrm{d}\tau}. \tag{3.92}$$

Logic should not be sought in the use of lowercase and uppercase letters, as none is present. The sole criterion is to minimize notation conflicts with other parts of the textbook. The equation of motion in special relativity is written in one of the forms that replicate classical mechanics:

► $$\frac{\mathrm{d}P^{\mu}}{\mathrm{d}\tau} = F^{\mu}, \tag{3.93}$$

$$m_0 \frac{\mathrm{d}U^{\mu}}{\mathrm{d}\tau} = F^{\mu}. \tag{3.94}$$

The proper time and the rest mass must appear on the left-hand side to remain a four-vector transforming via the Lorentz matrix. Consequently, the four-force on the right-hand side is also a four-vector. A correctly chosen relation yields motion in agreement with observation, representing the sole criterion for the four-force selection. Section I.1.6.3 of "*Selected Topics I*" showed that for a charged particle in electric and magnetic fields, the Lorentz force acting on the object is given by relation (1.267):

► $$F^{\mu} = QF^{\mu\nu}U_{\nu}. \tag{3.95}$$

It must always be realized that the four-dimensional notation provides information not only on the temporal change of momentum (the spatial part) but also on the temporal change of the object energy (the time part).

All four-vectors introduced thus far also possess a magnitude, analogously to standard three-dimensional vectors where the magnitude is defined as the square root of the scalar product of the vector with itself. Given that for four-vectors, the scalar product of a vector with itself

$$V \cdot V = V_{\mu}V^{\mu} = \eta_{\mu\nu}V^{\mu}V^{\nu} = -\left(V^0\right)^2 + \left(V^1\right)^2 + \left(V^2\right)^2 + \left(V^3\right)^2 \tag{3.96}$$

can be negative (the time part dominates), problems could arise during square root extraction. Therefore, the magnitude of a four-vector is understood directly as the scalar product (3.96), representing the square of the conventional magnitude. This value can be negative (*timelike*, the four-vector pointing along the time axis), positive (*spacelike*, the four-vector pointing along spatial axes), or *zero* (the four-vector pointing along the light cone). The magnitude of the four-velocity is now determined:

► $$U_{\mu}U^{\mu} = \eta_{\mu\nu}U^{\mu}U^{\nu} = \eta_{\mu\nu}\frac{\mathrm{d}x^{\mu}}{\mathrm{d}\tau}\frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau} = \frac{\eta_{\mu\nu}\mathrm{d}x^{\mu}\mathrm{d}x^{\nu}}{\mathrm{d}\tau^2} = \frac{\mathrm{d}s^2}{\mathrm{d}\tau^2} = \frac{-c^2\mathrm{d}\tau^2}{\mathrm{d}\tau^2} = -c^2. \tag{3.97}$$

The four-velocity is thus a timelike vector. Analogously, the magnitude of the four-momentum of a moving or stationary object is found:

► $$P_{\mu}P^{\mu} = m_0^2 U_{\mu}U^{\mu} = -m_0^2 c^2. \tag{3.98}$$

The rest mass of photons must be zero; otherwise, they could not propagate at the speed of light. Therefore, the magnitude of their four-momentum is zero, representing a null four-vector pointing along the surface of the future light cone.

The four-velocity is always orthogonal to the four-acceleration. To demonstrate this, the four-velocity magnitude is differentiated with respect to the proper time:

$$U_\mu U^\mu = -c^2 \quad \Rightarrow \quad \frac{\mathrm{d}}{\mathrm{d}\tau}\left(U_\mu U^\mu\right) = 0 \quad \Rightarrow$$

$$a_\mu U^\mu + a^\mu U_\mu = 0 \quad \Rightarrow \quad 2a^\mu U_\mu = 0 \quad \Rightarrow$$

► $$a^\mu U_\mu = 0\,. \tag{3.99}$$

The four-vectors defined in this section are also utilized in general relativity, and thus even in cases where the scalar product is given by a general metric $g_{\alpha\beta}$.

Energy in Special Relativity

As already demonstrated, the four-momentum can be expressed in two ways. The first is based on the Emmy Noether theorem. Spacetime translations define conservation laws for quantities termed energy and momentum, forming a four-vector called the four-momentum. The second expression is obtained by multiplying the four-velocity by the rest mass; see (3.88). Comparing both relations yields:

$$\begin{pmatrix} E/c \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} \gamma m_0 c \\ \gamma m_0 \mathbf{v} \end{pmatrix}. \tag{3.100}$$

The spatial parts yield the relation for the relativistic momentum, whereas the time part yields the previously found expression for the energy:

► $$E = mc^2\,; \qquad m \equiv \gamma m_0\,. \tag{3.101}$$

This relation implies that energy and matter are merely two manifestations of the same entity. A change in energy always yields a change in the mass of the object:

► $$\Delta m = \frac{\Delta E}{c^2}\,. \tag{3.102}$$

This relation forms the basis of all nuclear power engineering. Iron-like nuclei are the most tightly bound, possessing the lowest binding energy per nucleon. Energy can therefore be obtained either by splitting heavier nuclei into smaller ones or by fusing nuclei lighter than iron into heavier ones. The second reaction occurs naturally in the Sun, where protons fuse into helium nuclei. Energy is produced during nuclear reactions, which according to relation (3.102) implies that the sum of the particle masses after the reaction is lower than before. This is frequently referred to as the mass deficit. A portion of the matter is converted into energy. If a liter of water is heated, for instance by 20 °C, energy is supplied, and the mass of the liter of water increases by $\Delta E/c^2$. However, such a mass change is undetectable by conventional instruments.

The relation (3.101) for the energy carries another interesting consequence. The ratio of the relativistic energy to the rest energy of the object is expressed as:

$$\frac{mc^2}{m_0 c^2} = \frac{\gamma m_0 c^2}{m_0 c^2} = \gamma\,. \tag{3.103}$$

The Lorentz contraction factor equals the ratio of the actual energy of the particle to its rest energy. It thus plays the role of the particle energy relative to its rest energy. If, for example, $\gamma = 10$, the particle possesses an energy equal to tenfold its rest energy.

Special relativity provides another crucial relation for the energy, following immediately from the magnitude of the four-momentum:

$$P_\mu P^\mu = -m_0 c^2 \quad \Rightarrow \quad -\frac{E^2}{c^2} + \mathbf{p}^2 = -m_0 c^2 \quad \Rightarrow$$

$$\blacktriangleright \qquad E^2 = p^2 c^2 + m_0^2 c^4. \tag{3.104}$$

This relation is sometimes termed the Pythagorean theorem of energy. The energy consists of two parts: the rest energy (given by rest mass) and the kinetic energy (given by momentum). In quantum theory, this relation is key to constructing a relativistic equation for an elementary particle. The result is the Klein-Gordon equation for a spin-0 particle and the Dirac equation for a spin-½ particle (electron). For particles with zero rest mass (the photon), a simple relation between energy and momentum is obtained:

$$\blacktriangleright \qquad E = pc. \tag{3.105}$$

The photon energy is determined by its momentum rather than its mass. Nevertheless, a relativistic mass is attributed to photons, given by a direct conversion from energy:

$$\blacktriangleright \qquad m_\gamma = \frac{E}{c^2} = \frac{\hbar\omega}{c^2}. \tag{3.106}$$

3.1.4 Relativistic Paradoxes

Time dilation and length contraction remain a source of distrust toward special relativity among those who acknowledge only everyday phenomena. These effects are associated with velocities so high that they are not commonly encountered; however, this does not imply they do not exist. Without taking them into account, no accelerator, laser, or GPS navigation would function. Special relativity also introduces seemingly peculiar situations termed paradoxes, which frequently involve astonishment that two different calculations yield disparate results. The solution usually lies in the fact that each calculation applies to a different inertial coordinate system, and an observer cannot be in both simultaneously for an actual discrepancy to occur. Another paradox type relates to assumptions contradicting special relativity, such as non-inertial coordinate systems, or utilizing ideally rigid rods that do not exist in reality. The objective of this textbook is certainly not the resolution of individual paradoxes, so the discussion is limited to a few illustrative examples.

Train and Station Paradox

Consider a train stationary at a station, which is precisely as long as the station building. The following day, the identical train travels along the tracks but does not stop at this station. Due to length contraction, an observer in the station building observes the passing train to be shorter than the building. However, to an observer on the train, the station building appears to move past, and since length contraction occurs, the station building will be shorter than the train.

How is this possible? A train cannot be simultaneously shorter, longer, and equal in length to the station building. The solution to the paradox is straightforward. The pri-

mary error in reasoning concerns the notion that it is always the identical observer measuring the train length in all situations. This is incorrect. If the train is moving, an observer cannot be simultaneously on the train and at the station, meaning two different observers measure two different results. The concept of length is relative, and measurements from different systems yield different results. The reason is the requirement of simultaneous measurement of the positions of both the front and rear parts of the train. The observer at the station records and evaluates the arrival of light (otherwise the train would not be observed) from the front and rear of the train. Light propagates at a finite velocity, so the signal propagation time must be included in the measurement. A careful analysis reveals that events that are simultaneous from the perspective of the station (measuring the positions of the front and rear of the train) are not simultaneous from the perspective of the observer on the train. For the observer on the train, the station measurements occur at different moments, and the train moves in the meantime, yielding a different train length. Result: the paradox does not occur. Incomparable quantities are being compared: measurements performed by different observers and, moreover, measurements that are simultaneous from the perspective of one coordinate system but not the other.

Numerous variants of the paradox just described exist, formulated somewhat more sophistically, yet their solution is analogous.

Rod and Slot Paradox

Consider a perfectly rigid rod flying horizontally at a high velocity above a street slot (without a grate), whose proper length equals the dimension of the slot. From the perspective of the rod, the slot opening contracts in the direction of flight, and the rod does not fall into the slot. From the perspective of the slot, the rod dimension contracts, and it falls into the slot. The paradox is somewhat more serious this time. The falling of the rod into the slot is an objective fact that should not depend on the choice of coordinate system. What is the solution this time?

The solution must be divided into several parts. For the rod to move downward and fall at all, the application of some force, however small, is required to cause the downward motion of the rod. Natural gravity could be used, but this would lead to the immediate objection that general relativity must be applied to solve gravitational problems. This is easily avoided by tapping both ends of the rod simultaneously so that it moves downward. The deceptive nature of this task is already known from the previous paradox. What is simultaneous in one system is not simultaneous from the perspective of the other system, so tapping both ends of the rod simultaneously in both systems is impossible. If the tapping at both ends is not simultaneous, the rod tilts, and its fall into the slot ceases to be paradoxical. Consider an alternative approach where the rod is tapped only once, precisely in the middle. Even this is unrealistic. The center of the rod must be measured relative to its ends, from which the signal travels to the observer for a certain time. Result: the center measured in one system is not the center in the other system. The tapping on the rod will be asymmetric from the perspective of one system, and the rod tilts again.

The greatest problem lies in the assumption of a perfectly rigid rod. If one end were tapped with a hammer, the other end would have to move immediately, implying an infinite velocity of information propagation from one end of the rod to the other. An infinite velocity of interaction propagation is incompatible with special relativity; thus, a perfectly rigid rod does not exist. Every rod is in fact “soft”; if any part is tapped with a

hammer, the entire rod does not move as a result of the blow, but initially only the location that was tapped, generating an elastic wave by which information begins to propagate to other parts of the rod. Consequently, at a sufficiently high velocity, the rod always falls into the slot; it either tilts or deforms due to the acting forces to such an extent that it falls through the slot.



Fig. 3.8: Rod and slot paradox.

Twin Paradox

Among the best-known and most frequently discussed paradoxes is the twin paradox. Consider twins born on the Earth. One spends his entire life on the Earth. The other brother becomes an astronaut and departs in a rocket on a long journey. He travels in a rocket equipped with state-of-the-art technology, reaching nearly the speed of light. The traveler reaches the destination, conducts research on a distant planet, and then flies back to the Earth. After landing, he meets his brother. Which of them will be older? From the perspective of the observer on the Earth, the brother in the rocket traveled at a high velocity, time dilation occurred, and upon meeting, the traveler in the rocket will be younger. However, the traveler reasons identically: the brother on the Earth initially moved away from him at a high velocity and, in the second part of the journey, approached him. Consequently, time dilation occurred for the brother on the Earth, and upon meeting, the one who did not leave the native soil will be younger. What is the solution now? The situations of both brothers are only seemingly symmetric. The brother living on the Earth will have a suitcase packed with things ready on the shelf above his bed. What if he also departed into space someday? His suitcase remains on the shelf for the entire duration of the experiment. The brother on the rocket will have a similar suitcase above his bed. It falls on his head for the first time during launch, when extreme acceleration is perceptible in the rocket, the second time during braking at the destination planet, the third time during launch from it, and the fourth time during landing on the Earth. The bruised head of the traveler undoubtedly concludes that the reference frame in which he is located is very far from an inertial system. The problem can therefore not be resolved from the perspective of special relativity. Within special relativity, the twins would never meet again. Two inertial systems can meet only once in special relativity. The twin paradox concerns a non-inertial system in which a change in the rate of time and all biological processes actually occurs (due to the presence of acceleration). The traveling twin in fact ages less than the one on the Earth.

3.2 Introduction to General Relativity

3.2.1 Equivalence Principle

The mass of a body initially appears straightforward, but its definition involves significant complexity. Mass can be defined in two ways, and whether these two masses are equal represents a fundamental principle. Until recently, a problematic standard of mass stored at the International Bureau of Weights and Measures in Sèvres near Paris was utilized. During each cleaning process, the number of atoms in the standard decreased; hence, the definition of the kilogram constantly changed. Since May 20, 2019, the kilogram has therefore been defined by fixing the value of the Planck constant, which dimensionally contains the kilogram. This step replicates the 1983 definition of the meter, when the meter was defined by fixing the speed of light.

Inertial and Gravitational Mass

The mass of a body can be defined either by its inertial effects or by its gravitational interaction with other bodies.

Inertial Mass

According to the Newton motion law, the acceleration of a body is directly proportional to the acting force and inversely proportional to its mass. This mass is termed inertial mass and can be defined, for instance, as follows: *Inertial mass is the capacity of a body to maintain its state of motion (rest or uniform linear motion)*.

A ball possesses a low inertial mass. It is easily caught (altering its motion to rest) and equally easily thrown (altering its rest to motion). Conversely, a train possesses a high inertial mass from a human perspective – a train can neither be stopped nor set in motion with bare hands. Inertial mass can be measured, for example, directly from the Newton law using the acceleration imparted to a specific body by a given force.

Gravitational Mass

According to the gravitational law, all bodies attract each other directly proportionally to their masses and inversely proportionally to the square of their distance. This mass is termed gravitational mass and can be defined, for instance, as follows: *Gravitational mass is the capacity of bodies to attract each other*.

The Sun possesses a greater gravitational mass than a small stone. Gravitational mass can be measured by weighing, wherein the attraction of objects by the Earth is compared to the attraction of a reference weight.

* * *

If one and the same object is chosen as the standard for both masses, one kilogram of inertial mass equals one kilogram of gravitational mass. However, will this hold true for a larger quantity of matter? Will the inertial mass of an arbitrary object equal its gravitational mass? Are the inertial and gravitational effects identical?

Equivalence Principle

Consider simple motions in a gravity field, described by the equation:

$$m_s \ddot{\mathbf{x}} = \mathbf{F}\,, \tag{3.107}$$

where $\mathbf{F}$ is the gravitational force given by the expression $\mathbf{F} = m_g\mathbf{g}$, and the position vector of the observed object is denoted by $\mathbf{x}$. In this relation, mass expresses the capacity of a body to be attracted by the Earth; hence, gravitational mass is involved

$$m_s \ddot{\mathbf{x}} = m_g \mathbf{g}\,. \tag{3.108}$$

If the masses are equal (their proportionality would suffice), they can be canceled, and the resulting motion will not depend on the mass of the body. A glass marble will move along the identical path as a flying brick; a small screw will fall from a tower to the ground in the identical time as a concrete panel. This assertion naturally does not apply to a feather, where air resistance, neglected in equation (3.108), exerts a decisive influence. The first logical reasoning that motion is independent of mass originates from Galileo Galilei. Consider a falling brick; if a third thereof is broken off and both parts are connected by a thin wire of negligible mass, both parts will fall identically to the original brick. If the wire is cut, the motion remains unaltered, and both parts fall identically again, implying that free fall is independent of mass. This concerns not only a local gravity field but also motion in the vicinity of the Sun, whose gravitational mass is denoted by M_g. The equation of motion for bodies in the vicinity of the Sun will be:

$$m_s \ddot{\mathbf{x}} = -G\frac{m_g M_g}{r^2}\frac{\mathbf{x}}{r}\,. \tag{3.109}$$

If the inertial and gravitational masses of a test body are identical, they cancel, and the motion of bodies around the Sun again does not depend on their mass. A planet and a screw released from a space shuttle will move around the Sun along the identical trajectory, provided their initial positions and velocities were identical. This leads to the formulation of the equivalence principle.

Weak Equivalence Principle

The inertial and gravitational masses of bodies are proportional. When a common standard is chosen for both masses, equality is established. Inertial and gravitational effects are identical during motion. Accelerated motion cannot be distinguished from the local effect of gravity. Consider a rapidly accelerating automobile that presses the occupant into the seat in the identical manner as a gravitational interaction would.

Strong Equivalence Principle

The electromagnetic field is a carrier of energy, momentum, and angular momentum, mediating the transfer of these quantities. According to the relation $E = mc^2$, a certain mass corresponds to the energy of the field. The assumption that this mass also possesses both inertial and gravitational effects (i.e., the electromagnetic field is capable of gravitational attraction) constitutes the substance of the *strong equivalence principle*.

Very Strong Equivalence Principle

The very strong equivalence principle assumes that the gravitational field itself is a carrier of energy, and thus of mass, possessing inertial and gravitational effects that cannot be distinguished from each other.

Experiments

The equivalence principle is a fundamental assertion in any form, leading numerous physicists have to attempt its experimental verification. The first precise experiment was conducted by Hungarian physicist Loránd Eötvös. Refining the method from 1885 onward, Loránd Eötvös, Jenő Fekete, and Dezső Pekár achieved verification of the equivalence between inertial and gravitational mass in 1909 with a relative precision of 5×10^{-9}. The experiment utilized two bodies of different materials placed at the ends of a horizontal rod suspended by a fiber (a torsion balance). The bodies are acted upon by the gravitational force, determined by the gravitational mass of the suspended bodies, and by the centrifugal force of the Earth rotation, determined by their inertial mass. If both masses were not proportional, a torsion force would arise, twisting the rod between the bodies. However, no such twisting occurred despite multiple repetitions of the experiment with various bodies at different locations. The Eötvös experiment was later replicated and refined by other authors. In 1964, Robert Dicke, Robert Krotkov, and Peter Roll modified the experiment to measure the acceleration difference caused by the Sun as well. For a gold/aluminum body combination, they achieved a relative precision of 10^{-11} in verifying the equivalence principle. In 1996, the equivalence between inertial and gravitational mass was verified via laser ranging from the Moon (the Earth–Moon system in the field of the Sun) by Jean Dicke with a relative precision of 4×10^{-13}. In 2012, a group of scientists from the University of Washington led by Eric Adelberg conducted experiments with torsion balances (beryllium/aluminum, beryllium/titanium), verifying the equivalence principle with a relative precision of 3×10^{-14}.

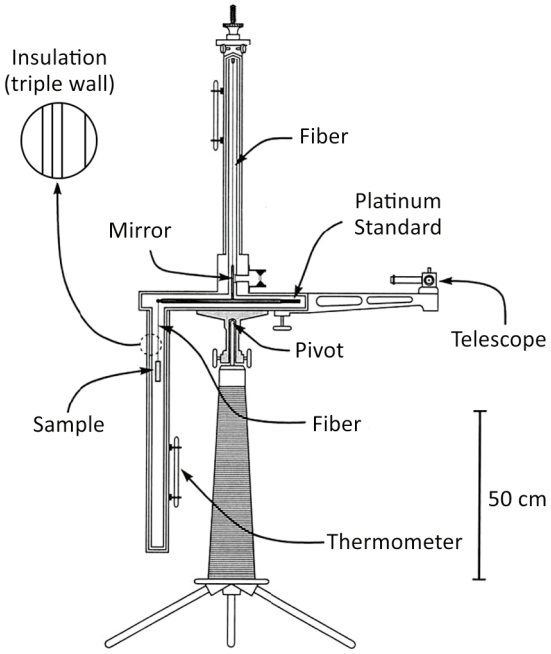


Fig. 3.9: Sketch of the original Eötvös balance. Torsion motion was observed via a mirror and a telescope. The original is stored in the museum of the Geophysical Observatory in Tihany. Drawing by Fischbach and Talmadge.

The gravitational experiment STEP (*Satellite Test of the Equivalence Principle*), prepared at Stanford University to verify the principle in orbit with a relative precision of 10^{-18} was canceled due to a lack of funding. In 2016, the French satellite Microscope was launched, utilizing specialized accelerometers. In 2022, the published results showed that the verification precision reached 1.5×10^{-15}. Since 2013, the AEgIS experiment at the European Organization for Nuclear Research (CERN) has been determining whether gravity acts identically on matter and antimatter. According to the experimental results, the equivalence principle holds for both matter and antimatter.

Local Inertial Frame

Falling Elevator

Consider riding in an elevator when the cable snaps. Although unpleasant, an individual dedicated to physics can enjoy an inertial frame where special relativity holds for a short duration. If a cigarette butt drops from the smoker's mouth in astonishment, the equivalence principle implies that the individual, cigarette, and elevator fall in the gravity field identically. The cigarette thus remains suspended. Placing keys in mid-air causes them to remain stationary, and if thrown, they move along a straight line at a constant velocity. The law of inertia therefore holds; if no force acts on a body, it remains at rest or moves uniformly in a straight line. The equivalence principle implies that weightlessness is experienced and gravitational effects vanish. A gravity field can be eliminated by an appropriately accelerated motion, and the converse is also true. However, the size of the elevator must not be excessive; if comparable to the Earth, the inhomogeneity of the gravity field would be perceived and deviations from the law of inertia discovered. The elevator must therefore be sufficiently small. Furthermore, experiments must not be conducted for too long; after an extended period, a slight convergence of the falling bodies would be observed, as everything falls toward the center of the Earth along non-parallel trajectories. The laws of special relativity hold in a freely gravitating cage of small dimensions where experiments are performed for a short duration. Such a cage represents an ideal inertial coordinate system, termed a *local inertial frame* (LIF).

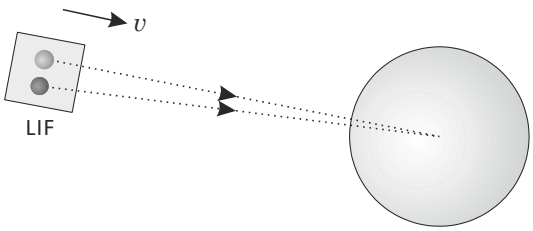


Fig. 3.10: Freely gravitating cage and local inertial frame.

Special and General Relativity

The equivalence principle carries another serious consequence. Consider the motion of bodies in the vicinity of the Sun. Their trajectories do not depend on the mass of the bodies, but are given solely by the initial position and velocity. They are therefore determined strictly by the central body (the Sun) and not by the orbiting bodies. This leads directly to the notion that the Sun has distorted spacetime around itself, and other bodies move along the straightest possible paths in this distorted world. This conceptualization of gravity was introduced by Albert Einstein in 1915, referred to as the geometric theory

of gravity or general relativity. This theory relates the existence of time and space to the presence of bodies. Every body, not only the Sun, distorts spacetime around itself, thereby contributing its small share to the existence of these entities. They are no longer external parameters of motion; without bodies, time and space would not exist. The fundamental task of general relativity is to describe how bodies distort spacetime and how they move within this world.

It can frequently be useful to resolve a given problem from the perspective of an LIF where special relativity holds, and then transform the found solution into a general coordinate system. Although not straightforward, this approach is usually very illustrative. A correct solution must determine the metric tensor from the distribution of matter and energy, and ascertain the motion of bodies from the knowledge of the metric tensor. General relativity is seen below to be historically the first theory that contains both the motion law and the field law within a single set of equations.

3.2.3 Covariant Derivative

When differentiating a four-vector, the basis vectors also alter from event to event in spacetime, which requires writing:

$$\frac{\partial \boldsymbol{\mathcal{A}}}{\partial x^{\beta}} = \partial_{\beta}\left(A^{\alpha}\mathbf{e}_{\alpha}\right) = A^{\alpha}{}_{,\beta}\,\mathbf{e}_{\alpha} + A^{\alpha}\mathbf{e}_{\alpha,\beta} \tag{3.110}$$

The derivative of a basis vector $\mathbf{e}_{\alpha,\beta}$ must also represent a vector, which can be expressed as a linear combination of the basis vectors:

$$\mathbf{e}_{\alpha,\beta} = \Gamma^{\delta}{}_{\alpha\beta}\,\mathbf{e}_{\delta} \tag{3.111}$$

The coefficients of this linear combination are termed *affine connection coefficients*. The expression for the derivative of a four-vector can be easily rearranged into the form:

$$\frac{\partial \boldsymbol{\mathcal{A}}}{\partial x^{\beta}} = \left(A^{\alpha}{}_{,\beta} + \Gamma^{\alpha}{}_{\xi\beta}A^{\xi}\right)\mathbf{e}_{\alpha}\,. \tag{3.112}$$

The expression in the brackets is termed the covariant derivative; the first part represents a standard partial derivative, while the second part is caused by the variation of basis vectors from place to place, i.e., by spacetime curvature. This derivative is denoted by a semicolon:

► $$A^{\alpha}{}_{;\beta} \equiv A^{\alpha}{}_{,\beta} + \Gamma^{\alpha}{}_{\xi\beta}A^{\xi}\,. \tag{3.113}$$

For a scalar function, this derivative coincides with a standard derivative, as a scalar does not require basis vectors for its expression and nothing alters:

► $$\phi_{;\beta} = \phi_{,\beta}\,. \tag{3.114}$$

Properties of Affine Connection Coefficients

The affine connection is not a tensorial quantity; its indices do not transform according to tensor rules. In practice, this implies that these indices cannot be raised or lowered. Conversely, the indices $A_{;\mu}$ generated by covariant differentiation possess a tensorial character and can be raised and lowered. This does not hold for a standard partial de-

rivative $A_{,\mu}$, where the component of the derivative given by spacetime curvature is absent. The prototype of a four-vector localized at a point in spacetime is an infinitesimal coordinate displacement:

$$\mathrm{d}\boldsymbol{\xi} = \mathrm{d}x^{\alpha}\mathbf{e}_{\alpha}. \tag{3.115}$$

The basis vectors and their derivatives can therefore be formally written as:

$$\mathbf{e}_{\alpha} = \frac{\partial \boldsymbol{\xi}}{\partial x^{\alpha}} = \boldsymbol{\xi}_{,\alpha}; \qquad \mathbf{e}_{\alpha,\beta} = \boldsymbol{\xi}_{,\alpha\beta}. \tag{3.116}$$

In a spacetime free of discontinuities and torsion, the second derivatives are interchangeable, i.e., $\mathbf{e}_{\alpha,\beta} = \mathbf{e}_{\beta,\alpha}$ holds, and the lower indices of the affine connection coefficients must therefore be symmetric:

$$\Gamma^{\alpha}{}_{\beta\gamma} = \Gamma^{\alpha}{}_{\gamma\beta}. \tag{3.117}$$

In such a case, these coefficients are termed *Christoffel symbols*.

Covariant Derivative of a Second-Order Tensor

The components of second-order tensors transform as products of two vectors; therefore, the covariant derivative of a prototype second-order tensor $T^{\alpha\beta} \equiv A^{\alpha}B^{\beta}$ is calculated:

$$\begin{aligned} T^{\alpha\beta}{}_{;\gamma} &= \left(A^{\alpha}B^{\beta}\right)_{;\gamma} = A^{\alpha}{}_{;\gamma}B^{\beta} + A^{\alpha}B^{\beta}{}_{;\gamma} = \\ &= \left(A^{\alpha}{}_{,\gamma} + \Gamma^{\alpha}{}_{\xi\gamma}A^{\xi}\right)B^{\beta} + A^{\alpha}\left(B^{\beta}{}_{,\gamma} + \Gamma^{\beta}{}_{\xi\gamma}B^{\xi}\right) = \\ &= \left(A^{\alpha}B^{\beta}\right)_{,\gamma} + \Gamma^{\alpha}{}_{\xi\gamma}A^{\xi}B^{\beta} + \Gamma^{\beta}{}_{\xi\gamma}A^{\alpha}B^{\xi} \end{aligned}$$

Thus, for a second-order tensor, each component must be differentiated separately

► $$T^{\alpha\beta}{}_{;\gamma} = T^{\alpha\beta}{}_{,\gamma} + \Gamma^{\alpha}{}_{\xi\gamma}T^{\xi\beta} + \Gamma^{\beta}{}_{\xi\gamma}T^{\alpha\xi}. \tag{3.118}$$

This rule is easily generalized for an n-th order tensor:

$$\begin{aligned} &T^{\alpha_1\alpha_2\ldots\alpha_N}{}_{;\gamma} = \\ &= T^{\alpha_1\alpha_2\ldots\alpha_N}{}_{,\gamma} + \Gamma^{\alpha_1}{}_{\xi\gamma}T^{\xi\alpha_2\ldots\alpha_N} + \Gamma^{\alpha_2}{}_{\xi\gamma}T^{\alpha_1\xi\ldots\alpha_N} + \cdots + \Gamma^{\alpha_N}{}_{\xi\gamma}T^{\alpha_1\alpha_2\ldots\xi}. \end{aligned} \tag{3.119}$$

Covariant Derivative of a Covariant Component

Given that indices cannot be lowered or raised for Christoffel symbols, the derivative of covariant components must be determined differently. The magnitude of a four-vector does not depend on the choice of coordinates; hence, its covariant derivative is zero:

$$\begin{gathered} \left(A^{\alpha}A_{\alpha}\right)_{;\beta} = 0 \qquad \Rightarrow \qquad A^{\alpha}{}_{;\beta}A_{\alpha} + A^{\alpha}A_{\alpha;\beta} = 0 \qquad \Rightarrow \\ \left(A^{\alpha}{}_{,\beta} + \Gamma^{\alpha}{}_{\xi\beta}A^{\xi}\right)A_{\alpha} + A^{\alpha}A_{\alpha;\beta} = 0 \qquad \Rightarrow \\ A^{\alpha}{}_{,\beta}A_{\alpha} + \Gamma^{\alpha}{}_{\xi\beta}A^{\xi}A_{\alpha} + A^{\alpha}A_{\alpha;\beta} = 0 \end{gathered}$$

In the first expression, the derivative is transferred to the second factor using the product rule, and in the second expression, the summation indices are swapped:

$$-A^{\alpha}A_{\alpha,\beta}+\Gamma^{\xi}{}_{\alpha\beta}A^{\alpha}A_{\xi}+A^{\alpha}A_{\alpha;\beta}=0 \qquad \Rightarrow$$

$$\left[-A_{\alpha,\beta}+\Gamma^{\xi}{}_{\alpha\beta}A_{\xi}+A_{\alpha;\beta}\right]A^{\alpha}=0\,.$$

Since this equality must hold for an arbitrary A_{α}, the square bracket must equal zero, which readily determines the derivative of the covariant vector component:

► $$A_{\alpha;\beta}=A_{\alpha,\beta}-\Gamma^{\xi}{}_{\alpha\beta}A_{\xi}\,. \tag{3.120}$$

A minus sign precedes the Christoffel symbol. For a second-order tensor, individual indices are again differentiated:

► $$T_{\alpha\beta;\gamma}=T_{\alpha\beta,\gamma}-\Gamma^{\xi}{}_{\alpha\gamma}T_{\xi\beta}-\Gamma^{\xi}{}_{\beta\gamma}T_{\alpha\xi}\,. \tag{3.121}$$

Covariant Derivative of the Metric

In this section, it is demonstrated that the metric with arbitrarily placed indices behaves as a constant with respect to the covariant derivative; the covariant derivative of the metric is always zero. It is sufficient to realize that the metric with contravariant indices represents the inverse matrix to the metric with covariant indices:

$$g^{\alpha\beta}g_{\beta\gamma}=\delta^{\alpha}{}_{\gamma}\,.$$

On the right is the Kronecker delta, representing the elements of the identity matrix, which remains identical in all bases and its derivative is zero:

$$\left(g^{\alpha\beta}g_{\beta\gamma}\right)_{;\delta}=0 \qquad \Rightarrow \qquad g^{\alpha\beta}{}_{;\delta}g_{\beta\gamma}+g^{\alpha\beta}g_{\beta\gamma;\delta}=0\,.$$

All indices possess a tensorial character, allowing the application of all rules for raising, lowering, and swapping indices:

$$g^{\alpha}{}_{\gamma;\delta}+g^{\alpha}{}_{\gamma;\delta}=0 \qquad \Rightarrow \qquad 2g^{\alpha}{}_{\gamma;\delta}=0 \qquad \Rightarrow$$

► $$g_{\alpha\beta;\gamma}=0\,. \tag{3.122}$$

Christoffel Symbols and the Metric

The metric completely determines the properties of curved spacetime. Christoffel symbols can therefore be expressed using the metric. Writing the covariant derivative in relation (3.122) according to relation (3.121) immediately yields:

$$g_{\alpha\beta,\gamma}=\Gamma^{\xi}{}_{\alpha\gamma}g_{\xi\beta}+\Gamma^{\xi}{}_{\beta\gamma}g_{\alpha\xi}\,. \tag{3.123}$$

Algebraic substitution then suffices to demonstrate that Christoffel symbols can be determined from a known metric via a simple relation:

$$\Gamma^{\gamma}{}_{\alpha\beta} = \frac{1}{2} g^{\gamma\xi} \left(g_{\xi\alpha,\beta} + g_{\xi\beta,\alpha} - g_{\alpha\beta,\xi} \right). \tag{3.124}$$

In physics, an abbreviated notation is frequently utilized:

$$\Gamma^{\gamma}_{\alpha\beta} \equiv \Gamma^{\gamma}{}_{\alpha\beta} \tag{3.125}$$

Properties of the Covariant Derivative

The covariant derivative accounts both for the variation of the vector components themselves and for the variation of the basis vectors. Curvature manifests in the covariant derivative via a term involving Christoffel symbols, which are determinable from the metric. For the product and sum of derivatives, the identical rules as for standard partial derivatives apply. Indices generated by covariant differentiation possess the character of tensorial indices and can be raised and lowered. For Christoffel symbols, the indices can be completed in an unambiguous manner. In the differentiation of upper (contravariant) indices, a plus sign precedes the Christoffel symbol, while in the differentiation of lower (covariant) indices, a minus sign appears. Covariant derivatives of the components of higher-order tensors are performed for each component separately. The covariant derivative of a scalar function coincides with the partial derivative. The covariant derivative of the metric is zero. These rules are summarized in a simple table:

$$\left(A^{\alpha} + B^{\alpha} \right)_{;\beta} = A^{\alpha}{}_{;\beta} + B^{\alpha}{}_{;\beta},$$

$$\left(A^{\alpha} B^{\beta} \right)_{;\gamma} = A^{\alpha}{}_{;\gamma} B^{\beta} + A^{\alpha} B^{\beta}{}_{;\gamma},$$

$$A^{\alpha}{}_{;\beta} \equiv A^{\alpha}{}_{,\beta} + \Gamma^{\alpha}{}_{\xi\beta} A^{\xi},$$

$$A_{\alpha;\beta} = A_{\alpha,\beta} - \Gamma^{\xi}{}_{\alpha\beta} A_{\xi},$$

$$T^{\alpha\beta}{}_{;\gamma} = T^{\alpha\beta}{}_{,\gamma} + \Gamma^{\alpha}{}_{\xi\gamma} T^{\xi\beta} + \Gamma^{\beta}{}_{\xi\gamma} T^{\alpha\xi},$$

$$T_{\alpha\beta;\gamma} = T_{\alpha\beta,\gamma} - \Gamma^{\xi}{}_{\alpha\gamma} T_{\xi\beta} - \Gamma^{\xi}{}_{\beta\gamma} T_{\alpha\xi}, \tag{3.126}$$

$$T^{\alpha}{}_{\beta;\gamma} = T^{\alpha}{}_{\beta,\gamma} + \Gamma^{\alpha}{}_{\xi\gamma} T^{\xi}{}_{\beta} - \Gamma^{\xi}{}_{\beta\gamma} T_{\alpha\xi}$$

$$\phi_{;\beta} = \phi_{,\beta},$$

$$g_{\alpha\beta;\gamma} = 0,$$

$$\Gamma^{\gamma}{}_{\alpha\beta} = \Gamma^{\gamma}{}_{\beta\alpha},$$

$$\Gamma^{\gamma}{}_{\alpha\beta} = \frac{1}{2} g^{\gamma\xi} \left(g_{\xi\alpha,\beta} + g_{\xi\beta,\alpha} - g_{\alpha\beta,\xi} \right).$$

3.2.4 Geodesic Equation

A particle moves along the straightest possible path in curved spacetime, termed a *geodesic*. The geodesic equation replaces the motion equations in general relativity.

Total Derivative of a Vector Field

Consider a vector field $A^{\mu}(x^{\alpha})$, present in spacetime, varying with both time and space. A curve stretched between events A and B, for instance the worldline of a particle, is also given. This curve is parametrized by a parameter τ. In the case of a particle worldline, this represents the proper time of the particle.

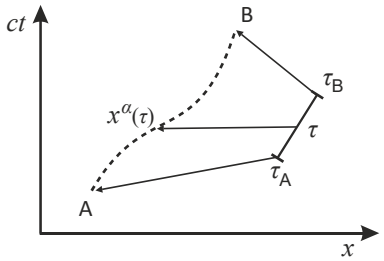


Fig. 3.11: Parametrization of a spacetime curve between events A and B using the parameter τ.

The change of the vector field along the curve is analyzed:

$$\frac{\mathrm{d}A^{\mu}}{\mathrm{d}\tau} = \frac{\partial A^{\mu}}{\partial x^{\nu}} \frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau} = A^{\mu}{}_{,\nu} \frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau} . \tag{3.127}$$

This represents the correct change in flat spacetime. In curved spacetime, the components of the vector field must be differentiated covariantly, accounting for the changes of the basis vectors so that the derivatives form a second-order tensor. Therefore, a total derivative is introduced via the definition

► $$\frac{\mathrm{D}A^{\mu}}{\mathrm{D}\tau} \equiv A^{\mu}{}_{;\nu} \frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau} . \tag{3.128}$$

The quantity $A^{\mu};\nu$ now forms the components of a second-order tensor, and the curvature is included in the derivative. The relation between the standard and total derivatives is determined by expanding the covariant derivative according to relation (3.113):

$$\frac{\mathrm{D}A^{\mu}}{\mathrm{D}\tau} \equiv \left[A^{\mu}{}_{,\nu} + \Gamma^{\mu}{}_{\xi\nu} A^{\xi} \right] \frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau} = A^{\mu}{}_{,\nu} \frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau} + \Gamma^{\mu}{}_{\xi\nu} A^{\xi} \frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau} . \tag{3.129}$$

The first term is the standard derivative (3.127), which yields:

► $$\frac{\mathrm{D}A^{\mu}}{\mathrm{D}\tau} = \frac{\mathrm{d}A^{\mu}}{\mathrm{d}\tau} + \Gamma^{\mu}{}_{\xi\nu} A^{\xi} \frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau} . \tag{3.130}$$

The interpretation is apparent: the first term represents the standard derivative, while the second term is caused by spacetime curvature.

Parallel Transport of a Vector Along a Curve

In flat spacetime, *parallel transport* is defined as the sequential shifting of a vector along a curve. If a vector is transported along a closed curve, it returns to the initial point unaltered after traversing the curve. In curved space, the process is more complex. Consider, for example, the parallel transport of a vector along the surface of a sphere where poles and the equator are designated. The curve first proceeds from the pole along a meridian to the equator, then continues along the equator, and finally returns to the pole along a different meridian, thus forming a closed curve. At the pole, the process begins with a vector tangent to the curve, pointing toward the equator. After each parallel displacement along the meridian, the vector must be projected into the tangent plane at that point, as it must transform as an infinitesimal vector. In this manner, the vector is brought to the equator, transported along it, and finally brought back to the pole along another meridian, where it must again be projected into the tangent plane. The result is remarkably interesting: the vector transported along the closed curve altered its direction during the journey.

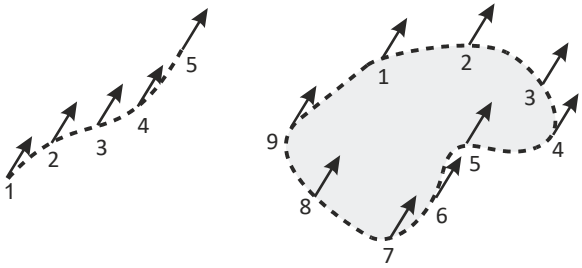


Fig. 3.12: Parallel transport of a vector in a plane along an open and a closed curve. The curve lies in the plane, as do the transported vectors.

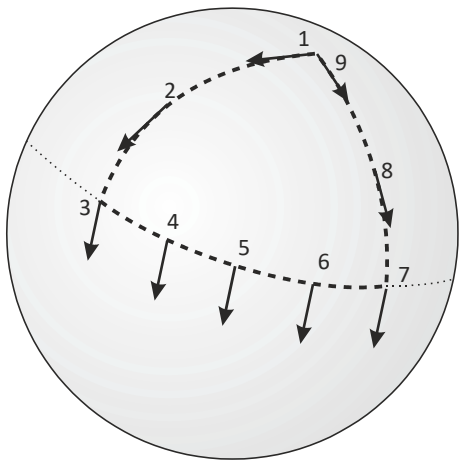


Fig. 3.13: Parallel transport of a vector along a closed curve on the surface of a sphere.

In flat spacetime, parallel transport is described by a simple equation, as the vector does not change during transport along a curve:

$$\frac{\mathrm{d}A^{\mu}}{\mathrm{d}\tau} = 0\,. \tag{3.131}$$

In curved spacetime, the equation remains equally simple, provided the changing curvature is accounted for:

$$\frac{\mathrm{D}A^{\mu}}{\mathrm{D}\tau} = 0\,. \tag{3.132}$$

Expanding this equation using the covariant derivative and Christoffel symbols via (3.129) yields:

► $$\frac{\mathrm{d}A^{\mu}}{\mathrm{d}\tau} + \Gamma^{\mu}{}_{\xi\nu} A^{\xi} \frac{\mathrm{d}x^{\nu}}{\mathrm{d}\tau} = 0 \tag{3.133}$$

This represents the required equation for the parallel transport of a vector field A^{μ} along a curve $x^{\nu}(\tau)$. If spacetime is flat and the Christoffel symbols vanish, this equation coincides with relation (3.131) for flat spacetime.

Geodesic Equation

When observing any phenomenon, a guideline can always be provided by how the process transpires in a local inertial frame (LIF), i.e., in a frame that freely falls in the gravitational fields of other bodies for a short duration within a small region of space. In such a frame, the laws of special relativity hold, a free point mass moves uniformly in a straight line, and the metric tensor is given by the Minkowski metric. If the coordinates in the LIF are denoted by ξ^{α}, the interval equals:

$$\mathrm{d}s^2 = \eta_{\alpha\beta}\, \mathrm{d}\xi^{\alpha} \mathrm{d}\xi^{\beta}; \qquad \eta_{\alpha\beta} \equiv \mathrm{diag}(-1,+1,+1,+1)\,. \tag{3.134}$$

In a general coordinate system S, the following holds:

$$\mathrm{d}s^2 = g_{\alpha\beta}\, \mathrm{d}x^{\alpha} \mathrm{d}x^{\beta}\,, \tag{3.135}$$

where the metric describes the spacetime curvature, and the coordinates in S are x^{α}.



Fig. 3.14: Worldline in an LIF.

The particle worldline is first constructed in an LIF. The following figure displays the worldlines of a stationary particle (a), a moving particle (b), and a photon (c). The corresponding four-velocity always points in the direction of all worldlines:

$$U^{\mu} \equiv \frac{\mathrm{d}\xi^{\mu}}{\mathrm{d}\tau} = \frac{\mathrm{d}\xi^{\mu}}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}\tau} = \begin{pmatrix} \gamma c \\ \gamma \mathbf{v} \end{pmatrix} \tag{3.136}$$

The four-vectors of a stationary particle, a particle moving at a velocity v, and a photon can be sketched in two dimensions (time and one spatial coordinate); their 2D four-velocities are sequentially:

$$\begin{pmatrix} \gamma c \\ 0 \end{pmatrix}; \quad \begin{pmatrix} \gamma c \\ \gamma v \end{pmatrix}; \quad \begin{pmatrix} \gamma c \\ \gamma c \end{pmatrix} \tag{3.137}$$

For velocities low compared to the speed of light, the time component of the four-velocity is substantially larger than the spatial components (in 2D, a single spatial component is present).

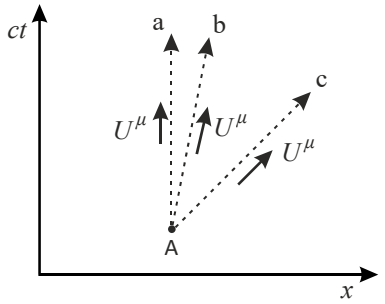


Fig. 3.15: Direction of four-velocity and worldline.

The figure implies that in special relativity (in an LIF), worldlines can be constructed by sequentially transporting the four-velocity. It is sufficient to know the initial event (A) and shift the four-velocity by a small segment along the worldline, yielding:

$$\frac{\mathrm{d}U^\mu}{\mathrm{d}\tau} = 0\,. \tag{3.138}$$

In an LIF, free point masses move linearly since no force acts upon them; thus, a simple relation holds:

$$\frac{\mathrm{d}^2\xi^\mu}{\mathrm{d}\tau^2} = 0 \tag{3.139}$$

Given that $U^\mu = \mathrm{d}\xi^\mu/\mathrm{d}\tau$, the relation $\mathrm{d}U^\mu/\mathrm{d}\tau = 0$ is immediately obtained, meaning that equation (3.138) describes the motion of a free body in an LIF. In the coordinate system S, where a general metric $g_{\mu\nu}$ is present, the equation for the parallel transport of the four-velocity takes the form:

$$\frac{\mathrm{D}U^\mu}{\mathrm{D}\tau} = 0\,. \tag{3.140}$$

Expanding this equation via the definition of the total derivative (3.130) yields:

$$\frac{\mathrm{d}U^\mu}{\mathrm{d}\tau} + \Gamma^\mu{}_{\xi\nu} U^\xi \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = 0\,. \tag{3.141}$$

Utilizing the definition of the four-velocity yields:

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu{}_{\xi\nu} \frac{\mathrm{d}x^\xi}{\mathrm{d}\tau} \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = 0\,. \tag{3.142}$$

Appropriate index relabeling provides the required geodesic equation in the form:

► $$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu{}_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau} \frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = 0 . \tag{3.143}$$

For light, the parameter of the worldline cannot be the proper time, but must be, for instance, the proper length of the curve, yielding:

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu{}_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\beta}{\mathrm{d}\lambda} = 0 . \tag{3.144}$$

The geodesic equation can be written without parametrization as a differential form:

► $$\mathrm{d}^2 x^\mu + \Gamma^\mu{}_{\alpha\beta}\, \mathrm{d}x^\alpha \mathrm{d}x^\beta = 0 . \tag{3.145}$$

Note: The geodesic equation can also be derived from a variational principle. From theoretical mechanics, $L\mathrm{d}t$ must represent a scalar, and thus a function of $\mathrm{d}s^2$ in relativity. For the expression to be proportional to time, a square root must be involved, yielding:

$$L\mathrm{d}t \approx \sqrt{-\mathrm{d}s^2} \approx \mathrm{d}\tau .$$

The interval is negative for causally linked events; hence, a minus sign appears inside the square root. The geodesic equation can then be derived from the Hamilton variation principle:

► $$\delta \int_A^B \sqrt{-\mathrm{d}s^2} = 0 \tag{3.146}$$

This derivation can be found in specialized literature on general relativity.

3.2.5 Newtonian Limit

How does the geodesic equation, representing a fundamental pillar of general relativity, relate to the Newton motion equations? It is demonstrated below that in the limit of weak fields (small curvature) and low velocities, the geodesic equation transforms into the standard motion equations in a gravitational field. The Newton motion equation in a gravitational field takes the form:

$$m\ddot{\mathbf{x}} = -\nabla W_\mathrm{p} , \tag{3.147}$$

where W_p is the potential energy, depending linearly on the mass \(m\) of the test particle. The gravitational field potential can also be introduced:

$$\phi \equiv W_\mathrm{p}/m , \tag{3.148}$$

which is independent of the body mass. The equation of motion simplifies to:

$$\ddot{\mathbf{x}} = -\nabla\phi , \tag{3.149}$$

reflecting the fact that the motion of bodies in a gravitational field does not depend on their mass (a consequence of the equivalence principle). In relativity, it is advantageous to introduce a dimensionless potential ϕ^*:

$$\begin{aligned} &W_{\mathrm{p}}; && \left[W_{\mathrm{p}}\right]=\mathrm{J}, \\ &\phi \equiv W_{\mathrm{p}}/m; && [\phi]=\mathrm{J/kg}=(\mathrm{m/s})^2, \\ &\phi^* \equiv \phi/c^2; && \left[\phi^*\right]=1. \end{aligned} \tag{3.150}$$

The geodesic equation is now written under the following assumptions:

1) the metric is independent of time and is diagonal;
2) the body moves at a low velocity compared to the speed of light;
3) the gravitational field is weak, i.e., the spacetime curvature is small.

Ad 1. This assumption is not essential but simplifies calculations. A suitable choice of coordinate mesh can ensure that the metric is diagonal. If the coordinate system is attached to the body generating the field, the metric is stationary, i.e.:

$$g_{\alpha\beta,0}=0. \tag{3.151}$$

Ad 2. A low velocity of the body relative to the speed of light implies that the coordinate time and proper time need not be distinguished, and the time component of the four-velocity is significantly larger than the spatial components, i.e.:

$$\frac{\mathrm{d}t}{\mathrm{d}\tau}=\gamma=1/\sqrt{1-v^2/c^2}\approx 1, \tag{3.152}$$

$$\left|\frac{\mathrm{d}x^k}{\mathrm{d}\tau}\right| \ll \frac{\mathrm{d}x^0}{\mathrm{d}\tau}, \qquad k=1,2,3. \tag{3.153}$$

Ad 3. A weak gravitational field implies that the metric deviates only slightly from the Minkowski metric, i.e.:

$$g_{\alpha\beta}=\eta_{\alpha\beta}+h_{\alpha\beta}, \tag{3.154}$$

where $h_{\alpha\beta}$ represent small deviations from the Minkowski metric. Small in the sense that their squares are neglected in all expressions. The metric thus possesses the final form:

$$g_{\alpha\beta}=\begin{pmatrix} -1+h_{00} & 0 & 0 & 0 \\ 0 & 1+h_{11} & 0 & 0 \\ 0 & 0 & 1+h_{22} & 0 \\ 0 & 0 & 0 & 1+h_{33} \end{pmatrix}, \tag{3.155}$$

The inverse matrix is:

$$g^{\alpha\beta}=\begin{pmatrix} (-1+h_{00})^{-1} & 0 & 0 & 0 \\ 0 & (1+h_{11})^{-1} & 0 & 0 \\ 0 & 0 & (1+h_{22})^{-1} & 0 \\ 0 & 0 & 0 & (1+h_{33})^{-1} \end{pmatrix}. \tag{3.156}$$

Given that the perturbations h are small, a first-order Taylor expansion is performed, utilizing $(1+x)^n \approx 1+nx$. The final form of the contravariant metric components is:

$$g^{\alpha\beta} = \begin{pmatrix} -1-h_{00} & 0 & 0 & 0 \\ 0 & 1-h_{11} & 0 & 0 \\ 0 & 0 & 1-h_{22} & 0 \\ 0 & 0 & 0 & 1-h_{33} \end{pmatrix}. \tag{3.157}$$

Multiplying the matrices $g_{\alpha\beta}$ and $g^{\alpha\beta}$ together yields the identity matrix, provided the squares of the perturbations h are neglected.

* * *

The discussion now returns to the geodesic equation. Due to assumption (3.153), only the time terms remain in the second term, i.e.:

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu{}_{00} \frac{\mathrm{d}x^0}{\mathrm{d}\tau} \frac{\mathrm{d}x^0}{\mathrm{d}\tau} = 0\,. \tag{3.158}$$

Expressing the derivatives of the time components of the event from (3.152) yields:

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + c^2 \Gamma^\mu{}_{00} = 0\,. \tag{3.159}$$

To expand the geodesic equation, the corresponding Christoffel symbols must be determined from relation (3.124):

$$\Gamma^\gamma{}_{00} = \frac{1}{2} g^{\gamma\xi} \left(g_{\xi 0,0} + g_{\xi 0,0} - g_{00,\xi} \right).$$

According to assumption (3.151), the time derivatives are zero, leaving only:

$$\Gamma^\gamma{}_{00} = -\frac{1}{2} g^{\gamma\xi} g_{00,\xi}\,.$$

The four components of the Christoffel symbols are therefore of interest. In the calculation, the vanishing of non-diagonal terms and time derivatives is utilized:

$$\Gamma^0{}_{00} = -\frac{1}{2} g^{0\xi} g_{00,\xi} = -\frac{1}{2} g^{00} g_{00,0} - \frac{1}{2} g^{01} g_{00,1} - \cdots = 0\,,$$

$$\Gamma^1{}_{00} = -\frac{1}{2} g^{1\xi} g_{00,\xi} = -\frac{1}{2} g^{11} g_{00,1} = -\frac{1}{2}(1-h_{11})(-1+h_{00})_{,x} \approx -\frac{\partial}{\partial x}\left(\frac{h_{00}}{2}\right),$$

$$\Gamma^2{}_{00} = \cdots - \frac{\partial}{\partial y}\left(\frac{h_{00}}{2}\right),$$

$$\Gamma^3{}_{00} = \cdots - \frac{\partial}{\partial z}\left(\frac{h_{00}}{2}\right).$$

In the geodesic equation (3.158), the proper time can be replaced by the coordinate time, as they do not differ in the approximation of the Newtonian limit. The time component is satisfied trivially, and the spatial components yield:

$$\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}t^2} = \nabla\left(\frac{c^2 h_{00}}{2}\right). \tag{3.160}$$

A form identical to the Newton motion equations (3.149) is obtained. The perturbations h are dimensionless; the square of the speed of light provides the right-hand side of equations (3.160) with the correct dimension. The geodesic equation coincides with the Newton equations assuming that:

► $$h_{00} = -\frac{2\phi}{c^2} = -2\phi^* . \tag{3.161}$$

The metric is clearly related to the field potential. Metric derivatives, such as Christoffel symbols, relate to the field intensity (the spatial derivative of the potential). In the Newtonian limit, the interval takes the form:

$$\mathrm{d}s^2 \approx g_{00}\mathrm{d}(ct)^2 + g_{11}\mathrm{d}x_1^2 + g_{22}x_2^2 + g_{33}\mathrm{d}x_3^2 = -\left(1+2\phi^*\right)\mathrm{d}t^2 + \cdots \tag{3.162}$$

Nothing is known about the spatial part of the metric yet, but the time part is clearly deformed by the presence of the massive body distorting the local spacetime (expressed in Newtonian physics as the presence of a gravitational potential). For a spherically symmetric field source:

$$\phi = -\frac{GM}{r} . \tag{3.163}$$

The metric is:

$$\mathrm{d}s^2 \approx -\left(1-\frac{2GM}{c^2 r}\right)\mathrm{d}t^2 + \cdots \tag{3.164}$$

At large distances from the body ($r \to \infty$), the metric transitions into the Minkowski metric. The greatest curvature occurs at the surface of the body, and the deviation from the Minkowski metric can be expressed according to (3.162) using the dimensionless potential ϕ^* at the surface of the body

$$\phi^*(R) = -\frac{GM}{c^2 R} . \tag{3.165}$$

The values are $\phi^* \sim 10^{-39}$ for a proton, 10^{-9} for the Earth, 10^{-6} for the Sun, 10^{-4} for a white dwarf, 10^{-1} for a neutron star, and 1 for a black hole. The Newtonian limit cannot be applied to a neutron star or a black hole, and the approximation $g_{\alpha\beta} \approx \eta_{\alpha\beta} + h_{\alpha\beta}$ is incorrect.

3.3 Time Dilation and Gravitational Redshift

Consider a coordinate system rigidly attached to the source of a gravitational field (for instance, the Earth). In this system, the rate of stationary clocks (located at point A, e.g., on a shelf) is observed. The metric in this coordinate system does not depend on time, and the coordinates can be chosen such that the metric is diagonal (which is not necessary, though it simplifies the calculations).

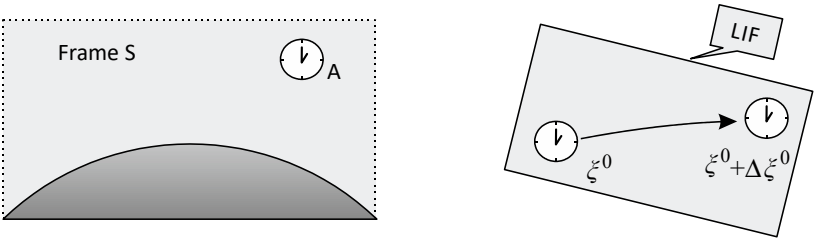


Fig. 3.16: Time dilation – single clock.

For the interval between two ticks of the clock in the coordinate system S, where the clock is stationary ($\Delta x^k = 0$), the following holds:

$$\Delta s^2 = g_{\alpha\beta}\,\Delta x^\alpha \Delta x^\beta = g_{00}c^2\Delta t^2 + g_{kl}\Delta x^k \Delta x^l = g_{00}c^2\Delta t^2 . \tag{3.166}$$

This clock can be observed in an LIF, where it moves and the following applies:

$$\Delta s^2 = \eta_{\alpha\beta}\,\Delta\xi^\alpha \Delta\xi^\beta = -\left(\Delta\xi^0\right)^2 + \left(\Delta\xi^1\right)^2 + \left(\Delta\xi^2\right)^2 + \left(\Delta\xi^3\right)^2 . \tag{3.167}$$

Three distinct times must be distinguished: 1) the coordinate time t, running in S attached to the clock and the Earth or another field source – such a time is chosen arbitrarily; 2) the time ξ^0 running in the LIF, i.e., the time in the Minkowski metric; 3) the proper (natural) time, for which the following holds:

$$\Delta s^2 = -c^2\Delta\tau^2 . \tag{3.168}$$

Relations (3.166) and (3.168) yield:

$$-c^2\Delta\tau^2 = g_{00}c^2\Delta t^2 \qquad \Rightarrow$$

$$\Delta\tau = \sqrt{-g_{00}(A)}\;\Delta t . \tag{3.169}$$

The proper time is the sole "natural" time at the clock; the reciprocal of the duration $\Delta\tau$ between two ticks of the clock represents the proper frequency of the clock, i.e.:

$$\Delta\tau = \frac{1}{\nu_0} = \sqrt{-g_{00}(A)}\;\Delta t . \tag{3.170}$$

Two Clocks

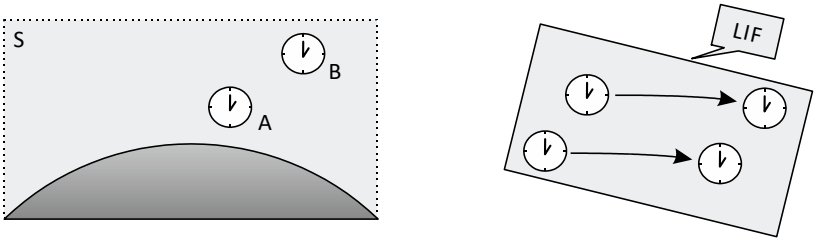


Fig. 3.17: Time dilation – two clocks.

Consider two stationary clocks in the system S whose rates are compared. An electromagnetic signal is emitted from clock A to point B. The first photon coincides with a tick of clock A, and the second with the subsequent tick.



Fig. 3.18: Two stationary clocks.

The propagation of the photon is governed by a simple equation:

$$\mathrm{d}s^2 = g_{\alpha\beta}\,\mathrm{d}x^{\alpha}\mathrm{d}x^{\beta} = 0\,. \tag{3.171}$$

This quadratic equation for dt yields a time-independent result, as the metric coefficients in S are not functions of time. Consequently, a photon travels from point A to point B in the identical duration at any time. Thus, the following must hold:

$$\Delta t_A = \Delta t_B = \Delta t\,. \tag{3.172}$$

For the observed tick of clock A, monitored at point B, relation (3.170) yields:

$$\frac{1}{\nu_B} = \sqrt{-g_{00}(B)}\;\Delta t_B\,. \tag{3.173}$$

If a signal is emitted from clock B to point A, the following is observed there:

$$\frac{1}{\nu_A} = \sqrt{-g_{00}(A)}\;\Delta t_A\,. \tag{3.174}$$

Given that $\Delta t_A = \Delta t_B = \Delta t$, dividing the last two relations yields:

► $$\frac{\nu_B}{\nu_A} = \frac{\sqrt{-g_{00}(A)}}{\sqrt{-g_{00}(B)}}\,. \tag{3.175}$$

Relation (3.175) represents a completely general expression for the change in the rate of time caused by the presence of a gravitational field (spacetime curvature). This relation can also be interpreted as the frequency shift of a photon propagating from point A to point B (the so-called gravitational redshift). In the weak-field limit (small spacetime curvature), relation (3.162) allows writing:

$$\frac{\nu_B}{\nu_A} = \frac{\sqrt{1+2\phi^*(A)}}{\sqrt{1+2\phi^*(B)}} = \left(1+2\phi^*(A)\right)^{1/2}\left(1+2\phi^*(B)\right)^{-1/2} \approx$$

$$\approx 1-\left(\phi^*(B)-\phi^*(A)\right) = 1-\Delta\phi^*$$

Therefore, for the frequency shift of a photon propagating between locations with different gravitational potentials, a simple relation holds in the Newtonian limit:

$$\frac{\nu_B}{\nu_A} = 1-\frac{\Delta\phi}{c^2}. \tag{3.176}$$

Multiplying by ν_A and rearranging the terms yields the relative frequency change:

$$\frac{\Delta\nu}{\nu} = -\frac{\Delta\phi}{c^2}. \tag{3.177}$$

The relative frequency change is thus related to the change in the gravitational potential.

3.3.1 Different Views on Time Dilation

Explanation from the LIF Perspective

Consider a photon propagating in a gravity field from point A to a higher point B. The points are not excessively distant, so they can be enclosed within a single elevator that begins falling at the moment the photon is emitted from point A. The elevator represents a local inertial frame (LIF) in which the photon moves along a straight line at speed c.

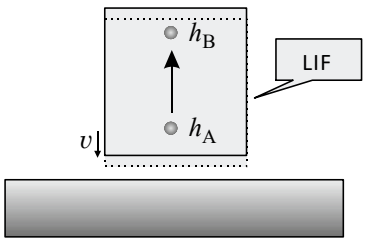


Fig. 3.19: Explanation from the LIF perspective.

At the moment the photon reaches B, the elevator is already moving at a velocity:

$$v = g\,\Delta t = g\,\Delta h/c. \tag{3.178}$$

An observer at height h_B observes the photon arriving from the moving elevator and thus perceives a frequency due to the Doppler effect:

$$\nu_B = \left(1 - \frac{v}{c}\right)\nu_A = \left(1 - \frac{g\Delta h}{c^2}\right)\nu_A . \tag{3.179}$$

Considering that the potential of the gravity field is $\phi = gh$, the following holds:

$$\frac{\nu_B}{\nu_A} = 1 - \frac{\Delta\phi}{c^2}, \tag{3.180}$$

which is the identical relation derived from the metric in the Newtonian limit.

Explanation from the Energy Conservation Perspective

In a gravity field, the sum of potential energy and photon energy is constant, i.e.:

$$m_A g h_A + \hbar\omega_A = m_B g h_B + \hbar\omega_B . \tag{3.181}$$

The relativistic mass of the photon is determined from its energy ($m = E/c^2 = \hbar\omega/c^2$):

$$\frac{\hbar\omega_A}{c^2} g h_A + \hbar\omega_A = \frac{\hbar\omega_B}{c^2} g h_B + \hbar\omega_B \quad \Rightarrow$$

$$\nu_A\left(1 + \frac{g h_A}{c^2}\right) = \nu_B\left(1 + \frac{g h_B}{c^2}\right) \quad \Rightarrow \tag{3.182}$$

$$\frac{\nu_B}{\nu_A} = \left(1 + \frac{g h_A}{c^2}\right)\cdot\left(1 + \frac{g h_B}{c^2}\right)^{-1} \cong 1 - \frac{g(h_B - h_1)}{c^2} = 1 - \frac{\Delta\phi}{c^2} .$$

The familiar relation following from the Newtonian limit is obtained once again.

◖ **Example 3.2:** Determine the wavelength change for a photon with $\lambda_0 = 500$ nm emitted from the surface of a star and traveling to a large distance therefrom (to infinity):

$$\frac{\Delta\lambda}{\lambda} \approx -\frac{\Delta\nu}{\nu} = \frac{\Delta\phi}{c^2} .$$

In this example, a local gravity field cannot be used. Since the photon reaches a large distance, the Newton gravitational potential must be utilized:

$$\lambda = \lambda_0(1 + \Delta\phi/c^2) = \lambda_0\left[1 - G\frac{M}{c^2}\left(\frac{1}{\infty} - \frac{1}{R}\right)\right] = \lambda_0\left(1 + \frac{GM}{Rc^2}\right). \tag{3.183}$$

	M	R	$\Delta\lambda/\lambda = GM/Rc^2$	λ (nm)
Sun	1 M_S	700,000 km	2×10^{-6}	500.001
White dwarf	1 M_S	10,000 km	1.5×10^{-4}	500.075
Neutron star	1 M_S	20 km	0.074	537

◗

3.3.2 Pound-Rebka Experiment

The first measurement of gravitational redshift was performed in 1960 by Robert Pound and Glen Rebka at Harvard University. For the measurement, they utilized a tower that remains part of the Jefferson Laboratory today. In their original paper, Pound and Rebka state that the distance between the emitter and receiver (detector) was 74 feet, corresponding to a height of 22.6 meters. Over such a small height difference, general relativity predicts a relative frequency change of:

$$\frac{\Delta\nu}{\nu} = -\frac{\Delta\phi}{c^2} = -\frac{g(h_2 - h_1)}{c^2} \approx 2{,}5\times10^{-15} \tag{3.184}$$



Fig. 3.20: Pound-Rebka experiment.

Measuring such a minute frequency change required extraordinary skill combined with considerable experience. Given that the measured frequency difference was:

$$\Delta\nu \approx 2{,}5\times10^{-15}\nu_0\,, \tag{3.185}$$

It was necessary to find a source with the highest possible frequency. Ultimately, radioactive cobalt ^{57}Co embedded in iron ^{57}Fe was utilized. The iron ^{57}Fe emitted gamma photons with a precisely defined energy of 14.4 keV (frequency 3.5×10^{18} Hz). An absorber consisting again of a ^{57}Fe layer served as the detector, resonantly absorbing photons of the identical frequency. Whether photons were absorbed in the detector or passed through was determined using a NaI(Tl) scintillation crystal and a photomultiplier. The crystal had a diameter of 7.5 cm and a thickness of 6 mm. The source and detector were thus tuned to the identical frequency, i.e., the detector was capable of absorbing only photons with a frequency precisely equal to the emitted frequency. In normal atoms, recoil during photon absorption in the detector would affect the received frequency; however, in crystals, due to the Mössbauer effect, the entire crystal absorbs the recoil, leaving the frequency of the absorbed photons unaltered. The sole frequency

change occurred via gravitational shift (red if the source was at the bottom and the detector at the top, and blue in the inverted configuration). Consequently, the detector should not absorb photons with an altered frequency.

At this point, the Doppler effect enters the scenario. The photon source was attached to a speaker membrane, oscillating it vertically back and forth at a frequency ranging from 10 to 50 Hz. Through the Doppler effect, the frequency of the emitted photons varied periodically. At a specific phase, the resulting shift compensated for the gravitational shift, and the detector absorbed photons with an unaltered frequency (or rather, altered twice – once by the gravitational shift and back by the Doppler shift). The entire method represents a modified Mössbauer spectroscopy, enabling precise determination of the frequency shift. To prevent undesirable photon scattering in the atmosphere, the photons traveled between the source and detector through a Mylar tube (40 cm in diameter) filled with helium. The photograph shows Robert Pound at the lower end of the tube. The experimental result was positive; Pound and Rebka confirmed gravitational red and blue shifts with a relative precision of 0.1, i.e., 10 %. Subsequent modifications of the experiment achieved a verification precision of general relativity of 0.01. This represented one of the "great" tests of general relativity, detecting the change in the rate of time caused by the presence of the Earth.

3.3.3 Hafele-Keating Experiment

Another notable experiment investigating the change in the rate of time caused by the gravity of the Earth was designed by Joseph Hafele and Richard Keating in 1971. Cesium clocks were utilized to measure time. Control clocks were placed at the USNO (*United States Naval Observatory*). With additional clocks, they circumnavigated the Earth in an eastward direction, and with a final set in a westward direction. Commercial airliners were used for the circumnavigation, transferring the clocks from plane to plane.

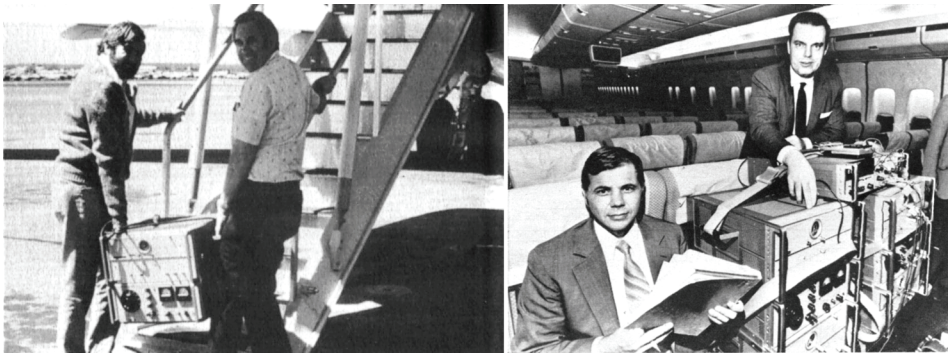


Fig. 3.21: Hafele and Keating transporting cesium clocks on a trip.

The time on the clocks moving at an altitude of ten kilometers was then compared with the time on the control clocks. The resulting value was governed by both special relativity effects (time dilation) and general relativity effects. The airplane represents a moving system, so the following holds:

$$\Delta s^2 = g_{\alpha\beta}\,\Delta x^{\alpha}\Delta x^{\beta} = g_{00}c^2\Delta t^2 + g_{kl}\Delta x^k \Delta x^l = \left(g_{00}c^2 + g_{kl}v^k v^l\right)\Delta t^2 \qquad (3.186)$$

Expressing the interval on the left via proper time and the component g_{00} via the Newtonian limit yields:

$$-c^2\Delta\tau^2 = \left(-(1+2\phi^*)c^2 + v^2\right)\Delta t^2 \tag{3.187}$$

and therefore:

$$\frac{1}{\nu} = \Delta\tau = \sqrt{1+\frac{2\phi}{c^2}-\frac{v^2}{c^2}}\,\Delta t \tag{3.188}$$

The change in the rate of time is thus caused both by time curvature due to gravity (the ϕ term) and by special-relativistic time dilation caused by the motion of the airplane (the v term). Comparing the rate of clocks on the airplane and on the Earth surface (moving at velocity v_Z yields:

$$\frac{\Delta\tau_A}{\Delta\tau_E} = \frac{\sqrt{1+\frac{2\phi_A}{c^2}-\frac{v_A^2}{c^2}}}{\sqrt{1+\frac{2\phi_E}{c^2}-\frac{v_E^2}{c^2}}} \approx 1+\frac{\phi_L-\phi_E}{c^2}-\frac{v_A^2-v_E^2}{2c^2}. \tag{3.189}$$

Substituting the gravitational potential of the Earth provides a simple relation:

$$\frac{\Delta\tau_A}{\Delta\tau_E} = 1-\frac{GM_E}{(R_E+h)c^2}+\frac{GM_E}{R_E c^2}-\frac{v_A^2-v_E^2}{2c^2}. \tag{3.190}$$

or equivalently:

$$\Delta\tau_A-\Delta\tau_E = \left(-\frac{GM_E}{(R_E+h)c^2}+\frac{GM_E}{R_E c^2}-\frac{v_A^2-v_E^2}{2c^2}\right)\Delta\tau_E. \tag{3.191}$$

The airplanes flew at an average altitude of 8,900 m, and the circumnavigation lasted approximately three days. Calculations based on the actual flight path and specific airplane velocities yielded a time difference between the clocks on the airplane and on the Earth of –40±23 ns for the eastward flight and 275 ± 21 ns for the westward flight. The measured values were –59±10 ns for the eastward direction and +273±10 ns for the westward direction. The experiment confirmed the predictions of general relativity with a relative precision of about 10 % (10^{-1}). In 1976, the experiment was replicated by the University of Maryland, confirming general relativity with a precision of 1 % (10^{-2}).

Note: In the preceding calculations, the coordinate system was assumed to be rigidly attached to the field source. For a rotating Earth, this does not hold; however, effects caused by the Earth rotation are of a higher order than those utilized in the Newtonian limit.

◖ Example 3.3: Global Positioning System (GPS)

Determine the maximum discrepancy between the automobile position obtained via GPS and its actual position over 24 hours, assuming no correction is applied for the differing rates of time on the satellite and in the automobile. Assume the automobile drives along the equator. The GPS satellite orbits at an altitude h = 20,200 km. The

radius of the Earth is $R_E = 6{,}371$ km, the mass of the Earth is $M_E = 5.97\times10^{24}$ kg, and $G = 6.67\times10^{-11}$ Nm^{2}kg^{-2}.

Solution: The proper velocity of the automobile is negligible relative to the rotation of the Earth. First, the orbital velocity of the Earth surface at the equator is determined:

$$v_E = R_E\omega = R_E\ 2\pi/T \approx 463 \text{ m/s}\,. \tag{3.192}$$

Next, the orbital velocity of the satellite is determined from the equality of the inertial centrifugal force and the gravitational force:

$$\frac{mv_S^2}{R+h} = G\frac{mM_E}{(R+h)^2} \qquad \Rightarrow \qquad v_S = \sqrt{\frac{GM_E}{R+h}} \approx 3{,}871 \text{ m/s}\,. \tag{3.193}$$

Relation (3.191) is now applied, substituting the satellite velocity v_S for the airplane velocity v_A. Over one day, i.e., for $\Delta\tau_E = 24$ h $= 86{,}400$ s, the difference between the two times is:

$$\Delta\tau_S - \Delta\tau_E = 38\ \mu\text{s}\,. \tag{3.194}$$

Multiplying this value by the speed of light reveals the positioning error:

$$\Delta l = \left(\Delta\tau_S - \Delta\tau_E\right)c \doteq 11 \text{ km}\,. \tag{3.195}$$

This calculation demonstrates that without accounting for general-relativistic effects, operating the GPS positioning system would be completely impossible.

◗

3.3.4 Gravity Probe A

The first high-precision experiment measuring gravitational redshift was the ballistic flight of the Gravity Probe A spacecraft in 1976. Experts from the SAO (*Smithsonian Astrophysical Observatory*) and NASA participated in preparing the experiment, with the scientific team led by Martin Levine and Robert Vessot. The probe had a mass of 100 kilograms and was launched from Wallops Island (Virginia) aboard a Scout launch vehicle to an altitude of 10,000 km. The probe intentionally did not reach escape velocity; upon reaching maximum altitude, it fell back toward the Earth and splashed down in the Atlantic Ocean. A hydrogen maser onboard served as a radio signal source with a precise frequency (acting as an accurate clock). Using a transponder, the signal was received on the Earth surface throughout the flight. After subtracting the Doppler effect, only the gravitational blueshift caused by the signal traveling from the probe to the Earth remained. For the first time, the prediction of general relativity was verified with a relative precision of 0.01 % (10^{-4}).

The Gravity Probe A spacecraft was developed by experts from Stanford University. Their laboratory also produced Gravity Probe B, tasked with measuring frame-dragging by the rotating Earth, and the STEP (*Satellite Test of Equivalence Principle*) probe for verifying the equivalence principle. Gravity Probe B launched in 2004, though its measurements were largely degraded by the electromagnetic field of the solar wind. The STEP project was canceled.

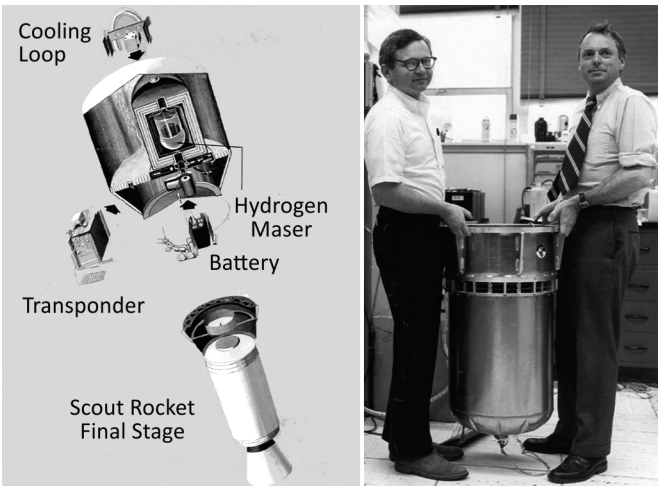


Fig. 3.22: Gravity Probe A. On the left is a schematic of the probe; on the right, Martin Levine and Robert Vessot carry the hydrogen maser.

3.3.5 Berkeley Experiment

The latest method for measuring gravitational redshift is entirely revolutionary, measuring the gravitational shift via a quantum effect over an altitude difference of a mere 0.1 mm. The central figure of this novel method is Steven Chu, Nobel laureate for laser cooling. Chu served for many years as director of the renowned LBNL (*Lawrence Berkeley National Laboratory*). He conceived the idea of utilizing de Broglie waves instead of electromagnetic waves to measure redshift, as measuring time via electromagnetic oscillations or de Broglie waves is equivalent. These waves possess a substantially higher frequency; for instance, a cesium atom cooled via the Chu method exhibits a de Broglie wave frequency of 3×10^{25} Hz.

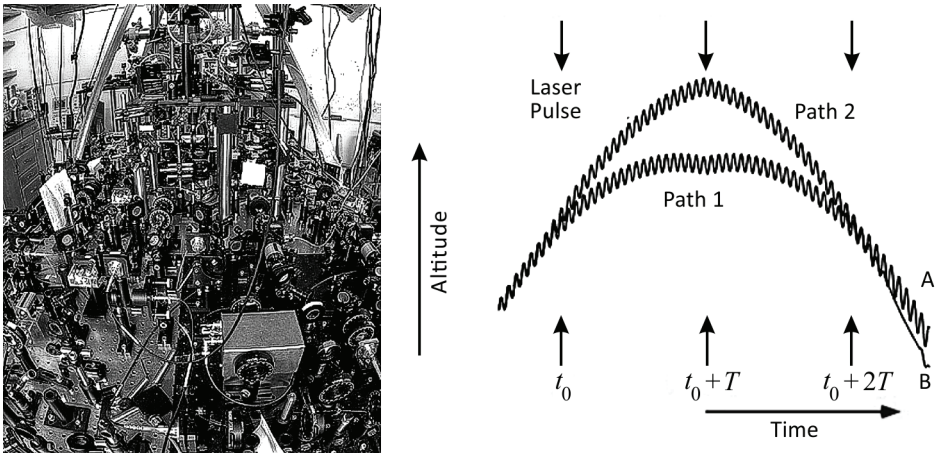


Fig. 3.23: Optical bench where the Berkeley experiment was conducted and its schematic.

This concept was pursued by Achim Peters (Humboldt University) and Holger Müller (UCB), who reinterpreted Peters' 1997 experiments in February 2010. In 1997, Peters cooled cesium atoms via the Chu method to a few millionths of a kelvin, imparted a vertical impulse using a laser, and observed their subsequent free fall. While the 1997 experiments aimed to verify the equivalence principle, the identical setup can measure gravitational redshift. A laser pulse acting on a cloud of cooled cesium atoms prepares them in a superposition of two states: one representing undeflected atoms and the other representing atoms deflected vertically by approximately 0.1 mm. Time flows differently for cesium atoms in the deflected state than for undeflected ones. Over approximately 0.3 s of free fall, the elapsed time in the two states differs by 10^{-20} s. Although an unimaginably brief interval, the high frequency of de Broglie waves makes it measurable via wave interference between the two states. Three laser pulses suffice: the first imparts a vertical impulse with a 50 % probability, placing the atoms in a superposition of undeflected (lower path) and deflected (higher path) states. The second laser pulse directs the atoms on the higher path toward those on the lower path. Upon meeting, interference occurs between the de Broglie waves of both states. A third laser pulse measures the phase shift between the states, effectively functioning as an atom interferometer.

The experimental results are remarkable – gravitational redshift was measured with a relative precision of 7×10^{-9}, four orders of magnitude more precise than the Gravity Probe A measurements.

3.4 Einstein Field Equations

3.4.1 Riemann Curvature Tensor

If a metric is found in some manner and consists of complex functions, it is impossible to determine at first glance whether the "wildness" of the metric is caused by the curvature of the chosen coordinate system or by genuine spacetime curvature. In this chapter, a quantity is constructed that unambiguously determines the curvature of spacetime.

How to Test Curvature?

In general relativity, spacetime curvature expresses the gravitational interaction between bodies. As established previously, a local inertial frame (LIF) – a small cage freely falling for a short duration – can always be chosen where the manifestations of gravitational interaction vanish. A four-dimensional curved world cannot be drawn on a two-dimensional sheet of paper. For simplicity, it is replaced in the figure by a curved surface where one coordinate represents time and the other represents space. The local inertial frame then corresponds to a tangent plane drawn at a given point on the surface.

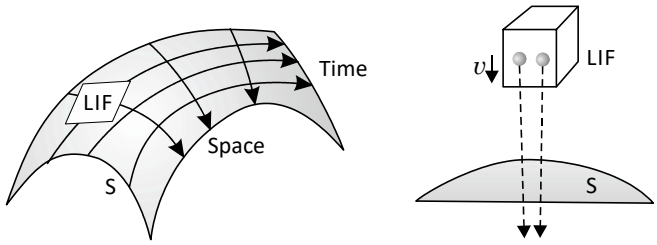


Fig. 3.24: Relationship between an LIF and general coordinates.

It is clearly visible that the tangent plane of the LIF (spacetime hypersurface) adheres to the actual curved surface S (curved spacetime) only in the vicinity of the chosen point. Therefore, the LIF must be local – small in space and observed over a short duration ("small" in time) – for this approximation to hold. In a cage freely falling toward the Earth over a sufficiently long period, it becomes apparent that this is not a true inertial frame, as bodies slightly converge toward each other. In mathematical terms, the trajectory on the tangent plane has moved too far from the point of contact with the curved surface. If the surface S were flat, the tangent plane of the LIF would coincide with it, enabling the construction of a global inertial frame (GIF); i.e., spacetime would be flat and describable by the Minkowski metric.

In Newtonian physics, the convergence of spheres in an LIF is governed by inhomogeneities of the gravitational field, i.e., the first derivatives of the field and the second derivatives of the potential. In the language of the metric, the gravitational potential corresponds to the metric, the field intensity to the first derivatives of the metric (Christoffel symbols), and the field inhomogeneities to the second derivatives of the metric. Spacetime curvature itself must therefore be describable via a quantity contain-

ing the second derivatives of the metric (otherwise, experiments would be conducted at the level of the LIF, i.e., the tangent hypersurface).

Consider a two-dimensional coordinate grid on a curved surface. If a vector is transported parallelly from point A to point B first along one axis and then along the second axis, the result can be compared with an experiment performed in the reverse order:

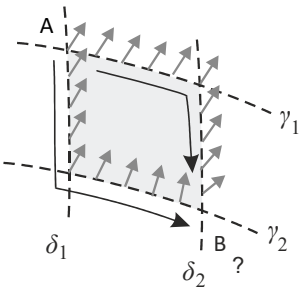


Fig. 3.25: Spacetime curvature test.

If the vector transported along both paths coincides, the surface may be flat; if not, it is certainly curved. Since parallel transport is described by the covariant derivative, the following difference must be examined:

$$A^{\beta}{}_{;\gamma\delta} - A^{\beta}{}_{;\delta\gamma} \,. \tag{3.196}$$

If a zero result is obtained for all possible values of γ and δ, spacetime is flat; if even a single value is nonzero, spacetime is genuinely curved, making it impossible to introduce a global inertial frame. The key to the question in the section heading is thus the evaluation of the difference (3.196). Given that this represents a tensorial quantity, all indices can be lowered to calculate the difference:

$$A_{\beta;\gamma\delta} - A_{\beta;\delta\gamma} = ? \tag{3.197}$$

Only the first part of the expression is written out, as the second part carries the opposite sign and swapped indices δ and γ, formally denoted as $\gamma \leftrightarrow \delta$:

$$A_{\beta;\gamma\delta} - A_{\beta;\delta\gamma} = \left[A_{\beta;\gamma} \right]_{;\delta} \; - \; \{\gamma \leftrightarrow \delta\} \tag{3.198}$$

On the right-hand side, the covariant derivative with respect to δ is applied to the two-index tensor $A_{\beta;\gamma}$:

$$A_{\beta;\gamma\delta} - A_{\beta;\delta\gamma} = \left[A_{\beta;\gamma} \right]_{,\delta} - \Gamma^{\xi}_{\beta\delta} A_{\xi;\gamma} - \Gamma^{\xi}_{\gamma\delta} A_{\beta;\xi} \; - \; \{\gamma \leftrightarrow \delta\} \tag{3.199}$$

The third term on the right-hand side is symmetric in the indices γ and δ; therefore, an identical term with the opposite sign appears inside the curly brackets and cancels it out:

$$A_{\beta;\gamma\delta} - A_{\beta;\delta\gamma} = \left[A_{\beta;\gamma} \right]_{,\delta} - \Gamma^{\xi}_{\beta\delta} A_{\xi;\gamma} \; - \; \{\gamma \leftrightarrow \delta\} \tag{3.200}$$

The two remaining covariant derivatives are now expanded:

$$A_{\beta;\gamma\delta} - A_{\beta;\delta\gamma} = \left[A_{\beta,\gamma} - \Gamma^{\xi}_{\beta\gamma} A_{\xi} \right]_{,\delta} - \Gamma^{\xi}_{\beta\delta} \left[A_{\xi,\gamma} - \Gamma^{\eta}_{\xi\gamma} A_{\eta} \right] \; - \; \{\gamma \leftrightarrow \delta\} \tag{3.201}$$

Taking the partial derivatives of the first term and expanding the second term yields:

$$\begin{aligned} &A_{\beta;\gamma\delta} - A_{\beta;\delta\gamma} = \\ &= A_{\beta,\gamma\delta} - \Gamma^{\xi}_{\beta\gamma,\delta} A_{\xi} - \Gamma^{\xi}_{\beta\gamma} A_{\xi,\delta} - \Gamma^{\xi}_{\beta\delta} A_{\xi,\gamma} + \Gamma^{\xi}_{\beta\delta} \Gamma^{\eta}_{\xi\gamma} A_{\eta} \;\; - \{\gamma \leftrightarrow \delta\} \end{aligned} \tag{3.202}$$

The first term on the right-hand side is symmetric in the indices γ and δ, and a term with the opposite sign again appears inside the curly brackets. The identical cancellation applies to the sum of the third and fourth terms on the right-hand side, leaving:

$$A_{\beta;\gamma\delta} - A_{\beta;\delta\gamma} = -\Gamma^{\xi}_{\beta\gamma,\delta} A_{\xi} + \Gamma^{\xi}_{\beta\delta} \Gamma^{\eta}_{\xi\gamma} A_{\eta} \;\; - \{\gamma \leftrightarrow \delta\} \tag{3.203}$$

This yields the result:

$$A_{\beta;\gamma\delta} - A_{\beta;\delta\gamma} = \Gamma^{\xi}_{\beta\delta,\gamma} A_{\xi} - \Gamma^{\xi}_{\beta\gamma,\delta} A_{\xi} + \Gamma^{\xi}_{\beta\delta} \Gamma^{\eta}_{\xi\gamma} A_{\eta} - \Gamma^{\xi}_{\beta\gamma} \Gamma^{\eta}_{\xi\delta} A_{\eta} \tag{3.204}$$

To factor out the vector A, the summation index in the first two terms is changed to η:

$$A_{\beta;\gamma\delta} - A_{\beta;\delta\gamma} = \left[\Gamma^{\eta}_{\beta\delta,\gamma} - \Gamma^{\eta}_{\beta\gamma,\delta} + \Gamma^{\xi}_{\beta\delta} \Gamma^{\eta}_{\xi\gamma} - \Gamma^{\xi}_{\beta\gamma} \Gamma^{\eta}_{\xi\delta} \right] A_{\eta} \tag{3.205}$$

The resulting expression is written as:

► $$A_{\beta;\gamma\delta} - A_{\beta;\delta\gamma} = R^{\eta}{}_{\beta\gamma\delta} A_{\eta} \,; \tag{3.206}$$

► $$R^{\eta}{}_{\beta\gamma\delta} \equiv \Gamma^{\eta}_{\beta\delta,\gamma} - \Gamma^{\eta}_{\beta\gamma,\delta} + \Gamma^{\xi}_{\beta\delta} \Gamma^{\eta}_{\xi\gamma} - \Gamma^{\xi}_{\beta\gamma} \Gamma^{\eta}_{\xi\delta} \,, \tag{3.207}$$

where $R^{\eta}{}_{\beta\gamma\delta}$ is the Riemann curvature tensor. It represents the key quantity for determining whether spacetime is flat or curved.

Flatness of Spacetime

In physics, an abbreviated notation is usually utilized for the Riemann curvature tensor:

$$R^{\eta}_{\beta\gamma\delta} \equiv R^{\eta}{}_{\beta\gamma\delta} \,. \tag{3.208}$$

This is analogous to the Christoffel symbols. As is apparent from definition (3.207), the Riemann tensor contains second derivatives of the metric tensor, since the Christoffel symbols themselves contain first derivatives. From the definition it is not immediately obvious that this quantity represents a tensor, as Christoffel symbols do not obey tensor transformation rules. Nevertheless, the tensorial character of $R^{\eta}{}_{\beta\gamma\delta}$ is clearly visible from relation (3.206). Individual components can therefore be raised and lowered arbitrarily, allowing the introduction of the Riemann tensor with covariant components:

► $$R_{\alpha\beta\gamma\delta} \equiv g_{\alpha\eta} R^{\eta}{}_{\beta\gamma\delta} \,. \tag{3.209}$$

The commutator of the second covariant derivatives (3.206) for a contravariant vector component can also be readily written:

► $$A^{\beta}{}_{;\gamma\delta} - A^{\beta}{}_{;\delta\gamma} = R_{\eta}{}^{\beta}{}_{\gamma\delta} A^{\eta} \,; \tag{3.210}$$

Contractions of the Riemann tensor are also highly useful:

► $$R_{\alpha\beta} \equiv R^{\xi}{}_{\alpha\xi\beta}, \tag{3.211}$$

► $$R \equiv R^{\xi}{}_{\xi}. \tag{3.212}$$

The former is termed the *Ricci tensor* and the latter the *scalar curvature*. If R or $R_{\alpha\beta}$ is nonzero, spacetime is not flat. If they are zero, the space may be flat, though not necessarily. For a complete assessment of spacetime curvature, all components of the Riemann curvature tensor must be taken into account. The following statements are equivalent and represent different expressions of spacetime flatness.

Equivalent Conditions for Spacetime Flatness

- All components of the Riemann tensor are zero.
- Spacetime is flat.
- A global inertial coordinate system exists.
- The Minkowski metric can be introduced in spacetime.
- A vector remains unaltered under parallel transport along any curve.

Riemann Tensor Properties

The Riemann tensor has four indices, each assuming four values (one temporal and three spatial). yielding 256 components in total. Various symmetries hold among these components, leaving only 20 independent components. The Riemann tensor is antisymmetric in both the first and second pairs of indices, symmetric with respect to swapping the two pairs, and the sum of cyclic permutations of the last three indices equals zero (all properties can be demonstrated directly from the definition):

$$R_{\alpha\beta\gamma\delta} = -R_{\beta\alpha\gamma\delta}, \tag{3.213}$$

$$R_{\alpha\beta\gamma\delta} = -R_{\alpha\beta\delta\gamma}, \tag{3.214}$$

$$R_{\alpha\beta\gamma\delta} = +R_{\gamma\delta\alpha\beta}, \tag{3.215}$$

$$R_{\xi\alpha\beta\gamma} + R_{\xi\beta\gamma\alpha} + R_{\xi\gamma\alpha\beta} = 0. \tag{3.216}$$

Viewing the first two indices as a 4×4 antisymmetric matrix, only the elements above the diagonal are independent. Zeros must lie on the diagonal, while elements below the diagonal are determined from those above. The total number of independent elements associated with the first pair of indices is therefore 6. Analogous reasoning applies to the second pair of indices. Replacing the first pair with a single index taking six values and doing the same for the second pair allows viewing the Riemann tensor as a 6×6 matrix containing 36 independent values. According to (3.215), however, this matrix is symmetric, meaning elements below the diagonal can be calculated from those above. Consequently, there are only 6 + 5 + 4 + 3 + 2 + 1 = 21 independent elements. Taking into account that relation (3.216) also holds among them, the Riemann tensor possesses at most 20 independent values. It can be shown that there are precisely 20.

Gravity Probe B

The ballistic spacecraft Gravity Probe A, which tested the change in the rate of time using a hydrogen maser time standard, has already been mentioned. At Stanford University, however, another probe – Gravity Probe B – was developed to test, using gyroscopes: a) the curvature of spacetime around the Earth; b) frame-dragging by the rotating Earth. The spin axis of a rotating gyroscope represents a parallel-transported vector. Inside the probe, the gyroscopes were suspended in housings by a magnetic field and arranged so that they moved along a geodesic (along with the entire spacecraft). If the Earth genuinely curves spacetime, the Riemann tensor is nonzero and the gyroscope spin axis should shift by 6.6″ per year. Even smaller is the frame-dragging effect of the rotating Earth, amounting to a mere 0.042″ per year; along a polar orbit, this angular shift is perpendicular to the shift caused by spacetime curvature.

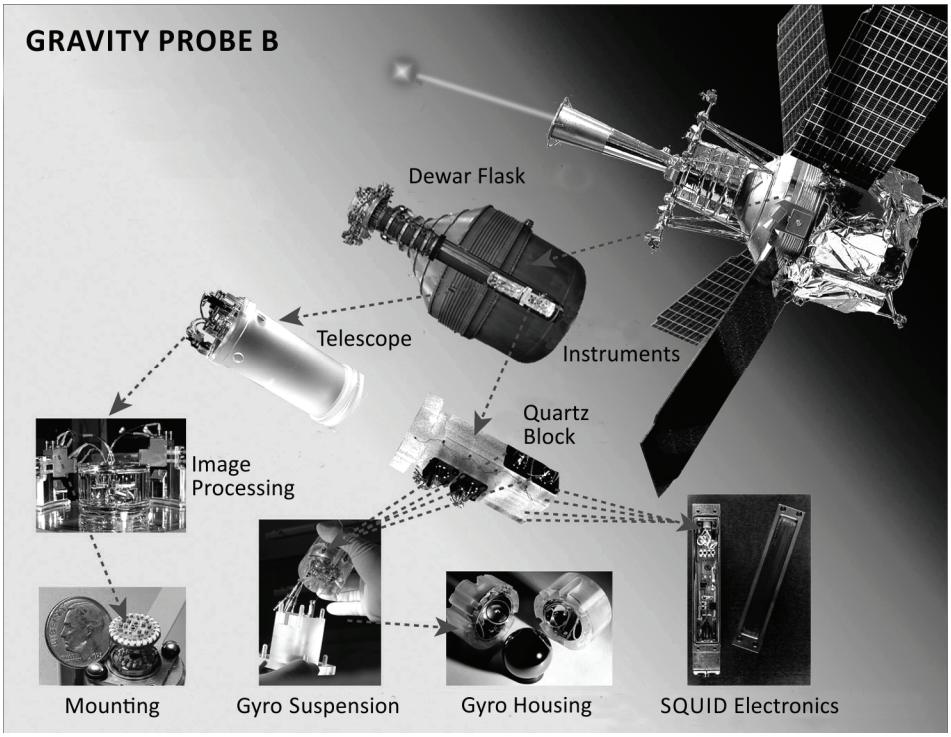


Fig. 3.26: Gravity Probe B.

The spacecraft had a mass of 3,400 kg and was launched into a polar orbit on April 20, 2004, from Vandenberg Air Force Base in southern California, orbiting the Earth at an altitude of 640 km. The core of the spacecraft was a Dewar flask containing 2.5 m³ of helium, cooling the experimental chamber to a temperature of 1.8 K for one year. Built into the axis of the Dewar flask within a lead shield was a housing containing the measuring apparatus, consisting of two rigidly joined parts: a 53 cm long fused-quartz block with four gyroscopes and a guide telescope. Solar panels and a telescope sunshade were mounted on the exterior of the Dewar flask.

The heart of the spacecraft consisted of four gyroscopes, each comprising a housing and a rotor. Each gyroscope was placed in a separate cylindrical cavity within the fused-quartz block. The gyroscope rotors resembled ping-pong balls, measuring 3.8 cm in diameter, and extraordinary care was devoted to their fabrication. Made of fused Brazilian quartz, their surfaces were polished to a precision of a mere forty atomic layers (0.01 μm). A thin layer of niobium, which became superconducting at the operating temperature, was coated onto the rotors.

The rotation of the superconducting material (niobium on the rotor surfaces) generated a nonzero magnetic moment (the London effect), measured by a superconducting loop embedded in the gyroscope housing. The loop was connected to SQUID measuring electronics, enabling precise determination of the gyroscope spin axis.

The fused-quartz block housing the gyroscopes was rigidly attached to a small guide telescope with an aperture diameter of 14 cm and a length of 36 cm. Throughout the observation period, the telescope tracked a selected guide star (IM Pegasi), toward which the spin axes of all gyroscopes were initially aligned, with two rotating in one direction and the remaining two in the opposite direction.

The spacecraft was the culmination of four decades of meticulous preparation. Testing general relativity using rotating gyroscopes was proposed in 1959 by Leonard Schiff of Stanford University, which also became the primary institution behind launching the probe into orbit. All data transmitted by the spacecraft were processed at the university campus computing center. State-of-the-art technologies were employed in preparing the spacecraft, which was assembled in a Class 10 clean room according to U.S. standards, as even micrometer-sized dust grains could damage the precision-machined components. The spacecraft successfully measured the shift in the gyroscope axes caused by spacetime curvature; however, due to repeated measurement interruptions caused by solar flares, it failed to accurately measure the shift induced by frame-dragging from the rotating Earth. Project funding was terminated on May 21, 2008.

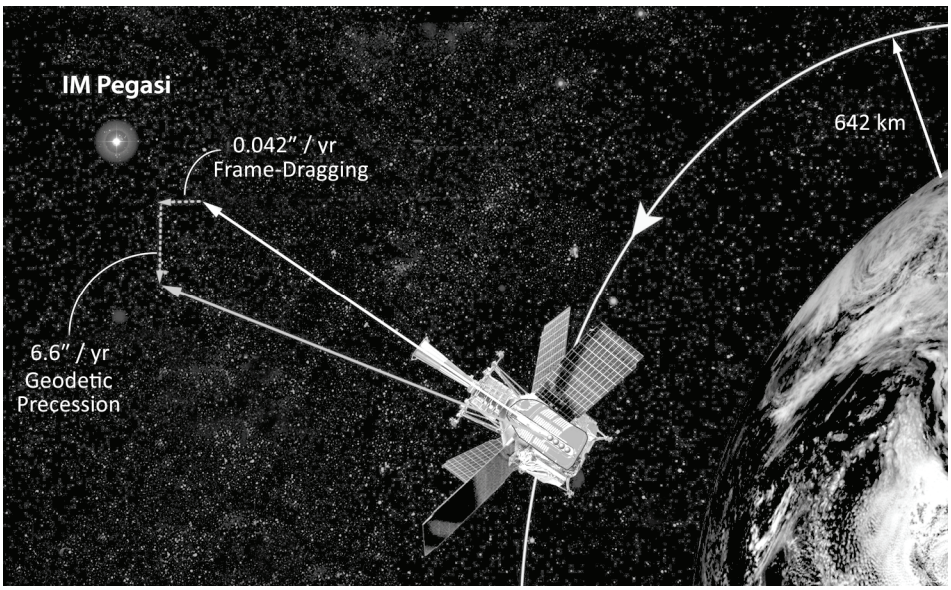


Fig. 3.27: Gravity Probe B deflection measurements.

3.4.2 Continuity Equation in Classical Physics

The Einsteinian law of gravitation addresses the question of how matter curves spacetime and how matter moves within this curved spacetime. The preceding section demonstrated how to describe curved spacetime. Attention now turns to describing the distribution of mass and energy in spacetime. The continuity equation, expressing the conservation law of a physical quantity in differential form, has already been encountered. These relations must be generalized to the entire spacetime. Therefore, the derivation of the conservation laws of energy and momentum in a flowing ideal fluid free of external forces is briefly reviewed.

Consider the flow of an additive quantity A (scaling with the amount of substance, e.g., mass, energy, momentum, or charge). The flow of an additive quantity is described by four quantities (density and flux), where $\mathbf{u}(t, \mathbf{x})$ denotes the flow velocity:

$$\rho_A \equiv \lim_{\Delta V \to 0} \frac{\Delta A}{\Delta V}; \qquad \mathbf{j}_A \equiv \rho_A \mathbf{u}\,. \tag{3.217}$$

As will be demonstrated later in this section, these four quantities form a relativistic four-vector and transform via the Lorentz matrix; ρ_A is termed the density, and $\mathbf{j}_A$ is termed the flux of quantity A. What is the physical meaning of flux? Like any vector, it possesses both magnitude and direction. The direction is apparent, coinciding with the velocity direction, i.e., the flow direction of quantity A. To determine the magnitude of the flux, consider an imaginary surface ΔS placed perpendicular to the flow; over a time interval Δt, the flowing substance shifts by Δl:

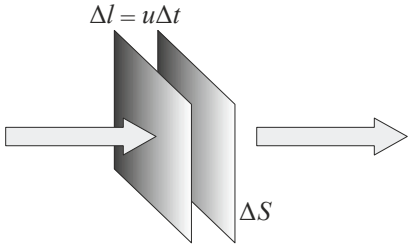


Fig. 3.28: Physical meaning of the flux of an additive quantity.

The magnitude of the flux is now determined directly from its definition:

$$j_A = \rho_A u = \frac{\Delta A}{\Delta V}\frac{\Delta l}{\Delta t} = \frac{\Delta A}{\Delta S \Delta l}\frac{\Delta l}{\Delta t} = \frac{\Delta A}{\Delta S\,\Delta t}\,. \tag{3.218}$$

The magnitude of the flux thus represents the amount of quantity A flowing through a unit area per unit time. For instance, charge flux carries the unit $\mathrm{Cm^{-2}s^{-1}}$. If quantity A is neither lost nor created during the flow, the temporal decrease of the quantity within an arbitrary volume must equal its flux across the surface bounding this volume:

$$-\frac{\mathrm{d}}{\mathrm{d}t}\int_V \rho_A \,\mathrm{d}V \;=\; \oint_{\partial V} \mathbf{j}_A \cdot \mathrm{d}\mathbf{S}$$

The boundary of volume V is denoted by ∂V. The dimension of the integral follows from context rather than the number of integral symbols, as is customary in introductory

physics courses. The dot product on the right accounts for the fact that the entire substance flows perpendicular to the surface, whereas zero substance crosses it when flowing parallel to the surface. Utilizing the Gauss divergence theorem converts the surface integral into a volume integral, combining both integrals:

$$\int_V \left(\frac{\partial \rho_A}{\partial t} + \operatorname{div} \mathbf{j}_A \right) \mathrm{d}V \;=\; 0 \;.$$

The time derivative has been moved inside the integral; however, since $\rho_A = \rho_A(t, \mathbf{x})$ offers a choice of variables during differentiation, a partial derivative must be used. This relation must hold for an arbitrary volume during flow, which is possible only if the integrand vanishes (up to sets of lower dimension than the domain of integration):

► $$\frac{\partial \rho_A}{\partial t} + \operatorname{div} \mathbf{j}_A \;=\; 0 \tag{3.219}$$

The derived relation is termed the continuity equation; the right-hand side equals zero if quantity A is conserved during flow, and is nonzero if it is not conserved. The continuity equation thus expresses the conservation law of the additive quantity A.

Conservation of Mass and Energy in Classical Physics

For the (rest) mass density, the continuity equation takes the form:

► $$\frac{\partial \rho}{\partial t} + \partial_k \left(\rho u_k \right) = 0 \,. \tag{3.220}$$

The identical equation holds for energy density ($E = mc^2$, i.e., $\rho_\mathrm{E} = \rho_\mathrm{m} c^2$), differing only by the constant c^2. Introducing the four-vector

$$J^\mu \equiv \begin{pmatrix} \rho c \\ \rho \mathbf{u} \end{pmatrix} \tag{3.221}$$

allows writing the mass-energy conservation law in a concise form:

$$\partial_\mu J^\mu = 0 \,; \qquad J^\mu{}_{,\mu} = 0 \,, \tag{3.222}$$

demonstrating that J^μ forms a four-vector describing the flow of mass and energy.

Conservation of Momentum in Classical Physics

For an object of mass m, the Newton motion equation holds:

$$m \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} \;=\; \mathbf{F} \,. \tag{3.223}$$

For a flowing medium, the force density is introduced:

$$\mathbf{f} \equiv \lim_{\Delta V \to 0} \frac{\Delta \mathbf{F}}{\Delta V} \,. \tag{3.224}$$

In terms of densities, the velocity **v** of a single particle transitions into the flow velocity field **u**(*t*, **x**):

$$\rho \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \mathbf{f} . \tag{3.225}$$

Expanding the derivative for the field **u**(*t*, **x**) yields the equation:

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho (\mathbf{u} \cdot \boldsymbol{\nabla}) \mathbf{u} = \mathbf{f} . \tag{3.226}$$

It remains to determine the force density. In conservative fields, the force points toward the minimum of potential energy:

$$\mathbf{F} = - \boldsymbol{\nabla} W_{\mathrm{p}} , \tag{3.227}$$

or in terms of densities:

$$\mathbf{f} = - \boldsymbol{\nabla} w_{\mathrm{p}} . \tag{3.228}$$

The pressure energy is:

$$W_{\mathrm{p}} = \int F_{\mathrm{p}} \mathrm{d}l = \int pS \, \mathrm{d}l = \int p \, \mathrm{d}V , \tag{3.229}$$

hence the pressure energy density is $w_{\mathrm{p}} = p$, and the force density caused by pressure is:

$$\mathbf{f} = - \boldsymbol{\nabla} p . \tag{3.230}$$

The equation of motion (3.226) with the pressure-induced force density therefore takes the form:

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho (\mathbf{u} \cdot \boldsymbol{\nabla}) \mathbf{u} = - \boldsymbol{\nabla} p . \tag{3.231}$$

The time evolution of the momentum density (i.e., the law of momentum conservation) is now determined:

$$\frac{\partial}{\partial t} (\rho u_k) = \frac{\partial \rho}{\partial t} u_k + \rho \frac{\partial u_k}{\partial t} .$$

Substituting the continuity equation (3.220) for the time change of density and the equation of motion (3.231) for the time change of momentum yields:

$$\frac{\partial}{\partial t} (\rho u_k) = - \partial_l (\rho u_l) \, u_k - \rho (u_l \partial_l) \, u_k - \partial_k p .$$

Moving all terms to the left-hand side and rearranging gives:

$$\frac{\partial}{\partial t} (\rho u_k) + \frac{\partial (\rho u_l)}{\partial x_l} u_k + (\rho u_l) \frac{\partial u_k}{\partial x_l} + \frac{\partial p}{\partial x_k} = 0 \quad \Rightarrow$$

$$\frac{\partial}{\partial t} (\rho u_k) + \frac{\partial}{\partial x_l} \left(\rho u_l u_k \right) + \frac{\partial p}{\partial x_k} = 0 \quad \Rightarrow$$

$$\frac{\partial}{\partial t} (\rho u_k) + \frac{\partial}{\partial x_l} \left(\rho u_k u_l + p \delta_{kl} \right) = 0 .$$

The law of momentum conservation has been obtained. Inside the parentheses within the spatial derivatives lies the momentum flux, also termed the stress tensor. Since momentum itself is a vector quantity, its flux forms a second-order tensor. The symmetry of the stress tensor with respect to index swapping ensures the conservation of angular momentum in the flowing fluid. The stress tensor consists of two parts: a tensorial part associated with fluid flow and a scalar part formed by normal pressure acting equally in all directions. The law of momentum conservation can be written in component form:

$$\frac{\partial}{\partial t}(\rho u_k)+\frac{\partial}{\partial x_l}(T_{kl})=0\,; \qquad T_{kl}\equiv \rho u_k u_l + p\delta_{kl}\,. \tag{3.232}$$

or in invariant form:

$$\frac{\partial}{\partial t}(\rho \mathbf{u})+\nabla\cdot\mathbf{T}=0\,. \tag{3.233}$$

Once again, the form of the continuity equation is obtained. The temporal part contains the momentum density, while the spatial part contains the momentum flux. The latter is now a second-order tensor, and the result of the divergence operation is a vector quantity, as already seen in the chapter *Electromagnetic Field*.

3.4.3 Stress-Energy Tensor

Energy and momentum form a four-vector; therefore, the law of energy conservation (3.220) and the law of momentum conservation (3.232) belong inseparably together. Expanding them by components yields:

$$\begin{aligned}
&\partial_t(\rho)+\partial_x(\rho u_x)+\partial_y(\rho u_y)+\partial_z(\rho u_z)=0\,,\\
&\partial_t(\rho u_x)+\partial_x(\rho u_x^2+p)+\partial_y(\rho u_x u_y)+\partial_z(\rho u_x u_z)=0\,,\\
&\partial_t(\rho u_y)+\partial_x(\rho u_y u_x)+\partial_y(\rho u_y^2+p)+\partial_z(\rho u_y u_z)=0\,,\\
&\partial_t(\rho u_z)+\partial_x(\rho u_z u_x)+\partial_y(\rho u_z u_y)+\partial_z(\rho u_z^2+p)=0\,.
\end{aligned} \tag{3.234}$$

The first row represents the law of energy conservation, while the remaining rows represent the components of the law of momentum conservation. Writing these conservation laws in matrix form gives:

$$\begin{pmatrix}\frac{\partial}{\partial t} & \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z}\end{pmatrix}\cdot\begin{pmatrix}\rho & \rho u_x & \rho u_y & \rho u_z\\ \rho u_x & \rho u_x u_x+p & \rho u_x u_y & \rho u_x u_z\\ \rho u_y & \rho u_y v_x & \rho u_y u_y+p & \rho u_y u_z\\ \rho u_z & \rho u_z u_x & \rho u_z u_y & \rho u_z u_z+p\end{pmatrix}=0 \tag{3.235}$$

This notation represents four equations, obtained by multiplying the row matrix on the left by the columns of the square matrix on the right. The symmetric 4×4 matrix on the right is termed the *stress-energy tensor*, denoted in relativity by $T^{\alpha\beta}$. Quantities in the first column represent the mass-energy density ρ and momentum density $\rho\mathbf{u}$. In the first row, the identical quantity $\rho\mathbf{u}$ represents the mass flux. Adjusting the matrix notation so

that all elements possess identical dimensions (by appropriately introducing the speed of light c) into both matrices) yields:

$$\left(\frac{\partial}{\partial ct} \quad \frac{\partial}{\partial x} \quad \frac{\partial}{\partial y} \quad \frac{\partial}{\partial z}\right).\begin{pmatrix} \rho c^2 & \rho c u_x & \rho c u_y & \rho c u_z \\ \rho c u_x & \rho u_x u_x + p & \rho u_x u_y & \rho u_x u_z \\ \rho c u_y & \rho u_y u_x & \rho u_y u_y + p & \rho u_y u_z \\ \rho c u_z & \rho u_z u_x & \rho u_z u_y & \rho u_z u_z + p \end{pmatrix} = 0\,. \quad (3.236)$$

The component T^{00} was originally ρ (mass density) and now equals ρc^2 (energy density). In a local inertial frame falling instantaneously with the matter, the matrix on the right-hand side takes the form $T^{\alpha\beta} = \mathrm{diag}(\rho c^2, p, p, p)$. The trace of the matrix (sum of diagonal terms) is $T^{\alpha}{}_{\alpha} = -\rho c^2 + 3p$. The sign change in the temporal component resulted from lowering the index. The trace is an important matrix invariant, and this trace must remain unchanged in the relativistic form as well. Introducing c into the terms has no effect on the resulting equations (the conservation laws of mass-energy and momentum). The spatial part of the tensor has the form:

$$T_{kl} = \rho u_k u_l + p\delta_{kl}\,. \quad (3.237)$$

In special relativity, lower and upper indices need not be distinguished in the spatial part. A direct generalization to the metric of general relativity appears to be:

$$T^{\alpha\beta} \overset{?}{=} \rho U^{\alpha} U^{\beta} + p g^{\alpha\beta}\,. \quad (3.238)$$

Performing the Newtonian limit verifies whether the matrix in expression (3.236) is recovered. Evaluating the four-velocity and the metric in the Newtonian limit yields:

$$U^{\alpha} = \frac{\mathrm{d}x^{\alpha}}{\mathrm{d}\tau} = \begin{pmatrix} \gamma c \\ \gamma \mathbf{u} \end{pmatrix} \to \begin{pmatrix} c \\ \mathbf{u} \end{pmatrix}, \quad (3.239)$$

$$g^{\alpha\beta} \to \eta^{\alpha\beta} = \mathrm{diag}(-1,+1,+1,+1)\,. \quad (3.240)$$

For the spatial components, this gives:

$$T^{kl} = \rho U^k U^l + p g^{kl} \to \rho u_k u_l + p\delta_{kl}\,. \quad (3.241)$$

Once again, the fact that there is no difference between contravariant and covariant components in the spatial part in the Newtonian limit has been utilized. The mixed spatiotemporal components are:

$$T^{0l} = \rho U^0 U^l + p g^{0l} \to \rho c u_l. \quad (3.242)$$

Both expressions transition in the Newtonian limit into the correct tensor terms of expression (3.236). However, an issue arises with the T^{00} component:

$$T^{00} = \rho U^0 U^0 + p g^{00} \to \rho c^2 - p. \quad (3.243)$$

Such a term does not correspond to the non-relativistic expression and would furthermore violate the correct value of the trace of the stress-energy tensor, which is an invariant. The correct generalization for relativity is the expression:

► $$T^{\alpha\beta} = \left(\rho + \frac{p}{c^2}\right) U^\alpha U^\beta + p g^{\alpha\beta}, \tag{3.244}$$

where ρ is the rest mass density. In the low-velocity limit, $p/c^2 << \rho$, meaning it does not affect the Newtonian limit. In definition (3.244), the pressure term – arising from the chaotic motion of particles – also contributes additively to the rest energy density. The T^{00} component and the trace of the tensor under definition (3.244) yield the correct values in an LIF ($\gamma = 1$) instantaneously comoving with the matter:

$$T^{00} = \left(\rho + \frac{p}{c^2}\right) U^0 U^0 + p g^{00} = \left(\rho + \frac{p}{c^2}\right) \gamma^2 c^2 - p = \rho c^2, \tag{3.245}$$

$$\mathrm{Tr}(T) \equiv T^\alpha{}_\alpha = \left(\rho + \frac{p}{c^2}\right) U^\alpha U_\alpha + p g^\alpha{}_\alpha = \left(\rho + \frac{p}{c^2}\right)(-c^2) + 4p = 3p - \rho c^2. \tag{3.246}$$

The distribution of matter and energy in space is thus described by the stress-energy tensor (3.244). The conservation law of energy and momentum can be written in a compact matrix form:

$$\partial_\beta T^{\alpha\beta} = 0. \tag{3.247}$$

The corresponding covariant notation in general relativity is:

► $$T^{\alpha\beta}{}_{;\beta} = 0. \tag{3.248}$$

For $\alpha = 0$, the law of energy conservation is obtained, and for $\alpha = 1, 2, 3$, the conservation laws of the individual momentum components follow. However, equation (3.248) implies

$$T^{\alpha\beta}{}_{,\beta} + \Gamma^\alpha{}_{\xi\beta} T^{\xi\beta} + \Gamma^\beta{}_{\xi\beta} T^{\alpha\xi} = 0. \tag{3.249}$$

The first expression is the continuity equation, whereas the remaining terms are caused by curvature and violate the conservation law. Consequently, the conservation law of energy and momentum of matter does not hold in general relativity. Matter transfers energy and momentum to the gravitational field, i.e., to spacetime curvature. Only for the combined system of field + particles (curvature + particles) would the conservation law of energy and momentum remain valid.

3.4.4 Einstein Field Equations

The new gravitation law should take the form:

$$G(\text{spacetime curvature}) = F(\text{matter and energy distribution}), \tag{3.250}$$

since the matter-energy distribution is assumed to cause spacetime curvature, and conversely, spacetime curvature alters it. Because this distribution is described by the stress-energy tensor, quantities with at most two indices characterizing spacetime curvature must be utilized on the left-hand side:

$$G(R^{\alpha\beta}, R, g^{\alpha\beta}) = F(T^{\alpha\beta}). \tag{3.251}$$

Assuming that the sought law possesses the simplest possible form leaves:

$$c_1 R^{\alpha\beta} + c_2 R g^{\alpha\beta} + c_3 g^{\alpha\beta} = c_4 T^{\alpha\beta} . \tag{3.252}$$

The new law of gravitation contains four undetermined constants. One of them can be eliminated by dividing the entire equation by the constant c_1:

$$R^{\alpha\beta} + d_1 R g^{\alpha\beta} + d_2 g^{\alpha\beta} = d_3 T^{\alpha\beta} . \tag{3.253}$$

The constant d_1 is determined from condition (3.248), i.e., $T^{\alpha\beta}{}_{;\beta} = 0$. Differentiating equation (3.253) covariantly with respect to β yields:

$$R^{\alpha\beta}{}_{;\beta} + d_1 \left(R g^{\alpha\beta} \right)_{;\beta} + d_2 g^{\alpha\beta}{}_{;\beta} = d_3 T^{\alpha\beta}{}_{;\beta} .$$

The right-hand side is zero, and the last term on the left-hand side vanishes (the covariant derivative of the metric is always zero, see (3.126). The equation reduces to:

$$R^{\alpha\beta}{}_{;\beta} + d_1 \left(R g^{\alpha\beta} \right)_{;\beta} = 0 .$$

From this condition, the constant d_1 can be determined. Calculation yields $d_1 = -½$, and therefore:

$$R^{\alpha\beta} - ½ R g^{\alpha\beta} + d_2 g^{\alpha\beta} = d_3 T^{\alpha\beta} . \tag{3.254}$$

Newtonian Limit

In the limit of weak fields and low velocities, the new law of gravitation must reduce to the classical equation for the gravitational potential. For the potential energy of two electrostatically interacting charges and two gravitationally interacting masses, the following formulas hold:

$$W_{\mathrm{ES}} = \frac{qQ}{4\pi\varepsilon_0 r} ; \qquad W_{\mathrm{G}} = -G\frac{mM}{r} . \tag{3.255}$$

The opposite sign corresponds to the fact that two positive charges repel each other, whereas two positive masses attract each other. Dividing by the properties of the test particle q in electrostatics and m in gravitation) yields the potentials:

$$\phi_{\mathrm{ES}} = \frac{Q}{4\pi\varepsilon_0 r} ; \qquad \phi_{\mathrm{G}} = -G\frac{M}{r} . \tag{3.256}$$

The defining equations are identical up to constants:

$$\frac{1}{4\pi\varepsilon_0} \quad \leftrightarrow \quad -G . \tag{3.257}$$

The equations for the potentials must be analogous (in electrostatics, this is the well-known Poisson equation):

$$\Delta\phi_{\mathrm{ES}} = -\frac{\rho_Q}{\varepsilon_0} ; \qquad \Delta\phi_{\mathrm{G}} = +4\pi G\rho_{\mathrm{M}} . \tag{3.258}$$

In the Newton limit, where

$$g_{00} = -\left(1 + 2\phi/c^2\right), \tag{3.259}$$

equation (3.254) is expected to reduce to the equation for the gravitational potential:

$$\Delta\phi = +4\pi G\rho\,. \tag{3.260}$$

Expanding the "^{00}" component of equation (3.254) yields second derivatives of the potential on the left-hand side and a term proportional to density appears on the right-hand side. Requiring the Einstein field equations (3.254) to reduce into the Newton law (3.260) yields $d_2 = 0$ and $d_3 = 8\pi G/c^4$. The resulting formula thus takes the form:

► $$R^{\alpha\beta} - \tfrac{1}{2}Rg^{\alpha\beta} = \frac{8\pi G}{c^4}T^{\alpha\beta}\,. \tag{3.261}$$

It is occasionally written as:

► $$G^{\alpha\beta} = \frac{8\pi G}{c^4}T^{\alpha\beta}\,; \qquad G^{\alpha\beta} \equiv R^{\alpha\beta} - \tfrac{1}{2}Rg^{\alpha\beta}\,, \tag{3.262}$$

where $G^{\alpha\beta}$ is termed the Einstein tensor.



Fig. 3.29: The Einstein field equations are even immortalized at the train cemetery in Bolivia. The opposite sign on the right is due to the different convention $\eta_{\mu\nu}$ = diag (1, –1, –1, –1).

Significance of the New Law

The new law of gravitation represents 10 differential equations for the metric. The number of equations is determined by the fact that both sides are symmetric tensors, i.e., $G^{\alpha\beta} = G^{\beta\alpha}$ and $T^{\alpha\beta} = T^{\beta\alpha}$. The equations are nonlinear. Albert Einstein presented the law in a lecture before the Prussian Academy of Sciences in the autumn of 1915. Histori-

cally, it represents the first theory containing both field equations and equations of motion. That these are field equations is apparent at first glance: knowing the distribution of matter and energy allows determining the spacetime metric. However, the law of gravitation also inherently contains the geodesic equation (unlike classical electrodynamics, where Maxwell equations serve as the field equations, while particle motion is calculated separately from the Lorentz force equation).

Differentiating the law of gravitation (3.261) covariantly with respect to β (summing over this index, as the covariant derivative incorporates spacetime curvature) yields zero on the left-hand side (due to the choice of the constant in the second term), while the right-hand side provides the conservation laws of energy and momentum:

$$T^{\alpha\beta}{}_{;\beta} = 0 \tag{3.263}$$

or equivalently:

$$\left[\left(\rho + p/c^2\right)U^\alpha U^\beta + pg^{\alpha\beta}\right]_{;\beta} = 0 \tag{3.264}$$

Consider now massive incoherent dust with zero pressure:

$$\left[\rho U^\beta U^\alpha\right]_{;\beta} = 0 \tag{3.265}$$

Differentiating this expression as a product yields:

$$\left(\rho U^\beta\right)_{;\beta} U^\alpha + \rho U^\beta U^\alpha{}_{;\beta} = 0$$

The first term vanishes because $(\rho U^\beta)_{;\beta} = 0$ (the continuity equation for matter). What remains is:

$$U^\alpha{}_{;\beta}U^\beta = 0 \quad \Rightarrow \quad \left[U^\alpha{}_{,\beta} + \Gamma^\alpha{}_{\xi\beta}U^\xi\right]U^\beta = 0 \quad \Rightarrow$$

$$U^\alpha{}_{,\beta}U^\beta + \Gamma^\alpha{}_{\xi\beta}U^\xi U^\beta = 0 \quad \Rightarrow \quad U^\alpha{}_{,\beta}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} + \Gamma^\alpha{}_{\xi\beta}U^\xi U^\beta = 0 \quad \Rightarrow$$

$$\frac{\mathrm{d}U^\alpha}{\mathrm{d}\tau} + \Gamma^\alpha{}_{\xi\beta}U^\xi U^\beta = 0 \quad \Rightarrow \quad \frac{\mathrm{d}^2x^\alpha}{\mathrm{d}\tau^2} + \Gamma^\alpha{}_{\xi\beta}\frac{\mathrm{d}x^\xi}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = 0\,.$$

This demonstrates that the geodesic equation is embedded within the Einstein field equations. In principle, the Einstein field equations could also contain a term proportional to the metric on the left-hand side:

► $$R^{\alpha\beta} - \tfrac{1}{2}Rg^{\alpha\beta} + \Lambda g^{\alpha\beta} = \frac{8\pi G}{c^4}T^{\alpha\beta}\,. \tag{3.266}$$

Such a term does not affect the conservation laws (the covariant derivative of the metric is zero), yet in the Newtonian limit it modifies the equation for the gravitational potential. Omitting additive constants, this equation assumes the form:

$$\Delta\phi + 2\Lambda\phi = 4\pi G\rho \tag{3.267}$$

For small Λ, the modification of the solution manifests only over large distances. Consequently, the added term is termed the cosmological term, and the constant Λ is termed the *cosmological constant*. Like his contemporaries, Albert Einstein believed in a static universe. However, his equations without the cosmological term led to the spontaneous

collapse of a universe homogeneously filled with incoherent dust. Therefore, in 1917, Einstein added the cosmological term to his equations to counterbalance collapse; for a specific value of Λ, it acted repulsively, opposing gravitational attraction. Without the cosmological term, the equations permit only collapsing or expanding solutions (depending on initial conditions). In 1922, Russian mathematician and meteorologist Alexander Friedmann demonstrated that the cosmological term could not salvage the vision of a static universe, as the equilibrium between attractive gravitation and the repulsive cosmological term is unstable. Any arbitrary perturbation ultimately triggers either collapse or expansion of the universe. Einstein acknowledged the validity of Friedmann's reasoning only in 1923. When Edwin Hubble observationally confirmed the expansion of the universe in 1929 (following Georges Lemaître's 1927 work), Einstein removed the cosmological term from his equations, declaring it the greatest blunder of his life. Depending on the value of Λ, the cosmological term either decelerates (compensates) or accelerates expansion. A detailed analysis of the equations with the cosmological term was conducted by Dutch astronomer Willem de Sitter. In 1998, Adam Riess and Saul Perlmutter discovered the accelerated expansion of the universe, which cannot be driven by attractive gravitational interaction and may instead stem from quantum processes occurring in the vacuum. Accelerated expansion prompted the reintroduction of the cosmological term into the Einstein field equations. The dimension of the cosmological constant is m^{-2}:

$$[\Lambda] = \frac{1}{\mathrm{m}^2} . \tag{3.268}$$

According to current measurements of cosmic expansion, particularly from the SCP (*Supernova Cosmology Project*), the value of the cosmological constant is about:

$$\Lambda \approx 2{,}25 \times 10^{-52}\ \mathrm{m}^{-2} . \tag{3.269}$$

However, when not dealing with cosmological distances, there is no reason to include the cosmological term in the equations.



3.5 Gravitational Waves

Gravitational waves are periodic ripples in spacetime propagating at the speed of light. They possess two independent polarizations inclined at 45°. Their sources are either periodic processes from the rotating objects with quadrupole asymmetry, or catastrophic events (mergers of black holes or neutron stars, asymmetric stellar collapse, etc.).

3.5.1 Wave Equation

Assume a gravitational wave is a small perturbation of the Minkowski metric, i.e.:

► $$g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta}\,. \tag{3.270}$$

The perturbations $h_{\alpha\beta}$ are treated as first-order quantities, meaning their squares are neglected. First, the Christoffel symbols are obtained from formula (3.124):

$$\begin{gathered}
\Gamma^{\alpha}{}_{\beta\gamma} = \frac{1}{2} g^{\alpha\xi}\left(g_{\xi\beta,\gamma} + g_{\xi\gamma,\beta} - g_{\beta\gamma,\xi}\right), \\
\Gamma^{\alpha}{}_{\beta\gamma} = \frac{1}{2}\left(\eta^{\alpha\xi} + h^{\alpha\xi}\right)\left(h_{\xi\beta,\gamma} + h_{\xi\gamma,\beta} - h_{\beta\gamma,\xi}\right), \\
\Gamma^{\alpha}{}_{\beta\gamma} = \frac{1}{2}\eta^{\alpha\xi}\left(h_{\xi\beta,\gamma} + h_{\xi\gamma,\beta} - h_{\beta\gamma,\xi}\right), \\
\Gamma^{\alpha}{}_{\beta\gamma} = \frac{1}{2}\left(h^{\alpha}{}_{\beta,\gamma} + h^{\alpha}{}_{\gamma,\beta} - h_{\beta\gamma,}{}^{\alpha}\right).
\end{gathered} \tag{3.271}$$

Raising and lowering indices on quantities proportional to h needs only to be performed via the Minkowski metric, since squares of h are neglected. Next, the Riemann curvature tensor (3.207) is constructed:

$$\begin{gathered}
R^{\alpha}{}_{\beta\gamma\delta} = \Gamma^{\alpha}_{\beta\delta,\gamma} - \Gamma^{\alpha}_{\beta\gamma,\delta} + \Gamma^{\xi}_{\beta\delta}\Gamma^{\alpha}_{\xi\gamma} - \Gamma^{\xi}_{\beta\gamma}\Gamma^{\alpha}_{\xi\delta}\,, \\
R^{\alpha}{}_{\beta\gamma\delta} = \frac{1}{2}\left(h^{\alpha}{}_{\beta,\delta\gamma} + h^{\alpha}{}_{\delta,\beta\gamma} - h_{\beta\delta,}{}^{\alpha}{}_{\gamma}\right) - \frac{1}{2}\left(h^{\alpha}{}_{\beta,\gamma\delta} + h^{\alpha}{}_{\gamma,\beta\delta} - h_{\beta\gamma,}{}^{\alpha}{}_{\delta}\right), \\
R^{\alpha}{}_{\beta\gamma\delta} = \frac{1}{2}\left(h^{\alpha}{}_{\delta,\beta\gamma} - h_{\beta\delta,}{}^{\alpha}{}_{\gamma}\right) - \frac{1}{2}\left(h^{\alpha}{}_{\gamma,\beta\delta} - h_{\beta\gamma,}{}^{\alpha}{}_{\delta}\right).
\end{gathered} \tag{3.272}$$

Quadratic terms in Γ do not contribute, as they are proportional to squares of $h_{\alpha\beta}$. Next, the Ricci tensor is calculated:

$$\begin{gathered}
R_{\beta\delta} = R^{\xi}{}_{\beta\xi\delta} = \frac{1}{2}\left(h^{\xi}{}_{\delta,\beta\xi} - h_{\beta\delta,}{}^{\xi}{}_{\xi}\right) - \frac{1}{2}\left(h^{\xi}{}_{\xi,\beta\delta} - h_{\beta\xi,}{}^{\xi}{}_{\delta}\right), \\
R_{\beta\delta} = \frac{1}{2}h^{\xi}{}_{\delta,\beta\xi} - \frac{1}{2}\Box h_{\beta\delta} - \frac{1}{2}s_{,\beta\delta} + \frac{1}{2}h_{\beta\xi,}{}^{\xi}{}_{\delta},
\end{gathered}$$

$$R^{\alpha\beta} = \frac{1}{2} h^{\xi\beta}{}_{,}{}^{\alpha}{}_{\xi} - \frac{1}{2}\Box h^{\alpha\beta} - \frac{1}{2} s_{,}{}^{\alpha\beta} + \frac{1}{2} h^{\alpha}{}_{\xi,}{}^{\xi\beta}, \tag{3.273}$$

where the trace of the metric perturbation tensor is denoted as:

$$s \equiv \mathrm{Tr}\,\vec{\mathbf{h}} = h^{\xi}{}_{\xi}. \tag{3.274}$$

Finally, the scalar curvature R remains:

$$R = R^{\eta}{}_{\eta} = \frac{1}{2} h^{\eta\xi}{}_{,\xi\eta} - \frac{1}{2}\Box s - \frac{1}{2}\Box s + \frac{1}{2} h^{\xi}{}_{\eta,}{}^{\eta}{}_{\xi} = h^{\eta\xi}{}_{,\eta\xi} - \Box s. \tag{3.275}$$

The Einstein field equations (3.266) are now constructed:

$$\frac{1}{2} h^{\xi\beta}{}_{,}{}^{\alpha}{}_{\xi} - \frac{1}{2}\Box h^{\alpha\beta} - \frac{1}{2} s_{,}{}^{\alpha\beta} + \frac{1}{2} h^{\alpha}{}_{\xi,}{}^{\xi\beta} - \frac{1}{2}\eta^{\alpha\beta} h^{\xi\eta}{}_{,\xi\eta} + \frac{1}{2}\eta^{\alpha\beta}\Box s = \frac{8\pi G}{c^4} T^{\alpha\beta}. \tag{3.276}$$

Gauge Freedom

When deriving equations for electromagnetic potentials, their non-uniqueness is utilized to impose the Lorenz gauge condition $A^{\alpha}{}_{,\alpha} = 0$. Under this condition, the equations for electromagnetic potentials assume a simple form, reducing to the standard wave equation $\Box A^{\alpha} = -\mu_0 J^{\alpha}$. Even then, electromagnetic potentials are not uniquely determined, permitting further constraints, such as setting the scalar potential to zero. Equation (3.276) behaves analogously. Here, gauge freedom entails, among other things, the freedom to choose the coordinate system. Instead of x^{α}, alternative coordinates $x^{\alpha} + \xi^{\alpha}(x)$ can be chosen such that $h^{\alpha\beta}{}_{,\beta} = 0$. This gauge eliminates the first, fourth, and fifth terms in equation (3.276). Analogous to the electromagnetic case, additional conditions can be imposed on the metric perturbations: the trace of the perturbation tensor can be required to vanish (eliminating the third and sixth terms in the equation), and – just as the scalar potential is set to zero in electricity and magnetism – the components $h^{0\alpha}$ can be required to vanish. The result is a wave equation supplemented by auxiliary conditions:

► $$\begin{gathered} \Box h_{\alpha\beta} = -\frac{16\pi G}{c^4} T_{\alpha\beta}; \\ h^{\alpha\beta}{}_{,\beta} = 0, \qquad s = h^{\alpha}{}_{\alpha} = 0, \qquad h^{0\alpha} = 0. \end{gathered} \tag{3.277}$$

3.5.2 Properties of Gravitational Waves

Every solution of the wave equation in vacuum can be written as a superposition of plane waves of the form:

$$\psi = A\,\mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{x} - \omega t]} = A\,\mathrm{e}^{\mathrm{i} k_{\mu} x^{\mu}}. \tag{3.278}$$

The physical solution of the wave equation is, of course, the real part of the resulting perturbation. Plane waves exhibit simple properties with respect to differentiation:

$$\partial^{\gamma}\psi = \partial^{\gamma} A\,\mathrm{e}^{\mathrm{i}[\mathbf{k}\cdot\mathbf{x} - \omega t]} = \partial^{\gamma} A\,\mathrm{e}^{\mathrm{i} k_{\mu} x^{\mu}} = \mathrm{i} k^{\gamma} A\,\mathrm{e}^{\mathrm{i} k_{\mu} x^{\mu}} = \mathrm{i} k^{\gamma}\psi.$$

Differentiation can therefore be performed symbolically:

► $$\partial^\gamma \psi = \mathrm{i} k^\gamma \psi \qquad \Rightarrow \qquad \partial^\gamma \to \mathrm{i} k^\gamma \tag{3.279}$$

Decomposing an actual wave into plane wavefronts is termed the Fourier transform. For linear equations, the superposition principle holds, allowing attention to focus solely on individual partial waves obeying the simple relation (3.279). Consider now a gravitational wave already generated by some mechanism and propagating through empty space. It then satisfies the wave equation

$$\Box h_{\alpha\beta} = 0 \tag{3.280}$$

and its solution can be constructed from plane waves:

► $$h_{\alpha\beta} = A_{\alpha\beta}\, \mathrm{e}^{\mathrm{i} k_\alpha x^\alpha} \,. \tag{3.281}$$

Substituting the plane wave (3.281) into the wave equation (3.280) yields the condition under which a plane wave solves the wave equation:

► $$k_\mu k^\mu = 0 \,. \tag{3.282}$$

This condition states that the wave four-vector is a null, i.e., in spacetime it points along the light cone. Condition (3.282) is simultaneously the dispersion relation (the relationship between angular frequency and wave vector). Expanding into components gives:

$$-\frac{\omega^2}{c^2} + k^2 = 0 \tag{3.283}$$

or equivalently:

► $$\omega = ck \,, \tag{3.284}$$

which is the identical dispersion relation obeyed by electromagnetic waves in vacuum. Like an electromagnetic wave, a gravitational wave propagates at the speed of light. The fundamental gauge condition leads to the relation

► $$h_{\alpha\beta} k^\beta = 0 \,, \tag{3.285}$$

demonstrating that gravitational waves are *transverse waves* (perturbations are perpendicular to the wave four-vector). Choosing the coordinate system such that the wave vector (the spatial part of the wave four-vector) points along the first spatial axis gives:

$$k^\alpha = \begin{pmatrix} \omega/c \\ k \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} k \\ k \\ 0 \\ 0 \end{pmatrix} . \tag{3.286}$$

The magnitude of the temporal component of the wave four-vector is readily calculated from relation (3.282) or from the fact that it is a null four-vector. Applying the gauge condition in the form (3.285), the remaining gauge conditions, tracelessness, and the symmetry of perturbations requires the plane wave amplitudes to satisfy:

$$\begin{aligned} A_{\alpha\beta} k^\beta = 0, \qquad & A_{0\beta} = 0, \\ A_{11} + A_{22} + A_{33} = 0, \qquad & A_{\alpha\beta} = A_{\beta\alpha}. \end{aligned} \tag{3.287}$$

Consequently, the amplitude possesses only two free parameters ensuring the satisfaction of the first equation in (3.287) along with the remaining conditions:

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & A & B \\ 0 & 0 & B & -A \end{pmatrix} \cdot \begin{pmatrix} k \\ k \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.288}$$

A gravitational wave can thus be decomposed into two independent polarizations:

$$h = h_+ + h_\times = A \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \mathrm{e}^{\mathrm{i}\, k_\alpha x^\alpha} + B \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \mathrm{e}^{\mathrm{i}\, k_\alpha x^\alpha}. \tag{3.289}$$

The real part contributes to the metric tensor, yielding:

$$g_+ = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1+h & 0 \\ 0 & 0 & 0 & 1-h \end{pmatrix}; \qquad h \equiv A\cos(kx - \omega t). \tag{3.290}$$

$$g_\times = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & h \\ 0 & 0 & h & 1 \end{pmatrix}; \qquad h \equiv B\cos(kx - \omega t). \tag{3.291}$$

The corresponding intervals are:

$$\mathrm{d}s_+^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + (1+h)\,\mathrm{d}y^2 + (1-h)\mathrm{d}z^2, \tag{3.292}$$

$$\mathrm{d}s_\times^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + 2h\,\mathrm{d}y\,\mathrm{d}z\,. \tag{3.293}$$

The wave propagates along the x-axis with deformations in the (yz)-plane, making it a transverse wave. The relative deformation along the y-axis in the first case gives:

$$y = \sqrt{1+h}\; y_0 \qquad \Rightarrow \qquad y \approx (1 + h/2)\, y_0 \qquad \Rightarrow$$

► $$\frac{\Delta y}{y_0} \approx \frac{h}{2}. \tag{3.294}$$

The perturbation h is thus of the same order of magnitude as the relative deformation of a body through which the gravitational wave passes. If a ring of point masses lies in the path of a gravitational wave, its deformation proceeds according to the following figure, which also clearly illustrates that the two polarizations are rotated by an angle of 45°.

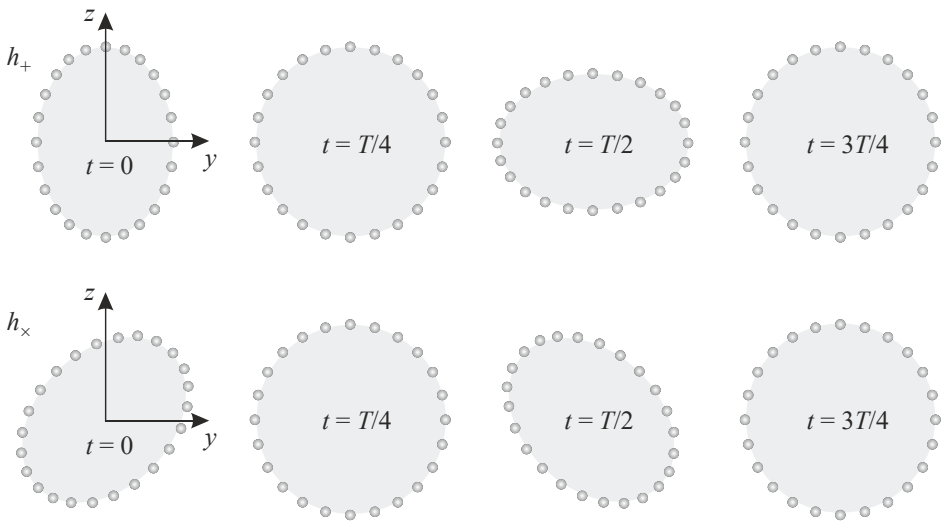


Fig. 3.30: Passage of a gravitational wave through a ring of point masses.

The source of an electromagnetic wave can be a dipole charge distribution. The electric dipole moment is given by the relation:

$$\mathbf{p}_\mathrm{E} = \sum_a q_a \mathbf{r}_a \,. \tag{3.295}$$

The intensity of dipole radiation is then:

$$I \sim \ddot{\mathbf{p}}_\mathrm{E}^2 \,. \tag{3.296}$$

For gravitational waves (gravitational radiation), a rigorous general-relativistic analysis would be required. Nevertheless, even a simplified consideration reveals that a rotating body with a dipole mass distribution symmetry around its rotation axis cannot serve as a source of gravitational waves. Introducing the gravitational dipole moment

$$\mathbf{p}_\mathrm{G} = \sum_a m_a \mathbf{r}_a \tag{3.297}$$

shows that its derivative is proportional to the constant center-of-mass velocity, so:

$$I \sim \ddot{\mathbf{p}}_\mathrm{G}^2 = 0 \,. \tag{3.298}$$

Only a quadrupole asymmetry in the matter distribution can act as a source of gravitational waves. In summary, the key properties of gravitational waves are:

- Gravitational waves propagate at the speed of light.
- Gravitational waves are transverse waves.
- Gravitational waves possess two independent polarizations rotated by 45°
- The source of gravitational waves is a quadrupole asymmetry of matter.

Comparison with Electricity and Magnetism

	Electricity and Magnetism	General Relativity
Quantity	A^{α}	$h^{\alpha\beta}$
Wave equation in a medium	$\Box A^{\alpha} = -\mu_0 J^{\alpha}$	$\Box h^{\alpha\beta} = -\frac{16\pi G}{c^4} T^{\alpha\beta}$
Fundamental gauge	$\partial_{\alpha} A^{\alpha} = 0$	$\partial_{\alpha} h^{\alpha\beta} = 0$
Additional gauges	$A^0 = 0$	$h^{0\alpha} = 0\,; \quad s = 0$
Wave equation in vacuum	$\Box A^{\alpha} = 0$	$\Box h^{\alpha\beta} = 0$
Vacuum solution	$A^{\alpha} = A^{\alpha}_{(0)}\, \mathrm{e}^{\mathrm{i} k_{\mu} x^{\mu}}$	$h_{\alpha\beta} = A_{\alpha\beta}\, \mathrm{e}^{\mathrm{i} k_{\mu} x^{\mu}}$
Solvability condition	$k_{\mu} k^{\mu} = 0$	$k_{\mu} k^{\mu} = 0$
FT gauge conditions	$k_{\alpha} A^{\alpha} = 0$	$k_{\alpha} h^{\alpha\beta} = 0$
Dipole moment	$\mathbf{p}_{\mathrm{E}} = \sum_a q_a \mathbf{r}_a$	$\mathbf{p}_{\mathrm{G}} = \sum_a m_a \mathbf{r}_a$
Dipole radiation intensity	$I \propto \ddot{\mathbf{p}}^2_{\mathrm{E}}$	$I \propto \ddot{\mathbf{p}}^2_{\mathrm{G}} = 0$

3.5.3 Indirect Observations of Gravitational Waves

Binary Neutron Stars

In 1974, the binary neutron star system PSR 1913+16, in which one component is a pulsar with a pulse period of 0.059 s, was discovered using the Arecibo radio telescope, then the largest in the world. The dimensions of both components and the entire system are so small that the system serves as a nearly ideal relativistic laboratory provided by nature. It consists of two neutron stars in close proximity, resulting in substantial curvature of space and time to which the components respond. Moreover, the space between the components is free of scattered material that would complicate the interpretation of measured quantities. The uniqueness of the system is easily illustrated: the masses of the components are slightly higher than of the Sun (1.44 M_{S} and 1.39 M_{S}). The size of each neutron star does not exceed a few tens of kilometers. At closest approach, the components separation is a mere 700,000 km – approximately equal to the radius of the Sun. The orbital period of the binary pulsar is 7 h 45 min.

In 1993, Russell Hulse and Joseph Taylor were awarded the Nobel Prize in Physics for their research on this unique system, which exhibited numerous phenomena predicted by general relativity. For instance, its periastron advance amounts to 4° per year (the identical effect shifts Mercury's perihelion by a mere 43″ per century). Other measured phenomena include the relativistic Doppler effect, gravitational redshift, time dilation from mutual orbital motion, and light deflection.

However, the most celebrated discovery was the orbital period decay corresponding to energy loss caused by the emission of gravitational waves. During each orbit, the two components spiral closer together by approximately 3 millimeters, causing the orbital period to decrease by 76 μs/yr. Russell Hulse and Joseph Taylor thus indirectly detected gravitational waves for the first time in history. This constituted a measurement of the consequences of gravitational wave emission rather than a direct detection.

Subsequently, binary neutron stars with even more favorable properties were discovered. For example, the binary pulsar PSR 1534+12 or the binary pulsar PSR J0737-3039A (discovered by the Lovell Telescope at Jodrell Bank in 2003, exhibiting a periastron advance of 17° per year) serve as exceptional relativistic laboratories for testing general relativity effects and, naturally, for the indirect detection of gravitational waves.

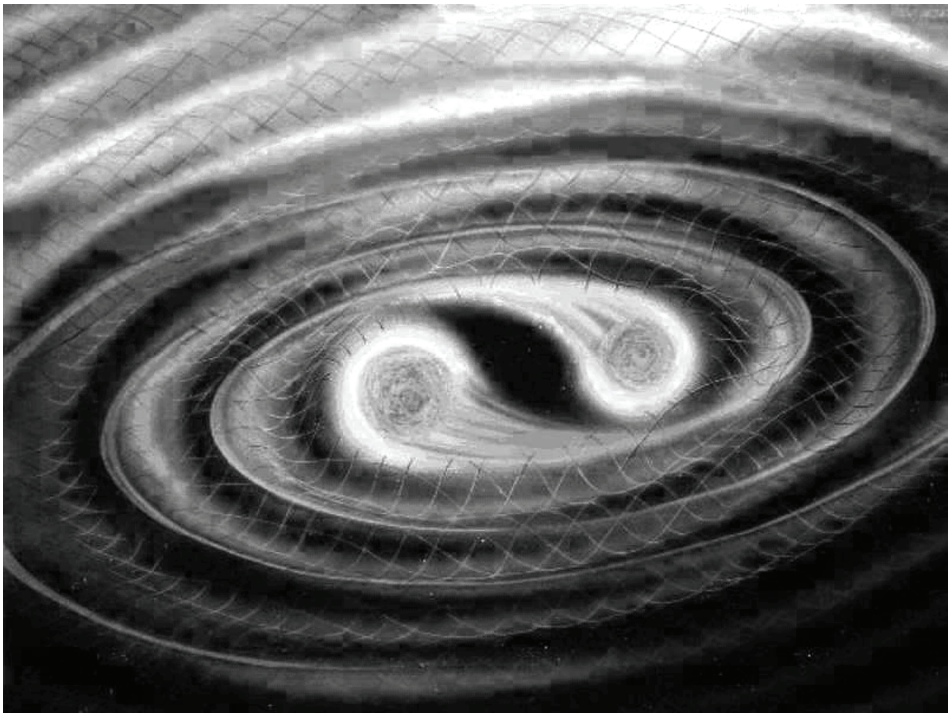


Fig. 3.31: Gravitational waves generated by a pair of mutually orbiting stars. The coordinate grid is merely a two-dimensional projection of the four-dimensional spacetime continuum.

3.5.4 Detection of Gravitational Waves

Direct detection of gravitational waves is far more challenging than that of electromagnetic waves. In an electromagnetic wave, oppositely charged particles oscillate in opposite directions, whereas a gravitational wave moves two nearby particles identically, requiring the detection of higher-order effects, such as the relative motion of two objects with respect to each other. Direct detection methods include resonant detectors, interferometric detectors, or monitoring the phase delay of a reference signal. In the future, quantum detectors based on alterations of the wave function of a microscopic object will certainly emerge. However, such detectors will operate only at extremely short wavelengths and will not be suitable for astrophysical sources of gravitational waves.

Resonant Detectors

The earliest well-known attempts to detect gravitational waves were conducted by Joseph Weber, who utilized two large aluminum cylinders. One was located at the University of Maryland near Washington, D.C., and the other at Argonne National Laboratory near Chicago. The cylinders were separated by approximately 1,000 km so that any potential gravitational wave detection could be confirmed from an independent site rather than being a local disturbance. They acted as natural oscillators tuned to a frequency of 1,660 Hz (the expected frequency of gravitational waves from binary sources lies in the kilohertz range), with a mass of 1.4 metric tons, a diameter of 66 cm, and a length of 153 cm. Each cylinder was suspended in vacuum by a metallic wire and mechanically isolated from its surroundings. Near the center, a piezoelectric sensor (visible in the photograph) was mounted and connected to electronic circuits sensitive to the fundamental oscillation frequency of the cylinder. Upon the passage of a gravitational wave, the cylinder was expected to resonate at its natural frequency.

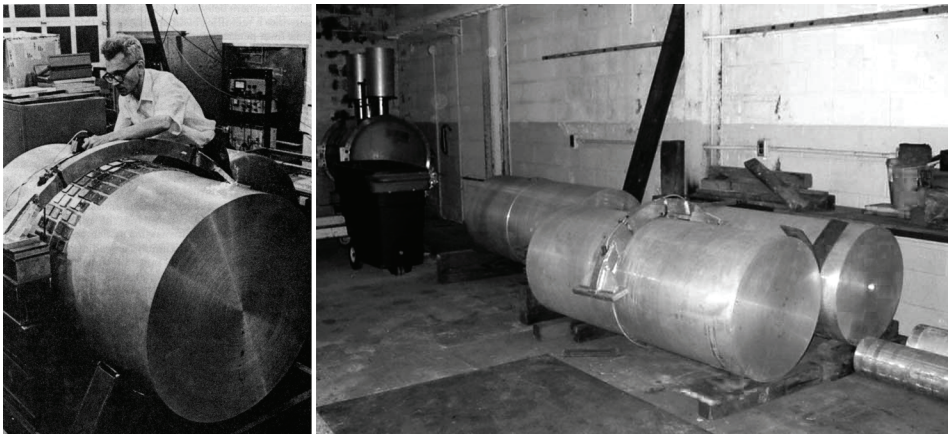


Fig. 3.32: Joseph Weber beside one of the resonant cylinders (left). Resonant cylinders being prepared for experiments (right).

The cylinders were commissioned in 1966, and in 1972 a single coincidence was recorded that was never repeated. Today, it is understood that the relative sensitivity $h \equiv \Delta l/l \sim 10^{-15}$ of this apparatus was insufficient to detect gravitational waves from typical sources, where $h \sim 10^{-23}$ is expected. The Weber bars suffered from an additional drawback – they operated at a single frequency and therefore could capture neither the spectrum of a gravitational wave nor the time evolution of its frequency. Joseph Weber believed he had discovered gravitational waves and published a paper on the findings, which drew substantial criticism from fellow scientists, particularly regarding errors in statistical analysis.

Notably, in the late 1940s (long before initiating his search for gravitational waves), Joseph Weber offered George Gamow assistance in searching for the relic radiation from the end of the Big Bang. For unknown reasons, Gamow declined the offer, even though Weber was engaged in microwave and radio wave detection at that time.

A successor to the Weber detector was the Allegro detector, operated by Louisiana State University in Baton Rouge from the early 1990s until 2008. Its aluminum cylinder had a length of 3 m, a mass of 2,300 kg, and a resonant frequency of 904 Hz.

Interferometric Detectors

Most contemporary systems for the direct detection of gravitational waves rely on laser interferometry. A laser beam is split by a semi-transparent mirror into two perpendicular arms terminated by precision-polished mirrors serving as test masses. The reflected beams return through the arms via a resonant cavity, interfere, and variations in the interference pattern are recorded electronically. The sensitivity of these instruments depends on the arm length. Because contraction occurs along one direction while elongation occurs along the other, the resulting sensitivity is substantial.

The largest system of this kind in the world is LIGO (*Laser Interferometer Gravitational-Wave Observatory*), a project established through a collaboration between Caltech and MIT. A pair of four-kilometer detectors was constructed 3,000 km apart. The first is located in Hanford, Washington, with its construction overseen by Caltech. The second detector is in Livingston, Louisiana, constructed under the auspices of MIT. Hanford also houses an additional two-kilometer detector dedicated to various tests. Two interferometers are built so that any potential gravitational wave detection can be confirmed via coincidence from two independent sites.

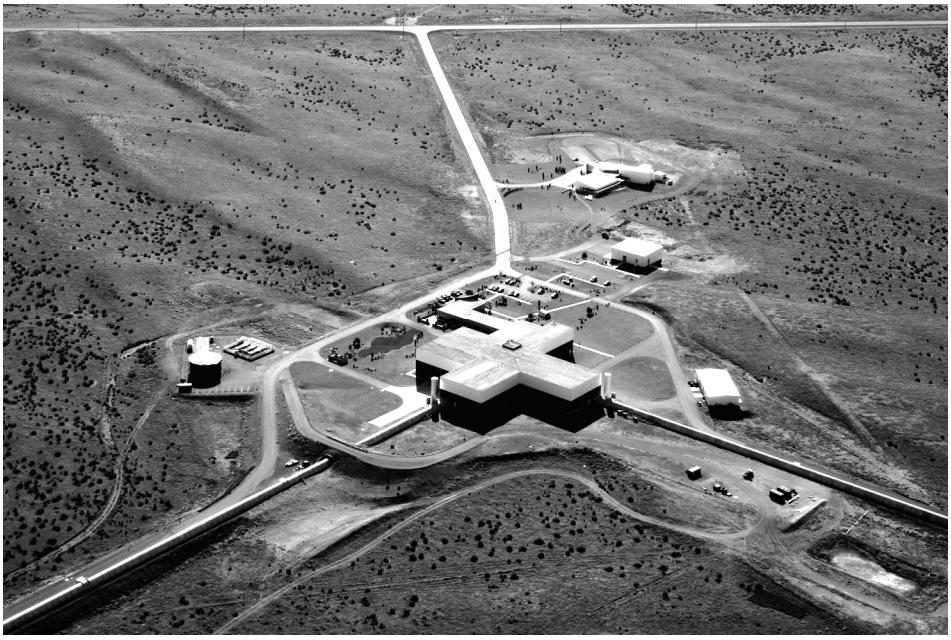


Fig. 3.33: LIGO – aerial view of the central section of the LIGO detector (Hanford, 2008).

The interferometer was commissioned in 2002 and extensively upgraded between 2010 and 2015. The current instrument parameters are: each arm has a length of 4 km, a diameter of 60 cm, and an operating pressure of 10^{-9} torr. The volume of the vacuum system is 9,000 m³ (making it the largest vacuum system in the world). The beam source, operating at a wavelength of 1.06 μm, is an Nd:YAG laser capable of reaching a power of up to 200 W. The positional measurement precision of the end test mirrors suspended on fused-silica fibers is 10^{-18} m, with each mirror having a mass of 40 kg. The sensitive frequency range of the instrument spans 10 Hz to 10 kHz, corresponding

to most expected sources of gravitational waves. The relative precision of length change measurements is $h \sim 10^{-22}$, and better than 10^{-23} in the hundreds-of-hertz band – eight orders of magnitude higher than that of the Joseph Weber cylinders.

Immediately after commissioning the upgraded detector, a gravitational wave signal from a binary black hole merger was detected on September 14, 2015. Prior to merging, the black holes had masses of approximately 29 M_S and 36 M_S ($\pm 4\ M_S$). The newly formed black hole is less massive than their sum, as roughly three solar masses were converted into gravitational waves. The signal arrived from a distance of 1.3 billion light-years. Directional localization is not highly precise, yielding only an estimated region on the southern sky near the Magellanic Clouds (though the clouds themselves are, of course, much closer). The entire pulse lasted less than three-tenths of a second. The signals from Hanford and Livingston exhibited similar waveforms, with a time delay of 7 ms between the two observatories. As the two components spiraled closer together, the frequency increased from tens of hertz to approximately 300 Hz at the moment of merger. Agreement between the signal and theoretical predictions from numerical simulations was also clearly evident.

For the first time, gravitational waves were directly detected, a binary black hole merger was observed, and intermediate-mass black holes were identified. Alongside the electromagnetic spectrum, neutrinos, and cosmic rays, gravitational waves represent the fourth window for observing phenomena in the universe.

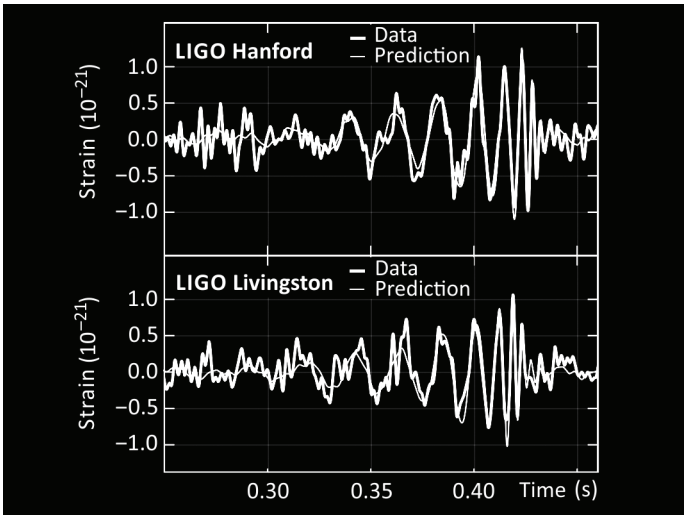


Fig. 3.34: LIGO – signal detected on September 14, 2015. At the moment of black hole merger, signal amplitude and frequency reach their maximum, decaying thereafter. The peak relative strain was 10^{-21}, occurring at the merger of the two black holes.

In Europe, the comparable Virgo detector is located near Cascina, Italy, 10 km from Pisa, famous for its leaning tower. Virgo features 3 km arms and underwent an upgrade similar to LIGO between 2010 and 2017, achieving comparable parameters. Virgo joined LIGO for a 25-day joint observing run in August 2017, capturing two signals. The first originated from another distant binary black hole merger. The second, on August 17, 2017, stemmed from a binary neutron star merger – a process also generating

intense broadband electromagnetic emission, from a gamma-ray burst caused by heavy-element decay in shock waves to radio emission from relativistic jets interacting with the interstellar medium. The resulting *kilonova* (a thousand times brighter than a nova) was observed simultaneously in both gravitational and electromagnetic windows for the first time. LIGO and Virgo are complemented by Japan's 3 km underground KAGRA (*KAmioka GRAvitational wave detector*) near the Super-Kamiokande neutrino detector. This fully cryogenic system (with cooled beam tubes and end mirrors) began operating in 2020 and joined joint observing runs in 2023.

By 2026, 390 gravitational-wave events had been detected, establishing their detection as a routine astronomical method for observing the universe.



Fig. 3.35: One of the arms of the European Virgo detector.

3.5.5 Future Detectors

Interferometric detectors have achieved one success after another and face a promising future. In India, another LIGO detector is actively under construction to provide observations from the Asian region. In Europe, plans are underway for an underground ten-kilometer detector named the Einstein Telescope, with candidate sites including the tripoint border of Germany, the Netherlands, and Belgium, Sardinia in Italy, and Lusatia in Germany, though the exact location has not yet been selected. Construction is scheduled to begin around 2028, with commissioning expected after 2035. Placing detectors underground reduces seismic noise, while cryogenic cooling suppresses thermal noise. Detectors of this class exhibit peak sensitivity in the hundreds of hertz and can observe gravitational waves from various binary systems. To detect relic gravitational waves – thought to have originated during the inflationary epoch or the very birth of the universe – detectors sensitive to much longer wavelengths, and thus with substantially longer arms, must be constructed. Such instruments can only be built in space. Another viable approach involves monitoring phase shifts in pulsar signals using radio telescope arrays.

LISA

Constructing a giant interferometer in space has been envisioned since the late 20th century. The project of three spacecraft shining laser beams at one another across millions of kilometers to track the position of a reflective cube – the core test mass oscillating with passing gravitational waves – was named LISA (*Laser Interferometer Space Antenna*). Originally slated for launch in 2016, the joint venture between NASA and the European Space Agency (ESA) faced numerous financial and technical hurdles. Following NASA's withdrawal in 2011 due to budget constraints, the project underwent several renamings, arm-length adjustments, and launch delays, appearing nearly doomed. Renewed confidence came from the European LISA Pathfinder probe, launched in December 2015. Serving as a technological precursor, LISA Pathfinder confirmed the viability of the core concept, demonstrating both ultra-precise spacecraft maneuvering and accurate tracking of the test cube's position. Following the landmark discovery of gravitational waves, momentum accelerated, and the giant LISA interferometer was formally approved in 2017 (and officially adopted for construction in 2024), with launch scheduled for 2035.

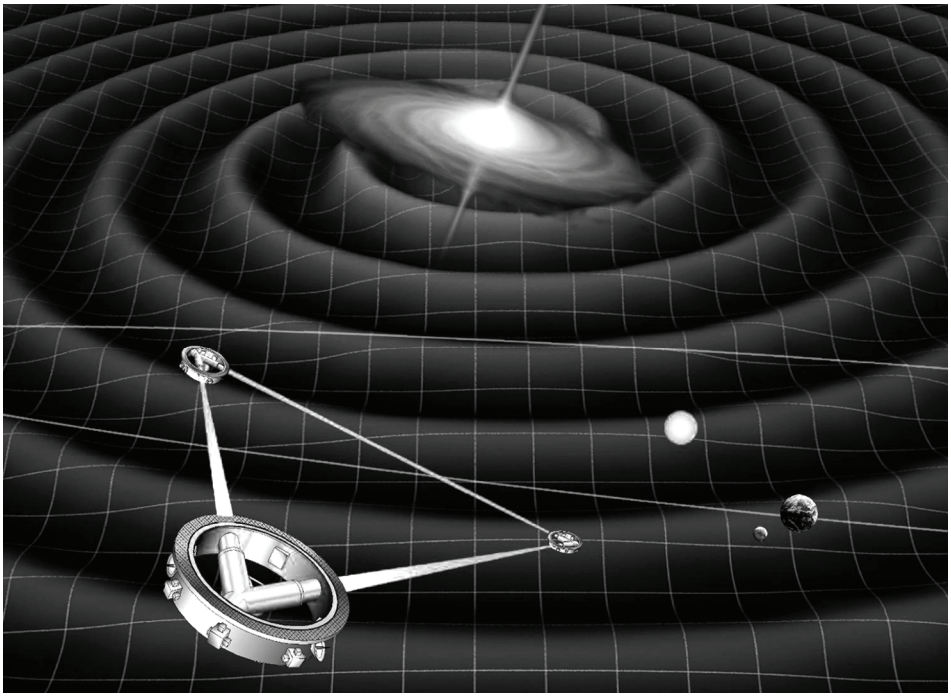


Fig. 3.36: Ambitious concept of the LISA space interferometer.

What is the current design of the giant LISA interferometer? Three identical spacecraft will fly in a triangular formation in an independent heliocentric orbit, trailing the Earth by 20°. They will exchange infrared laser beams across distances of 2.5 million kilometers. The laser light will reflect off free-floating cubes made of a gold-platinum alloy, each measuring 46 millimeters on a side and having a mass of 2 kilograms. Each spacecraft will house two such cubes flying freely through space and responding to gravitational waves. Surrounding each test cube is an electrode housing monitoring its position

relative to the enclosure. Ideally, the cube would follow its geodesic path alongside the spacecraft and housing, remaining centered inside. However, external non-gravitational forces act on the spacecraft, including solar radiation pressure, solar wind pressure, and asymmetric thermal radiation from the craft itself. Whenever a cube approaches the housing walls – separated by a mere two millimeters – the spacecraft must maneuver to keep the cube centered. Coarse maneuvering will be handled by cold-gas micro-thrusters, while the finest navigation will rely on colloidal micro-thrusters ejecting charged droplets accelerated by an electric field, whose recoil gently nudges the spacecraft in the desired direction.

Additional instrumentation includes an optical bench and two 30-centimeter telescopes per spacecraft to receive laser beams from the neighboring craft. Out of the 2 watts of laser power transmitted, a mere 700 picowatts – or 35 billionths of a percent – will reach the test cube of an adjacent spacecraft. LISA will be capable of detecting gravitational waves from stars inspiraling into supermassive black holes, white dwarf binaries, the galactic binary background noise, and mergers of supermassive black hole binaries. LISA will operate near the sensitivity threshold required to detect relic gravitational waves; with some fortune, such a detection might succeed, though cautious skepticism remains appropriate so that any positive result comes as a welcome surprise.

If LISA performs as intended, 12 identical spacecraft exchanging laser beams will be launched into heliocentric orbit in subsequent decades under the BBO (*Big Bang Observer*) project. This time, inter-spacecraft separations will be smaller – "only" 50,000 kilometers – with the requisite sensitivity achieved through the larger constellation size. Signal noise is projected to be just one percent of LISA's noise floor. BBO will consist of four triangular formations: one orbiting the Sun in the ecliptic plane (the plane of the Earth's orbit), a second inclined perpendicular to the ecliptic, and the remaining two combined into a six-pointed star (hexagram) configuration. This geometry will enable the detection of stochastic backgrounds expected from relic gravitational waves – BBO's primary scientific target – with sufficient angular resolution, sensitivity, and wavelength coverage to capture them. Thus, the future of interferometric detectors is mapped out decades ahead, promising remarkable discoveries.

Radio Telescope Detectors

Observing extremely long wavelengths requires instruments of immense dimensions. Modern civilization does not yet possess the resources to construct a detector the size of the Solar System or even the entire Galaxy. Nevertheless, one viable possibility presents itself. Our Galaxy contains numerous pulsars – neutron stars whose rotation axis does not coincide with their magnetic field axis, emitting highly regular pulses via the lighthouse effect. They serve as nearly ideal clocks distributed across the entire Galaxy. Because pulsar emission falls predominantly within the radio band, it can be observed with radio telescopes. Giant radio telescope arrays possess exceptional spatial and temporal resolution, enabling the detection of minute phase shifts in pulsar signals arriving from different regions of the Galaxy. Such phase shifts are also induced when a signal passes through a gravitational wave. Pairs of pulsars are monitored, with the two lines of sight replacing the interferometric arms of terrestrial detectors. Statistical analysis then reconstructs information about the gravitational waves traversed by the signals.

This method is termed PTA (*Pulsar Timing Array*). The first tests began at Australia's Parkes array in 2005; Australian dishes stood at the dawn of radio astronomy and maintained communications with Apollo Moon crews. In Europe, PTA utilizes

a network led by the 100-meter Effelsberg telescope in western Germany, alongside the 76-meter Lovell Telescope at Jodrell Bank (near Manchester, England), and dishes in Westerbork, the Netherlands, and Nançay, France. In North America, PTA is pursued by NANOGrav (*North American Nanohertz Observatory for Gravitational Waves*), using instruments like the steerable 100×110-meter Green Bank Telescope; it also formerly included the famous Arecibo Observatory, irreparably damaged by its 2020 platform collapse. Soon, the SKA (*Square Kilometre Array*), under construction in Australia and South Africa with a one-square-kilometer collecting area, will join the search – the only contemporary instrument capable of detecting relic gravitational waves, holding strong promise for success.

In June 2023, the statistical observation of a gravitational-wave background was announced by PTA collaborations in North America, Europe, and Australia – the culmination of a quarter-century of data collection. In the future, isolating signals from individual sources should become feasible. For further details, see [63].

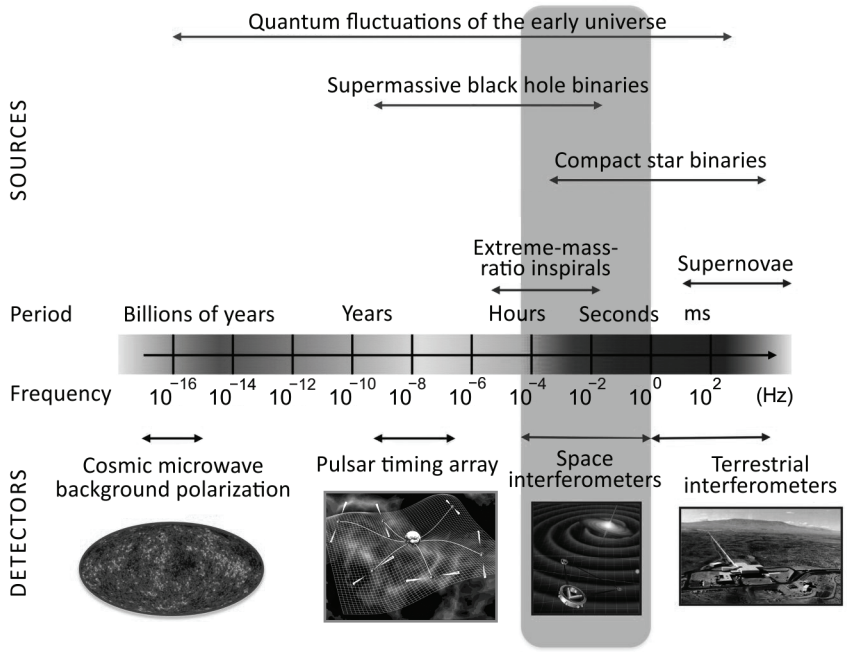


Fig. 3.37: Various detectors and sources of gravitational waves as a function of frequency. Another indirect method of searching for gravitational waves detects polarization changes in the cosmic microwave background induced by passing relic gravitational waves (far left).

3.6 Spherical Source of Gravitation

The first complete solution of the new Einstein field equations described the curvature of spacetime around a perfectly symmetric spherical body. This solution was found by the German mathematician and physicist Karl Schwarzschild in 1916, the very year general relativity was published (Einstein had already presented it orally in 1915). The solution applies to an external observer located at a fixed distance from the object where the matter distribution vanishes – i.e., empty space where the stress-energy tensor is zero, leaving zero on the right-hand side of the Einstein law of gravitation.

3.6.1 Schwarzschild Solution

Metric Around a Spherically Symmetric Body

The figure shows the worldline of the center parameterized by its proper time τ. For $\tau = \text{const}$, a spatial hypersurface is depicted, representing the three-dimensional world surrounding the center. The true radial coordinate is denoted by ρ and increases along the drawn coordinate lines on which the angular variables θ and φ remain constant.

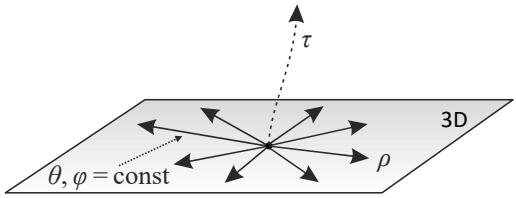


Fig. 3.38: Choice of coordinates.

Assume a diagonal metric of the form:

$$\mathrm{d}s^2 = -c^2\alpha(\tau,\rho)\,\mathrm{d}\tau^2 + \beta(\tau,\rho)\,\mathrm{d}\rho^2 + \gamma(\tau,\rho)\left[\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right]. \qquad (3.299)$$

The radial coordinate ρ is defined such that for constant τ, θ, and φ, the following holds:

$$\mathrm{d}s^2 = \mathrm{d}\rho^2\,. \qquad (3.300)$$

Consequently, $\beta = 1$, and the most general form of the interval in spherical geometry contains two unknown functions α and γ:

$$\mathrm{d}s^2 = -c^2\alpha(\tau,\rho)\,\mathrm{d}\tau^2 + \mathrm{d}\rho^2 + \gamma(\tau,\rho)\left[\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right]. \qquad (3.301)$$

Inside the square brackets are the standard angular terms adopted from spherical coordinates. While radial distances can be measured using the radial coordinate ρ, the surface area of a sphere will not equal $4\pi\rho^2$. Instead of the coordinates $(\tau, \rho, \theta, \varphi)$, a new coordi-

nate grid (t, r, θ, φ) is introduced, where t is the standard coordinate time and r is a spherical coordinate chosen so that the surface area of a sphere continues to satisfy:

$$S = 4\pi r^2 \neq 4\pi\rho^2 . \tag{3.302}$$

Thus, ρ represents the true radial distance, whereas r is the circumference-derived radial coordinate entering the angular element, which in the new coordinates takes the form:

$$\mathrm{d}s^2 = -c^2 A(t,r)\,\mathrm{d}t^2 + B(t,r)\mathrm{d}r^2 + r^2\left[\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right]. \tag{3.303}$$

In these coordinates, the interval contains two unknown functions A and B to be determined from the Einstein field equations. The signs are chosen to match the Minkowski metric, to which the metric around the body must reduce at large distances (Sylvester law of inertia for quadratic forms). The sought metric therefore has the form:

$$g_{\alpha\beta} = \begin{pmatrix} -c^2 A(t,r) & 0 & 0 & 0 \\ 0 & B(t,r) & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta \end{pmatrix}. \tag{3.304}$$

The subsequent procedure is lengthy yet relatively straightforward. First, the inverse metric is determined:

$$g^{\alpha\beta} = \begin{pmatrix} -c^{-2} A^{-1}(t,r) & 0 & 0 & 0 \\ 0 & B^{-1}(t,r) & 0 & 0 \\ 0 & 0 & r^{-2} & 0 \\ 0 & 0 & 0 & r^{-2}\sin^{-2}\theta \end{pmatrix}. \tag{3.305}$$

Next, the Christoffel symbols are obtained from relation (3.124), the Riemann curvature tensor from (3.207), the Ricci tensor and scalar curvature via contraction, and finally the Einstein field equations are assembled:

$$R^{\alpha\beta} - \frac{1}{2} R g^{\alpha\beta} = 0 . \tag{3.306}$$

The right-hand side vanishes outside the central body. At first glance, this appears to represent 16 differential equations for the metric; however, due to symmetry, there are only 10 equations. Several of these are identities of the form $0 = 0$, leaving only 4 nontrivial equations and only two independent ones. Analysis of these equations reveals that the functions A and B depend solely on the coordinate r and not on t. Every solution outside a spherically symmetric body is therefore time-independent (Birkhoff theorem). Among other consequences, this implies that a spherically symmetric body cannot generate gravitational waves. For a single dependent variable, the gravity law reduces to ordinary differential equations whose solution takes the simple form:

$$A(r) = c_1 + \frac{c_2}{r}; \qquad B(r) = \frac{1}{A(r)}., \tag{3.307}$$

where c_1 and c_2 are integration constants determined from the Newtonian limit – far from the body. The following must hold:

$$g_{00} = -c^2\left(1+\frac{2\phi}{c^2}\right) = -c^2\left(1-\frac{2GM}{c^2 r}\right). \tag{3.308}$$

From this follows:

$$c_1 = 1; \qquad c_2 = -\frac{2GM}{c^2}. \tag{3.309}$$

The result is therefore:

$$\mathrm{d}s^2 = -c^2\left(1-\frac{2GM}{c^2 r}\right)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{\left(1-\frac{2GM}{c^2 r}\right)} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right). \tag{3.310}$$

This metric was derived by Karl Schwarzschild in 1916 and can be written compactly as:

► $$\mathrm{d}s^2 = -c^2\left(1-r_\mathrm{g}/r\right)\mathrm{d}t^2 + \frac{\mathrm{d}r^2}{\left(1-r_\mathrm{g}/r\right)} + r^2\mathrm{d}\Omega^2; \qquad r_\mathrm{g} \equiv \frac{2GM}{c^2}; \qquad \mathrm{d}\Omega^2 \equiv \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2. \tag{3.311}$$

The metric holds outside a spherically symmetric body for a stationary observer at a constant distance from the body. Schwarzschild derived it during World War I while serving on the Eastern Front. He contracted a severe autoimmune disease in Russia and died shortly after returning to Germany. A lunar crater is named in his honor.

Selected Properties of the Schwarzschild Metric

Taking the limit $r \to \infty$ reduces the Schwarzschild metric to the Minkowski metric, as spacetime becomes flat sufficiently far from the central body.

Radial Distance

From the metric, a simple relation holds between the areal radial coordinate r (determined from the surface area of a sphere via $S = 4\pi r^2$) and the true radial distance ρ:

$$\mathrm{d}\rho = \frac{\mathrm{d}r}{\left(1-r_g/r\right)^{1/2}}. \tag{3.312}$$

The distance determined by direct measurement does not equal the radius calculated from the surface area of the corresponding sphere due to the spatial curvature surrounding the central body.

Metric Divergence

In the limit $r \to r_\mathrm{g}$, the metric coefficient associated with the radial coordinate diverges. For ordinary bodies, this situation cannot occur. The quantity r_g, termed the Schwarzschild radius, equals 3 km for the Sun and only 9 mm for the Earth. To observe this divergence, the entire mass of the Sun would have to be compressed into a sphere of radius 3 km, and that of the Earth into a region of radius 9 mm. Such an object is termed a black hole, and as will be shown later, nothing – not even light – can escape from the region inside its Schwarzschild radius. Black holes represent the final evolu-

tionary stage of very massive stars (e.g., five or ten times the mass of the Sun). Supermassive black holes (with masses ranging from millions to billions of solar masses) are also found at the centers of most galaxies. The divergence of the metric is not a genuine physical singularity, but rather an artifact of the chosen coordinate system; adopting alternative coordinates (covering space with different coordinate hypersurfaces) removes the divergence. This can be demonstrated using *isotropic coordinates* by transforming from (t, r, θ, φ) to (t, R, θ, φ) via:

$$r = R\left(1+\frac{r_g}{4R}\right)^2; \tag{3.313}$$

$$\mathrm{d}r = \left(1+\frac{r_g}{4R}\right)\left(1-\frac{r_g}{4R}\right)\mathrm{d}R\,. \tag{3.314}$$

Differentiating the first expression directly confirms the relation for the differential. From the definition of the coordinate R, it also follows that:

$$r = r_\mathrm{g} \qquad \Leftrightarrow \qquad R = \frac{r_\mathrm{g}}{4}\,. \tag{3.315}$$

Substituting for r and $\mathrm{d}r$ into the Schwarzschild metric yields:

► $$\mathrm{d}s^2 = -c^2\frac{\left(1-r_g/4R\right)^2}{\left(1+r_g/4R\right)^2}\mathrm{d}t^2 + \left(1+r_g/4R\right)^4\left[\mathrm{d}R^2 + R^2\mathrm{d}\Omega^2\right]. \tag{3.316}$$

At first glance, no divergence occurs at $R = r_\mathrm{g}/4$ in these coordinates; this is therefore referred to as a removable (coordinate) singularity. An observer falling into a black hole would notice nothing unusual at the Schwarzschild radius itself (aside from potential tidal forces), yet after crossing it, no return path exists. From the perspective of an external observer, this boundary acts as a horizon beyond which nothing can be seen. Isotropic coordinates were introduced by the German mathematician Hermann Weyl in 1917. Transformations to other coordinate systems also exhibit intriguing properties; for instance, the metric remains completely invariant under the transformation:

$$R \to \tilde{R} \equiv r_g^2/16R\,. \tag{3.317}$$

Coordinate Swap

If a body were confined within a region smaller than its Schwarzschild radius, the metric coefficient of the radial coordinate would become negative and that of the temporal coordinate positive – space and time effectively swap roles. This arises because time must be a quantity flowing in a single direction (a monotonic function). Inside the Schwarzschild radius, however, the sole monotonic quantity is the radial distance, which continuously decreases; thus, the radial coordinate assumes the role of time. Nevertheless, an external observer receives no information about events inside the Schwarzschild radius, and an observer inside cannot transmit any signal to the outside. They represent two causally disconnected worlds separated by the horizon. The interior solution is no longer static, containing r as a timelike coordinate.

Laplace Solution

The earliest known considerations regarding the possible existence of objects from which not even light could escape likely originate from the English natural philosopher John Michell as early as 1783, who referred to them as *dark stars*. In 1798, the French mathematician and physicist Pierre-Simon Laplace derived the dimensions of such a hypothetical body based on Newton mechanics using the escape velocity formula:

$$v = \sqrt{2GM/r}\,, \tag{3.318}$$

Substituting the speed of light c for the escape velocity yields the radius such a body would have to possess:

$$r = 2GM/c^2 = r_\mathrm{g}\,, \tag{3.319}$$

which is the relation obtained by Karl Schwarzschild from the general relativity.

Derivation Using an LIF

The Schwarzschild metric can also be derived (albeit not entirely rigorously) by considering a local inertial frame (LIF) falling toward the central body. Consider a coordinate system S of an observer stationary relative to the central body – clearly a non-inertial frame. Imagine another frame, an LIF falling freely from infinity toward the central body. Its instantaneous speed relative to the central body at distance r equals:

$$v = \sqrt{2GM/r}\,, \tag{3.320}$$

Within the falling LIF, the special relativity hold, and the interval equals (θ, φ = const):

$$\mathrm{d}s^2 = -c^2\mathrm{d}t_\mathrm{LIF}^2 + \mathrm{d}r_\mathrm{LIF}^2\,. \tag{3.321}$$

Observing rods and clocks in frame S from the LIF reveals length contraction and time dilation; conversely, from frame S, rods in the LIF appear contracted and time intervals measured by clocks in the LIF appear dilated:

$$\mathrm{d}r = \sqrt{1-v^2/c^2}\,\mathrm{d}r_\mathrm{LIF};\tag{3.322}$$

$$\mathrm{d}t = \frac{\mathrm{d}t_\mathrm{LIF}}{\sqrt{1-v^2/c^2}}\,. \tag{3.323}$$

Substituting the velocity from (3.320), determining $\mathrm{d}r_\mathrm{LIF}$ and $\mathrm{d}t_\mathrm{LIF}$, and inserting them into interval (3.321) yields:

$$\mathrm{d}s^2 = -c^2\left(1-\frac{2GM}{c^2 r}\right)\mathrm{d}t^2 + \mathrm{d}r^2\Big/\left(1-\frac{2GM}{c^2 r}\right), \tag{3.324}$$

which is the Schwarzschild metric for constant angular coordinates.

Radial Propagation of Light

Suppose that at a distance r_0 from the central body and at time t_0, photons are emitted radially (one outward and one inward toward the central body):



Fig. 3.39: Radial propagation of light.

The photon equation of motion $ds^2 = 0$ for radial motion ($\varphi = \theta = \text{const}$) in the Schwarzschild metric yields:

$$\begin{aligned} -c^2(1-r_g/r)\,dt^2 + \frac{dr^2}{(1-r_g/r)} = 0 \quad &\Rightarrow \\ \frac{dr^2}{(1-r_g/r)^2} = c^2 dt^2 \quad &\Rightarrow \\ \frac{dr}{(1-r_g/r)} = \pm c\,dt \quad &\Rightarrow \\ \int_{r_0}^{r} \frac{dr}{(1-r_g/r)} = \pm \int_{t_0}^{t} c\,dt\,. & \end{aligned} \tag{3.325}$$

Integrating this equation (by multiplying numerator and denominator by r and decomposing into partial fractions, i.e., subtracting and adding r_g) gives:

$$r - r_0 + r_g \ln\frac{r-r_g}{r_0-r_g} = \pm c(t-t_0) \tag{3.326}$$

or equivalently:

$$r = r_0 \pm c(t-t_0) - r_g \ln\frac{r-r_g}{r_0-r_g}\,. \tag{3.327}$$

The signs correspond to photons traveling outward (+) or inward (−) toward the black hole. The logarithmic term arises from spacetime curvature. Far from the black hole (r, $r_0 \to \infty$), the logarithm on the right-hand side vanishes, recovering the special-relativistic result. The coordinate travel time of a photon from r_0 to r is:

$$\Delta t \equiv t - t_0 = \pm\frac{1}{c}\left(r - r_0 + r_g \ln\frac{r-r_g}{r_0-r_g} \right). \tag{3.328}$$

The upper sign applies to an outgoing photon and the lower to an ingoing one. Consider now the travel time of a photon reaching the Schwarzschild radius ($r_0 \to r_g$). With the negative sign, the logarithm diverges to $-\infty$ and the coordinate time interval becomes infinite. To an external observer, the infall of photons or any material body takes infinite coordinate time – time dilation pushed to its extreme. Conversely, any falling object crosses the horizon in a finite proper time in its own frame.

Now examine the reverse scenario. For a photon emitted from the event horizon of a black hole, $r_0 \to r_g$, the positive sign applies, the logarithm diverges to $+\infty$, and the travel time is again infinite – even for any finite destination radius r. Escaping from the

horizon, even by an infinitesimal distance, would take a photon an infinite amount of time. In other words, no photon can escape from the horizon or the region beneath it. Provided neither the initial nor the final point lies on the horizon (at the Schwarzschild radius), the photon travel time remains finite.

3.6.2 Embedding Diagrams

Visualizing a four-dimensional curved world directly is extremely difficult. Therefore, a mathematical technique exists that maps this world onto a two-dimensional curved surface, known as an embedding diagram. In the first step, two of the four coordinates are set to constants, leaving only differentials and functions of the remaining two coordinates in the interval. The resulting interval is interpreted as the squared distance on a two-dimensional surface embedded in three-dimensional Euclidean space. This procedure is demonstrated in the Schwarzschild geometry, where the temporal axis ($t = t_0$ and the polar angle ($\theta = \pi/2$, restricting attention to the equatorial plane) are held constant:

$$\mathrm{d}s^2 = \frac{\mathrm{d}r^2}{\left(1 - r_\mathrm{g}/r\right)} + r^2 \mathrm{d}\varphi^2 \,. \tag{3.329}$$

In principle, the temporal axis and one spatial coordinate could also be retained (for instance, in the case of gravitational waves, where temporal oscillation is essential). Writing the general equation of a two-dimensional surface in Cartesian space gives:

$$z = f(x, y) \,. \tag{3.330}$$

In the case of a spherically symmetric body, only surfaces of revolution $z = f(r)$ are of interest. Utilizing standard polar or cylindrical coordinates, the line element on the curved surface in Cartesian coordinates satisfies:

$$\begin{aligned} \mathrm{d}s^2 &= \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \,, \\ \mathrm{d}s^2 &= \mathrm{d}x^2 + \mathrm{d}y^2 + f'^2\, \mathrm{d}r^2 \,, \\ \mathrm{d}s^2 &= \mathrm{d}r^2 + r^2 \mathrm{d}\varphi^2 + f'^2\, \mathrm{d}r^2 \,, \\ \mathrm{d}s^2 &= \left(1 + f'^2\right) \mathrm{d}r^2 + r^2 \mathrm{d}\varphi^2 \,, \end{aligned} \tag{3.331}$$

where the prime denotes differentiation with respect to r. Relations (3.330) and (3.331) are unrelated to the curved spacetime of general relativity; they merely express the line element of a surface of revolution $z = f(r)$ in 3D Euclidean space. Assuming now that (3.329) corresponds to a surface in Cartesian space (x, y, z), comparing (3.329) with (3.331) yields the condition:

$$1 + f'^2 = \frac{1}{1 - r_\mathrm{g}/r} \,. \tag{3.332}$$

From this, it readily follows that:

$$f' = \sqrt{\frac{r_\mathrm{g}}{r - r_\mathrm{g}}} \,. \tag{3.333}$$

Integration yields:

$$f(r) = \int \sqrt{\frac{r_\mathrm{g}}{r - r_\mathrm{g}}}\,\mathrm{d}r\,, \tag{3.334}$$

$$f(r) = \sqrt{4r_\mathrm{g}\left(r - r_\mathrm{g}\right)}\,. \tag{3.335}$$

This is the equation of a 3D surface of revolution whose curvature matches the spatial geometry around a black hole. For $r \to \infty$, the surface properly flattens into Euclidean space, whereas for $r \to r_\mathrm{g}$, the surface terminates and cannot be extended further (it does not exist for $r < r_\mathrm{g}$). Conversely, at $r = r_\mathrm{g}$, the surface exhibits no physical singularity, contrary to what the form of the Schwarzschild metric might suggest. From isotropic coordinates, it is already known that this coordinate singularity is removable.

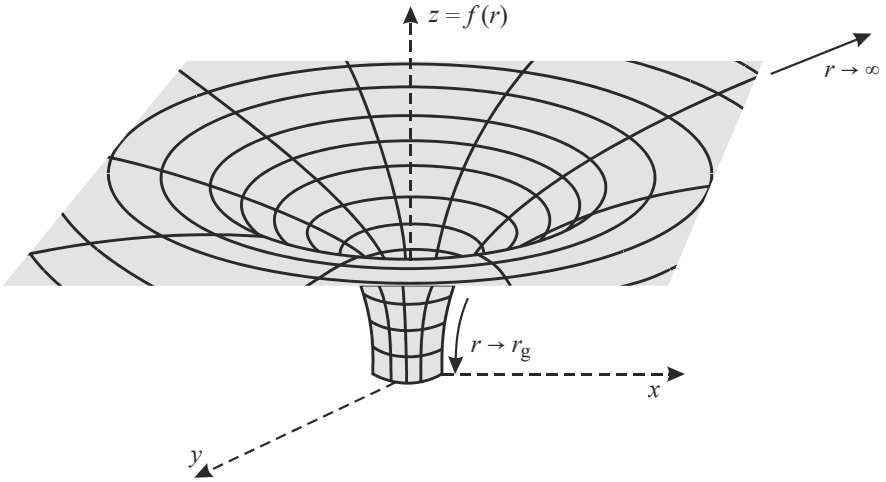


Fig. 3.40: Embedding diagram for the Schwarzschild geometry.

Einstein–Rosen Bridge

Substituting isotropic coordinates (3.313) into relation (3.335) describes the identical surface as:

$$f(r) = \sqrt{4r_\mathrm{g}R^2\left(1 - \frac{r_\mathrm{g}}{4R}\right)^2}\,. \tag{3.336}$$

Recall that $R = r_\mathrm{g}/4$ corresponds to the event horizon $r = r_\mathrm{g}$ in the original coordinates. This surface is no longer restricted to $R > r_\mathrm{g}/4$, but can be plotted across the entire domain $R \in (0, \infty)$. In both limits $R \to \infty$ and $R \to 0$, the surface asymptotically flattens into Euclidean space. Such a solution resembles a bridge connecting two flat spacetimes. It was studied in detail in 1935 by Albert Einstein and the American-Israeli physicist Nathan Rosen, and is therefore termed the *Einstein–Rosen bridge* (later referred to as a *wormhole*). In principle, an object falling into a black hole could emerge in another region of spacetime termed a *white hole*. This spacetime shortcut would enable travel between two distant regions without exceeding the speed of light.

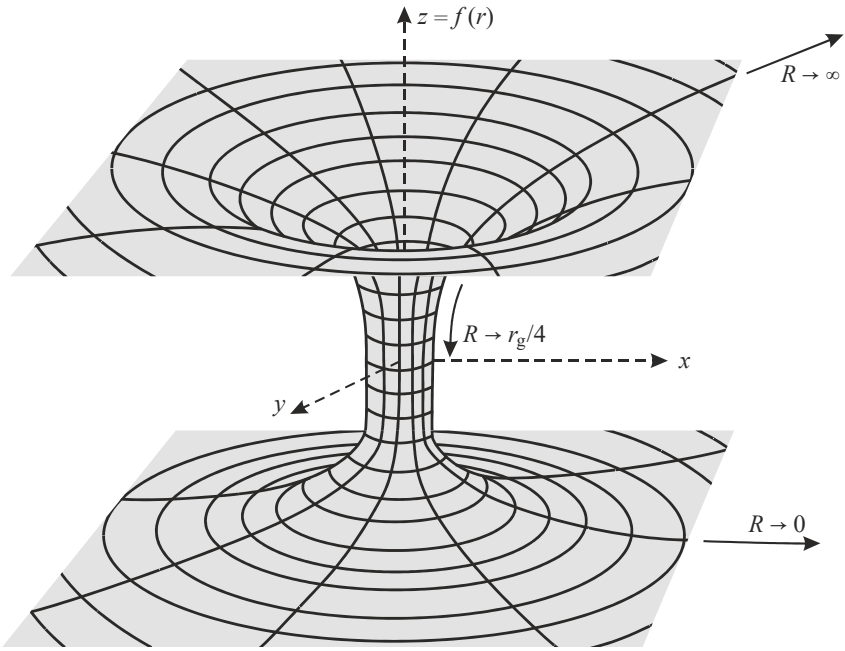


Fig. 3.41: Einstein–Rosen bridge as an embedding diagram.

However, the embedding diagram shown represents a solution at a fixed instant of time. Its dynamical evolution (see Fig. 3.43) was analyzed by American physicists Robert Fuller and John Archibald Wheeler in 1962. They demonstrated that the solution is unstable and pinches off faster than even the fastest particle – a photon – can traverse it.

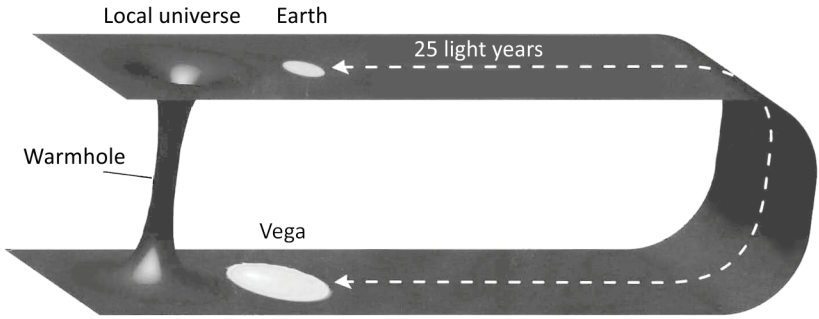


Fig. 3.42: Einstein–Rosen bridge as a spacetime shortcut.

Wormholes proved so captivating that physicists were reluctant to abandon them. Quantum processes might conceivably allow a particle to tunnel from one region to another; however, within classical general relativity, a wormhole built upon the Schwarzschild metric is non-traversable. Numerous attempts have been made with alternative metrics that yield traversable wormholes, albeit under nonphysical conditions. One example is the simple metric:

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}\rho^2 + (a^2 + \rho^2)\left[\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right], \tag{3.337}$$

where t is coordinate time, ρ is the proper radial distance, θ and φ are spherical angular coordinates, and a is a constant. Evidently, as $\rho \to \pm\infty$, the metric approaches flat Minkowski spacetime. Dynamical analysis confirms that this represents a traversable wormhole. Computing the Christoffel symbols, the Riemann curvature tensor, and assembling the Einstein field equations reveals that this metric requires a nonzero stress-energy tensor with negative energy density.

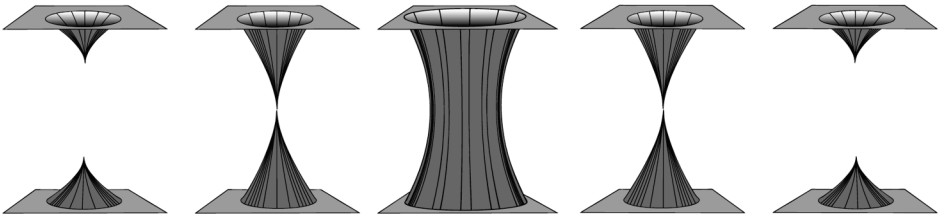


Fig. 3.43: Dynamics of the Einstein–Rosen bridge.

In 2012, a group led by American theorist Kip Thorne developed an alternative technique for visualizing curved spacetime beyond embedding diagrams, employing two families of curves termed *vortex lines* and *tendex lines*. Vortex lines depict the twisting (frame-dragging) of spacetime, whereas tendex lines illustrate its stretching and compressing. These lines vividly capture the behavior of curved spacetime.

3.6.3 Black Holes

A remarkable prediction of general relativity was the existence of black holes, now known in vast numbers. They form in the final stages of massive stars when gravitational collapse is halted neither by electron degeneracy pressure (yielding a white dwarf) nor by neutron degeneracy pressure (producing a neutron star). Stellar-mass black holes typically span 5–10 M_S, ($M_S = 2\times10^{30}$ kg is the solar mass). Supermassive black holes reside at the centers of galaxies, with masses ranging from 10^6–10^{10} M_S. The black hole at the center of our Galaxy has a mass of 4×10^6 M_S, the black hole at the center of galaxy M87 in the constellation Virgo has a mass of 6.5×10^9 M_S, and the black hole at the center of the giant Phoenix galaxy cluster is estimated to reach even 10^{11} M_S. The origin of supermassive black holes at galactic centers is not yet fully understood.

A black hole is not a dead, inactive object. Beyond the innermost stable circular orbit, objects orbit it just like any other body of equivalent mass; for instance, an entire star cluster orbits the supermassive black hole at the center of our Galaxy. Closer in, gas and dust spiral inward toward the black hole. Internal friction heats the orbiting material into a plasma state, forming an *accretion disk* that radiates intensely across the entire electromagnetic spectrum, including X-rays. Surrounding the accretion disk is a dusty gas torus. Strong magnetic fields within the accretion disk form magnetic funnels along the black hole's rotation axis (a rotating black hole is described by the Kerr metric), capturing a fraction of the infalling particles outside the Schwarzschild radius and ejecting them as two relativistic jets. These jets radiate intensely, interact with the surrounding medium, and produce characteristic radio lobes at their extremities.

The first stellar-mass black hole (Cygnus X-1) was identified by the English astronomer Paul Murdin in 1971 in the constellation Cygnus near the star γ Cyg, where an intense X-ray source had been known since 1964. It proved to be a black hole of at least 8 M_S (modern estimates are around 21 M_S) at a distance of roughly 6,000–7,000 ly. The first supermassive galactic black holes were discovered by the Hubble Space Telescope in 1992, including the black hole at the center of galaxy NGC 4261, which has a mass of approximately 400×10^6 M_S, an accretion disk mass of 10^5 M_S, jet lengths of 88,000 ly, a disk diameter of 60 au (or core region), and lies at a distance of 100×10^6 ly. The first direct image of the immediate surroundings of a black hole (at the center of galaxy M87) was captured by the *Event Horizon Telescope* (EHT) array in 2019.

Upon formation, a black hole retains information solely about its mass, angular momentum, and electric charge: M, b, Q = const. All other attributes of matter (dipole and quadrupole moments, various quantum numbers) are lost upon crossing the event horizon. This theorem was formulated by Brandon Carter, Werner Israel, David C. Robinson, and Stephen Hawking, and is commonly termed the *no-hair theorem* – meaning a black hole retains virtually no distinguishing features from its progenitor state. Based on these three parameters, black holes are classified into:

1. *Schwarzschild black holes*: Possess nonzero mass, zero angular momentum and zero electric charge. Any collapsed non-rotating object becomes a Schwarzschild black hole.
2. *Kerr black holes*: Possess nonzero mass and angular momentum. Resulting from the collapse of rotating objects, they are characterized by an ergosphere – the region between the static limit (inside which neither light nor particles can move against the direction of rotation) and the event horizon. The metric of these objects was first derived by the New Zealand mathematician Roy Kerr.
3. *Reissner–Nordström black holes*: Possess nonzero electric charge; the metric around such an object was derived by the German physicist Hans Reissner and the Finnish physicist Gunnar Nordström. If an electrically charged black hole also rotates, it is termed a Kerr–Newman black hole.

Astrophysical black holes are predominantly Kerr (rotating) black holes, whereas no black hole with a significant net electric charge has ever been observed. The Schwarzschild solution asymptotically approaches flat spacetime. In 1933, British mathematician and cosmologist George McVittie generalized it into a metric approaching an expanding universe at large distances. In 1993, the Irish mathematician and theoretical physicist Brian Nolan demonstrated that a central singularity is not mandatory within the McVittie solution. In 2019, it was proposed that the interior medium preventing singularity formation could consist of quantum vacuum fluctuations. However, all such solutions necessarily possess a time-varying mass, because vacuum fluctuations do not dilute with cosmic expansion; rather, their total energy (and thus mass) grows with increasing volume. In an expanding universe, energy conservation does not hold globally, as time-translation symmetry is broken. The growth of black hole mass driven by cosmic expansion is termed *cosmological coupling* of black holes.

In 2023, a team of nineteen researchers led by Duncan Farrah suggested that the mass growth of black holes driven by ordinary (decelerating) cosmic expansion could even account for the accelerated expansion of the universe; see. Using five galaxy samples from all-sky surveys, the team sought observational evidence for black hole mass growth, finding results not inconsistent with this hypothesis. Nevertheless, what fraction of cosmic acceleration – if any – black holes might actually drive remains uncertain.

Stellar-mass black holes exhibit extreme average densities, unlike supermassive black holes at galactic centers. The average density of a black hole can be estimated by substituting the Schwarzschild radius for its characteristic size:

$$\rho = \frac{M}{V} = \frac{M}{4\pi r_{\mathrm{g}}^3/3} = \frac{3M}{4\pi(2GM/c^2)^3} = \frac{3c^6}{32\pi G^3}\frac{1}{M^2}. \tag{3.338}$$

This represents the average density as defined for an external observer, who never crosses the event horizon to experience this density directly. The average density of very massive black holes can be remarkably low; their extreme gravitational effects stem from their total mass and compact horizon size rather than high density.

According to Stephen Hawking, matter inside a black hole need not remain trapped forever. Quantum processes can enable particles to tunnel out of a black hole into surrounding space – equivalently viewed as particle–antiparticle pair creation just outside the event horizon, where one particle falls in, while the other escapes. Both descriptions (tunneling and pair creation) are equivalent. For stellar-mass and supermassive black holes, this process is extraordinarily slow, taking many orders of magnitude longer than the current age of the universe. Estimated evaporation times are listed below:

Object	Mass	Horizon	Density	Evaporation Time
Rock	1 kg	10^{-27} m	10^{76} g cm^{-3}	10^{-21} s
Earth	6×10^{24} kg	9 mm	10^{27} g cm^{-3}	10^{46} yr
Sun	2×10^{30} kg	3 km	10^{16} g cm^{-3}	10^{62} yr
Galactic core	10^{8} M_S	2 au	1.82 g cm^{-3}	10^{86} yr
Galaxy	10^{11} M_S	30 ly	10^{-6} g cm^{-3}	10^{95} yr

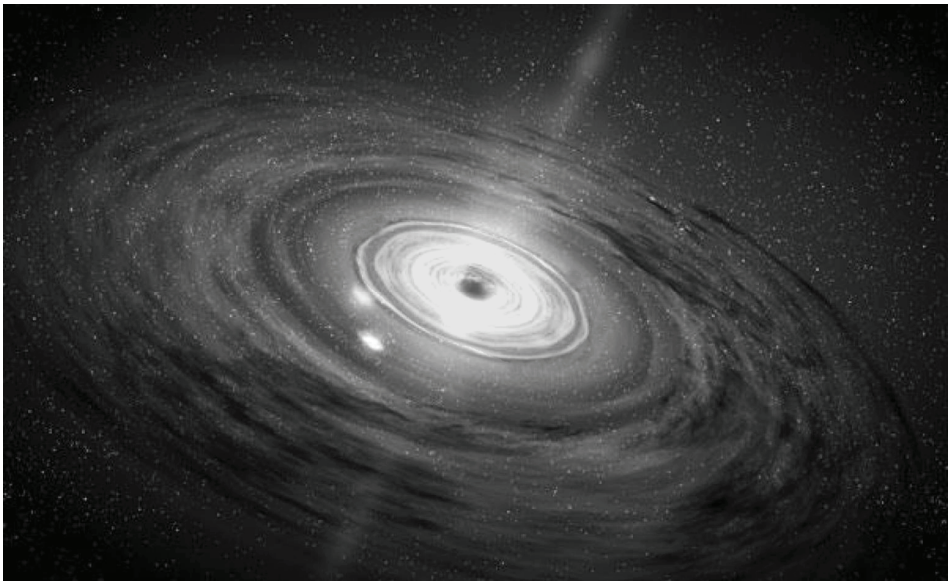


Fig. 3.44: Typical black hole with a dusty gas torus (dark region), accretion disk (bright region), and two relativistic jets (faint in visible light). Credit: NASA.

3.6.4 Motion in the Schwarzschild Geometry

Lagrangian

The motion of particles should be computed from the geodesic equation, which can be recast into the form of the Lagrange equations:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial L}{\partial \dot{x}^{\mu}} - \frac{\partial L}{\partial x^{\mu}} = 0 \,. \tag{3.339}$$

Here, the overdot denotes differentiation with respect to the proper time τ, which parameterizes the particle's worldline $x^{\mu}(\tau)$. For photons, an affine parameter λ is used to parameterize the worldline. The Lagrange equations then take the form

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial L}{\partial \dot{x}^{\mu}} - \frac{\partial L}{\partial x^{\mu}} = 0 \,, \tag{3.340}$$

where the overdot denotes differentiation with respect to λ. The Lagrangian that yields the correct equations of motion represents a generalization of kinetic energy:

$$L = \frac{1}{2} g_{\mu\nu} \frac{\mathrm{d}x^{\mu}}{\mathrm{d}\lambda}\frac{\mathrm{d}x^{\nu}}{\mathrm{d}\lambda} = \frac{1}{2}\frac{\mathrm{d}s^2}{\mathrm{d}\lambda^2} \,. \tag{3.341}$$

For a massive particle, proper time serves as the parameter, whereas for light, an affine parameter is used. Potential energy does not appear in the Lagrangian because gravitational interaction is absorbed into spacetime curvature – that is, into the metric coefficients $g_{\mu\nu}$. Particle mass is likewise absent from the Lagrangian: gravitational motion is independent of mass (provided the mass is not large enough to significantly curve spacetime itself). Omitting the numerical factor ½, the Lagrangian used to compute particle motion becomes:

$$L = \frac{\mathrm{d}s^2}{\mathrm{d}\lambda^2} \,. \tag{3.342}$$

An alternative Lagrangian is also occasionally employed:

$$\tilde{L} = \sqrt{\frac{-\mathrm{d}s^2}{\mathrm{d}\lambda^2}} \,. \tag{3.343}$$

Provided the parameterization is chosen such that $\mathrm{d}s^2/\mathrm{d}\lambda^2$ is constant (which proper time satisfies since $\mathrm{d}s^2/\mathrm{d}\tau^2 = -c^2\mathrm{d}\tau^2/\mathrm{d}\tau^2 = -c^2$), both functions are constant scalars and yield identical equations of motion. Under this assumption, the square of the Lagrangian indeed produces the exact same equations as the Lagrangian itself:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial L^2}{\partial \dot{x}^{\mu}}\right) - \frac{\partial L^2}{\partial x^{\mu}} = 0 \quad \Rightarrow$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(2L\frac{\partial L}{\partial \dot{x}^{\mu}}\right) - 2L\frac{\partial L}{\partial x^{\mu}} = 0 \quad \Rightarrow$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial L}{\partial \dot{x}^{\mu}}\right)-\frac{\partial L}{\partial x^{\mu}}=0\,.$$

Throughout the following derivations, the simpler Lagrangian (3.342) is used.

Effective Potential and Particle Motion

Particle motion is evaluated in the equatorial plane $\theta=\pi/2$. The interval and the Lagrangian take the forms:

$$\mathrm{d}s^2=-c^2\left(1-\frac{r_{\mathrm{g}}}{r}\right)\mathrm{d}t^2+\frac{\mathrm{d}r^2}{1-\dfrac{r_{\mathrm{g}}}{r}}+r^2\mathrm{d}\varphi^2\,, \tag{3.344}$$

$$L=-c^2\left(1-\frac{r_{\mathrm{g}}}{r}\right)\dot{t}^2+\frac{\dot{r}^2}{1-\dfrac{r_{\mathrm{g}}}{r}}+r^2\dot{\varphi}^2\,. \tag{3.345}$$

The equations of motion describing a particle's worldline $x^{\mu}(\tau)$ in the Schwarzschild geometry are given by:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial L}{\partial \dot{t}}-\frac{\partial L}{\partial t}=0\,;\\
&\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial L}{\partial \dot{\varphi}}-\frac{\partial L}{\partial \varphi}=0\,;.\\
&\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\partial L}{\partial \dot{r}}-\frac{\partial L}{\partial r}=0\,;
\end{aligned} \tag{3.346}$$

The objective is to determine the functions $t(\tau)$, $\varphi(\tau)$, and $r(\tau)$. The first two equations lead to conservation laws, which is immediately evident from the Lagrangian, as t and φ are cyclic coordinates (do not appear in L). The first two Lagrange equations yield:

$$\frac{\partial L}{\partial \dot{t}}=\text{const}\,;\qquad\Rightarrow\qquad\left(1-\frac{r_{\mathrm{g}}}{r}\right)\dot{t}=k\,; \tag{3.347}$$

$$\frac{\partial L}{\partial \dot{\varphi}}=\text{const}\,;\qquad\Rightarrow\qquad r^2\dot{\varphi}=\ell\,. \tag{3.348}$$

The first equation expresses the generalized conservation of energy, while the second represents the conservation of angular momentum. Here, k and ℓ are constants, with ℓ representing the specific angular momentum (angular momentum per unit mass). The radial Lagrange equation is a second-order differential equation, as radial momentum is not conserved. However, this equation can be replaced by the condition $\mathrm{d}s^2=-c^2\mathrm{d}\tau^2$, which immediately gives the evolution of the radial coordinate:

$$\frac{\mathrm{d}s^2}{\mathrm{d}\tau^2}=-c^2\qquad\Rightarrow \tag{3.349}$$

$$-c^2\left(1-r_{\mathrm{g}}/r\right)\dot{t}^2+\frac{\dot{r}^2}{1-r_{\mathrm{g}}/r}+r^2\dot{\varphi}^2=-c^2\,. \tag{3.350}$$

Substituting $\dot{t}$ from (3.347) and $\dot{\varphi}$ from (3.348) yields:

► $$\dot{r}^2 + V_{\text{eff}}(r) = c^2(k^2 - 1)\,; \tag{3.351}$$

► $$V_{\text{eff}}(r) \equiv \frac{\ell^2}{r^2} - \frac{\ell^2 r_{\text{g}}}{r^3} - c^2\frac{r_g}{r}\,. \tag{3.352}$$

Relation (3.351) is the relativistic analog of classical energy conservation. Unlike to classical mechanics, the effective potential contains terms proportional to $1/r$, $1/r^2$, and $1/r^3$. The final equation for the evolution of the radial coordinate $r(\tau)$ follows directly:

$$\dot{r} = \pm\sqrt{c^2(k^2 - 1) - V_{\text{eff}}(r)}\,. \tag{3.353}$$

Motion is clearly restricted to regions where $c^2(k^2–1) \geq V_{\text{eff}}$. The shape of the effective potential curve depends on the specific angular momentum ℓ. For intermediate values of ℓ, the effective potential exhibits two extrema; the first is a maximum corresponding to an unstable circular orbit.

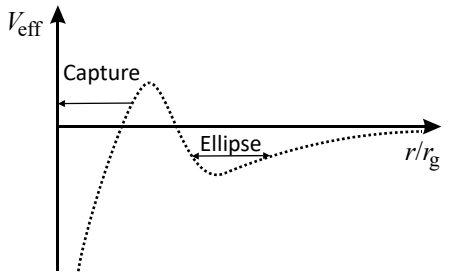


Fig. 3.45: Effective potential for an intermediate value of angular momentum.

Hot plasma typically orbits a black hole, forming an accretion disk whose inner edge lies at the innermost stable circular orbit. The second extremum corresponds to a stable circular orbit near the black hole. From the shape of the potential, bound elliptical (precessing) orbits clearly exist here as well, which is not the case for small values of angular momentum. Finding the extrema from the condition:

$$\frac{\mathrm{d}}{\mathrm{d}r}\left[\frac{\ell^2}{r^2} - \frac{\ell^2 r_{\text{g}}}{r^3} - c^2\frac{r_g}{r}\right] = 0 \quad \Rightarrow \quad c^2 r_{\text{g}} r^2 - 2\ell^2 r + 3\ell^2 r_{\text{g}} = 0\,. \tag{3.354}$$

readily yields both roots:

$$r_{1,2} = \frac{\ell^2 \pm \sqrt{\ell^4 - 3c^2\ell^2 r_{\text{g}}^2}}{c^2 r_{\text{g}}}\,. \tag{3.355}$$

The smaller root corresponds to the unstable maximum, whereas the larger root represents a stable circular orbit, which exists provided that:

$$\ell^2 \geq 3c^2 r_{\mathrm{g}}^2 \,. \tag{3.356}$$

For the limiting value of specific angular momentum, the radius of the innermost stable circular orbit is given by

$$r_2 = \frac{\ell^2}{c^2 r_{\mathrm{g}}} = \frac{3c^2 r_{\mathrm{g}}^2}{c^2 r_{\mathrm{g}}} = 3r_{\mathrm{g}} . \tag{3.357}$$

Thus, the *innermost stable circular orbit* for massive particles lies at $r = 3r_{\mathrm{g}}$.

Propagation of Light

Light propagation in the background of the Schwarzschild geometry is parameterized by an affine parameter λ. The Lagrangian retains the form (3.345), where the overdot now denotes differentiation with respect to the affine parameter λ. The equations for t and φ again yield the conservation laws (3.347) and (3.348), from which the parametric evolution of these coordinates is obtained:

► $$\frac{\mathrm{d}t}{\mathrm{d}\lambda} = \frac{k}{1 - r_{\mathrm{g}}/r} ; \tag{3.358}$$

► $$\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = \frac{\ell}{r^2} . \tag{3.359}$$

Instead of the final Lagrange equation for the radial coordinate, the simpler null condition for light propagation can be used:

$$\mathrm{d}s^2 = 0 , \tag{3.360}$$

which, upon parameterization by the affine parameter λ, leads to:

$$-c^2 \left(1 - \frac{r_{\mathrm{g}}}{r}\right) \dot{t}^2 + \frac{\dot{r}^2}{1 - \dfrac{r_{\mathrm{g}}}{r}} + r^2 \dot{\varphi}^2 = 0 . \tag{3.361}$$

Substituting $\dot{t}$ from (3.358) and $\dot{\varphi}$ from (3.359) yields:

$$\dot{r}^2 + V_{\mathrm{eff}}(r) = c^2 k^2 ; \tag{3.362}$$

$$V_{\mathrm{eff}}(r) \equiv \frac{\ell^2 \left(1 - r_{\mathrm{g}}/r\right)}{r^2} . \tag{3.363}$$

The effective potential has a slightly simpler form than that for a massive particle. From (3.362), the radial evolution equation is:

► $$\frac{\mathrm{d}r}{\mathrm{d}\lambda} = \pm\sqrt{c^2 k^2 - V_{\mathrm{eff}}(r)} . \tag{3.364}$$

Using equations (3.358), (3.359), and (3.364), the worldline of a light ray in the Schwarzschild metric can be computed via an appropriate numerical method. Relation (3.362) is again analogous to energy conservation, and equation (3.364) shows that motion is restricted to regions where $c^2 k^2 \geq V_{\mathrm{eff}}$.

Circular Photon Orbit

A circular photon orbit is found from the condition

$$\frac{\mathrm{d}V_{\mathrm{eff}}}{\mathrm{d}\,r} = 0\,, \tag{3.365}$$

which yields:

$$r = \frac{3}{2} r_{\mathrm{g}}\,. \tag{3.366}$$

This trajectory is unstable, as the effective potential reaches a maximum at this radius.

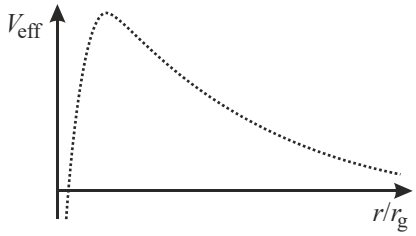


Fig. 3.46: Effective potential for photons.

In the Schwarzschild geometry, photons at a distance of 1.5 r_{g} from the center move along unstable circular orbits (the photon sphere). Under any slight perturbation, they either spiral into the black hole or escape outward. The actual trajectory of photons passing near a massive body, such as the Sun, is analyzed next.

Bending of Light

Photon motion near a central mass M is described using polar coordinates $r(\lambda)$ and $\varphi(\lambda)$, where r is the distance from the central mass and φ is the angle measured from the periastron (point of closest approach). The closest approach distance is denoted by R_0 and the impact parameter by b.

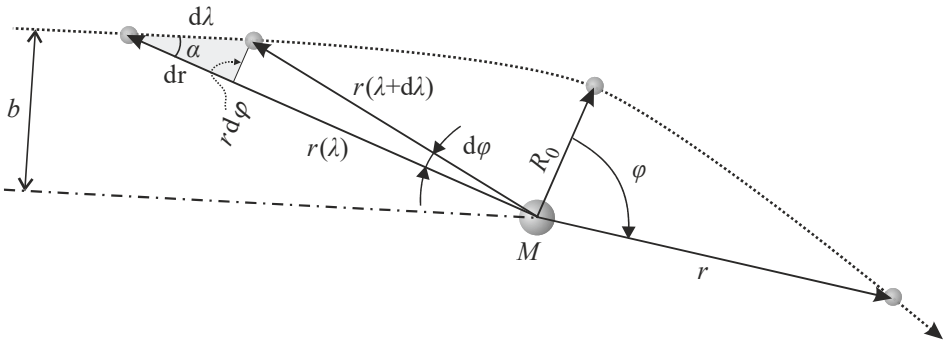


Fig. 3.47: Bending of light near a massive object.

From the highlighted gray triangle in the figure, the angle α satisfies:

$$\cos\alpha = \frac{\mathrm{d}r}{\mathrm{d}\lambda}, \tag{3.367}$$

$$\sin\alpha = \frac{r\,\mathrm{d}\varphi}{\mathrm{d}\lambda}. \tag{3.368}$$

If only the spatial trajectory of the photon is of interest, the equations of motion are:

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = \frac{\ell}{r^2} \tag{3.369}$$

$$\frac{\mathrm{d}r}{\mathrm{d}\lambda} = \pm\sqrt{c^2k^2 - V_{\mathrm{eff}}(r)}\,. \tag{3.370}$$

The physical meaning of the integration constants k and ℓ is determined directly from the equations of motion. Because these constants are conserved, any point along the trajectory may be chosen for their evaluation. Evaluating them when the photon is far away and deflection has not yet occurred ($t \to -\infty$, $r \to \infty$) gives:

$$c^2k^2 = \dot{r}^2 + V_{\mathrm{eff}}(r) = \left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + V_{\mathrm{eff}}(r) = \cos^2\alpha + V_{\mathrm{eff}}(r) \underset{r\to\infty}{\to} 1+0=1\,. \tag{3.371}$$

$$\ell = r^2\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = r\frac{r\,\mathrm{d}\varphi}{\mathrm{d}\lambda} = r\sin\alpha = b. \tag{3.372}$$

The constant ℓ thus carries a dual interpretation: it represents both the angular momentum per unit mass and the impact parameter b. The resulting equations of motion for the photon are:

► $$\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = \frac{b}{r^2}, \tag{3.373}$$

► $$\frac{\mathrm{d}r}{\mathrm{d}\lambda} = \pm\sqrt{1 - V_{\mathrm{eff}}(r)}\,. \tag{3.374}$$

While these equations can be solved numerically, an analytical solution is derived below in the weak-field (small-deflection) limit. First, an important relation connecting b, M, and the closest approach distance R_0 is established. At the point of closest approach, the radial distance reaches a minimum, meaning $\mathrm{d}r/\mathrm{d}\lambda = 0$, and equation (3.374) immediately yields:

$$V_{\mathrm{eff}}(R_0) = 1\,, \tag{3.375}$$

Substituting the effective potential (3.363) gives the relation:

► $$b^2\left(1 - r_{\mathrm{g}}/R_0\right) = R_0^2\,, \tag{3.376}$$

where b is the impact parameter, R_0 is the closest approach distance, and the Schwarzschild radius r_{g} encodes the mass of the deflecting body. To determine the total deflection angle alone, solving the full system of equations (3.373) and (3.374) is unnecessary; taking their ratio yields:

$$\frac{\mathrm{d}\varphi}{\mathrm{d}r} = \pm\frac{b/r^2}{\sqrt{1-V_{\mathrm{eff}}(r)}}\,. \tag{3.377}$$

Separating variables and integrating gives:

$$\Delta\varphi \equiv \varphi_{\infty} - \varphi_{-\infty} = 2\int_{R_0}^{\infty} \frac{b/r^2}{\sqrt{1-V_{\mathrm{eff}}(r)}}\,\mathrm{d}r\,. \tag{3.378}$$

Because the trajectory is symmetric about the point of closest approach, integration is performed from R_0 to infinity and the result is doubled. Substituting the effective potential yields:

$$\Delta\varphi = 2\int_{R_0}^{\infty} \frac{1/r^2}{\sqrt{b^{-2} - \dfrac{1}{r^2} + \dfrac{r_g}{r^3}}}\,\mathrm{d}r\,. \tag{3.379}$$

Since R_0 and b are interdependent, eliminating one simplifies the expression. Solving (3.376) for b^{-2} gives:

$$\Delta\varphi = 2\int_{R_0}^{\infty} \frac{1/r^2}{\sqrt{R_0^{-2} - r_{\mathrm{g}}R_0^{-3} - \dfrac{1}{r^2} + \dfrac{r_g}{r^3}}}\,\mathrm{d}r\,. \tag{3.380}$$

Introducing the substitution

$$u \equiv 1/r \tag{3.381}$$

transforms the integral into its final exact form:

► $$\Delta\varphi = 2\int_{0}^{1/R_0} \frac{\mathrm{d}u}{\sqrt{R_0^{-2} - r_{\mathrm{g}}R_0^{-3} - u^2 + r_g u^3}}\,. \tag{3.382}$$

Although this integral can be evaluated numerically, an analytical approximation is readily obtained for small central masses. The change in angle depends on the central mass via the integral:

$$\Delta\varphi = 2\int_{0}^{1/R_0} F(u, r_g)\,\mathrm{d}u\,, \tag{3.383}$$

where the mass dependence enters through the Schwarzschild radius r_{g}. For small masses, a Taylor expansion is performed:

$$\Delta\varphi = 2\int_{0}^{1/R_0} F(u,0)\,\mathrm{d}u + 2\int_{0}^{1/R_0} F'(u,0)\,r_g\,\mathrm{d}u + \cdots, \tag{3.384}$$

where the prime denotes partial differentiation with respect to r_{g}. The zeroth-order term corresponds to a massless central body, where the light ray travels along an undeflected straight line, whereas the second term (first-order correction) is linear in the central mass. Differentiating and setting $r_{\mathrm{g}} = 0$ yields:

$$\Delta\varphi = 2\int_0^{1/R_0} \frac{\mathrm{d}u}{\sqrt{R_0^{-2}-u^2}} + r_g \int_0^{1/R_0} \frac{R_0^{-3}-u^3}{\left[R_0^{-2}-u^2\right]^{3/2}}\,\mathrm{d}u + \cdots. \tag{3.385}$$

Simplifying the integrands gives:

$$\Delta\varphi = 2R_0\int_0^{1/R_0} \frac{\mathrm{d}u}{\sqrt{1-(R_0u)^2}} + r_g \int_0^{1/R_0} \frac{1-(R_0u)^3}{\left[1-(R_0u)^2\right]^{3/2}}\,\mathrm{d}u + \cdots. \tag{3.386}$$

Substituting $\xi = R_0u$ in the first integral and $\sin\xi = R_0u$ in the second yields:

$$\Delta\varphi = \int_0^1 \frac{\mathrm{d}\xi}{\sqrt{1-\xi^2}} + \frac{r_g}{R_0}\int_0^{\pi/2}\left[\frac{1}{\cos^2\xi} - \frac{\sin^3\xi}{\cos^2\xi}\right]\mathrm{d}\xi \;+\cdots. \tag{3.387}$$

The first integral evaluates to arcsin ξ, the first term of the second integral yields tan ξ, and the antiderivative of the final term is found via the substitution $t = \cos\xi$:

$$\Delta\varphi = \left[\arcsin\xi\right]_0^1 + \frac{r_g}{R_0}\left[\tan\xi - \frac{1}{\cos\xi} - \cos\xi\right]_0^{\pi/2} + \cdots \tag{3.388}$$

To avoid individually diverging terms at the upper limit, the first two terms of the second bracket are combined:

$$\Delta\varphi = \pi + \frac{r_g}{R_0}\left[\frac{\sin\xi - 1}{\cos\xi} - \cos\xi\right]_0^{\pi/2} + \cdots \tag{3.389}$$

Evaluating the limits gives:

$$\Delta\varphi = \pi + 2r_g/R_0 + \cdots \tag{3.390}$$

The first term corresponds to an undeflected straight ray (zero central mass), while the second term represents the gravitational deflection of a ray passing at closest approach distance R_0. The total deflection angle of a light ray passing a massive body in the weak-field limit is therefore:

► $$\chi = \frac{2r_g}{R_0} = \frac{4GM}{c^2R_0}\,. \tag{3.391}$$

Relation (3.391) is a landmark prediction of general relativity; calculations based on freely falling local inertial frames or Newtonian mechanics yield only half of this value. For incoming light from distant stars grazing the surface of the Sun (R_0 = 700,000 km, $M = 2\times10^{30}$ kg), this yields:

$$\chi = 1{,}745''\,. \tag{3.392}$$

Stellar positions near the solar disk can be photographed during a total solar eclipse and compared with positions measured when the Sun is absent from that sky region (for example, six months later at night). The first such measurement was conducted by Sir Arthur Stanley Eddington in 1919 during a solar eclipse expedition split between the island of Príncipe off the coast of West Africa and Sobral in Brazil, confirming general relativity to within a relative accuracy of 20 %.

3.6.5 Gravitational Lenses

An intervening massive object located between an observer and a background source – such as a star or a galaxy – can bend light, acting as a gravitational lens. The concept of gravitational lensing originated with the Russian physicist Orest Khvolson, who considered the phenomenon in 1924 and was the first to publish on its possible existence. In 1936, the Czech astronomer and educator František Link performed preliminary calculations on gravitational lensing, publishing his work in a French journal that went largely unread by the general relativity community. During the 1930s, the Czech engineer Rudi W. Mandl also pondered gravitational lenses and brought the idea to the attention of Albert Einstein. In 1936, Einstein calculated that such a background source would appear as multiple images, or, in the ideal case of perfect alignment, as a ring. At Mandl's urging, Einstein published these calculations in the journal Science in late 1936, though Einstein himself doubted that such an effect would ever be observable. As his research notebooks reveal, Einstein had already explored the idea of gravitational lensing as early as 1912, but left his calculations unpublished at the time.

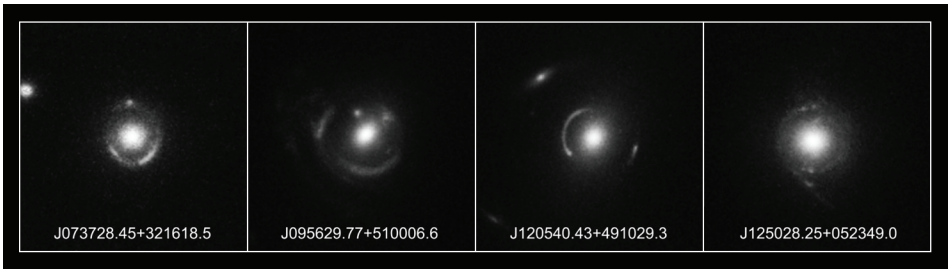


Fig. 3.48: Einstein rings photographed by the Hubble Space Telescope.
The bright object at the center is the intervening lensing galaxy.

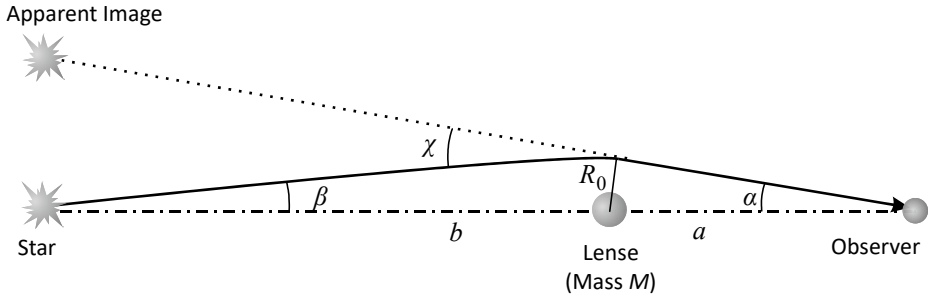


Fig. 3.49: Geometry for calculating the radius of an Einstein ring.

According to the figure, a background star aligned directly behind a lensing object of mass M appears as a ring with angular radius α. Because all angles are small, the following relations hold:

$$\alpha = \frac{R_0}{a}, \tag{3.393}$$

$$\beta = \frac{R_0}{b}, \tag{3.394}$$

For the deflection angle, the following equations apply:

$$\chi = \frac{4GM}{c^2 R_0}, \tag{3.395}$$

$$\chi = \alpha + \beta. \tag{3.396}$$

These four equations form a system for the variables α, β, χ, R_0. Successively eliminating β, χ and R_0 yields the final expression for the angular radius of the Einstein ring:

► $$\alpha = \sqrt{\frac{b}{a(a+b)}\frac{4GM}{c^2}}. \tag{3.397}$$

When the source, lens, and observer are not collinear, or when the lensing mass distribution is asymmetric, multiple images (such as an Einstein cross; see book cover) or distorted arcs appear. The first gravitational lens was discovered in 1979 during observations of the quasar QSO 0957+561, which had a magnitude of 17, a cosmological redshift of $z = 1.405$, and appeared split into a double image due to gravitational lensing. It was discovered by an Anglo-American team (Dennis Walsh, Robert Carswell, and Ray Weyman) using the 2.1-meter telescope at Kitt Peak National Observatory in Arizona. The first Einstein ring was discovered in 1988 in the radio source designated MG 1131+0456. Images of background galaxies sheared into characteristic arcs by intervening galaxy clusters are also common. Today, gravitational lensing is utilized to map the distribution of dark matter from the distortion of distant galaxy images, as well as to detect exoplanets, which – together with their host stars – produce characteristic light-curve perturbations when passing in front of a distant background star (the gravitational microlensing effect).

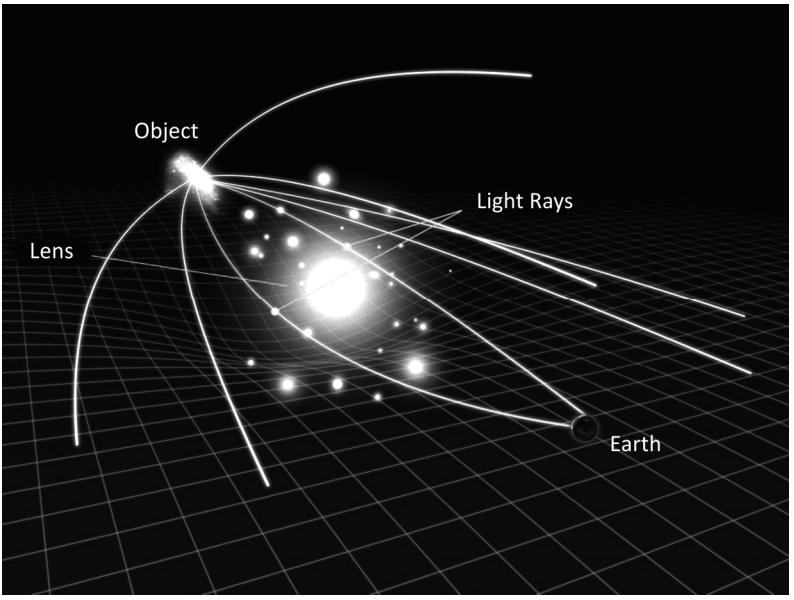


Fig. 3.50: Gravitational lensing by an intervening galaxy cluster. Credit: NASA/ESA.

Microlensing

In Einstein's time, observational technology had not yet reached a level capable of detecting gravitational lenses. Today, lensing caused by much smaller bodies than intervening galaxies can be observed – for example, when a foreground star passes in front of a more distant star, amplifying its light. Although rare, such events become observable when monitoring vast numbers of stars in large-scale survey projects. Gravitational lensing induced by a star or an even smaller object is termed *gravitational microlensing*. The first microlensing event was discovered in 1989 with the Canada–France–Hawaii Telescope, when a star in our Galaxy amplified the light of the distant quasar QSO 2237+0305 (the Einstein Cross). Most commonly, however, gravitational microlensing occurs when two stars at different distances within our Galaxy align along the observer's line of sight.

Gravitational microlensing is not a permanent phenomenon. Because stars move through the Galaxy, the foreground lens remains aligned with the background source for only a brief interval, producing a characteristic peak in the background star's light curve. This amplification occurs regardless of the electromagnetic waveband in which the lensing object radiates – in fact, the lens need not emit any light at all. Consequently, gravitational microlensing enables the identification of faint or dark objects such as white dwarfs, brown dwarfs, planets, black holes, or other non-luminous compact bodies. It is therefore used to estimate the abundance of dark matter in the Galactic halo, where such objects are collectively referred to as MACHOs (*Massive Astrophysical Compact Halo Objects*).

If the intervening lensing star hosts a companion, such as a planet, it induces an additional amplification of the background star's light, creating a secondary local peak on the light curve. Microlensing can thus be utilized to discover extrasolar planets.

The term “microlensing” was introduced by the Polish astronomer Bohdan Paczyński, who conceived the first dedicated microlensing survey, OGLE (*Optical Gravitational Lensing Experiment*). Established in 1992 and led by Andrzej Udalski of the University of Warsaw, this international project utilizes a 1.3-meter telescope at Las Campanas Observatory in Chile to search for microlensing events toward the Galactic bulge. The survey demonstrated that microlensing is an effective tool for finding exoplanets (co-discovering the very first microlensing exoplanets). While OGLE detects roughly 500 microlensing events annually, planetary detections remain relatively rare, yielding several dozen exoplanet discoveries over the course of the project. Today, numerous other survey programs actively search for microlensing events.

Microlensing can also occur in binary star systems. In 1973, the Swiss astronomer André Maeder predicted that when one component of a binary star transits across the other, gravitational lensing could actually increase the total observed brightness of the system. This effect occurs when a small, compact object passes in front of the disk of its companion: instead of the expected dimming caused by mutual eclipse (where one body blocks light from the other), a favorable geometric configuration produces the opposite effect. In 2014, such a system was first detected in data from the Kepler space telescope (which searches for exoplanets via the transit method). The object, designated KOI-3278, was initially flagged as an exoplanet candidate orbiting a Sun-like star located 2,600 light-years away in the constellation Lyra, exhibiting periodic light-curve dips every 88.18 days. A detailed analysis by Ethan Kruse and Eric Agol of the University of Washington in Seattle revealed periodic 0.1 % brightenings lasting 5 hours. Through detailed modeling, they concluded that the orbiting companion is a white dwarf: the

total flux dims when the white dwarf passes behind the star, and is gravitationally lensed (amplified) when it passes in front. Four decades after André Maeder's prediction, a self-lensing eclipsing binary system was finally confirmed.

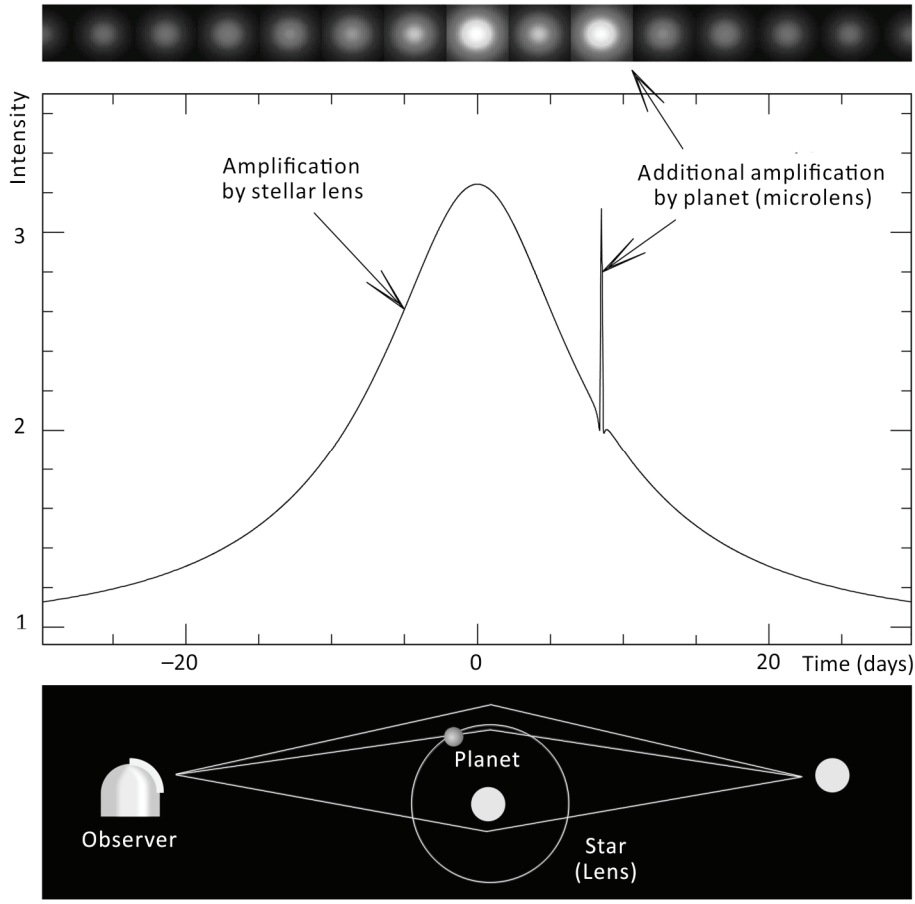


Fig. 3.51: Light curve of a microlensing event produced by a star with an orbiting planet.

3.6.6 Frame-Dragging by a Rotating Body

Another intriguing problem in general relativity is describing the behavior of spacetime in the vicinity of a rotating sphere. Such a solution was first sought by the Austrian physicist Hans Thirring in 1918, later assisted in the calculations by the Austrian mathematician Josef Lense. They initially solved the problem for a thin rotating spherical shell with an empty interior. Remarkably, the interior metric coincided with the flat Minkowski metric, yet to the astonishment of both scientists, the interior spacetime rotated. This phenomenon represents a manifestation of the Mach principle, according to which the inertia of individual bodies is determined by the mass distribution of the universe; the rotating shell influenced the state of spacetime inside it. Their solution was constructed within linearized gravity using a metric of the form:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \,. \tag{3.398}$$

Lense and Thirring subsequently derived the solution for a solid rotating sphere, demonstrating that exterior spacetime is dragged along by the rotating body and rotates in its vicinity. Today, this phenomenon is known as the Lense–Thirring effect (or frame-dragging) and resembles the dragging of a viscous fluid around a spinning object.

Analogy with Electrostatics

A strong analogy exists between electrostatic and gravitational fields, whose governing equations share an identical form. The relations for force, field intensity, potential energy, and potential for a test particle (q, m) in the field of a central body (Q, M) are:

$$\mathbf{F} = \frac{qQ}{4\pi\varepsilon_0 r^2}\frac{\mathbf{r}}{r}, \qquad \mathbf{F}_{\mathrm{G}} = -G\frac{mM}{r^2}\frac{\mathbf{r}}{r}; \tag{3.399}$$

$$\mathbf{E} = \frac{Q}{4\pi\varepsilon_0 r^2}\frac{\mathbf{r}}{r}, \qquad \mathbf{E}_{\mathrm{G}} = -G\frac{M}{r^2}\frac{\mathbf{r}}{r}; \tag{3.400}$$

$$W = \frac{qQ}{4\pi\varepsilon_0 r}, \qquad W_{\mathrm{G}} = -G\frac{mM}{r}; \tag{3.401}$$

$$\phi = \frac{Q}{4\pi\varepsilon_0 r}, \qquad \phi_{\mathrm{G}} = -G\frac{M}{r}. \tag{3.402}$$

When the field is generated by a system of particles, the potentials for discrete and continuous distributions satisfy the familiar relations (where $\mathbf{r}_a$ denotes the positions of discrete sources, $\mathbf{r}'$ is the integration variable over the source volume, $\mathbf{r}$ is the observer's position, ρ_{M} is the mass density, and ρ_{Q} is the charge density):

$$\phi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\sum_a \frac{Q_a}{|\mathbf{r}-\mathbf{r}_a|}, \qquad \phi_{\mathrm{G}}(\mathbf{r}) = -G\sum_a \frac{M_a}{|\mathbf{r}-\mathbf{r}_a|}; \tag{3.403}$$

$$\phi(\mathbf{r}) = \frac{1}{4\pi\varepsilon_0}\int_V \frac{\rho_{\mathrm{Q}}(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}, \qquad \phi_{\mathrm{G}}(\mathbf{r}) = -G\int_V \frac{\rho_{\mathrm{M}}(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}. \tag{3.404}$$

In the static case or for slowly moving sources, the Laplace–Poisson field equation suffices. When source motions become appreciable, the Laplacian operator is replaced by the d'Alembert wave operator:

$$\Delta\phi = -\frac{\rho_{\mathrm{Q}}}{\varepsilon_0}, \qquad \Delta\phi_G = 4\pi G\rho_{\mathrm{M}}; \tag{3.405}$$

$$\Box\phi = -\frac{\rho_{\mathrm{Q}}}{\varepsilon_0} \qquad \Box\phi_G = 4\pi G\rho_{\mathrm{M}}. \tag{3.406}$$

The solutions to the field equations (3.405) are given by (3.404). When solving the wave equations (3.406), retarded time must be used in the solutions (3.404) – that is, the observer's time must be corrected for the finite propagation delay of the signal from the source to the observer.

Gravitoelectric Field

For a rotating body, a diagonal metric can no longer be assumed. Gauge freedom still permits imposing the Lorenz-like gauge condition $h^{\alpha\beta}{}_{,\beta}=0$, but no further simplification is possible: neither the trace of the perturbations $s = h^{\alpha}{}_{\alpha}$ nor the mixed components h_{0k} vanish. The linearized field equation (3.277) now takes the more general form:

$$\begin{gathered}\Box\left(h_{\alpha\beta}-\frac{1}{2}s\,\eta_{\alpha\beta}\right)=-\frac{16\pi G}{c^4}T_{\alpha\beta}\,;\\ h^{\alpha\beta}{}_{,\beta}=0\,,\qquad s\equiv h^{\alpha}{}_{\alpha}\,.\end{gathered}\tag{3.407}$$

On the right-hand side, the stress-energy tensor of incoherent dust is considered:

$$T_{\alpha\beta}=\rho_{\mathrm{M}}\,U_{\alpha}U_{\beta}\,.\tag{3.408}$$

For a rotating body, the spatial components of the four-velocity are nonzero. Returning briefly to the Schwarzschild metric of a non-rotating body, its line element in the linear approximation (3.398) reads:

$$\begin{gathered}\mathrm{d}s^2=-c^2\left(1-\frac{r_g}{r}\right)\mathrm{d}t^2+\frac{\mathrm{d}r^2}{1-\dfrac{r_g}{r}}+r^2\mathrm{d}\Omega^2;\\ \mathrm{d}s^2\approx\left(-1+\frac{r_g}{r}\right)\mathrm{d}(ct)^2+\left(1+\frac{r_g}{r}\right)\mathrm{d}r^2+r^2\mathrm{d}\Omega^2\end{gathered}\tag{3.409}$$

The metric perturbations $h_{\mu\nu}$ satisfy:

$$h_{00}=h_{11}=\frac{r_g}{r}=-\frac{2\phi_{\mathrm{G}}}{c^2}\,.\tag{3.410}$$

A solution to equations (3.407) is now sought in isotropic coordinates, where:

$$h_{00}=h_{11}=h_{22}=h_{33}\equiv h_0\,.\tag{3.411}$$

The trace of the perturbation tensor is then:

$$s=h^{\alpha}{}_{\alpha}=\eta^{\alpha\beta}h_{\alpha\beta}=-h_0+h_0+h_0+h_0=2h_0\,.\tag{3.412}$$

From equation (3.407), the 00-component yields:

$$\begin{gathered}\Box\left(h_{00}-\frac{1}{2}s\,\eta_{00}\right)=-\frac{16\pi G}{c^4}\rho_{\mathrm{M}}U_0U_0\,;\\ \Box\left(h_0+h_0\right)=-\frac{16\pi G}{c^4}\rho_{\mathrm{M}}c^2\,;\\ \Box\,h_0=-\frac{8\pi G}{c^2}\rho_{\mathrm{M}}.\end{gathered}\tag{3.413}$$

In the Newton limit, $h_0=-2\phi_{\mathrm{G}}/c^2$, yielding the equation for the gravitational potential:

$$\Box\,\phi_G=4\pi G\rho_{\mathrm{M}}\,.\tag{3.414}$$

Thus, the Einstein Gravity Law recovers the exact classical field equation for the gravitational potential. For slowly moving sources in the linear approximation, the solution for a given mass distribution can be written immediately in the form (3.404):

$$\blacktriangleright \qquad h_0 = h_{00} = h_{11} = h_{22} = h_{33} = -\frac{2\phi_{\mathrm{G}}(\mathbf{r})}{c^2} = \frac{2G}{c^2}\int_V \frac{\rho_{\mathrm{M}}(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\,. \tag{3.415}$$

Because of its close analogy with the electric scalar potential, the perturbation h_0 is termed the gravitoelectric potential, and its associated field is called the *gravitoelectric field*. The corresponding line element is:

$$\mathrm{d}s^2 = \left(-1+h_0\right)\mathrm{d}(ct)^2 + \left(1+h_0\right)\mathrm{d}\mathbf{r}^2 + 2h_{0k}\,\mathrm{d}(ct)\,\mathrm{d}x_k\,. \tag{3.416}$$

The factor of 2 multiplying the off-diagonal terms reflects the symmetry of the perturbation tensor, $h_{0k} = h_{k0}$.

Gravitomagnetic Field

Proceeding now to the calculation of the off-diagonal perturbations h_{0k}, the Einstein field equations (3.407) for slowly moving field sources give:

$$\begin{aligned} \Box\, h_{0k} &= -\frac{16\pi G}{c^4}T_{0k}\,, \\ \Box\, h_{0k} &= -\frac{16\pi G}{c^4}\rho_{\mathrm{M}}U_0U_k\,, \\ \Box\, h_{0k} &= +\frac{16\pi G}{c^4}\rho_{\mathrm{M}}\,c\,v_k\,. \end{aligned} \tag{3.417}$$

The field equation for the off-diagonal perturbations therefore takes the form:

$$\begin{aligned} \Box\, h_k &= +\frac{16\pi G}{c^3}j_k^{(\mathrm{M})}\,; \\ h_k \equiv h_{0k}\,; \qquad j_k^{(\mathrm{M})} &\equiv \rho_{\mathrm{M}}\,v_k\,, \end{aligned} \tag{3.418}$$

where $\mathbf{j}_{\mathrm{M}}$ denotes the mass current density. For slow motions, the wave operator reduces to the Laplacian operator, allowing the equations to be written in vector form:

$$\Delta\mathbf{h} = \frac{16\pi G}{c^3}\mathbf{j}_{\mathrm{M}}\,. \tag{3.419}$$

An analogy arises with magnetostatics, where the magnetic vector potential satisfies:

$$\Delta\mathbf{A} = -\mu_0\mathbf{j}_{\mathrm{Q}}\,, \tag{3.420}$$

$$\mathbf{A}(\mathbf{r}) = \frac{\mu_0}{4\pi}\int_V \frac{\mathbf{j}_{\mathrm{Q}}(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\,. \tag{3.421}$$

In gravitation, mass current density replaces electric current density, while the equations remain completely analogous. Consequently, the perturbations h_0 and h_{k} play the exact

same role in gravitation as the scalar and vector potentials do in electromagnetism. By analogy with (3.404), the solution to the Poisson equation can be written as:

$$\mathbf{h}(\mathbf{r}) = -\frac{4G}{c^3}\int_V \frac{\mathbf{j}_{\mathrm{M}}(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}. \tag{3.422}$$

Dimensional analysis confirms that this expression is dimensionless, consistent with the definition of the linear perturbations $h_{\mu\nu}$ as dimensionless quantities. The complete metric for a slowly rotating central body can therefore be written in its final form:

$$\mathrm{d}s^2 = (-1+h_0)\,\mathrm{d}(ct)^2 + (1+h_0)\,\mathrm{d}\mathbf{r}^2 + 2(\mathbf{h}\cdot\mathrm{d}\mathbf{r})\,\mathrm{d}(ct);$$

► $$h_0(\mathbf{r}) = \frac{2G}{c^2}\int_V \frac{\rho_{\mathrm{M}}(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}, \qquad . \tag{3.423}$$

$$\mathbf{h}(\mathbf{r}) = -\frac{4G}{c^3}\int_V \frac{\rho_{\mathrm{M}}(\mathbf{r}')\mathbf{v}(\mathbf{r}')\,\mathrm{d}^3\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}.$$

The cross-term coupling the temporal coordinate with the spatial ones is responsible for the dragging of spacetime by the rotating body. By analogy with magnetism, the spacetime curvature governed by the perturbation **h** is termed the *gravitomagnetic field*.

Lense–Thirring Effect

Just as the magnetic field is defined in terms of the magnetic vector potential via:

$$\mathbf{B} = \mathrm{curl}\,\mathbf{A}, \tag{3.424}$$

the gravitomagnetic field can be introduced as:

$$\mathbf{K} = \mathrm{curl}\,\mathbf{h}. \tag{3.425}$$

In a magnetic field, magnetic dipoles undergo precession at an angular frequency proportional to the magnetic field strength. Exploiting the analogy between magnetic and gravitomagnetic fields maps the gravitational problem directly onto magnetostatics. Accordingly, the precession angular frequency of a gyroscope satisfies:

$$\boldsymbol{\omega}_{\mathrm{LT}} = \frac{c}{2}\mathrm{curl}\,\mathbf{h} \tag{3.426}$$

Evaluating the integrals in (3.423) yields a simple expression for the instantaneous precession of a gyroscope near a rotating sphere:

$$\boldsymbol{\omega}_{\mathrm{LT}} = \frac{2G}{c^2}\frac{J}{r^3}\left[-\boldsymbol{\Omega} + \frac{3}{r^2}(\boldsymbol{\Omega}\cdot\mathbf{r})\mathbf{r}\right]. \tag{3.427}$$

where J is the moment of inertia of the rotating sphere. Assuming a sphere of uniform density, its angular momentum can be written as:

$$J = \frac{2}{5} MR^2. \tag{3.428}$$

The precession angular frequency then becomes:

$$\boldsymbol{\omega}_{\mathrm{LT}} = \frac{4GM}{5c^2}\frac{R^2}{r^3}\left[-\boldsymbol{\Omega} + \frac{3}{r^2}(\boldsymbol{\Omega}\cdot\mathbf{r})\mathbf{r}\right] = \frac{4}{5}\frac{r_{\mathrm{g}}}{R}\left(\frac{R}{r}\right)^3\left[-\boldsymbol{\Omega} + \frac{3}{r^2}(\boldsymbol{\Omega}\cdot\mathbf{r})\mathbf{r}\right]. \tag{3.429}$$

The instantaneous precession frequency depends on latitude and decreases as the inverse cube of the distance from the rotating body. A satellite in a polar orbit does not return to the exact same point above the equator after completing one revolution; because rotating Earth drags spacetime along with it, the orbit fails to close, shifting in the equatorial region by on the order of millimeters per revolution. The spin axis of a gyroscope undergoes a corresponding precession. Lense–Thirring frame-dragging was measured from the orbital precession of the LAGEOS 1 and LAGEOS 2 satellites (an acronym for *LAser GEOdynamics Satellites*). Developed jointly by the Italian Space Agency (ASI) and NASA primarily to measure Earth's shape, LAGEOS 1 was launched in 1976 and LAGEOS 2 in 1992. Both satellites are passive laser retroreflectors constructed as brass spheres with an aluminum shell, measuring 60 cm in diameter and weighing approximately 400 kg. Using these satellites, the Lense–Thirring effect was confirmed to within 5% accuracy. In 2004–2005, NASA's Gravity Probe B mission attempted to measure the effect with higher precision using four onboard gyroscopes; however, unexpected electrostatic patch effects on the rotor coatings severely disturbed the spin-axis measurements, complicating the analysis. Today, the Lense–Thirring effect can also be studied using satellites of the European Galileo positioning system.

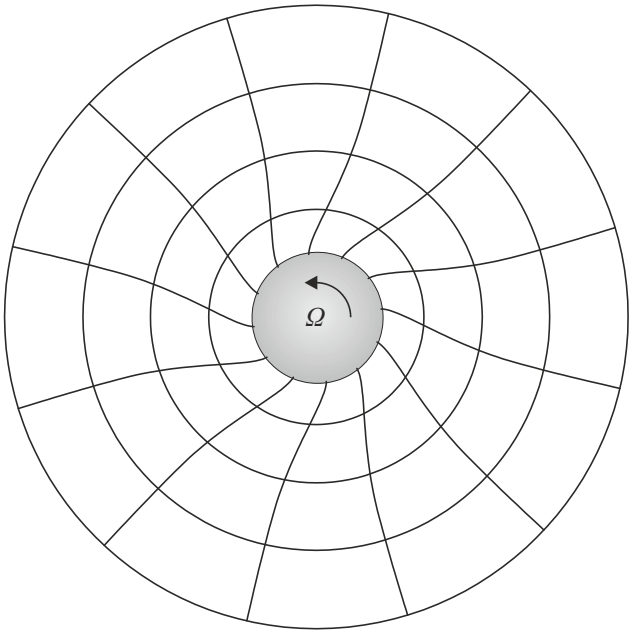


Fig. 3.52: Deformation of spacetime around a rotating body. The off-diagonal components of the metric are responsible for this effect.

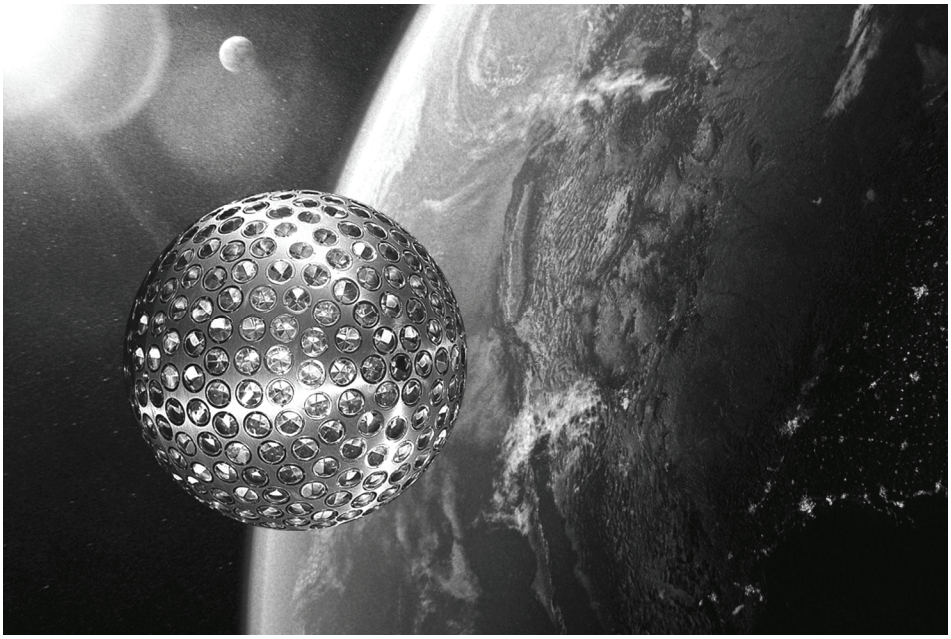


Fig. 3.53: The LAGEOS satellite. Credit: NASA/ASI.

3.7 Brief Introduction to Cosmology

3.7.1 Friedmann Metric

In 1922, the Russian physicist Alexander Friedmann derived from the Einstein Gravity Law the metric of a universe filled with incoherent dust of uniform density. Its spatial part proved to be analogous to the metric on the surface of a sphere, multiplied by a time-dependent scale factor $a(t)$:

► $$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\mathrm{d}\Omega^2\right]; \tag{3.430}$$

► $$\left(\frac{\mathrm{d}a}{\mathrm{d}t}\right)^2 - \frac{8}{3}\pi G\rho a^2 = -c^2k\,. \tag{3.431}$$

The parameter k denotes the Gaussian curvature, which can be positive (as on the surface of a sphere), negative (as on a horse saddle), or zero. In the limit $a \to 1$ and $k \to 0$, the Friedmann metric reduces to the metric of flat spacetime. A similar solution was independently discovered by the Belgian priest and physicist Georges Lemaître and others, leading to the designation FLRW (Friedmann–Lemaître–Robertson–Walker) metric. Crucially, the observer is an integral part of the universe; thus, the Friedmann metric describes the universe in the *comoving coordinates*.

The Friedmann equation for the scale factor has two left-hand terms: a kinetic one driving expansion and a gravitational one favoring contraction. In the observable universe, both terms are large yet nearly equal; consequently, the right-hand side is very small and the spatial curvature lies close to zero.

While the relations for the metric and the scale factor follow directly from the field equations of general relativity, the Friedmann metric notably resembles the metric on the surface of a four-dimensional sphere, and the evolution equation for the scale factor is analogous to energy conservation in a Newtonian homogeneous, isotropic universe.

Metric on the Surface of a 4D Sphere

The surface of a sphere of radius R centered at the origin of Cartesian coordinates satisfies $x^2 + y^2 + z^2 = R^2$, the analogous equation for a four-dimensional sphere reads:

$$x^2 + y^2 + z^2 + w^2 = R^2, \tag{3.432}$$

where w is the fourth spatial coordinate and all four dimensions are mutually orthogonal. By the Pythagorean theorem, the line element in these coordinates is:

$$\mathrm{d}l^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + \mathrm{d}w^2. \tag{3.433}$$

The variable w can be eliminated from this expression, since three coordinates suffice to parameterize the three-dimensional hypersurface. The total differential of (3.432) gives:

$$2x\,\mathrm{d}x + 2y\,\mathrm{d}y + 2z\,\mathrm{d}z + 2w\,\mathrm{d}w = 0 \qquad \Rightarrow$$
$$\mathrm{d}w = -\frac{x\,\mathrm{d}x + y\,\mathrm{d}y + z\,\mathrm{d}z}{w}. \tag{3.434}$$

Substituting this result into (3.433) and expressing w from (3.432) yields:

$$\mathrm{d}l^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 + \frac{\left(x\,\mathrm{d}x + y\,\mathrm{d}y + z\,\mathrm{d}z\right)^2}{R^2 - x^2 - y^2 - z^2}. \tag{3.435}$$

Rewriting this metric in local spherical coordinates r, θ, φ via:

$$x = r\cos\varphi\sin\theta,$$
$$y = r\sin\varphi\sin\theta, \tag{3.436}$$
$$z = r\cos\theta$$

gives the corresponding differentials:

$$\mathrm{d}x = \mathrm{d}r\cos\varphi\sin\theta - r\sin\varphi\,\mathrm{d}\varphi\sin\theta + r\cos\varphi\cos\theta\,\mathrm{d}\theta,$$
$$\mathrm{d}y = \mathrm{d}r\sin\varphi\sin\theta + r\cos\varphi\,\mathrm{d}\varphi\sin\theta + r\sin\varphi\cos\theta\,\mathrm{d}\theta, \tag{3.437}$$
$$\mathrm{d}z = \mathrm{d}r\cos\theta - r\sin\theta\,\mathrm{d}\theta.$$

Substituting (3.436) and (3.437) into the individual terms of (3.435) sequentially yields:

$$\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\varphi^2,$$
$$x\,\mathrm{d}x + y\,\mathrm{d}y + z\,\mathrm{d}z = r\,\mathrm{d}r, \tag{3.438}$$
$$R^2 - x^2 - y^2 - z^2 = R^2 - r^2.$$

Thus, the metric (3.435) of the three-dimensional surface of a four-dimensional sphere expressed in spherical coordinates takes the form:

$$\mathrm{d}l^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\varphi^2 + \frac{r^2\mathrm{d}r^2}{R^2 - r^2} =$$
$$= \frac{\mathrm{d}r^2}{1 - r^2/R^2} + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\,\mathrm{d}\varphi^2. \tag{3.439}$$

Introducing the Gaussian curvature parameter $k = 1/R^2$ casts metric (3.439) into the form:

$$\mathrm{d}l^2 = \frac{\mathrm{d}r^2}{1 - kr^2} + r^2\mathrm{d}\Omega^2. \tag{3.440}$$

The metric on the surface of a 4D sphere is therefore identical in form to the spatial part of the Friedmann metric for a homogeneous, isotropic universe, which is further multiplied by the scale factor $a(t)$ encoding cosmic dynamics.

Newtonian Law of Conservation of Energy

An analogy to the Friedmann equation for the scale factor also exists in classical mechanics. Consider two galaxies separated by a distance R. The motion of galaxy B relative to A depends solely on the matter within a sphere centered at A with radius R = AB. Given a total mass M and average density ρ in this region, energy conservation for galaxy B with M substituted yields:

$$\frac{1}{2}m\left(\frac{\mathrm{d}R}{\mathrm{d}t}\right)^2 - \frac{GmM}{R} = E\,; \qquad M = \rho V = \frac{4}{3}\pi R^3\rho \qquad \Rightarrow \tag{3.441}$$

$$\left(\frac{\mathrm{d}R}{\mathrm{d}t}\right)^2 - \frac{8}{3}\pi G\rho R^2 = \frac{2E}{m}\,. \tag{3.442}$$

This equation is formally identical to the Friedmann equation for the scale factor. Instead of the scale factor a, it features the distance R, and the right-hand side represents twice the total energy per unit mass rather than spatial curvature. Clearly, analogous equations emerge from both general relativity and classical physics.

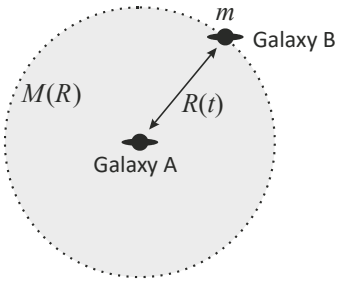


Fig. 3.54: Deriving the law of conservation of energy in a homogeneous, isotropic universe.

Expansion of the Universe

The redshifts of so-called extragalactic nebulae (now known to be galaxies) were measured at the Lowell Observatory by the American astronomer Vesto Slipher between 1912 and 1915. Although he was only a small step away from discovering the expansion of the universe, he unfortunately did not interpret his measurements. The Belgian priest Abbé Lemaître – in full, Georges Henri Joseph Édouard Lemaître – who lectured at the Catholic University of Leuven, went much further. Working from general relativity (similarly to Friedmann), he derived the expansion of the universe and demonstrated, two years prior to Hubble, that the recession velocity of objects measured from a given location increases linearly with distance:

$$v \sim d\,. \tag{3.443}$$

He published his findings in 1927 in the Belgian journal *Annales de la Société Scientifique de Bruxelles*. Based on Slipher's redshift measurements, Lemaître even estimated the value of the Hubble constant before Hubble introduced it. Because he lacked accurate distance measurements, his estimate (like Hubble's later value) came out to roughly 500 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$ – an order of magnitude higher than today's accepted value. Lemaître

envisioned that the universe originated from a small, hot, dense region, which he termed the "primeval atom" or "cosmic egg." He can therefore be regarded not only as the discoverer of the expansion of the universe, but also as the originator of the Big Bang theory. Because the local journal in which he published was not widely read among astronomers, many remained unaware of his results. The existence of galactic spectral redshifts due to cosmic expansion was also independently predicted from general relativity by the American mathematician and physicist Howard Percy Robertson in 1928. In 1929, Edwin Hubble published a paper pointing out, just as Lemaître had, the direct proportionality between the recession velocity of objects and their distance. Today, this fundamental relation is known as the Hubble–Lemaître law. The constant of proportionality is called the Hubble constant, i.e.,

► $$v = Hd\,; \qquad H \approx 70\ \mathrm{km\,s^{-1}Mpc^{-1}}\,. \tag{3.444}$$

The Hubble constant is a spatial constant, meaning we assume it is uniform throughout the universe at any given moment. Over time, however, it has changed throughout the history of the universe. A rigorous analysis of all possible modes of cosmic expansion based on mathematical group theory was later carried out by the English mathematician Arthur Geoffrey Walker in 1936. The Hubble constant can also be expressed as the ratio of the time derivative of the scale factor to the scale factor itself:

► $$H = \frac{\dot{a}}{a}\,. \tag{3.445}$$

The Friedmann equation for the scale factor (3.431) can be readily rewritten in the form:

► $$H^2 - \frac{8}{3}\pi G\rho = -\frac{c^2 k}{a^2}\,. \tag{3.446}$$

If the first (kinetic) term dominates, the universe will expand forever, and the curvature on the right-hand side must be negative. If the negative (gravitational) term dominates, the universe will eventually recollapse into a small spatial region, and its curvature is positive. If both terms are exactly balanced, the expansion will continue indefinitely while asymptotically decelerating toward zero (analogous to a stone thrown upward at escape velocity). This limiting case corresponds to zero curvature and the so-called critical density:

► $$\rho_\mathrm{c} = \frac{3H^2}{8\pi G}. \tag{3.447}$$

Based on the currently accepted value of the Hubble constant, the critical density corresponds to only a few protons per cubic meter. The actual density of the universe is very close to this critical value.

Today, the Hubble constant is measured primarily via two fundamental methods. The first relies on analyzing fluctuations in the cosmic microwave background (CMB). Data from the Planck space observatory yield a value of 67.4 ± 0.5 km s^{-1} Mpc^{-1} (an uncertainty of less than 1 %). The second method is based on local distance measurements using Type Ia supernovae calibrated with Cepheid variable stars. Projects in this category – such as the SH0ES program using the Hubble Space Telescope and the James Webb Space Telescope (JWST) – yield values around 73.0 to 73.5 km s^{-1} Mpc^{-1} with an uncertainty of roughly 1.1–1.4 %. The persistent discrepancy between these

early- and late-universe measurements, which now exceeds the 5σ statistical threshold, is known as the Hubble tension and remains one of the greatest unresolved puzzles in modern cosmology.

Other independent measurement techniques also exist. Wendy Freedman's group at the University of Chicago, utilizing high-resolution JWST observations of red giant and carbon stars, obtained a value of 70.4 km s^{-1} Mpc^{-1} (with an uncertainty of 3 %), lying right in the middle of the two extremes. Regardless of the method used, all measured values cluster around 70 km s^{-1} Mpc^{-1}. Another promising way to determine distances – crucial for pinning down the Hubble constant – involves kilonovae, the colossal gravitational-wave and optical transients produced by the merger of two neutron stars. Yet another approach relies on observing multiple images of supernovae and quasars produced by strong gravitational lensing. By measuring the time delays between brightness variations in the individual lensed images, astronomers can directly constrain the parameters of cosmic expansion. This phenomenon was predicted and the method proposed as early as 1964 by the Norwegian theorist Sjur Refsdal. Magnified multiple images of a distant supernova were first observed in 2014 and 2015, and modern facilities such as the newly operational Vera C. Rubin Observatory are now set to dramatically expand these datasets and refine measurements of the Hubble constant.

3.7.2 Properties of the Friedmann Metric

Volume of the Universe

For a universe with positive spatial curvature ($k > 0$), setting $k = 1/R^2$ allows using the substitution:

$$r = R\sin\chi\,; \qquad \chi \in (0,\pi)\,. \tag{3.448}$$

The spatial part of the metric then takes the form:

$$\begin{aligned} \mathrm{d}l^2 &= a^2R^2\left(\mathrm{d}\chi^2 + \sin^2\chi\,\mathrm{d}\Omega^2\right) = \\ &= a^2R^2\mathrm{d}\chi^2 + a^2R^2\sin^2\chi\sin^2\theta\,\mathrm{d}\varphi^2 + a^2R^2\sin^2\chi\,\mathrm{d}\theta^2\,. \end{aligned} \tag{3.449}$$

Consequently, the spatial part of the metric tensor is:

$$\begin{pmatrix} a^2R^2 & 0 & 0 \\ 0 & a^2R^2\sin^2\chi\sin^2\theta & 0 \\ 0 & 0 & a^2R^2\sin^2\chi \end{pmatrix} \tag{3.450}$$

The total volume follows directly from the metric:

$$\begin{aligned} V &= \int\sqrt{\det g}\;\mathrm{d}\chi\,\mathrm{d}\theta\,\mathrm{d}\varphi = \\ &= a^3R^3\left(\int_0^{\pi}\sin^2\chi\,\mathrm{d}\chi\right)\cdot\left(\int_0^{2\pi}\mathrm{d}\varphi\right)\cdot\left(\int_0^{\pi}\sin\theta\,\mathrm{d}\theta\right) = . \\ &= a^3R^3\cdot\frac{\pi}{2}\cdot 2\pi\cdot 2 = 2\pi^2a^3R^3\,. \end{aligned} \tag{3.451}$$

A universe with positive curvature thus has a finite volume; it is closed upon itself and, although unbounded (having no edge), is finite in volume. The situation differs for a universe with negative curvature ($k < 0$; $R^2=1/|k|$), where the metric takes the form:

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1+(r/R)^2} + r^2\mathrm{d}\Omega^2\right] \tag{3.452}$$

Here, it is advantageous to use the substitution:

$$\begin{aligned} r &= R\sinh\chi\,; \\ \chi &\in (0,\infty)\,, \end{aligned} \tag{3.453}$$

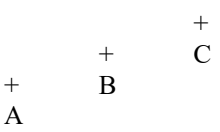
which transforms the metric into

$$\mathrm{d}l^2 = a^2R^2\left(\mathrm{d}\chi^2 + \sinh^2\chi\ \mathrm{d}\Omega^2\right). \tag{3.454}$$

Following the same integration procedure now yields $V \to \infty$. Within the Friedmann geometry, positively curved spaces have a finite volume, whereas models with negative or zero curvature have an infinite volume. However, this holds strictly for simply connected spaces ("without holes"). A space whose two-dimensional analog is a torus or double torus is not simply connected; relaxing simple connectivity allows a universe with constant negative curvature to possess a finite volume.

Cosmological Principle

The fact that cosmic expansion obeys the Hubble–Lemaître law (velocity proportional to distance) does not imply that Earth occupies the center of expansion. Rather, this linear relation ensures that every observer anywhere in the universe perceives the exact same isotropic expansion – a cornerstone known as the *cosmological principle*. Suppose three points A, B, and C lie along a straight line such that $AB = BC$.



Large-scale homogeneity of the universe implies $\mathbf{v}_{BA} = \mathbf{v}_{CB}$. For the relative velocity $\mathbf{v}_{CA}$, it follows that:

$$\mathbf{v}_{CA} = \mathbf{v}_C - \mathbf{v}_A = \mathbf{v}_C - \mathbf{v}_B + \mathbf{v}_B - \mathbf{v}_A = \mathbf{v}_{CB} + \mathbf{v}_{BA} = 2\,\mathbf{v}_{BA}. \tag{3.455}$$

This immediately yields the linear relation:

$$v = Hd\,. \tag{3.456}$$

A linear dependence of recession velocity on distance is the unique relation that holds simultaneously at every location in the universe. Any observer at any position observes expansion governed by the Hubble–Lemaître law, leaving no preferred center or location in the universe.

3.7.3 Cosmological Redshift

Consider the propagation of a light pulse through the universe and its evolution under the Friedmann metric (3.430). For light, $ds^2 = 0$; aligning the coordinate system so that angular terms vanish yields:

$$\frac{c\,dt}{a(t)} = \frac{dr}{\sqrt{1-kr^2}} \tag{3.457}$$

Suppose the pulse is emitted from coordinate r_e at time t_e with duration δt_e, and reaches an observer billions of years later at coordinate r_n at time t_n with observed duration δt_n.

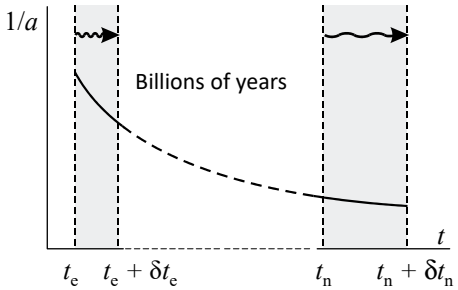


Fig. 3.55: Cosmological redshift.

The propagation of the leading edge of the pulse satisfies

$$\int_{t_e}^{t_n} \frac{c\,dt}{a(t)} = \int_{r_e}^{r_n} \frac{dr}{\sqrt{1-kr^2}} \tag{3.458}$$

while the trailing ("last") photon of the pulse satisfies:

$$\int_{t_e+\delta t_e}^{t_n+\delta t_n} \frac{c\,dt}{a(t)} = \int_{r_e}^{r_n} \frac{dr}{\sqrt{1-kr^2}}\,. \tag{3.459}$$

Subtracting these two relations yields the change in pulse duration:

$$\int_{t_e+\delta t_e}^{t_n+\delta t_n} \frac{c\,dt}{a(t)} - \int_{t_e}^{t_n} \frac{c\,dt}{a(t)} = 0\,. \tag{3.460}$$

Although both integrals span vast cosmic intervals over which the scale factor $a(t)$ changes substantially, interpreting the integral as the area under the curve $1/a(t)$ shows that the overlapping area cancels out, leaving:

$$\int_{t_n}^{t_n+\delta t_n} \frac{c\,dt}{a(t)} - \int_{t_e}^{t_e+\delta t_e} \frac{c\,dt}{a(t)} = 0 \tag{3.461}$$

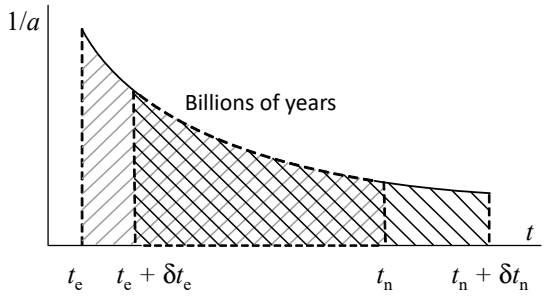


Fig. 3.56: Integration scheme for signal propagation.

These remaining integrals span only the brief duration of the pulse rather than the entire travel time. Because the pulse duration is negligibly small compared to the age of the universe, the scale factor $a(t)$ can be treated as constant during emission and likewise constant during reception:

$$\int_{t_n}^{t_n+\delta t_n} \frac{c\,\mathrm{d}t}{a_n} - \int_{t_e}^{t_e+\delta t_e} \frac{c\,\mathrm{d}t}{a_e} = 0\,. \tag{3.462}$$

Integrating constants yields a simple relation:

$$\begin{aligned} \frac{c\,\delta t_n}{a_n} - \frac{c\,\delta t_e}{a_e} &= 0 \quad \Rightarrow \\ \frac{c\,\delta t_e}{a_e} &= \frac{c\,\delta t_n}{a_n} \quad \Rightarrow \\ \frac{\lambda_e}{a_e} &= \frac{\lambda_n}{a_n}\,. \end{aligned} \tag{3.463}$$

An emitted pulse is thus stretched in direct proportion to the growth of the scale factor. In short, as radiation traverses the expanding universe, its wavelength stretches at the exact same rate at which space expands. To quantify this wavelength shift, the cosmological redshift z is defined as:

► $$z \equiv \frac{\lambda - \lambda_e}{\lambda_e} = \frac{a - a_e}{a_e}\,. \tag{3.464}$$

- The highest spectroscopically confirmed cosmological redshifts reach values around $z \approx 14$ (with photometric candidates extending slightly higher).
- Cosmological redshift is frequently quoted instead of object distance, which is practically advantageous because redshifts are directly measured observables.
- Wavelength stretching synchronous with the scale factor applies strictly to massless particles ($ds^2 = 0$); de Broglie matter waves of particles with nonzero rest mass do not obey this simple scaling.
- Cosmological redshift must not be confused with the Doppler effect; only to first order in a Taylor expansion do these phenomena yield the same formula.

3.7.4 Scale Factor

Scale Factor for Various Constituents

The energy or mass density of ordinary matter decreases as the inverse cube of the scale factor. Radiation energy density decreases as the inverse fourth power because:

$$\rho_{\text{rad}} = \frac{E}{V} \sim \frac{1/\lambda}{a^3} \sim \frac{1/a}{a^3} \sim \frac{1}{a^4} . \tag{3.465}$$

The energy density of quantum vacuum fluctuations is completely independent of the scale factor. Suppose the universe is filled with a general constituent scaling as:

$$\rho_{\text{E}} = \frac{K}{a^{\alpha}} , \tag{3.466}$$

Neglecting the small curvature term on the right-hand side of the Friedmann equation (3.431) yields:

$$\dot{a} \sim a^{\frac{2-\alpha}{2}} . \tag{3.467}$$

Separating variables gives:

$$\int \frac{\mathrm{d}a}{a^{(2-\alpha)/2}} = \int \mathrm{d}t . \tag{3.468}$$

Simple integration yields:

► $$a(t) \sim \begin{cases} t^{\frac{2}{\alpha}} , & \alpha \neq 0 , \\ \mathrm{e}^{\chi t} , & \alpha = 0 . \end{cases} \tag{3.469}$$

For a radiation-dominated universe (the early epoch, massless particles), $a(t) \sim t^{1/2}$; for a matter-dominated universe (from roughly 400,000 yr onward, massive particles), $a(t) \sim t^{2/3}$; and for a universe dominated by quantum vacuum fluctuations (roughly since 7 billion yr after the Big Bang), $a(t) \sim \mathrm{e}^{\chi t}$. From the definition $H \equiv \dot{a}/a$, the Hubble parameter H for each expansion regime is:

No expansion:	$a = 1$	$H = 0$
Radiation:	$a \sim t^{1/2}$	$H = 1/2t$
Matter:	$a \sim t^{2/3}$	$H = 2/3t$
Vacuum:	$a \sim \mathrm{e}^{\chi t}$	$H = \chi$

In radiation- and matter-dominated universes, the Hubble parameter decreases inversely with cosmic time. The reciprocal of the present Hubble constant is termed the *Hubble time* (an order-of-magnitude estimate of the age of the universe obtained by approximating expansion as linear).

Particle Horizon

The particle horizon is the maximum distance from which particles could have traveled to an observer during the entire age of the universe. This distance is also referred to as the *horizon of the observable universe*, as it is impossible to see beyond it even though matter exists at greater distances – its light simply has not yet had enough time to reach the observer. Considering photons (which travel at the maximum possible speed), light propagation in comoving coordinates is governed by:

$$-c^2\mathrm{d}t^2 + a^2(t)\frac{\mathrm{d}r^2}{1-kr^2} = 0\,. \tag{3.470}$$

Separating variables and integrating over the entire history of the universe yields:

$$\int_0^t \frac{c\mathrm{d}t'}{a(t')} = \int_{r_0}^{r_t} \frac{\mathrm{d}r}{\sqrt{1-kr^2}}\,. \tag{3.471}$$

The proper distance to the particle horizon is given by:

$$R_\mathrm{H}(t) = \int_{r_0}^{r_t} \mathrm{d}l = \int_{r_0}^{r_t} \frac{a(t)\mathrm{d}r}{\sqrt{1-kr^2}} = a(t)\int_{r_0}^{r_t} \frac{\mathrm{d}r}{\sqrt{1-kr^2}}\,. \tag{3.472}$$

Substituting the integral from (3.471) gives:

► $$R_\mathrm{H}(t) = a(t)\int_0^t \frac{c\mathrm{d}t'}{a(t')}\,. \tag{3.473}$$

The particle horizon depends on the history of the scale factor. Because the universe was mostly matter-dominated, taking an age of ≈ 14 billion years yields $R_\mathrm{H} = 3\,ct \approx$ ≈ 42 billion ly (a more precise ΛCDM integration gives approximately 46 billion ly).

Equation-of-State Parameter *w*

Consider the equation of state for a cosmic constituent whose energy density scales as K/a^α, while volume scales as $V \sim a^3$. Applying the first law of thermodynamics in the adiabatic approximation (the universe exchanges no heat with any exterior) gives:

$$\mathrm{d}U + p\,\mathrm{d}V = 0$$

$$\mathrm{d}\left(\rho_\mathrm{E}V\right) + p\,\mathrm{d}V = 0$$

$$\mathrm{d}\left(\frac{K}{a^\alpha}a^3\right) + p\,\mathrm{d}a^3 = 0$$

$$(3-\alpha)Ka^{2-\alpha}\mathrm{d}a + 3pa^2\mathrm{d}a = 0$$

$$(3-\alpha)\frac{K}{a^\alpha} + 3p = 0$$

$$p = \left(\frac{\alpha}{3} - 1\right)\frac{K}{a^{\alpha}},$$

$$p = w\rho_{\mathrm{E}}\,; \qquad w \equiv \left(\frac{\alpha}{3} - 1\right). \tag{3.474}$$

This yields a linear relation between pressure and energy density parameterized by w, one of the fundamental cosmological parameters. The results derived thus far are summarized in the following table:

Exponent α ($\rho \sim 1/a^{\alpha}$)	Scale Factor	Hubble Parameter	Particle Horizon	Equation of State	Parameter w	Constituent
$\alpha = 4$	$C\,t^{1/2}$	$\frac{1}{2t}$	$2ct$	$p = \rho/3$	$+\,1/3$	Radiation ($m_0 = 0$)
$\alpha = 3$	$C\,t^{2/3}$	$\frac{2}{3t}$	$3ct$	$p = 0$	0	Matter ($m_0 \neq 0$)
$\alpha = 2$	$C\,t$	$\frac{1}{t}$	$ct\,\ln(t/t_0)$	$p = -\rho/3$	$-1/3$	Curvature ($k < 0$)
$\alpha = 0$	$C\,\mathrm{e}^{\chi t}$	χ	$c\left(\frac{\mathrm{e}^{\chi t} - 1}{\chi}\right)$	$p = -\rho$	-1	Vacuum Energy

Of particular significance is the value of w for dark energy. According to general relativity, driving the observed accelerated expansion of the universe requires $w < -1/3$. For vacuum energy associated with quantum mechanical vacuum processes, $w = -1$ and the scale factor grows exponentially. If $w < -1$ ($\alpha < 0$, known as phantom energy), expansion would accelerate so violently that it would eventually tear apart bound structures down to atomic nuclei – a scenario termed the Big Rip. Observations from WMAP, CBI, 2dF, SDSS, Planck, and numerous other experiments indicate that the dark energy equation-of-state parameter lies very close to $w = -1$.

3.7.5 Cosmological Constant

In 1998, the accelerated expansion of the universe was discovered (by Adam Riess, Brian Schmidt, and Saul Perlmutter), prompting the reintroduction of the cosmological term with a nonzero cosmological constant Λ on the left-hand side of the Einstein field equations:

$$R^{\alpha\beta} - \tfrac{1}{2}Rg^{\alpha\beta} + \Lambda g^{\alpha\beta} = \frac{8\pi G}{c^4}T^{\alpha\beta}. \tag{3.475}$$

The fundamental origin of the cosmological term remains unknown, though it may be linked to quantum vacuum fluctuations satisfying:

$$\left\langle T^{\mu\nu} \right\rangle = \rho_{\text{vakua}}\, g^{\mu\nu} \,, \tag{3.476}$$

in which case the cosmological term on the left-hand side would originate from the stress-energy tensor of the quantum vacuum. The Friedmann equation for the scale factor then becomes:

$$\dot{a}^2 - \frac{8}{3}\pi G\rho a^2 - c^2 \frac{\Lambda}{3} a^2 = -c^2 k \,. \tag{3.477}$$

Dividing by a^2, introducing the Hubble parameter H, and separating total energy density into radiation and matter constituents yields:

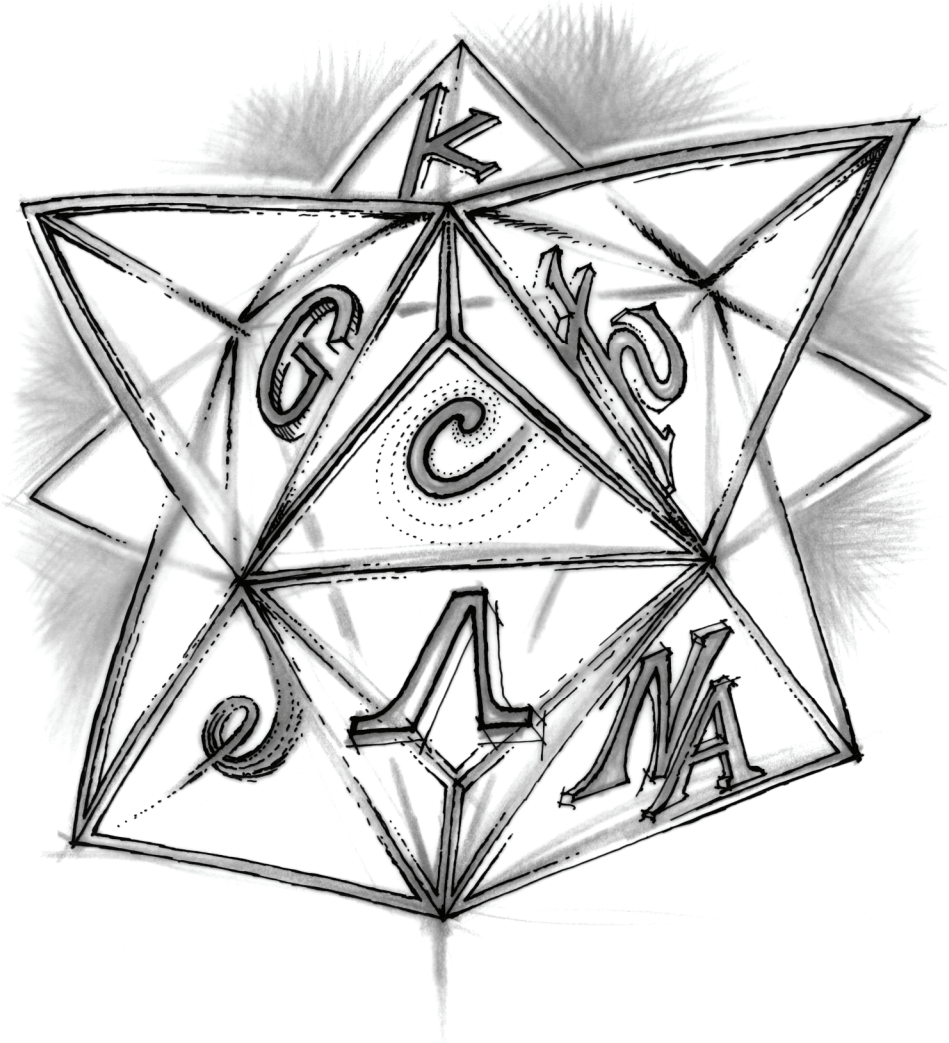
$$H^2 = \frac{8}{3}\pi G\rho_{\text{rad}} + \frac{8}{3}\pi G\rho_{\text{mat}} + c^2 \frac{\Lambda}{3} - \frac{c^2 k}{a^2} \,. \tag{3.478}$$

The terms on the right-hand side respond differently to cosmic expansion: the first scales as a^{-4}, the second as a^{-3}, the third as a^0 (the cosmological term), and the final curvature term as a^{-2} (which is nearly zero in our universe). Assuming $H \neq 0$, dividing the entire equation by H^2 yields the compact dimensionless relation:

$$\begin{aligned} 1 &= \Omega_{\text{rad}} + \Omega_{\text{mat}} + \Omega_\Lambda + \Omega_{\text{k}} \,; \\ \Omega_{\text{rad}} &\equiv \frac{8\pi G}{3H^2}\rho_{\text{rad}} \,; \\ \Omega_{\text{mat}} &\equiv \frac{8\pi G}{3H^2}\rho_{\text{mat}} \,; \\ \Omega_\Lambda &= \frac{c^2\Lambda}{3H^2} \,; \\ \Omega_{\text{k}} &= -\frac{c^2 k}{a^2 H^2} \,. \end{aligned} \tag{3.479}$$

According to current observations, the curvature contribution is negligible. The radiation density parameter today is approximately $\Omega_{\text{rad}} \approx 0.01$ %, while matter contributes $\Omega_{\text{mat}} \approx 32$ % (comprising roughly 5 % ordinary baryonic matter and 27 % dark matter), and the cosmological constant (dark energy) accounts for approximately $\Omega_\Lambda \approx 68$ %. The cosmological constant itself has a value of roughly $\Lambda \approx 1.1\times10^{-52}$ m^{-2} (corresponding to a vacuum energy density). The radiation-dominated era lasted until roughly 400,000 yr after the Big Bang, the matter-dominated era persisted until roughly half the current age of the universe, and the universe is currently in the dark-energy-dominated era. Detailed treatments of these cosmic constituents and their observational probes can be found in specialized textbooks on cosmology.

List of Symbols



a	scale factor, magnitude of acceleration, distance
a^{μ}	four-acceleration
$\mathbf{a}$	acceleration
a_{e}	classical electron radius
$\mathbf{a}_{\mathbf{k}}$	potential amplitude
$\hat{\mathbf{a}}_k^{\dagger}$	creation operator
$\hat{\mathbf{a}}_k$	annihilation operator
asinh	inverse hyperbolic sine
arctan	arctangent
atanh	inverse hyperbolic tangent
A	amplitude, dynamical variable, work
A^{μ}	four-potential
$\mathbf{A}$	vector potential
$\mathcal{A}$	four-vector
b	Wien displacement constant, distance
$\mathbf{b}$	angular momentum
$\mathbf{B}$	magnetic field
$\mathcal{B}$	radiation magnetic field
c	specific heat capacity, speed of light
cosh	hyperbolic cosine
cos	cosine
cot	cotangent
coth	hyperbolic cotangent
curl	curl operator
C	heat capacity
C	field circulation
$\mathbf{D}$	electric displacement
$\mathrm{D}/\mathrm{D}\tau$	absolute derivative
det	determinant
div	divergence
$\mathrm{d}s^2$	spacetime interval
e	Euler number
e	elementary charge
$\mathbf{e}_k$	unit vector
exp	exponential function
E	energy
E_{nN}	energy of state n for N systems
$\mathbf{E}$	electric field
$\mathcal{E}$	radiation electric field
f	force density, free energy density
F	Helmholtz free energy
F, $\mathbf{F}$	force
F_k	force components
$F^{\mu\nu}$	electromagnetic field tensor (field strength tensor)
g	Landé g-factor, gravitational acceleration
$\mathbf{g}$	particle momentum density, gravitational acceleration
g_n	degree of degeneracy (degeneracy)
$g_{\mu\nu}$	metric tensor
G	Gibbs potential (Gibbs free energy), gravitational constant, Green function
$G_{\mu\nu}$	Einstein tensor
h	Planck constant
$\hbar$	reduced Planck constant
$h_{\mu\nu}$	Minkowski metric perturbations
$\mathbf{H}$	magnetic field strength
H	enthalpy, Hamiltonian, Hubble constant
$\mathcal{H}$	radiation Hamiltonian
$\hat{H}$	energy operator
i	linear current density
I	electric current, intensity, radiation intensity
$\mathbf{j}$	flux
$\mathbf{j}_Q$	current density (charge current density)
j^{μ}	four-flux
J	total angular momentum, moment of inertia, coupling constant
J^{μ}	four-current density
k	curvature, wave number
$\mathbf{k}$	wave vector
k_{B}	Boltzmann constant
k^{μ}	four-wavevector

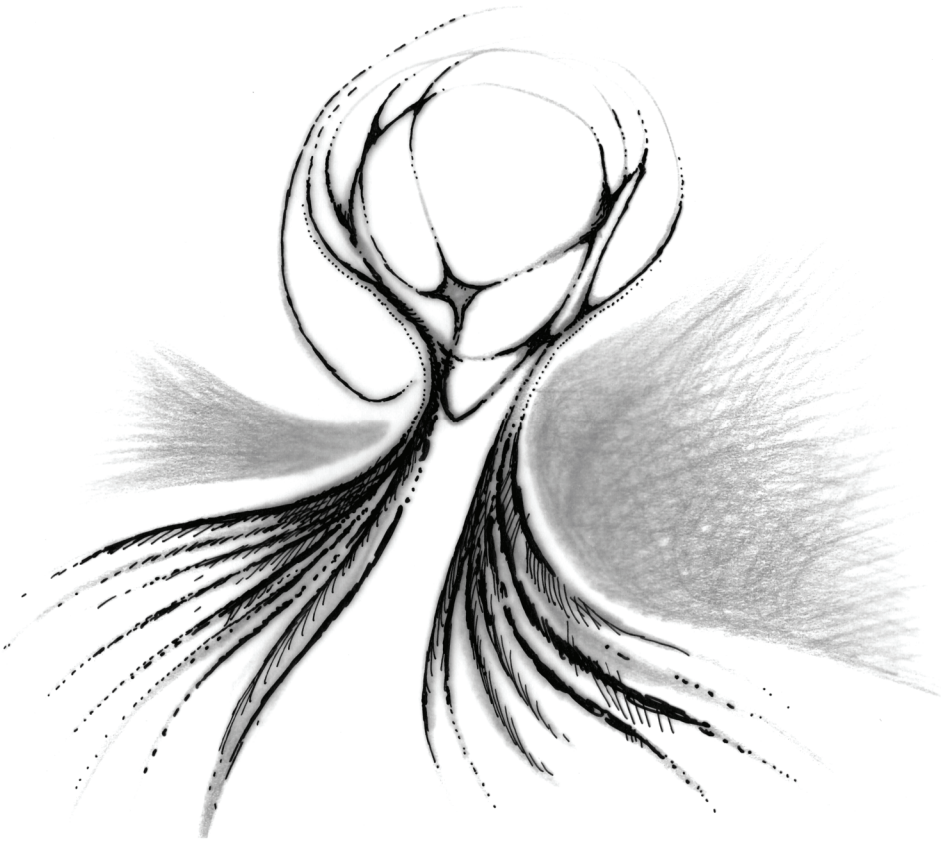
l distance
orbital quantum number
ℓ angular momentum per unit mass
ln natural logarithm
L Lagrangian, orbital angular momentum
$\mathscr{L}$ Lagrangian density, Laplace transform
$\mathbf{L}$ angular momentum vector
LIF local inertial frame
m mass, magnetic quantum number
m_s inertial mass
m_g gravitational mass
$\mathbf{M}$ torque, magnetization vector
M mass, magnetization
$\mathcal{M}_{kl}$ angular momentum density
$\mathcal{M}_U$ closed set
n principal quantum number, index of refraction, number density, quantum state
$\mathbf{n}$ normal vector
N number of particles, number of parameters, number of systems
N_k occupation numbers
$\mathbf{N}_k$ generators of the Lorentz group
$\mathcal{N}_{klm}$ angular momentum flux
$\mathbf{p}$ momentum vector
p momentum, pressure, generalized momentum
p_{nN} pressure of state n for N systems
$\mathbf{p}_E$ electric dipole moment
$\mathbf{p}_G$ gravitational dipole moment
$\mathbf{p}_M$ magnetic dipole moment
p^μ four-momentum
$\tilde{\mathbf{I}}$ identity matrix
$\mathbf{P}$ polarization vector
P polarization, power
P^μ four-momentum
$\mathbf{P}_k$ generalized field momentum in a cavity
$\mathscr{P}$ power density
$\mathcal{P}$ radiated power
P_{nk} momentum matrix
$\hat{\mathbf{P}}$ momentum operator
$\mathbf{q}$, q_k generalized coordinates
Q charge, number of states (Potts model), heat, generalized coordinate
$\mathbf{Q}_k$ generalized field coordinate in a cavity
Q_{kl} electric quadrupole moment
r radial distance, reflectivity amplitude, radius
r_g Schwarzschild radius
$\mathbf{r}$ position vector
R reflectivity, Ricci scalar, scalar curvature
$R_{\alpha\beta}$ Ricci tensor
$R^\alpha{}_{\beta\gamma\delta}$ Riemann curvature tensor
s entropy density
S entropy, surface, spin angular momentum
$\mathbf{S}$ Poynting vector
$\mathscr{S}$ radiation part of the Poynting vector
sinh hyperbolic sine
sin sine
t transmission amplitude, time, parameter, Hubbard coupling constant
T absolute temperature, kinetic energy, transmissivity, period, temperature
T_{kl} momentum flux tensor (stress tensor)
T_{kl}^{EM} Maxwell stress tensor
T_V vibrational temperature
T_R rotational temperature
Tr trace

$T^{\mu\nu}$	energy-momentum tensor (stress-energy tensor)
u	internal energy density, rapidity, velocity
$\mathbf{u}$	velocity field
U	voltage (potential difference), real part of a complex function, internal energy
U^{μ}	four-velocity
$\mathbf{v}$, v	velocity
v_{f}	phase velocity
v_{g}	group velocity
var	variance
V	volume, potential energy, imaginary part of a complex function
V_{eff}	effective potential
w	probability, equation of state parameter
w_{nN}	energy of state n for N systems
W	energy
W_{P}	potential energy
x_k	coordinates
$\mathbf{x}$	position vector
x^{μ}	spacetime event
$\hat{\mathbf{X}}$	position operator
$\hat{x}$	coordinate operator
y	coordinate
z	complex variable, single-particle partition function coordinate
Z	partition function, ionization state
α_k	constants of integration
β	reciprocal temperature $1/k_{\mathrm{B}}T$, relativistic velocity ratio v/c
γ	density of states (DOS), integration path (contour), Lorentz factor
$\boldsymbol{\gamma}$	field momentum density
Γ	statistical weight (factor)
$\Gamma^{\alpha}_{\beta\gamma}$	Christoffel symbols, affine connection
δ	Dirac delta function, Kronecker delta
δ_A	dispersion of A (relative fluctuation)
∂S	boundary of the surface S
$\partial\Omega$	boundary of the set Ω
Δ	finite increment, Laplacian (∇^2)
ε	energy of a single state, Levi-Civita tensor, permittivity
ε_0	vacuum permittivity
ε_{kl}	permittivity tensor
ζ	fugacity
$\eta_{\mu\nu}$	Minkowski metric
θ, ϑ	polar angle
θ	angle between acceleration and the normal
κ_{kl}	electric susceptibility
λ	eigenvalue, worldline parameter, wavelength
Λ	cosmological constant
$\boldsymbol{\Lambda}$	Lorentz transformation matrix
μ	chemical potential, permeability
μ_{kl}	permeability tensor
ν	frequency
ξ	auxiliary variable
Ξ	grand partition function
$\boldsymbol{\pi}$	total momentum density
ρ	density, probability density, proper radial distance, distance
ρ_{c}	critical density
ρ_E	energy density
ρ_M	mass density
ρ_Q	charge density
$\hat{\rho}$	density operator
σ	differential conductivity, surface charge density, Stefan-Boltzmann constant (blackbody radiation), cross section

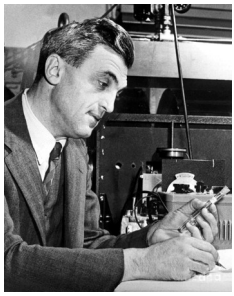
τ	common retarded time, proper time
φ	azimuthal angle, phase, angle
ϕ	phase, phase space, gravitational potential, scalar potential
ϕ^*	dimensionless gravitational potential
χ	deflection angle
χ_{kl}	magnetic susceptibility
ψ	field flux, wave function
ω	angular frequency
ω_c	cyclotron frequency
$\boldsymbol{\omega}_{LT}$	Lense-Thirring precession frequency
Ω	spacetime region, grandcanonical potential (grand potential or Landau potential), cosmological density parameter (omega parameter), spatial region, solid angle
$\boldsymbol{\Omega}$	angular rotation frequency
$\boldsymbol{\nabla}$	gradient
$\Box$	d'Alembertian



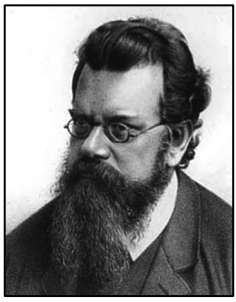
Biographical Index



Statistical Physics



Bloch, Felix (1905–1983), Bloch, Felix (1905–1983), a Swiss-born American physicist, shared the 1952 Nobel Prize in Physics with Edward Purcell for developing nuclear magnetic resonance (NMR) to measure nuclear magnetic fields. Bloch taught at the University of Leipzig until 1933, when Adolf Hitler rose to power, prompting his emigration to the United States, where he became a citizen in 1939. Joining the faculty at Stanford University in 1934, he proposed a method to split neutron beams into two components based on their spin orientations in a magnetic field. In 1939, he determined the magnetic moment of the neutron alongside Luis Alvarez. During World War II, he contributed to radar defense systems at Harvard and worked on the Manhattan Project. Returning to Stanford, he applied NMR to analyze the relationship between nuclear magnetic properties and crystal structures. This technique later became foundational for medical MRI. Bloch also served as the first Director-General of CERN (1954–1955).



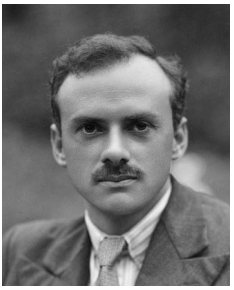
Boltzmann, Ludwig Eduard (1844–1906), an Austrian physicist, was a founding father of statistical mechanics. In 1872, he formulated the relationship between entropy and statistical probability. He developed the H-theorem to explain the irreversible increase of entropy and made pioneering contributions to the kinetic theory of gases. Boltzmann viewed the equipartition theorem as a cornerstone of kinetic theory and was a staunch defender of atomic theory. In 1869, he was appointed professor at the University of Graz. His profound legacy is honored through numerous eponymous concepts, including the Boltzmann transport equation for probability density evolution, the Boltzmann constant (k_B), the Maxwell–Boltzmann distribution, and a lunar crater named in his memory.



Bose, Satyendra Nath (1894–1974), an Indian physicist, specialized in quantum statistics and mathematical physics. Particles with integer spin, known as bosons, were named in his honor by Paul Dirac, as was the Bose-Einstein statistics that govern them. Bose also researched X-ray crystallography, ionospheric properties, and unified field theory. In 1924, after a journal rejected his breakthrough paper deriving Planck radiation law using identical photons, he sent it to Albert Einstein. Recognizing its genius, Einstein translated the paper into German and ensured its publication. Einstein extended these concepts to massive particles, leading to the prediction of the Bose-Einstein distribution and Bose-Einstein condensate.

Curie, Pierre (1859–1906), a French physicist and chemist, co-discovered the radioactive elements polonium and radium in pitchblende alongside his wife, Marie Curie, in 1898. He discovered that ferromagnetic materials lose their permanent magnetic properties above a critical threshold, now known as the Curie temperature. Furthermore, he discovered the piezoelectric effect, demonstrating that mechanical deformation of cer-

tain crystals generates an electric potential. Curie studied at the Sorbonne and later became a professor there. In 1903, he received the Davy Medal from the Royal Society and shared the Nobel Prize in Physics with Marie Curie and Henri Becquerel. The synthetic element curium was later named in honor of the Curies. Pierre Curie died tragically in a street accident after being struck by a horse-drawn carriage in Paris.



Dirac, Paul Adrien Maurice (1902–1984), an English physicist, was a chief architect of 20th-century quantum theory. He founded quantum electrodynamics and quantum field theory. In 1928, he derived the Dirac equation – a relativistic equation for the electron. He correctly interpreted negative energy states as antiparticles, predicting the positron, which Carl Anderson discovered in 1932. Independently of Enrico Fermi, he derived the statistical distribution for fermions (Fermi-Dirac statistics). He introduced bra-ket notation in quantum mechanics and proved the equivalence of Schrödinger's and Heisenberg's formulations. Dirac developed second quantization, enabling the transition from quantum particles to field theory. He predicted vacuum polarization, non-trivial dynamic vacuum properties from quantum effects, and magnetic monopoles. He formulated the many-body formalism in quantum theory. For his work on antiparticles and quantum field theory, he received the 1933 Nobel Prize in Physics. Later, he explored the implications of hypothetical variations in fundamental constants (gravitational, light speed, Planck). Throughout his life, he championed the mathematical beauty and simplicity of physical equations, being among the first to realize that nature's symmetries are the primary principle for constructing physical laws.



Doppler, Christian (1803–1853), an Austrian mathematician and physicist, spent part of his brief life as a professor at the Prague Polytechnic (predecessor to CTU in Prague) before lecturing at the Vienna Polytechnic. He is primarily famous for discovering the change in wave frequency during the relative motion of a source and an observer (the Doppler effect). In statistical physics, the Doppler effect causes the thermal Doppler broadening of spectral lines. With the advent of relativity, the relativistic Doppler effect was derived. He also published works on electricity and magnetism, studied the temporal variations of magnetic declination, and wrote several papers on optics and astronomy. A crater on the far side of the Moon and asteroid 3905 Doppler are named in his honor.

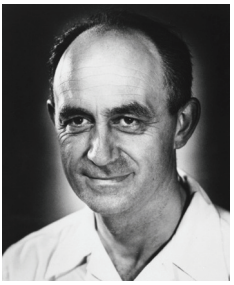


Dulong, Pierre Louis (1785–1838), a French physicist and chemist, investigated the elastic properties of steam, heat conduction, and specific heat capacities of gases. He also measured refractive indices of various gases. Collaborating closely with Alexis Petit, they experimentally discovered the Dulong–Petit law regarding the specific heat capacity of crystalline solids in 1819. He studied at the École Polytechnique, though poor health disrupted his studies. He discovered the highly explosive nitrogen trichloride, losing an eye and three fingers in an explosion. He kept the accident secret, which led to a subsequent, even more severe explosion in another laboratory. For the co-discovery of the Dulong–Petit law, he received a prize from the French Academy of Sciences.

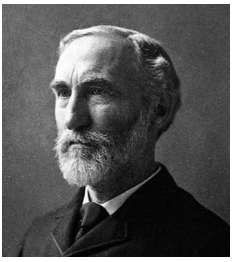


Ehrenfest, Paul (1880–1933), an Austrian physicist, focused heavily on statistical mechanics and thermodynamics. In 1900, he proposed a diffusion model and a statistical interpretation of the second law of thermodynamics. He first met Albert Einstein in Prague in 1912. Later, he worked on the quantum theory of rotating systems. Ehrenfest studied at the University of Vienna and completed his doctorate under Boltzmann in Göttingen in 1904. He worked at many prestigious institutions, most notably in Leiden as a professor of theoretical physics. His students included Hendrik Casimir and George Uhlenbeck, and he significantly influenced Ralph Kronig. The fundamental Ehrenfest theorems, which relate quantum mechanical expectation values to classical equations of modern motion, are named in his honor.

Einstein, Albert (1879–1955), see section *Relativity*, this textbook.



Fermi, Enrico (1901–1954), an Italian-American physicist, focused primarily on quantum theory and elementary particle physics. He named the small, neutral particle produced during beta decay the neutrino (Italian for "little neutral one"). Particles with half-integer spin that obey the Pauli exclusion principle are named fermions in his honor, and they follow the Fermi-Dirac statistical distribution. In 1942, Fermi designed and constructed the world's first nuclear reactor beneath the University of Chicago's stadium, using graphite blocks as a moderator. In 1943, he founded the Argonne National Laboratory. Fermi also investigated the acceleration mechanisms of cosmic rays, proposing a statistical acceleration of charged particles reflecting off magnetic mirrors, known today as the Fermi mechanism. In 1938, he received the Nobel Prize in Physics for discovering artificial radioactive elements produced by neutron bombardment. The Fermi Gamma-ray Space Telescope, launched in 2008, is named after him.



Gibbs, Josiah Willard (1839–1903), an American mathematician and theoretical physicist, made foundational contributions to thermodynamics and statistical mechanics. He formulated the concept of thermodynamic equilibrium using energy and entropy criteria. He also derived a simple rule governing the chemical equilibrium of multi-phase systems, known as the Gibbs phase rule. In mathematics, he founded vector analysis. Gibbs studied at Yale University, earning a doctorate in engineering, and later became a professor of mathematical physics there. His pioneering work seamlessly integrated mathematics, physics, and chemistry. Eponymous concepts honoring his legacy include the Gibbs phase rule and the Gibbs free energy.



Heisenberg, Werner (1901–1976), a German theoretical physicist, explored the fundamental principles of quantum theory. In 1925, he developed matrix mechanics, an alternative formulation for calculating quantum states compared to Schrödinger's wave mechanics. Heisenberg also derived the famous uncertainty relations, stating that the position and momentum of an object cannot be measured simultaneously with absolute precision, leading to unavoidable statistical fluctuations. This meas-

urement interference forms a cornerstone of modern physics. For his foundational role in quantum mechanics, he received the 1932 Nobel Prize in Physics. Heisenberg studied theoretical physics at Munich, earning his PhD under Arnold Sommerfeld in 1923, before becoming an assistant to Max Born in Göttingen. He spent three years in Copenhagen, contributing with Niels Bohr to the Copenhagen interpretation of quantum mechanics. He also proposed a successful model of ferromagnetism featuring two phase transitions. From 1927 to 1941, he was a professor of theoretical physics in Leipzig, from 1942 to 1945 director of the Kaiser Wilhelm Institute for Physics in Berlin, and from 1946 director of the Max Planck Institute for Physics in Göttingen (later Munich).

Hubbard, John (1931–1980), an English physicist, formulated the Hubbard model, the Bose-Hubbard model, and other theoretical frameworks describing interacting electrons. He studied at Imperial College London, where he earned his doctorate. Later, he worked at the IBM San Jose Research Laboratory in the United States and the Harwell Laboratory in the United Kingdom. His primary research field was condensed matter physics.



Ising, Ernst (1900–1998), a German physicist, focused primarily on phase transitions in ferromagnetic and paramagnetic materials. He developed the highly successful Ising model, which evaluates the phase transitions of ferromagnets on a lattice. In his doctoral thesis, Ising calculated the partition function of this model for a one-dimensional chain of two-state spins. Lars Onsager later solved the two-dimensional problem in 1944. Ising studied and worked at the University of Hamburg. In 1939, he fled Nazi persecution to Luxembourg, working as a shepherd and railroad laborer, and was later forced into military labor during the occupation. In 1947, he emigrated to the United States with his family, becoming a professor at Bradley University in Illinois, where he remained for the rest of his career.



Jeans, James (1877–1946), an English mathematician and astronomer, investigated a broad spectrum of physical problems. He rejected Laplace nebular hypothesis of solar system formation, proposing instead a tidal theory where a passing star drew matter from the Sun. This theory was later disproved. Jeans also researched thermodynamics and blackbody radiation, successfully deriving the Rayleigh-Jeans law for low-frequency radiation emission. Alongside Arthur Eddington, he co-founded British cosmology. He opposed the Big Bang theory, advocating for a steady-state universe model later refuted by the discovery of the cosmic microwave background. After 1928, Jeans became a successful popular science writer. He studied at Cambridge and later worked there and at Princeton.



Kosterlitz, John Michael (*1942), a British-American physicist, has dedicated his career to condensed matter physics, studying one- and two-dimensional spin systems, lattice models, and spin glasses. He studied at Cambridge University and earned his doctorate from Oxford. He held positions at various global universities, including Birmingham, Cornell, and Brown University, where he received a full professorship. He is now Professor Emeritus at Brown and has collaborated with Aalto University. In 1972, he and David Thouless discovered a novel “soft” phase

between high-temperature disordered and low-temperature ordered phases in certain ferromagnets, characterized by slowly varying spin directions forming vortices and spin waves. This transition is known as the Kosterlitz-Thouless (KT) transition. In the 1970s, they predicted surface superconductivity resulting from topological states in two-dimensional periodic structures. In the 1980s, Kosterlitz and Duncan Haldane developed description methods for phase transitions where symmetry breaking is not the defining characteristic. For his groundbreaking work, he received numerous awards, culminating in the prestigious 2016 Nobel Prize in Physics for topological phase transitions.



Landé, Alfred (1888–1976), a German-American physicist, dedicated his career to quantum physics. He is famous for the Landé *g*-factor, which determines state degeneracy and serves as a key variable in the anomalous Zeeman effect and magnetic resonance. Landé also focused on spectroscopy. In his later years, much like Einstein, he strongly disagreed with the Copenhagen interpretation of quantum mechanics and sought to develop his own conceptual framework for quantum theory. Landé studied at the University of Munich. In 1929, he lectured at Ohio State University, and during his second visit from 1930 to 1931, he made the decision to remain permanently in the United States.

Liouville, Joseph (1809–1882), a French mathematician, worked in number theory, complex analysis, differential geometry, topology, theoretical physics, and astronomy. He contributed to solving integral equations using eigenvalues, studied the integrability of systems of equations in detail, and demonstrated that phase space volume is conserved during the time evolution of Hamiltonian systems (Liouville theorem). Numerous mathematical structures, objects, and a lunar crater are named in his honor.

Maxwell, James Clerk, see section *Electromagnetic Field*, this textbook.



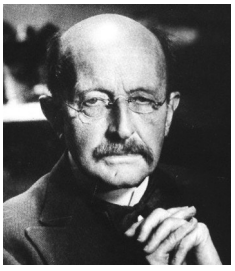
Onsager, Lars (1903–1976), a Norwegian-American chemist and theoretical physicist, won the 1968 Nobel Prize in Chemistry. He investigated the theory of electrolytes and the Brownian motion of ions. Passionate about statistical mechanics and thermodynamics, he examined particle diffusion driven by temperature and concentration gradients, alongside entropy production in these processes. He generalized his findings into the Onsager reciprocal relations, which describe the symmetries between thermodynamic fluxes and forces. He held positions at Brown University, Johns Hopkins University, and Yale University.



Pauli, Wolfgang (1900–1958), an Austrian-German-American physicist, formulated the Pauli exclusion principle in 1925, which asserts that no two identical fermions can occupy the same quantum state simultaneously. This principle accounts for the diverse properties of chemical elements and the structure of matter. He contributed significantly to quantum mechanics and formulated the Pauli equation, the first quantum wave equation to incorporate spin. In the 1930s, he predicted the existence of the neutrino. For his groundbreaking discovery of the exclusion principle, he received the 1945 Nobel Prize in Physics. This

principle fundamentally governs the statistical behavior of particles at low temperatures. Pauli was born in Vienna, and his godfather was Ernst Mach. His paternal grandparents came from a Jewish family in Prague. He published his first scientific paper on general relativity at age 18. He earned his doctorate in Munich under Arnold Sommerfeld, with a dissertation on the quantum properties of ionized molecular hydrogen. Pauli spent a year at the University of Göttingen as an assistant to Max Born. He also worked at the Institute for Theoretical Physics in Copenhagen (now the Niels Bohr Institute), the University of Hamburg, and ETH Zurich. He served as a visiting professor at the University of Michigan in 1931 and at Princeton in 1935. In 1939, due to the political climate in Europe, Pauli moved to the United States, accepting a professorship in theoretical physics at Princeton. He became a naturalized U.S. citizen after World War II.

Petit, Alexis Thérèse (1791–1820), a French physicist, focused primarily on the efficiency of steam engines. He discussed heat engines and thermodynamic efficiency with Sadi Carnot. Collaborating closely with Pierre Louis Dulong, they experimentally discovered the relation for the molar heat capacity of crystalline solids, known today as the Dulong–Petit law. Petit studied at the École Polytechnique, entering at the lowest possible age of 16, and became a professor there at just 23 years old.

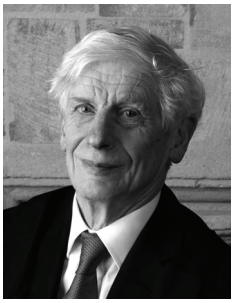


Planck, Max (1858–1947), a German physicist, formulated the blackbody radiation law by assuming that energy is emitted in discrete packets, or quanta, proportional to the frequency. He introduced this quantum hypothesis purely as a mathematical construct to solve the equations, initially remaining skeptical of its physical interpretation. In 1918, he received the Nobel Prize for his quantum theory, which was successfully applied by Einstein to the photoelectric effect and by Bohr to his atomic model. Planck was deeply involved in thermodynamics, and one of the formulations of the second law of thermodynamics bears his name. He criticized the probabilistic interpretation of entropy. In 1900, he introduced the universal gas constant and Avogadro number. Eponymous concepts include the Planck scales, the Max Planck Society research network, a lunar crater, and the highly successful European space probe analyzing the ancient cosmic microwave background radiation.

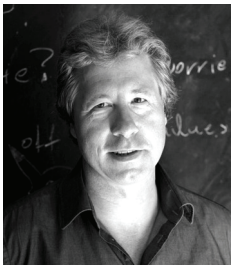
Potts, Renfrey Burnard (1925–2005), an Australian mathematician, developed the Potts model of ferromagnetism, which incorporates an arbitrary number of spin orientations, generalizing the Ising model. Potts possessed a remarkably broad research scope; he actively engaged in robotics and contributed significantly to transportation engineering, focusing on vehicle tracking and traffic flow optimization. He studied at the University of Adelaide and Oxford University, and received the Australian Mathematical Society Medal for his contributions.

Rayleigh, John William Strutt (1842–1919), see section *Electromagnetic Field*.

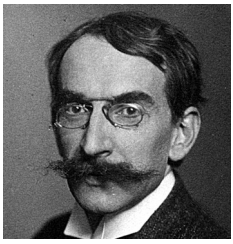
Stefan, Jožef (1835–1893), a Slovenian physicist, mathematician, and poet, studied mathematics and physics at the University of Vienna. Drawing from experiments on the Dulong–Petit law, he deduced the law governing the total energy flux from a blackbody, using it to estimate the surface temperature of the Sun. He solved temperature distribution problems during phase transitions (the Stefan problem) and addressed other topics bridging mathematics, physics, and chemistry. Ludwig Boltzmann was his student. Eponymous concepts include the Stefan–Boltzmann law, Stefan flow, and the Stefan problem. Stefan graduated from Vienna University and received the Lieben Prize.



Thouless, David James (1937–2019), an English theoretical physicist, graduated from the University of Cambridge and earned his doctorate at Cornell University under Hans Bethe. He completed a postdoctoral fellowship at UC Berkeley. From 1965 to 1978, he returned to England as a professor at the University of Birmingham, later accepting a professorship at the University of Washington in Seattle. His research explored the statistical properties of two-dimensional systems, lattice models, ferromagnets, superconductivity, and phase transitions. In 1972, he and Michael Kosterlitz discovered a novel "soft" phase between high-temperature disordered and low-temperature ordered phases in certain ferromagnets, characterized by slowly varying spin directions forming vortices and spin waves. This transition is called the Kosterlitz–Thouless transition. In the 1970s, they predicted surface superconductivity resulting from topological states in two-dimensional periodic structures. Thouless received numerous honors, culminating in sharing the 2016 Nobel Prize in Physics for theoretical discoveries of topological phase transitions and topological phases of matter.

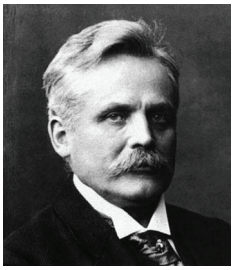


Verlinde, Erik Peter (*1962), a Dutch theoretical physicist, shocked the scientific community in 2009 by introducing his entropic gravity hypothesis. In this theory, gravity is not viewed as a fundamental force, but rather as an emergent entropic force driven by information gradients. The information itself is localized on a two-dimensional surface, invoking the holographic principle. From these first principles, he derived Newton laws of motion, the law of universal gravitation, and general relativity. In 2016, he introduced a volume component – vacuum fluctuations – as an additional entropy source, causing his predictions to diverge from standard general relativity. This theory aims to shed light on dark energy and dark matter. Verlinde compares gravity to pressure; individual molecules do not exert pressure, but their collective bulk behavior does. Similarly, gravity emerges from the quantum behavior of microscopic states. Although his hypothesis faces considerable skepticism, it remains highly influential. Verlinde studied at Utrecht University, worked at Princeton, and is currently a professor at the University of Amsterdam. The Verlinde algebra and Verlinde formula in conformal field theory are named in his honor.



Weiss, Pierre-Ernest (1865–1940), a French physicist, specialized in the study of magnetism. In 1907, he also developed the theory of magnetic domains, which are now named after him. He investigated mean-field theory, which led him to formulate the Curie–Weiss law describing the temperature dependence of magnetic susceptibility. Weiss studied at the University of Rennes, and later conducted pioneering research in Lyon, at ETH Zurich in Switzerland, and at the University of Strasbourg. In recognition of his scientific contributions, he was awarded the Legion of Honor in 1935.

Wien, Wilhelm (1864–1928), a German physicist, received the 1911 Nobel Prize in Physics for his displacement law governing blackbody radiation, which he discovered in 1893. This law demonstrates that the peak wavelength of emission decreases as the temperature of the body increases, meaning hotter objects radiate at shorter wavelengths



than cooler ones. Wien earned his doctorate from the University of Berlin in 1886 and shortly thereafter began investigating blackbody radiation. He attempted to derive a formula for blackbody radiation intensity as a function of temperature; while his formulation was successful for the short-wavelength spectrum, it failed in the long-wavelength region. The complete radiation law was later derived by Max Planck through the assumption of energy quantization, which laid the foundations of quantum mechanics. Wien was appointed professor at the University of Giessen in 1899 and at the University of Munich in 1920. His research also contributed to the development of cathode-ray tubes used to generate X-rays.

Electromagnetic Field



Ampère, André Maria (1775–1836), a French mathematician and physicist, demonstrated that a magnetic field surrounds a current-carrying conductor. By Experimenting with a compass needle, he established the relationship between current direction and magnetic flux lines, known as Ampère right-hand rule. He showed that parallel conductors attract each other when carrying current in the same direction and repel when currents flow in opposite directions. He explained magnetism via molecular electric currents and discovered that a current-carrying coil behaves like a bar magnet, naming his invention the solenoid. He also invented an electric telegraph. Although self-taught, he became a member of the French Academy of Sciences and a professor at the École Polytechnique. The unit of electric current (ampere, A), Ampère law, and his right-hand rule bear his honor.

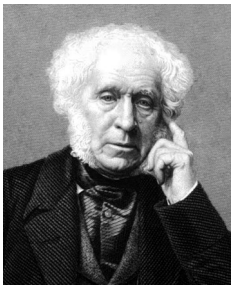


Anderson, Carl David (1905–1991), an American physicist, received the 1936 Nobel Prize in Physics for discovering the positron (positive electron), the first known particle of antimatter. Anderson earned his PhD in 1930 from the California Institute of Technology in Pasadena, where he worked with physicist Robert Andrews Millikan. From 1927, they studied X-ray photoelectrons released during collisions between atoms and high-energy photons, and in 1930, they began researching cosmic and gamma radiation. In a cloud chamber, Anderson photographed tracks of secondary cosmic-ray showers. Studying these photographs, he discovered numerous tracks whose shape implied they were caused by positively charged particles – yet particles much smaller than protons. In 1932, he published that the tracks were caused by positrons, which are positively charged particles with the same mass as electrons. The existence of the positron as the electron's antiparticle was predicted in 1928 by Paul Dirac. In 1936, Anderson discovered the muon (heavy electron), a particle 207 times heavier than the electron. He initially thought he had found the meson predicted by Japanese physicist Hideki Yukawa to bind protons and neutrons together in the atomic nucleus, but the muon turned out to interact only weakly, being just a heavier version of the electron. The actual meson was discovered in 1947 by British physicist Cecil Powell and named the pi-meson, or pion. Anderson spent his

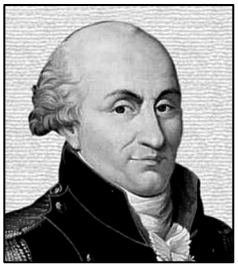
entire career at California Institute of Technology, joining the faculty in 1933 and serving as a professor until 1976. During World War II, he participated in rocket research.



Biot, Jean-Baptiste (1774–1862), a French physicist, astronomer, and mathematician, is best known for the Biot–Savart law, which calculates the magnetic field generated by a steady electric current. Together with Gay-Lussac, he used a hot-air balloon to investigate Earth's magnetic field at an altitude of five kilometers. In 1803, he analyzed meteorite fragments that fell near L'Aigle, concluding they originated from outer space. He also investigated the physical properties of cream of tartar. Among his prolific output of nearly three hundred papers, his most significant contributions focused on the polarization and birefringence of light. Biot studied at the prestigious university École Polytechnique and was awarded the highly prestigious Rumford Medal by the Royal Society of London.



Brewster, David (1781–1868), a Scottish physicist, mathematician, astronomer, and inventor, discovered that mechanical stress applied to an isotropic material can induce optical birefringence. He experimentally demonstrated that when reflected and refracted rays are perpendicular, the reflected light becomes perfectly linearly polarized. The angle of incidence at which this occurs is called the Brewster angle. Investigating biaxial crystals, he showed that two optical axes do not necessarily cause birefringence. He also invented the kaleidoscope. Brewster studied at the University of Edinburgh and later worked at the University of St Andrews. For his optics research, he received numerous accolades, including the Copley and Rumford Medals, the Keith Prize, and the Royal Medal.

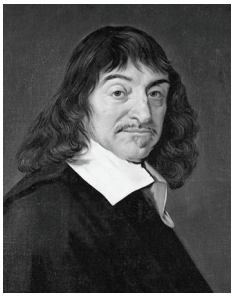


Coulomb, Charles-Augustin de (1736–1806), a French physicist and military engineer, conducted pioneering experiments using a torsion balance, an apparatus independently described by Joseph Priestley. His research proved that electrostatic and magnetic forces decrease with the square of the distance. For electrical phenomena, the relation $F = k\,qQ/r^2$ became known as Coulomb's law, though it was previously proposed by John Robison. Coulomb also investigated the mechanics of friction. The SI unit of electric charge, the coulomb, is named after him. During the French Revolution, he collaborated with the government to develop the metric system and also contributed to the design of retaining walls to prevent soil erosion. He graduated from the Royal Engineering School in Charleville-Mézières.

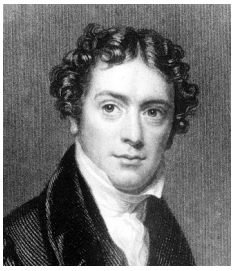


Debye, Peter Joseph William (1884–1966), a Dutch physicist and chemist, began his career studying eddy currents. He was the first to describe the behavior of asymmetric molecules using the dipole moment, analyzing them directly via X-ray diffractometry. In his honor, the molecular dipole moment is measured in debye units. His broad research extended Einstein specific heat theory to include low-frequency phonons, advanced the theory of electrolyte electrical conductivity, explored atomic shells, and explained the Compton effect. In 1936, he received the Nobel

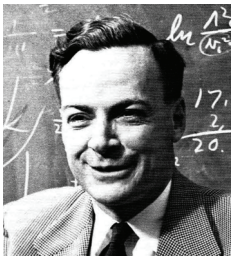
Prize in Chemistry for his contributions to molecular structure. In physics, the Debye screening length bears his name. Born Petrus Josephus Wilhelmus Debije, he later legally changed his name to Peter Joseph William Debye.



Descartes, René (1596–1650), a French philosopher, separated mathematics and physics from philosophy, becoming a founding father of modern science. Rejecting the work of his predecessors, he always forged his own path. He proposed that scientific problems must be analyzed by breaking them down into simpler components, solving them individually, and then synthesizing the final solution. He worked in numerous fields of mathematics and physics, introducing the Cartesian coordinate system and co-founding analytical geometry with Pierre de Fermat. In physics, he focused primarily on classical mechanics and optics. Regarding electromagnetic and light fields, he attempted to derive the law of refraction using a mechanical model of spheres passing through a surface. Although his derived ratio of velocities before and after the surface was inverted compared to actual light refraction, treating this ratio as a material constant (the refractive index) kept the error hidden. Using this relationship, he developed a detailed theory of the rainbow and designed early optical devices, including a microscope. Descartes studied at several institutions, beginning at the University of Poitiers.



Faraday, Michael, (1791–1867), an English bookbinder who developed a passionate interest in electricity, earned the opportunity to work as a secretary and later as a laboratory assistant to Humphry Davy, where he conducted his pioneering experiments. In 1821, he published an overview of contemporary views on electricity and magnetism, which included Oersted's discoveries. He became one of the greatest experimentalists in history. Being self-taught, he occasionally struggled with the advanced mathematics in Ampère's works. The properties of magnetic force led Faraday to hypothesize that it acts in circular lines. He also discovered that rotating a magnet can generate an electric current, leading to the invention of the dynamo in 1821. He discovered electromagnetic induction in 1831 and formulated the laws of electrolysis in 1857. Between 1839 and 1855, he published his results in the three-volume Experimental Researches in Electricity. In 1845, he developed the concept of electric and magnetic fields, establishing the field formulation of electromagnetism. He also experimented with dielectrics in capacitors and visualized the magnetic field of a bar magnet using iron filings. Beyond electromagnetism, he explained the phenomenon of regelation in 1859, correctly proposing that pressure melts a thin layer of water between ice blocks, which then refreezes.



Feynman, Richard Philips (1918–1988), an American physicist, was a co-creator of quantum electrodynamics (QED). In 1965, he was awarded the Nobel Prize in Physics for this work, sharing it with Sin-Itiro Tomonaga and Julian Schwinger. He is most famous for introducing Feynman diagrams, which greatly simplify complex calculations in quantum field theory and visually represent quantum processes. In electromagnetism, the Feynman formula for calculating electric and magnetic fields,

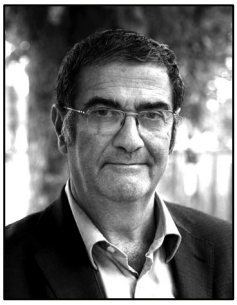
including radiation terms, is named in his honor. During World War II, he made significant contributions to the Manhattan Project at Los Alamos. Later in life, he served on the Rogers Commission, famously investigating the 1986 Space Shuttle Challenger disaster and demonstrating the failure of O-rings using a simple glass of ice water. Feynman was an exceptional educator whose lectures, transcribed by students, became the widely acclaimed textbook The Feynman Lectures on Physics. In his later years, he also wrote several popular, humorous autobiographical books about his life.



Fresnel, Augustin-Jean (1788–1827), a French physicist, independently performed the same double-slit experiment as Thomas Young. Fresnel demonstrated that only transverse light waves could account for double refraction. In deriving the Fresnel equations for reflection and refraction, he worked backward from established experimental observations to theory. He also developed a mathematical theory of refraction and polarization in anisotropic crystals, from which William Rowan Hamilton predicted conical refraction, discovered shortly thereafter. Fresnel graduated from the École Polytechnique. For his groundbreaking work, he received numerous honors, including the Grand Prix in Mathematical Sciences and the Rumford Medal, became a Foreign Member of the Royal Society, and is one of the 72 scientists whose names are inscribed on the Eiffel Tower. Eponymous concepts include the Fresnel biprism, the Fresnel lens, and the Fresnel equations.

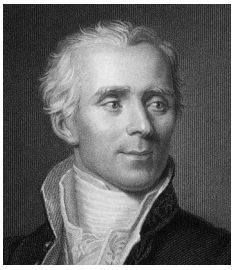


Gauss, Karl Friedrich (1777–1855), a German mathematician, is often referred to as the "*Prince of Mathematicians*." A child prodigy, at age three he corrected a calculation error in his father's payroll. In school, when tasked with summing numbers from 1 to 100, he quickly derived the formula for the sum of an arithmetic series. In statistics, the Gaussian (normal) distribution bears his name. In vector calculus, the divergence theorem (or Gauss theorem) for converting surface integrals into volume integrals honors his legacy. The gauss, a unit of magnetic flux density, is also named after him. Gauss made monumental contributions to number theory, integral and differential calculus, geometry, mathematical analysis, electrostatics, magnetostatics, astronomy, optics, and geodesy. He wrote his seminal work on number theory, *Disquisitiones Arithmeticae*, at just 21 years old.

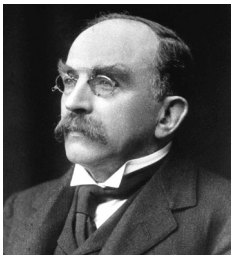


Haroche, Serge (*1944), a French physicist, was awarded the 2012 Nobel Prize in Physics for developing pioneering techniques that enable the non-destructive measurement and manipulation of individual quantum systems. Born to a Moroccan lawyer and a Ukrainian teacher, Haroche settled permanently in France in 1956. Today, he is a professor and chair of quantum physics at the Collège de France. Haroche specializes primarily in atomic physics and quantum optics. In the 1970s, he developed a novel method of laser spectroscopy based on quantum beats predicted by quantum electrodynamics. He subsequently focused on the quantum properties of Rydberg atoms. These giant atoms are exceptionally sensitive to microwaves, making them ideal for studying atom-photon interactions. Using a specialized Ramsey interferometer equipped with a superconducting microwave

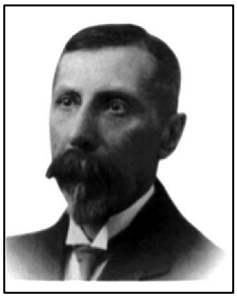
cavity developed by his team, Haroche and his colleagues successfully monitored the loss of quantum properties (decoherence) in Rydberg atoms. This cavity, containing just a few photons, is used for testing quantum decoherence and executing logic operations that utilize quantum information processing.



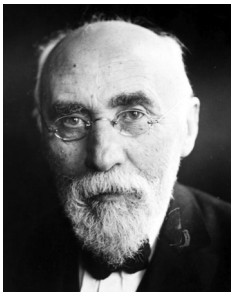
Laplace, Pierre Simon de (1749–1827), a French physicist and mathematician, finalized the foundation of mathematical astronomy by condensing the work of his predecessors into the five-volume *Mécanique Céleste* (1799–1825). Using classical mechanics, he was the first to calculate the radius of a body whose gravity is so intense that not even light can escape—a concept known today as the Schwarzschild radius. Laplace systematized and advanced probability theory in his *Essai Philosophique sur les Probabilités* (1814). He was the first to evaluate the Gaussian integral as the square root of π and investigated the integral transform now known as the Laplace transform, which Heaviside later perfected. Laplace also contributed to Lavoisier calorimetric theory, using a custom-designed calorimeter to determine the specific heats of numerous substances. He discovered and introduced the gravitational potential, demonstrating that it satisfies the Laplace equation in a vacuum. This equation applies equally to electrostatics and magnetostatics. After Napoleon appointed him Minister of the Interior, he was dismissed shortly after with the comment that he "*brought the spirit of the infinitely small into government*."



Larmor, Joseph sir (1857–1942), an Irish physicist and mathematician, explained the FitzGerald contraction independently of Lorentz and predicted time dilation. He modeled the luminiferous aether as a perfectly incompressible, elastic fluid, linking this to Kelvin mechanical aether model composed of a gyroscopic lattice. Larmor calculated the power radiated by an accelerated electric charge (the Larmor formula) and explained the splitting of spectral lines in a magnetic field. He was among the first to suggest that geomagnetic storms are linked to solar flares and are caused by streams of electrons ejected from the Sun. His research spanned electrodynamics, thermodynamics, and the electron theory of matter. He studied in Belfast and Cambridge, becoming a professor at Cambridge in 1903. Eponymous concepts honoring his legacy include the Larmor radius of the circular motion of a charged particle in a magnetic field and the Larmor formula for dipole radiation power.



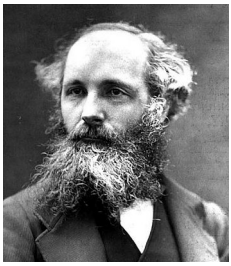
Liénard, Alfred-Marie (1869–1958), a French physicist and engineer, graduated from the École Polytechnique and became a professor of electrical engineering at the École des Mines in Paris three years later. He focused primarily on mechanics, electricity, and magnetism. Independently of Emil Wiechert, he derived the electromagnetic potentials for a moving point charge by accounting for signal retardation. Today, these are known as the Liénard–Wiechert potentials. He also conducted research in elasticity, strength of materials, thermodynamics, and hydrodynamics. For his scientific achievements, he received the Poncelet Prize and the Legion of Honour. Liénard also served as president of both the French Mathematical Society and the French Society of Electrical Engineers.



Lorentz, Hendrik Antoon (1853–1928), a Dutch physicist, began his studies at the University of Leiden at age sixteen, graduated at twenty-two, and became a professor of theoretical physics there two years later. Inspired by Hertz's experiments, he abandoned Helmholtz electromagnetic theory and embraced Maxwell's formulation as a monumental success, completely rebuilding contemporary physics on its foundation. He was the first to propose that the sources of the electromagnetic field are oscillating charged particles within atoms (electrons) and derived Maxwell's equations by modeling matter as a system of oscillators. He predicted that a strong magnetic field would split spectral lines, a phenomenon experimentally verified by his student Pieter Zeeman; both shared the 1902 Nobel Prize in Physics. He discovered the coordinate transformation under which Maxwell equations remain invariant. Independently of FitzGerald, he demonstrated that the null result of the Michelson–Morley experiment stems from the contraction of the interferometer arms. These contributions laid the foundation for special relativity. The Lorentz force acting on a moving charge and the Lorentz transformation bear his name.

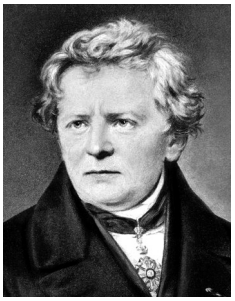


Lorenz, Ludvig Valentin (1829–1891), a Danish physicist and mathematician, investigated the optical and thermal properties of materials. For transparent media, he independently of Lorentz derived the mathematical expression relating the refractive index to material density, known as the Lorentz–Lorenz equation. He also determined the temperature dependence of electrical and thermal conductivity, establishing the Wiedemann–Franz–Lorenz law. He formulated the Lorenz gauge condition for electromagnetic potentials, which enabled him to calculate the properties of light waves, including the speed of light, and investigated light scattering. Years before Oliver Heaviside, he derived the telegrapher's equation but left it unpublished, though his remarkable research earned him acclaim from the Danish Academy.



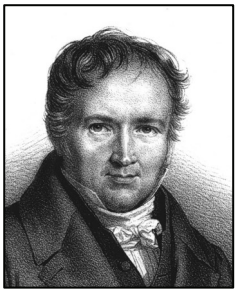
Maxwell, James Clerk (1831–1879), a Scottish mathematician and physicist, formulated the mathematical and physical theory of the electromagnetic field. Seeking to present electromagnetic phenomena more elegantly than his predecessors Faraday, Kelvin, and Ampère, he deduced that light consists of transverse modes driven by electric and magnetic effects. He published his unified framework in his 1873 masterpiece *A Treatise on Electricity and Magnetism*, introducing the relations now known as Maxwell's equations, though their modern form was later sculpted by Oliver Heaviside and Heinrich Hertz. Maxwell's theory directly predicts electromagnetic waves and naturally leads to special relativity, which obeys the same spacetime transformations. The existence of these waves, predicted by Maxwell, was experimentally proven by Heinrich Hertz after Maxwell's death. Maxwell also correctly determined that Saturn's rings are composed of small rocky particles. Alongside Clausius, he developed the kinetic theory of gases, a study that led him in 1867 to formulate the Maxwell demon paradox. He demonstrated that the second law of thermodynamics is a purely statistical law describing large ensembles of particles. In 1871, he became the first director of the famous and now prestigious Cavendish Laboratory in Cambridge. Eponymous concepts honoring his legacy include Maxwell equations, the

Maxwell distribution, Maxwell demon, a unit of magnetic flux, an extensive mountain range on Venus, a large gap in Saturn's rings, and the James Clerk Maxwell Telescope (JCMT) for submillimeter astronomy on Mauna Kea, as well as the Maxwell Foundation which actively supports young talented researchers in mathematics and physics.

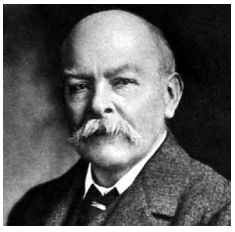


Ohm, Georg Simon (1789–1854), a German physicist, investigated electricity and deduced from experiments with his own electrochemical cells that the current through a conductor is directly proportional to voltage and cross-sectional area, and inversely proportional to length. He also worked in acoustics and optics. Largely self-taught, he earned his PhD from the University of Erlangen. Initially high school teacher, he published a paper in 1825 on current dependence on conductor length. A year later, he modeled electrical circuits based on Fourier's heat conduction framework. He definitively formulated his law, $I = U/R$, in 1827 during a research year in Berlin. Hoping for a university position that did not materialize, he remained in Berlin teaching mathematics. In 1833, he secured a post in Nuremberg, though it was not his dream university. In 1842, he became a Foreign Member of the Royal Society. He finally obtained a university professorship in Munich just five years before his death, becoming a full professor three years later. Beyond Ohm law, he formulated laws of physiological acoustics: Ohm's acoustic law states that the human ear perceives music as combinations of pure harmonic tones, while Ohm phase law asserts that the phase and waveform do not affect auditory perception. The SI unit of electrical resistance, the ohm, honors his legacy.

Planck, Max, see section *Statistical Physics*, this textbook.

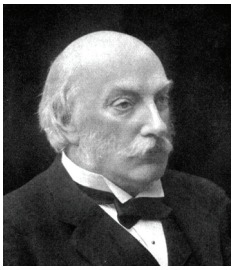


Poisson, Siméon Denis (1781–1840), a French mathematician, physicist, and geometer, was a student of Laplace. He generalized Laplace's equation in electromagnetism by introducing source terms, a relation now known as the Poisson equation. He demonstrated that the electrostatic potential must be constant on the surface of a conductor and formulated the theories of surface currents and bulk magnetization. He evaluated Fresnel's wave theory of light diffraction and published renowned works on probability. Poisson studied at the École Polytechnique, publishing two scientific papers and receiving special honors after just two years. He attended Joseph Lagrange's lectures on functional analysis. Immediately after graduation, he became an assistant, rising to extraordinary professor in 1802 and full professor in 1806. From 1808, he worked as an astronomer at the Bureau des Longitudes. Despite heavy administrative burdens in various prominent positions, he published over 300 mathematical papers.



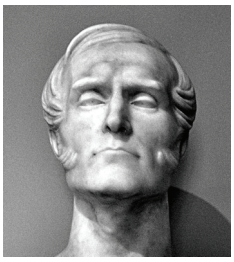
Poynting, John Henry (1852–1914), an English physicist, investigated energy radiation in electromagnetic waves. In 1844, he published the law of conservation of energy in electromagnetic fields, demonstrating that the directional energy flux density is given by the vector product of the electric and magnetic field strengths (the Poynting vector). This vector points in the direction of energy propagation and group velocity. In 1893, he

independently measured the gravitational constant. He also co-discovered the Poynting–Robertson effect in 1903, which describes the orbital drift of dust particles caused by solar radiation pressure. Poynting studied physics at Owens College, Manchester, under teachers including Osborne Reynolds, and later studied mathematics at Cambridge. In the late 1870s, he conducted research at the Cavendish Laboratory under James Clerk Maxwell. He served as a professor of physics at Mason Science College (now the University of Birmingham), holding the institution's first chair of physics until his death.



Rayleigh, John William Strutt, Third Baron (1842–1919), an English physicist, investigated acoustics and optics, focusing heavily on wave propagation in fluids. In 1904, he received the Nobel Prize in Physics for isolating the inert gas argon. Poor health prevented him from completing studies at Eton and Harrow, leading to four years of private tuition starting in 1857. He entered Trinity College, Cambridge, in 1861, graduating in 1865. He intensively researched Maxwell electromagnetic theory both experimentally and theoretically. In 1878, he published the two-volume *The Theory of Sound*, which became a foundational text in acoustics. He derived the mathematical law describing the wavelength dependence of light scattering in the atmosphere, becoming the first to explain the blue color of the sky (Rayleigh scattering). He also attempted to derive the blackbody radiation law; his formula (the Rayleigh–Jeans law) correctly described intensity for long wavelengths but diverged at short wavelengths, an anomaly known as the ultraviolet catastrophe. Max Planck resolved this for the entire spectrum in 1901.

Sahl, Ibn (c. 940–1000), a Persian mathematician and physicist, wrote a landmark treatise on optics around 984, building upon Ptolemy's Optics. In this manuscript, he investigated curved mirrors and lenses, and discovered the law of refraction. Using this law, he calculated the precise geometric shapes of lenses required to eliminate geometric aberrations. His work explored parabolic and elliptic mirrors, biconvex lenses, and other advanced optical elements. The law of refraction was later independently rediscovered in the Western world by René Descartes and Willebrord Snellius.



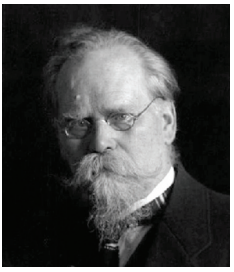
Savart, Félix (1791–1841), a French physicist and acoustician, originally studied medicine at a military hospital in Metz, earning his medical degree from the University of Strasbourg in 1816. However, fascinated by vibrating bodies and wave phenomena, he turned his career toward acoustics. He invented the Savart wheel – a rotating toothed metal disk used to produce sound at a precisely defined frequency proportional to its angular velocity. Moving to Paris in 1817, he constructed a specialized violin to experiment on how sound is transmitted from strings to the wooden body, and also studied human and animal vocalizations. In 1820, he became a physics teacher at the University of Paris, where he collaborated with Biot on experiments regarding magnetic fields generated by electric currents. This research led to the 1820 discovery of the Biot–Savart law, which calculates the magnetic field from current distributions, a relationship later given an elegant mathematical form by Laplace. In 1827, Savart became a member of the Paris Academy of Sciences and, a year later, a professor at the Collège de France. Eponymous concepts include the Biot–Savart law and the Savart wheel.



Snellius, Willebrord van Roijen (1580–1626), a Dutch physicist, astronomer, optician, and mathematician, rediscovered the law of refraction in 1621, which had been previously known to Persian mathematician Ibn Sahl in 984. He conducted detailed investigations into the loxodrome curve. He improved the polygon method for calculating π, computing it to seven significant figures at a time when it was generally known to only two. He proposed a method for measuring the Earth's circumference using triangulation between two cities located on the same latitude. Like his father, he studied at the University of Leiden, where he served as a professor of mathematics from 1613. The lunar crater Snellius is named in his honor.



Thomson, Joseph John, sir (1856–1940), an English physicist, discovered the electron within cathode rays in 1897. He calculated the electron's mass-to-charge ratio and designed an early mass spectrometer, which he used to discover stable isotopes. He proposed the plum pudding model of the atom, which was later superseded by his student Ernest Rutherford's discovery of the atomic nucleus in 1911. Thomson discovered the natural radioactivity of potassium and was the first to demonstrate that a hydrogen atom contains only a single electron. For many years, he investigated the electrical conductivity of gases. He derived the formula for the scattering of electromagnetic radiation by a free charged particle (Thomson scattering). Thomson studied engineering at Owens College in Manchester before graduating from Cambridge, where he became a professor of physics in 1884. He was an exceptionally gifted educator; seven of his research assistants won Nobel Prizes. His son, George Paget Thomson, also received the 1937 Nobel Prize for demonstrating the wave properties of the electron via crystal lattice diffraction. A Fellow of the Royal Society from 1884 (and its president from 1915 to 1920), Thomson won the 1906 Nobel Prize in Physics. He was knighted in 1908, received the Order of Merit in 1912, and served as Master of Trinity College, Cambridge, from 1918 until his death.



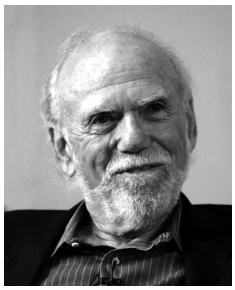
Wiechert, Emil Johann (1861–1928), a German physicist and geophysicist, contributed equally to both fields. In geophysics, he pioneered the theory that the Earth's interior consists of distinct layers and advocated for investigating the Earth's core using seismic waves. In physics, he discovered the electron in cathode rays independently of J. J. Thomson. He took a keen interest in Einstein's special relativity and, independently of Alfred-Marie Liénard, derived the electromagnetic potentials for a moving point charge by accounting for signal retardation (the Liénard–Wiechert potentials). He studied at the University of Königsberg and the University of Göttingen. A lunar crater on the Moon's far side is named in his honor.

Wien, Wilhelm, see section *Statistical Physics*, this textbook.

Relativity



Carswell, Robert (Bob) (*1946), an English astronomer, is the co-discoverer of the first gravitational lens. Carswell is a professor at the Institute of Astronomy at the University of Cambridge. His research focuses primarily on the observation and analysis of quasar spectra. He authored the VoigtFit Python package used for modeling absorption line profiles. In 1979, alongside Dennis Walsh and Ray Weyman, he discovered the first gravitational lens: the twin quasar QSO 0957+561A/B, whose nearly identical spectra revealed they were two images of the same object formed by an intervening gravitational lens. Carswell also conducts research in cosmology, particularly searching for potential variations in the fine-structure constant through observations of deep-space objects using the Zeeman effect. He is a fellow of both the International Astronomical Union and the Royal Astronomical Society.

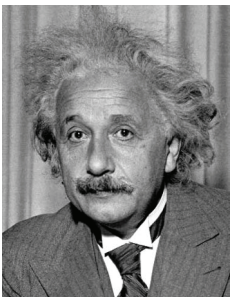


Barish, Barry Clark (*1936), an American physicist born in Nebraska, studied physics at the University of California, Berkeley. After graduating, he joined the California Institute of Technology (Caltech) in 1963 as an experimentalist. Initially focusing on particle accelerators, he designed experiments for Fermilab aimed at uncovering the quark structure of nucleons. His experiments provided some of the earliest observations of weak neutral currents. In the 1880s, he directed the MACRO experiment at Italy's underground Gran Sasso National Laboratory, hunting for exotic cosmic-ray particles like magnetic monopoles, which independently confirmed non-zero neutrino mass and neutrino oscillations. Barish became a full professor at Caltech in 1991 and joined the nascent LIGO project in 1994, becoming its director in 1997. Under his leadership, LIGO became operational, and the Advanced LIGO upgrade was planned and executed between 2010 and 2015, culminating in the 2015 discovery of gravitational waves. From 2005 to 2013, he also directed the International Linear Collider (ILC) global design effort. He was awarded the 2017 Nobel Prize in Physics for his decisive contributions to the LIGO detector and the observation of gravitational waves.

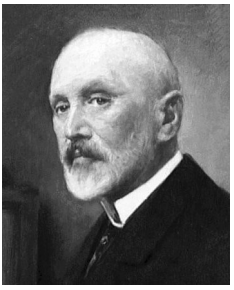


Dicke, Robert Henry (1916–1997), an American astronomer and physicist, made foundational contributions to cosmology, nuclear physics, and gravitation theory. In the late 1960s at Princeton, he led a research group calculating light element nucleosynthesis during the Big Bang and the decoupling of cosmic microwave background (CMB) radiation. Although his group was constructing an antenna to detect this relic radiation, they were scooped by Penzias and Wilson using the Bell Labs horn antenna. Dicke's team also performed high-precision tests of the equivalence principle. Dicke studied at Princeton and earned his PhD from the University of Rochester. During World War II, he worked at the MIT Radiation Laboratory, designing a sensitive microwave radiometer that placed an upper limit on the CMB temperature, before returning to Princeton. In his later years, he focused heavily

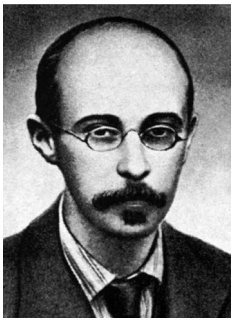
on general relativity and alternative scalar-tensor theories of gravity. He received the National Medal of Science and the Comstock Prize in Physics. Jim Peebles was the sole member of Dicke's group to receive the Nobel Prize for cosmology, awarded much later in 2019 for theoretical discoveries in physical cosmology.



Einstein, Albert (1879–1955), a German-American physicist, was the creator of the special and general theories of relativity, the scientist who explained the photoelectric effect and Brownian motion, and co-developed the Bose-Einstein statistics for integer-spin particles. He published special relativity in 1905, reconciling classical mechanics with Maxwell electrodynamics, which implied that the speed of light is independent of the motion of its source. This theory derived the famous mass-energy equivalence formula $E = mc^2$. Special relativity introduced length contraction, time dilation, and the understanding that space and time are not absolute. In 1905, Einstein also explained the photoelectric effect by assuming light consists of localized packets of energy, or quanta, building upon Max Planck blackbody radiation hypothesis. The ejection of electrons from a metal surface depends on the frequency of individual photons, not their intensity. The term photon was later coined by American chemist Gilbert Lewis in 1926. Einstein also explained Brownian motion in 1905. In 1916, he published his new theory of gravitation – general relativity – describing gravity as spacetime curvature and predicting gravitational waves and black holes. Massive bodies dictate this curvature and move through it along the straightest possible paths, called geodesics. Albert Einstein received the 1921 Nobel Prize in Physics, paradoxically for his explanation of the photoelectric effect rather than for general relativity, which remains his ultimate contribution to science.

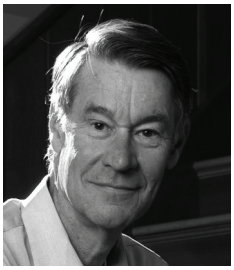


Eötvös, Loránd, baron (1848–1919), a Hungarian physicist, initially focused his research on capillarity and surface tension before turning his attention to gravitation. Using a highly precise torsion balance, he demonstrated that the ratio of gravitational mass to inertial mass is constant to a relative precision 5×10^{-9}. Today, this cornerstone of general relativity has been verified to a precision of 10^{-14}. He also invented the torsion pendulum. Eötvös was born during the Hungarian Revolution of 1848 to the rebel government's minister of education. He studied at the universities of Heidelberg and Königsberg, and after earning his doctorate, returned to Hungary to teach at the University of Budapest, which was renamed Eötvös Loránd University in his honor in 1950. Eötvös's law of surface tension and a lunar crater are also named after him. He was a nobleman whose full title was Baron Loránd Eötvös de Vásárosnamény.

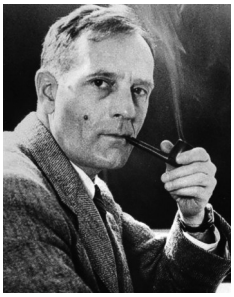


Friedmann, Alexander (1888–1925), a Russian mathematician, physicist, and meteorologist, discovered non-static solutions to Einstein's field equations of general relativity for a homogeneous medium in 1922, implying that the universe must either expand or contract. He formulated the first cosmological model originating from a Big Bang, independently of Lemaître or Gamow. He demonstrated that a universe with negative curvature has infinite volume only if it is simply connected, meaning its

geometry lacks topological “holes” like Swiss cheese. During World War World I, he served as a military pilot and taught at an aviation school, developing a lifelong passion for ballooning. During his final high-altitude flight, his balloon malfunctioned; Friedmann suffered severe exposure in the stratosphere, contracted pneumonia, and died shortly thereafter. The Friedmann metric and the lunar crater Fridman are named after him, with the crater's spelling tracking the original Russian form of his name.



Fuller, Robert Works (*1936), an American theoretical physicist, briefly conducted research in quantum theory, nuclear physics, and general relativity. In 1962, alongside John Archibald Wheeler, he investigated the stability of the Einstein–Rosen bridge, demonstrating that this wormhole solution is highly unstable. Fuller studied physics at Princeton University, where he earned his PhD, and subsequently lectured at Columbia University. Driven by the social unrest of the 1960s, Fuller shifted his focus toward politics and social reform. He spent the majority of his active career at Oberlin College, a liberal arts institution where he had previously completed his undergraduate studies. In 1970, he was appointed the tenth president of Oberlin College. His political endeavors eventually led him to the Oval Office, where he consulted with U.S. President Jimmy Carter.



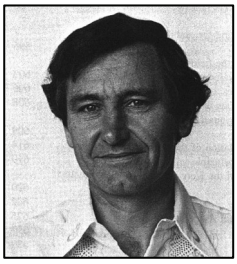
Hubble, Edwin Powel (1889–1953), an American astronomer, was a leading figure in 20th-century astronomy. Highly athletic, Hubble actively played basketball and boxed during his studies at the University of Chicago. He even sparred with champion Georges Carpentier, prompting promoters to urge him to turn pro, though he initially chose to study law and briefly operated a legal practice. In 1919, just three years after the publication of Einstein's general relativity, he secured a position at the Mount Wilson Observatory, home to the world's largest telescope at the time, the 100-inch Hooker telescope. In 1923, he analyzed the Andromeda Nebula, discovering to his surprise that it was composed of countless stars. He identified Cepheid variables within it to estimate its distance, becoming the first to discover an external galaxy and demonstrating that the Milky Way does not constitute the entire universe. In subsequent years, he discovered numerous other galaxies, vastly broadening our understanding of the known universe, and proposed a morphological classification system for them. In 1929, using spectral line shifts, he discovered that galaxies are receding from us at speeds proportional to their distance (the Hubble–Lemaître law). Hubble determined the proportionality constant to be 526 km s^{-1} Mpc^{-1}, a value now estimated at approximately 70 km s^{-1} Mpc^{-1} due to early uncertainties in cosmic distances. Redshifts of extragalactic nebulae had been measured by Vesto Slipher between 1912 and 1915. Slipher's data was interpreted as cosmic expansion by Belgian priest Georges Lemaître in 1927 – two years before Hubble – who also estimated the expansion coefficient at a value similar to Hubble's. It remains unknown if Hubble was aware of Lemaître's work. Hubble utilized Slipher's measurements alongside his own data and later became the first observer to use the 200-inch Hale Telescope at Mount Palomar. Hubble received the Newcomb Cleveland Prize, the Bruce Medal, and the Gold Medal of the Royal Astronomical Society. A lunar crater, asteroid 2069 Hubble, and the Hubble Space Telescope are named in his honor.



Hulse, Russel Alan (*1950), an American physicist and radio astronomer, co-discovered the binary neutron star system PSR 1913+16 alongside Joseph Taylor. On this unique relativistic system, they observed a decrease in the orbital period that precisely matched the energy loss predicted by general relativity through the emission of gravitational waves. This provided the first indirect evidence for the existence of gravitational waves. For this groundbreaking work, Russel Hulse and Joseph Taylor shared the 1993 Nobel Prize in Physics. Hulse earned his PhD in 1975 from the University of Massachusetts Amherst and subsequently worked at the Green Bank Observatory. He later spent many years conducting research at the Princeton Plasma Physics Laboratory. Since 2003, he has taught at the University of Texas.



Christoffel, Elwin Bruno (1829–1900), a German mathematician, made foundational contributions to tensor calculus, where the coefficients of an affine connection – the Christoffel symbols – bear his name. He also worked in differential equations, conformal mapping, potential theory, invariant theory, and orthogonal polynomials, while expanding his research into geodesics and shock waves. He studied at the University of Berlin where he was a student of Peter Gustav Lejeune Dirichlet, earning his PhD in 1856. In 1872, he became a professor at the University of Strasbourg, remaining there until retirement.



Kerr, Roy (1934), a New Zealand mathematician, discovered the first exact solution to Einstein's field equations of general relativity for a rotating, symmetric mass in 1963. This breakthrough described the geometry of spacetime surrounding rotating black holes. He graduated from the University of Canterbury and earned his PhD from the University of Cambridge. After working at Syracuse University, he joined the University of Texas at Austin in 1962, where he discovered the Kerr metric. He returned to the University of Canterbury in 1971.

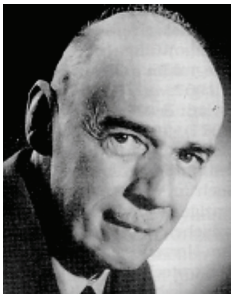


Khvolson, Orest Danilovich (1852–1934), a Russian and later Soviet physicist, was one of the first to propose and analyze the gravitational lensing phenomenon. He published his work in 1924, long before Einstein's brief 1936 note on the subject, though the first gravitational lens was not observed until 1979. Khvolson studied at the University of St. Petersburg, where he later became a professor. He published numerous papers across physics, including electricity, magnetism, photometry, and solar surface temperature measurements. In 1924, he demonstrated that a massive object could act as a lens, focusing light from a perfectly aligned background source into a ring-like structure, known today as an Einstein ring. Khvolson became an honorary member of the Soviet Academy of Sciences and received the Order of the Red Banner of Labor, a high state honor. The lunar crater Khvolson on the far side of the Moon is named in his memory.

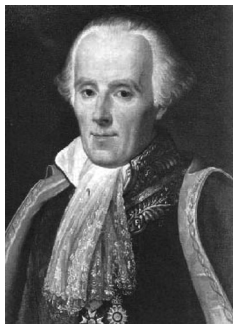
Laplace, Pierre Simon, see section *Electromagnetic Field.*, this textbook.



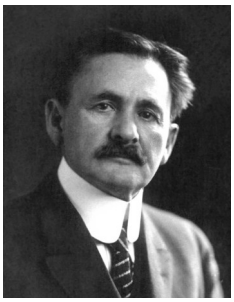
Lemaître, Georges Henri Joseph Édouard (1894–1966), a general relativity specialist, a Belgian priest, initially an abbé and later a monseigneur. Before his academic career, he served as an officer in World War I, receiving the Croix de Guerre. He later became a professor at the Catholic University of Leuven. He was familiar with the work of Alexander Friedmann, who used general relativity to show that a homogeneous universe could either expand or contract, and he knew of the redshifts of extragalactic nebulae measured by Vesto Slipher. Lemaître was the first to derive from general relativity, that in an expanding universe, the recession velocity of a galaxy is proportional to its distance (the Hubble–Lemaître law), and he estimated the proportionality constant from Slipher's data. His estimated value was roughly ten times larger due to early uncertainties in cosmic distances. Lemaître realized that an expanding universe must have originated from an exceptionally hot and dense state, which he termed the "primeval atom," laying the conceptual foundation of the hot early phase of the universe. He published his findings in French in 1927 in a local Brussels journal. Although it is often said his work was unknown in the English-speaking world, Lemaître explicitly shared his results with Albert Einstein and attended a conference alongside Edwin Hubble, who rediscovered the linear expansion relation two years later in 1929. The priority of discovering the expanding universe clearly belongs to Lemaître. Shortly before his death, the discovery of the cosmic microwave background in 1965 provided key evidence for his theory.



Lense, Josef (1890–1985), an Austrian mathematician, assisted Hans Thirring with spacetime metric calculations near a rotating mass. The resulting frame-dragging phenomenon, which they predicted in 1918, is called the Lense–Thirring effect. Lense served as a professor of applied mathematics at the Technical University of Munich, directing its Mathematical Institute from 1946 to 1951. Beyond mathematical physics, he worked in pure mathematics, including descriptive geometry, series expansions, and special functions, while also studying the history and evolution of mathematics and its relationship to other sciences.



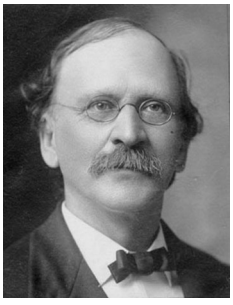
Michell, John (1724–1793), an English philosopher, physicist, and geophysicist, pioneered seismology, astrophysics, and optics, while researching magnetism and gravitation. His remarkably broad scope led him to propose the existence of objects whose gravity is too strong for light to escape (known today as black holes). He hypothesized that earthquakes propagate as seismic waves, demonstrated that binary stars are gravitationally bound, designed artificial magnets, built astronomical telescopes, and invented the torsion balance to weigh the Earth. The apparatus was later utilized by his friend Henry Cavendish to determine the gravitational constant. From an experimental standpoint, measuring the Earth's mass and finding this constant are identical tasks. Michell is considered the father of both seismology and magnetometry, and he applied statistical methods to astronomy. He graduated from Queens' College, Cambridge.



Michelson, Albert Abraham (1852–1931), a physicist born in Poland, immigrated to the United States with his parents at age three and entered the U.S. Naval Academy at seventeen. He quickly became a renowned American scientist. After spending two years gaining experience in optics in Germany and France (1880–1882), he returned to the United States. He invented the Michelson interferometer to study the Earth's motion through the luminiferous aether. Conducting numerous trials with Edward Morley, known as the Michelson–Morley experiments, they concluded that the speed of light is always invariant, regardless of the orientation, location, or timing of the apparatus. This definitively confirmed Maxwell's electrodynamics, which implies a different velocity addition rule than classical mechanics. For his experimental achievements, Michelson became the first American to receive the Nobel Prize in Physics in 1907.



Misner, Charles (1932–2023), an American physicist, pioneer in quantum gravity and numerical relativity. Alongside Kip Thorne and John Wheeler, he co-authored *Gravitation*, the definitive textbook on general relativity. His research explored cosmology, the horizon problem, and topology in relativity. He was a co-developer of the Hamiltonian formulation of Einstein's field equations – the ADM formalism (Arnowitt–Deser–Misner) serving for numerical relativity and alternative paths toward quantum gravity. He completed his undergraduate studies at the University of Notre Dame and earned his PhD from Princeton.



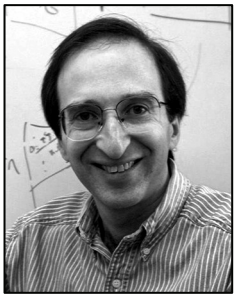
Morley, Edward (1838–1923), an American chemist, scientist and an exceptional optician, who worked closely with Michelson around 1887, when they conducted the famous experiment demonstrating that the speed of light is independent of the observer's frame of reference. Morley graduated from Williams College in 1860 and became a professor at Western Reserve College in 1869. He served as president of the American Association for the Advancement of Science (AAAS) in 1895 and as president of the American Chemical Society in 1899. He received numerous awards, and a lunar crater is named in his honor.



Nordström, Gunnar (1881–1923), a Finnish theoretical physicist, is best known for calculating the spacetime metric surrounding a spherically symmetric charged mass. The same solution to Einstein's field equations was independently discovered by Hans Reissner, Hermann Weyl, and George Barker Jeffery, and is known today as the Reissner–Nordström metric. Nordström also formulated an early alternative scalar theory of gravity, which failed because it could not predict the bending of light around massive objects. He shared a warm relationship of mutual respect with Einstein. In 1914, Nordström became the first to utilize an extra spatial dimension to unify electromagnetism and gravity, a mechanism later vital to Kaluza–Klein theory. Nordström earned his bachelor's degree from the Polytechnic Institute of Helsinki and his master's and PhD from the University of Helsinki. He also conducted research at Göttingen and Leiden.



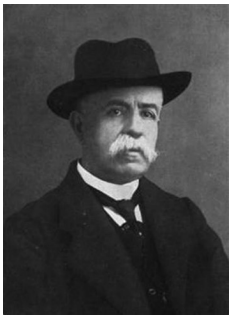
Paczyński, Bohdan (1940–2007), a Polish astronomer, investigated stellar evolution, accretion disks, gamma-ray bursts, and gravitational microlensing. Paczyński initiated the *Optical Gravitational Lensing Experiment* (OGLE), which became the definitive prototype for detecting gravitational microlenses; in fact, he coined the term "microlensing" itself. He also co-founded the *All Sky Automated Survey* (ASAS). Paczyński studied astronomy at the University of Warsaw. In 1981, he delivered a series of lectures at Caltech, subsequently relocating permanently to the United States, where he served as a professor of astrophysics at Princeton University until his death. For his groundbreaking work, he received numerous prestigious honors, including the Karl Schwarzschild Medal, the Eddington Medal, the Henry Draper Medal, the Gold Medal of the Royal Astronomical Society, the Bruce Medal, the Bruno Rossi Prize, and the Danny Heineman Prize.



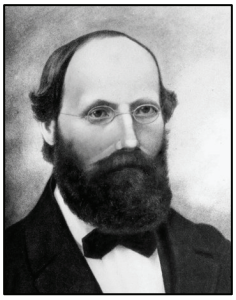
Perlmutter, Saul (*1969), an American astrophysicist, is a senior scientist at the Lawrence Berkeley National Laboratory and a professor of physics at the University of California, Berkeley. Perlmutter studied at Harvard University and earned his PhD from the University of California. He leads the *Supernova Cosmology Project* (SCP), which, independently of the team led by Adam Riess and Brian Schmidt, discovered the accelerated expansion of the universe in 1998 through observations of Type Ia supernovae. The SCP group published its historic findings in early 1999. The ongoing project has compiled a massive dataset suggesting that dark energy is most likely linked to the quantum properties of the vacuum. Perlmutter also led the development of the *SuperNova/Acceleration Probe* (SNAP), a space telescope concept designed to refine expansion models. This mission was later integrated into the *Joint Dark Energy Mission* (JDEM) and ultimately evolved into the *Nancy Grace Roman Space Telescope* project, scheduled for launch in 2027. Perlmutter has received numerous accolades, including the 2006 Shaw Prize in Astronomy, the 2007 Gruber Prize in Cosmology, and the 2015 Breakthrough Prize in Fundamental Physics, culminating in the 2011 Nobel Prize in Physics.



Reissner, Hans Jacob (1874–1967), a German aeronautical engineer, designed the world's first successful all-metal aircraft in 1912, utilizing corrugated duraluminum, while collaborating with Hugo Junkers, who was inspired by this partnership to enter aviation. For his pioneering wartime aircraft designs, Reissner was awarded the Iron Cross, Second Class. Theoretical physics was his great passion; he was the first to derive from Einstein's general relativistic field equations the spacetime metric surrounding a charged, spherically symmetric mass, elegantly combining gravitational theory with Maxwell's electromagnetism. This solution, independently co-derived by Gunnar Nordström, Hermann Weyl, and George Barker Jeffery, is known as the Reissner–Nordström metric. Reissner held professorships at RWTH Aachen University and the Technical University of Berlin. Due to his Jewish ancestry, he was forced to flee Nazi Germany, emigrating in 1938 to the United States, where he lectured at the Illinois Institute of Technology before spending the remainder of his career at the Polytechnic Institute of Brooklyn.



Ricci-Curbastro, Gregorio (1853–1925), an Italian mathematician, focused his research on tensor calculus, differential equations, and hypergeometric functions. Alongside his student Tullio Levi-Civita, he introduced tensor calculus into physics, providing the mathematical foundation for general relativity. The Ricci curvature tensor, a symmetric two-index tensor obtained by contracting the Riemann curvature tensor, is named in his honor. At age sixteen, he began studying mathematics at the University of Rome but left after a year due to political unrest. He resumed his studies in Pisa, earning his doctorate in 1875. From 1877 to 1878, he studied under Felix Klein at the Technical University of Munich. He returned to Pisa in 1879, and from 1880 until his death, he served as a professor of mathematics at the University of Padua. In his publications, he used the shortened name Ricci.

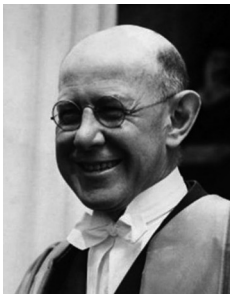


Riemann, Bernhard George Fridrich (1826–1866), a German mathematician who died in Italy, produced monumental works in differential geometry, complex analysis, and mathematical physics, laying the essential mathematical groundwork for general relativity. He also investigated Abelian functions and elliptic integrals, and was an exceptionally gifted educator whose immense influence permanently reshaped modern mathematics. Initially studying theology, Riemann entered the University of Göttingen in 1846, where his father secured sufficient funding for him to attend lectures by Carl Friedrich Gauss. In 1847, he moved to Berlin to study for two years before returning to Göttingen, where he was appointed extraordinary professor in 1857. Dozens of mathematical concepts bear his name, most notably non-Euclidean Riemannian geometry, which he began developing in 1854 and enabled the subsequent creation of general relativity, alongside the Riemann curvature tensor describing the properties of this geometry. Eponymous concepts also include the Cauchy–Riemann equations for the existence of a complex function derivative, the generalized Riemann integral, the Riemann–Stieltjes integral, the Riemann zeta function, the Riemann theta function, the Riemann sphere, and the Riemann surface. A lunar crater and asteroid 4167 Riemann are also named in his memory.



Riess, Adam Guy (*1969), an American physicist and astrophysicist, was born to parents who fled Nazi Germany for the United States in 1936. Riess graduated from the Massachusetts Institute of Technology (MIT) and earned his PhD from Harvard University in 1996. His doctoral dissertation focused on determining cosmic distances using Type Ia supernovae. Alongside Brian Schmidt, he co-led the *High-z Supernova Search Team*, tracking cosmic expansion through galaxies whose distances were calibrated via Type Ia supernovae. In 1998, their team announced the discovery of the universe's accelerated expansion. A group led by Saul Perlmutter independently published the same discovery nearly simultaneously. Riess is currently a Bloomberg Distinguished Professor at Johns Hopkins University and a distinguished astronomer at the Space Telescope Science Institute. He leads the SH0ES (*Supernovae and H_0 for the Equation of State of Dark Energy*)

project, which utilizes the Hubble Space Telescope to refine the value of the Hubble constant. Riess has received numerous accolades, most notably the 2006 Shaw Prize in Astronomy and the 2011 Nobel Prize in Physics.



Rosen, Nathan (1909–1995), an American-Israeli theoretical physicist, was born to a Jewish family in Brooklyn, New York. He studied physics at the Massachusetts Institute of Technology (MIT) and dedicated his career primarily to quantum mechanics and general relativity. He is most famous for formulating the EPR paradox (Einstein–Podolsky–Rosen paradox), which describes how the quantum entanglement of a particle pair allows a measurement on one particle to instantly reveal information about the other. He utilized this entanglement in the first correct description of electron states in a hydrogen molecule. In 1935, alongside Albert Einstein, he predicted the theoretical possibility of a spacetime bridge connecting two distant regions of space. Later termed a wormhole, this initial Einstein–Rosen bridge solution proved highly unstable against minor perturbations, collapsing too quickly for even a photon to pass through. However, physicists continue to search for spacetime metrics that could allow rapid transit between remote regions of the cosmos. Rosen traveled extensively, spending two years in the Soviet Union before settling in Israel, which he considered his true home. In Israel, he helped establish the Israel Academy of Sciences and Humanities, founded the Israel Physical Society (serving as its president from 1955 to 1957), and led the International Society on General Relativity and Gravitation as president from 1974 to 1977.

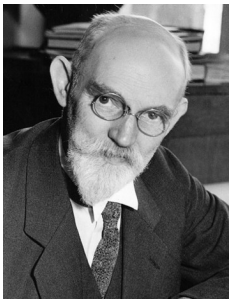


Schiff, Leonard Isaac (1915–1971), an American physicist, was the visionary behind orbital testing of general relativity. Schiff studied theoretical physics at Ohio State University and earned his PhD from the Massachusetts Institute of Technology (MIT). Throughout his career, he held positions at the University of California, Caltech, the University of Pennsylvania, and Los Alamos, before joining Stanford University in 1948, where he remained until his death. At Stanford, he authored a widely used textbook on quantum mechanics, first published in 1949. From the late 1950s, he advocated for testing general relativity using gyroscopes moving along geodesics in Earth orbit, which laid the groundwork for the Gravity Probe B satellite mission developed at Stanford. The satellite was launched in 2004, 33 years after his death; it successfully detected Earth's geodetic effect but failed to conclusively measure the frame-dragging (Lense–Thirring) effect due to underestimated interference from solar wind plasma. An exceptional educator, Schiff received the Oersted Medal and the Dinkelspiel Award, and a lecture hall at Stanford is named in his honor.



Schwarzschild, Karl (1873–1916), a German mathematician and physicist, studied in Munich, focusing on the origin of the solar system. He published his first scientific paper on celestial mechanics at just sixteen years old. He spent nine years at Göttingen, collaborating with Felix Klein, David Hilbert, and Hermann Minkowski. His primary research covered electrodynamics, optics, and stellar radiation. While serving on the Rus-

sian front during World War I, he engaged with the emerging quantum theory and general relativity. He discovered the first exact solution to Einstein's field equations for a spherically symmetric mass, which led to the theoretical discovery of black holes. This formulation bears his name (Schwarzschild solution, 1916), as does the event horizon boundary (Schwarzschild radius). Schwarzschild also derived the stability criterion for the onset of convection zones inside stars. The lunar crater Schwarzschild is named in his memory. He contracted a severe illness while in Russia, a painful autoimmune skin disease called pemphigus vulgaris, and died shortly after returning home.



De Sitter, Willem (1872–1934), a Dutch astronomer, physicist, and mathematician, discovered a solution to Einstein's field equations of general relativity for the case of a positive, non-zero cosmological constant and zero mass density. This framework is known today as the de Sitter universe, which expands exponentially and contains no matter, meaning the Hubble constant remains invariant throughout the expansion. In 1932, de Sitter and Einstein co-published a landmark paper proposing that the universe might contain a vast amount of non-luminous matter. De Sitter studied at the University of Groningen. He spent two years (1897 to 1899) working at the Royal Observatory at the Cape of Good Hope in South Africa. In 1908, he was appointed chair of astronomy at Leiden University, and in 1919, he became the director of the Leiden Observatory, a position he held until his death. The lunar crater De Sitter and asteroid 1686 De Sitter are named in his honor.



Slipher, Vesto Melvin (1875–1969), an American astronomer, specialized in the spectroscopy of galaxies and nebulae. He studied at Indiana University, earning his PhD in 1909, and subsequently joined the Lowell Observatory to conduct spectroscopic research. Initially focusing on determining planetary rotation periods and atmospheric compositions, he later shifted his attention to measuring the spectra of nebulae. At the time, other galaxies were classified as nebulae, or extragalactic nebulae once distance estimates were established. In 1912, Slipher discovered a blueshift in the Andromeda Nebula, indicating it is moving toward us. In subsequent years, he found that the vast majority of other objects exhibited redshifts. Although he was on the verge of discovering cosmic expansion, he left his measurements uninterpreted. Another prominent astronomer, Arthur Eddington, remained highly cautious, viewing the lack of data from the Southern Hemisphere as a major hurdle and suggesting the redshifts might stem from our own motion. The data was correctly interpreted as cosmic expansion by Belgian Georges Lemaître in 1927. His paper was published in a local Brussels journal and remained unknown to some cosmologists. Consequently, the discovery of the expanding universe is mistakenly credited to Edwin Hubble, who supplemented Slipher's measurements with his own data to arrive at the same conclusion.



Taylor, Joseph Hooton (*1941), an American physicist based at Princeton University, co-investigated the binary neutron star system PSR 1913+16, discovered by Arecibo in 1974. Monitoring this system alongside Russel Hulse, they observed numerous

relativistic effects, including a massive periastron precession of over four degrees per year. Their most significant finding was an orbital period decrease matching the energy loss predicted via gravitational waves, providing the first indirect evidence of their existence. For this discovery, Taylor and Hulse shared the 1993 Nobel Prize in Physics. Taylor earned his PhD from Harvard in 1968. After working at the University of Massachusetts, where he directed the Five College Radio Astronomy Observatory (FCRAO), he focused his career on radio astronomy at various prominent research centers.



Thirring, Hans (1888–1976), an Austrian theoretical physicist, was the first to derive the spacetime metric surrounding a rotating, spherically symmetric mass in 1918. Assisted with the complex calculations by mathematician Josef Lense, they demonstrated that spacetime is dragged along by a rotating body, a frame-dragging phenomenon known today as the Lense–Thirring effect. Thirring graduated from the University of Vienna. Renowned for his staunch pacifist views, he served for six years as a member of the upper chamber of the Austrian parliament. He received the Haitinger Prize for his scientific achievements. His son, Walter Thirring, also pursued a distinguished career in physics.

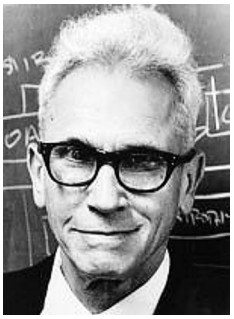


Thorne, Kip Stephen (*1940), an American theoretical physicist, specializes in gravitation theory, astrophysics, and cosmology. An expert on applying general relativity to astrophysical phenomena, he famously explored the theoretical possibility of traversable wormholes connecting distant regions of spacetime. He earned his bachelor's degree from Caltech and his master's and PhD from Princeton University under John Archibald Wheeler, becoming a full professor at Caltech in 1970. A dedicated educator who mentored generations of physicists, Thorne collaborated closely with Stephen Hawking and Carl Sagan. He authored numerous popular and academic books, including the popular science bestseller *Black Holes and Warped Spacetime*, and co-authored *Gravitation* (MTW after its authors Misner, Thorne, and Wheeler), which remains the definitive textbook on general relativity. He also contributed to the screenplays and scientific accuracy of several science fiction films. Thorne developed a novel method for visualizing warped spacetime using vortex and tendril lines as an alternative to traditional embedding diagrams. A visionary behind interferometric gravitational wave detectors, he co-founded the MARK experiments and the LIGO project, sharing the 2017 Nobel Prize in Physics for his decisive contributions to the LIGO detector and the observation of gravitational waves.

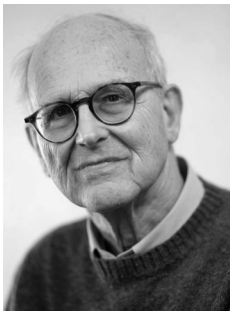


Walsh, Dennis (1933–2005), an English astronomer, conducted observations in both optical and radio spectra using a wide range of instruments. His optical research utilized the 2.54-meter Isaac Newton Telescope on La Palma, while his radio work relied on the 76-meter Lovell Telescope at Jodrell Bank near Manchester. In 1979, spectroscopic monitoring of quasars from the Jodrell Bank 966 MHz Survey led him to discover the first gravitational lens. He revealed that a pair of close quasars shared nearly identical spectra, proving they were twin images of a single object formed by an intervening gravitational lens. Walsh studied physics at the University of Man-

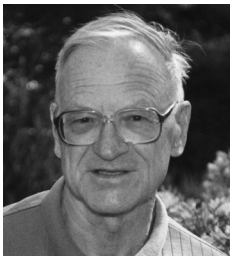
chester and earned his PhD at Jodrell Bank. From 1959, he researched the ionosphere and lectured at the University of Michigan. Returning to Manchester in 1967, he focused on satellite-based atmospheric research and served as treasurer of the Royal Astronomical Society for eight years.



Weber, Joseph (1919–2000), an American physicist, pioneered the development of the maser and laser principles and became the first to attempt the direct detection of gravitational waves. Weber studied electronics at the U.S. Naval Academy, serving during World War II, before earning his PhD from the Catholic University of America. In 1948, he joined the faculty at the University of Maryland, focusing on microwave spectroscopy in chemical kinetics. At a 1952 public conference, he outlined the concept of coherent amplification of microwaves and light. This principle was independently developed by Charles Townes, Nikolay Basov, and Aleksandr Prokhorov, who built functional prototypes and shared the 1964 Nobel Prize in Physics. By 1966, Weber began hunting for gravitational waves using massive resonant aluminum cylinders. To isolate random vibrations, he operated two detectors simultaneously, spaced 1,000 kilometers apart in Maryland and Argonne. Although he recorded a coincident signal, it was later deemed a statistical fluke due to data processing errors. He also designed a lunar gravitational wave detector deployed during the Apollo 17 mission in 1972.

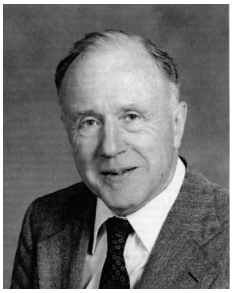


Weiss, Reiner (*1932), an American physicist born in Germany, fled rising anti-Semitic persecution with his Jewish family, relocating first to Prague and later to the United States following the geopolitical crisis in Czechoslovakia. He studied at the Massachusetts Institute of Technology (MIT), where he spent most of his career and now holds the status of professor emeritus. He developed critical technologies for the *Cosmic Background Explorer* (COBE) satellite, becoming the first to measure the cosmic microwave background (CMB) spectrum using a monolithic silicon bolometer. COBE discovered the temperature anisotropy and primordial fluctuations that hold vital clues about the early universe, which were later analyzed by the WMAP and Planck space probes. Weiss also became a foundational figure in the development of the LIGO gravitational wave detector, inventing the laser interferometry technique upon which the experiment is based. He also contributes to the Fermilab Holometer, a 40-meter interferometer probing microscopic quantum fluctuations. His accolades include the Gruber Cosmology Prize, the Einstein Prize, the Shaw Prize, the Kavli Prize, and the 2017 Nobel Prize in Physics.



Weymann, Ray (*1934), an American astronomer and astrophysicist, investigated galactic mass ejections and the redshifts of galaxies at cosmological distances. In 1979, he co-discovered the first gravitational lens. Alongside his colleagues, he captured the spectrum of a twin quasar using the University of Arizona's 2.3-meter telescope. The nearly identical spectra of both objects revealed they were twin images of a single quasar split by an intervening gravitational lens. He studied at Princeton Univer-

sity, where he earned his PhD. He is a member of the U.S. National Academy of Sciences and served as president of the Astronomical Society of the Pacific. He founded the *Climate Science Rapid Response Team* (CSRRT) to provide the media with accurate information on climate issues. Weymann's primary research institution was the Carnegie Institution of Washington (CIW), also known as the Carnegie Institution for Science.



Wheeler, John Archibald (1911–2008), a leading American theoretical physicist, dedicated his career to general relativity. Born in Florida, he studied physics at Johns Hopkins University, earning his PhD in 1933 before continuing his studies in Copenhagen under Niels Bohr. Reunited in the United States, he and Bohr co-developed the liquid drop model of nuclear fission. Wheeler then began teaching physics at Princeton University and later joined the University of Texas at Austin. His graduate students included Richard Feynman, Kip Thorne, and Charles Misner. During World War II, he participated in the development of the atomic bomb and subsequently the hydrogen bomb. Returning to Princeton after the war, he began teaching general relativity in 1953. Seeking to make the theory more accessible, he coined the terms “black hole” and “wormhole” for better conceptual understanding, and demonstrated that wormholes are unstable solutions to Einstein's equations. In 1973, Wheeler, Misner, and Thorne co-authored Gravitation, which remains the most comprehensive compilation of gravitational physics knowledge and research queries. He also explored quantum physics and its synthesis with general relativity. For his profound research, he received numerous accolades, including the Albert Einstein Award, the Enrico Fermi Award, the Franklin Medal, the National Medal of Science, the Oersted Medal, the Oppenheimer Prize, and the Wolf Prize.

Subject Index

absolute temperature 13, 28
accelerated expansion 332
affine connection coefficients 240
Aharonov-Bohm experiment 175
Ampère law 108–111
anharmonic oscillator 52
angular frequency 150
averaging 22
Berkeley experiment 260
Biot-Savart law 136
Boltzmann constant 28, 45
Boltzmann distribution 37
Boltzmann equation 29
Bose-Einstein distribution 63
bosons 61, 63
  ensemble of 63
boundary conditions 115
bremsstrahlung 187
  relativistic 195
Brewster law 167
cavity field 199–208
cavity potentials 199
charge conservation 100
chemical potential 13, 34, 57, 59
Christoffel symbols 241–243
circulation of a field 105, 119
classical velocity addition 210
compressibility 77, 79
conservative field 119, 126, 270
continuity equation 23, 100, 141, 268, 276
cosmological constant 332
cosmology 322–333
Curie law 92–93
cyclotron radiation 189
Debye length 124
derivative
  covariant 240
  total 24, 244
dispersion 79
dispersion relation 151–157, 280
divine intervention / miracle 12
Einstein-Rosen bridge 299–301
Einstein law of gravitation 262–277
electric dipole moment 83, 102, 129
electric dipole radiation 181
electric field energy 126
electromagnetic field 99–208
electromagnetic field tensor 172
electromagnetic field transformation 229
electrostatics 119–132
embedding diagrams 298–301
energy conservation 24, 88, 101, 141–142
enthalpy 14
entropic force 81–82
entropy 13, 28, 77–80
Eötvös experiment 238
equivalence principle 236–240
  strong 237
  very strong 237
  weak 237
ergodic problem 22
expansion function 330
Faraday law of induction 107–108
fermions 61–63, 65
  ensemble of 61–62, 65
field lines / Field at a sharp tip 125
field flux 105–106
first law of thermodynamics 13
fluctuations 77
  energy 78
  magnetization 89
  mean square 77, 89
  velocity 78
fhoton energy during expansion 75
fhotons, ensemble of 68
four-vectors 172, 219–222, 241
free energy 15
free space potentials 175
Fresnel equations 165
Friedmann metric 322
fugacity 60
Galilean transformation 210
gauge freedom 173
gauge transformation 133, 173
Gauss law for electrostatics 107
Gauss law for magnetism 105
geodesic equation 244
Gibbs canonical ensemble 26–32
Gibbs-Helmholtz equation
  1st type 15, 31

2nd type 16
Gibbs potential 16
grandcanonical ensemble 56–60
grandcanonical potential 16
gravitational lenses 312–315
gravitational redshift 252–261
gravitational waves 278–291
gravitoelectric field 317
gravitomagnetic field 318
Gravity Probe A 259–260
Gravity Probe B 239, 266–267, 320
Green function 120–122
Hafele-Keating experiment 257
Hamilton equations 18, 23, 154, 169
  for fields 203
Haroche resonant cavity 199, 208
Hharmonic oscillator 40, 42
Heisenberg model 96
Hertzian dipole 183
Hubbard model 97
ideal gas 33
integrating factor 13
interval 223
Ising model 93
Lagrangian 174, 304
Landé *g*-factor 90
Larmor formula
  differential 182
  integral 182
lattice models 93–98
laws of reflection and refraction 163–165
Lense-Thirring effect 319
LIC (Locally Inertial Coordinate / Light-Cone) 192
Liénard-Wiechert potentials 192
Liouville theorem 23
local inertial frame 239
London effect 267
Lorenz gauge condition 135, 175, 199, 200, 279
magnetic dipole moment 85, 90, 104, 137
magnetic field energy 141
magnetic resonance 91
magnetization 77, 86–89, 92, 104, 109–115, 133, 137
magnetostatics 133–140
mass
  gravitational 236
  inertial 236
Maxwell equations 100–118
Maxwell distribution 38
mechanical work 13
metric 219
Minkowski metric 219
moment
  dipole 83, 102, 129, 137
  gravitational dipole 282–283
  second central 77
Morse potential 44
multipole expansion
  radiative 178
  stationary 128, 136
Newtonian limit 248
parallel transport 245
partition function
  canonical 30
  grandcanonical 58
phase
  space 18
  velocity 152
  volume 18–21
phenomenology 12
Planck radiation law 68–71, 206–207
Poisson equation 119–124, 134, 175, 274
polarization 68, 84, 102, 113, 158
potentials
  ambiguity / gauge freedom 279
  four-potential 173, 179, 221, 226, 229
  retarded and advanced 175–177
  scalar 179, 199, 226, 279
  vector 133
Potts model 95
Pound-Rebka experiment 256
Poynting vector 142, 166, 181, 195
probability density 21
quadrupole 130
quadrupole moment 130–131, 180, 191
quantum rotator 47
quantum vibrator 42
radiation
  as dense as water 75
  blackbody / thermal (Planck) 68, 205
  higher moments 191
radiation pressure 146
rapidity 217
Rayleigh scattering 186
raytracing 169
redshift
  cosmological 328–329
relativistic velocity addition 214
relativity
  general 236–333
  special 210–235
Ricci tensor 278, 293
Riemann curvature tensor 262–267, 278, 293, 301
root-mean-square speed 39
rotating cylinder 37

scalar curvature 265
Schwarzschild metric 292
second law of thermodynamics 13
solar constant 74
solar power / luminosity 74
space contraction / Length contraction 214–216
specific heat capacity / Heat capacity
  diatomic gas 50–52
  rotational degrees of freedom 47–49
  translational degrees of freedom 51
  two-level system 54–55
  vibrational degrees of freedom 42–45
speed
  mean 39
  most probable 39
spin glasses 96
SQUID 267
state density / Density of states 25
static pressure 146
Stefan-Boltzmann law 72
stress-energy tensor
  (energy-momentum tensor) 271
  matter and energy 272–273
synchrotron radiation 195
t-J model 98
temperature
  planetary 74
  rotational 48–51
  vibrational 44, 48
thermal energy 13, 43, 48
thermodynamic potentials 14–17
thermodynamics 12–17
Thomson scattering 184
time dilation 214–216, 252–255
total internal reflection 168
w-parameter
  (equation of state parameter) 331
wave
  basic concepts 150
  in an anisotropic medium 159
  in a conductor 162
  in vacuum 156
  packet 153–154
  plane wave 151
  vector 150
wavefront 150
Wien displacement law 71, 73
Zeeman effect 91
ZQ model 95



References

Lecture Recordings

www.aldebaran.cz, section Study;
www.vsprednasky.cz;

Other Volumes of this Textbook

[1] P. Kulhánek: *Selected Topics in Theoretical Physics I (Theoretical Mechanics, Quantum Theory, Mathematics for Physics)*; AGA 2020, ISBN: 978-80-906638-2-4, online version (2026): https://www.aldebaran.cz/studium/TF1_en.pdf

[2] P. Kulhánek: *Selected Topics in Theoretical Physics III* (*Plasma Theory, Numerical Simulations)*; AGA 2020, ISBN 978-80-906638-4-8, online version (2026): https://www.aldebaran.cz/studium/TF3_en.pdf

Statistical Physics

[3] H. B. Callen: *Thermodynamics and an Introduction to Thermostatistics*; John Wiley & Sons, 1985

[4] M. Kardar: *Statistical Physics of Particles*; Cambridge University Press, 2007

[5] D. V. Schroeder: *An Introduction to Thermal Physics*; Oxford Univ. Press, 2021

[6] R. K. Pathria, P. D. Beale: *Statistical Mechanics*; Elsevier, 2021

[7] F. Bloch, J. Walecka: *Fundamentals of Statistical Mechanics*; World Scientic, 2000

[8] M. Dvorak, T. Ohno: *Liouville's Theorem*; PHGN 505 Report, Colorado School of Mines 2011

[9] P. Atkins, J. de Paula, J. Keeler: *Atkins' Physical Chemistry*; Oxford University Press, 2022

[10] S.: *Molecular Vibrational Spectroscopy*; University of Oxford, 2011

[11] L. Kadanoff: *Bosons and fermions*; Lecture Notes, Perimeter Institute statistical physics, 2009

[12] M. Towler: *Exchange, antisymmetry and Pauli repulsion*; TCM Group, Cavendish Laboratory, University of Cambridge; 2010

[13] J. P. Sethna. *Statistical Mechanics*: *Entropy, Order Parameters, and Complexity*, Oxford University Press, 2008

[14] Erik Verlinde: *On the origin of gravity and the laws of Newton*; Journal of High Energy Physics 04 (2011) 029; online: https://link.springer.com/content/pdf/10.1007/JHEP04(2011)029.pdf

[15] Erik Verlinde: *Emergent Gravity and the Dark Universe*; SciPost Phys. 2, 016 (2017); arXiv:1611.02269v2 [hep-th] 8 Nov 2016; online: https://arxiv.org/abs/1611.02269

[16] T. Padmanabhan: *Thermodynamical Aspects of Gravity: New Insights*; Reports on Progress in Physics 73 (2010) 046901

[17] R. Bauerschmidt: *Ferromagnetic spin systems*; Harvard 2016; online: http://www.math.harvard.edu/~brt/doc/spin.pdf

[18] M. Plischke, B. Bergersen: *Equilibrium Statistical Physics* (4th Edition); World Scientific, 2023

[19] J. Tobochnik, H. Gould: *Statistical and Thermal Physics: With Computer Applications* (2nd Edition); Princeton University Press, 2021

[20] S. R. A. Salinas: *Introduction to Statistical Physics* (2nd Edition); Springer, 2022

Electromagnetic Field

[21] James Clerk Maxwell: *A Treatise on Electricity and Magnetism I, II*; Oxford, Clarendon Press 1873

[22] Julius Adam Stratton: *Electromagnetic Theory*; McGRAw-HiLL Book Company 1941, další reedice Adams Press 2008

[23] Anthony L. Perat: *Physics of the Plasma Universe*; Springer 1991

[24] Lev Davidovich Landau, Evgeny Lifschitz: *The Classical Theory of Fields – Course of Theoretical Physics 2*; originál 1939, 1973, reedice 1994 Butterworth Heinemann

[25] Lev Davidovich Landau, Evgeny Lifshitz: *Electrodynamics of Continuous Media*, Volume 8 in Course of Theoretical Physics; Elsevier 1984

[26] John David Jackson: *Classical Electrodynamics* (3rd edition); Wiley 2021; ISBN 1119770769

[27] George B. Rybicki, Alan P. Lightman: *Radiative Processes in Astrophysics*; John Willey & Sons 1979.

[28] Thomas Howard Stix: *Waves in Plasmas*; Springer 2006

[29] Umran S. Inan, Robert A. Marshall: *Numerical Electromagnetics – The FDTD Method*; Cambridge University Press 2011

[30] Anupam Garg: *Classical Electromagnetism in a Nutshell*; Princeton University Press 2012

[31] Andrew Zangwill: *Modern Electrodynamics*; Cambridge University Press 2013

[32] David J. Griffith: *Introduction to Electrodynamics* (5th Edition); Cambridge University Press 2024

[33] Edward M. Purcell, David J. Morin: *Electricity and Magnetism* (3rd Edition); Cambridge University Press, 2013

[34] Richard P. Feynman, Robert B. Leighton, Matthew Sands: *The Feynman Lectures on Physics*, Volume II – Basic Books, 2010

[35] Melvin Schwartz: *Principles of Electrodynamics*; Dover Publications, 1987

[36] Paul Lorrain, Dale R. Corson, François Lorrain: *Electromagnetic Fields and Waves* (3rd Edition); W. H. Freeman, 1988

[37] Jakob Schwichtenberg: *No-Nonsense Electrodynamics: A Student Friendly Introduction*; No-Nonsense Books 2018

[38] John Pierrus: *Solved Problems in Classical Electromagnetism – Analytical and Numerical Solutions with Comments*; Oxford University Press 2018

[39] Ophelia Burgess: *Principles of Electromagnetism*; Larsen & Kefer 2019

[40] Jonathan W. Keohane, Joseph P. Foy: *An Introduction to Classical Electrodynamics*; Maricourt Academic Press 2019

[41] Xin Tao: *Classical Electrodynamics; University of Science and Technology of China*, 2019

[42] Robert Wald: *Advanced Classical Electromagnetism*; Princeton University Press, 2022

[43] Harald J.W. Müller-Kirsten: *Electrodynamics* (3rd Edition); World Scientific, 2026;

[44] Konstantin K. Likharev: *Classical Electrodynamics: Lecture Notes and Problems*; World Scientific, 2018

Relativity

[45] Ray d'Inverno, James Vickers: *Introducing Einstein's Relativity: A Deeper History*; Oxford University Press 2022

[46] N. M. J. Woodhouse: *Special Relativity*; Springer 2003 (Lecture Notes in Physics)

[47] Barbara Ryden: *Introduction to Cosmology* (2nd Edition); Cambridge University Press 2017

[48] Thomas A. Moore: *A General Relativity Workbook*; University Science Books 2012

[49] David J. Morin: *Special Relativity: For the Enthusiastic Beginner*; CreateSpace 2017

[50] Andrew Liddle: *An Introduction to Modern Cosmology* (3rd Edition); Wiley 2015

[51] N. David Mermin: *It's About Time: Understanding Einstein's Relativity*; Princeton University Press 2005

[52] Wolfgang Rindler: *Relativity: Special, General, and Cosmological* (2nd Edition); Oxford University Press 2006

[53] Pedro G. Ferreira: *The Perfect Theory: A Century of Geniuses and the Battle over General Relativity*; Mariner Books 2014

[54] James J.: Callahan: *The Geometry of Spacetime: an Introduction to Special and General Relativity*; Springer 2000

[55] James B. Hartle: *Gravity: an Introduction to Einstein's General Relativity*; Addison-Wesley 2003

[56] Sean M. Carroll: *Spacetime and Geometry: An Introduction to General Relativity*; Addison-Wesley 2004

[57] Ta-Pei Cheng: Relativity, *Gravitation and Cosmology: a Basic Introduction*; Oxford University Press 2005

[58] Øyvind Grøn, Sigbjorn Hervik: *Einstein's General theory of Relativity*; Springer 2007

[59] Bernard F. Schutz: *A First Course in General Relativity*; Cambridge University Press 2009

[60] Edwin F. Taylor, John Archibald Wheeler: *Spacetime Physics: Introduction to Special Relativity*; Freeman 2007

[61] Edwin F. Taylor, John Archibald Wheeler: *Exploring Black Holes: Introduction to General Relativity*; Addison Wesley 2010

[62] Charles W. Misner, Kip S. Thorne, John Archibald Wheeler: *Gravitation*; W. H. Freeman Princeton University Press 1973, 2017

[63] Robert M. Wald: *General Relativity*; University of Chicago Press 1984

[64] Steven Weinberg: *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity*; Wiley 1972

[65] Albert Einstein: Relativity: *The Special and the General Theory* (100th Anniversary Edition); Princeton University Press 2019

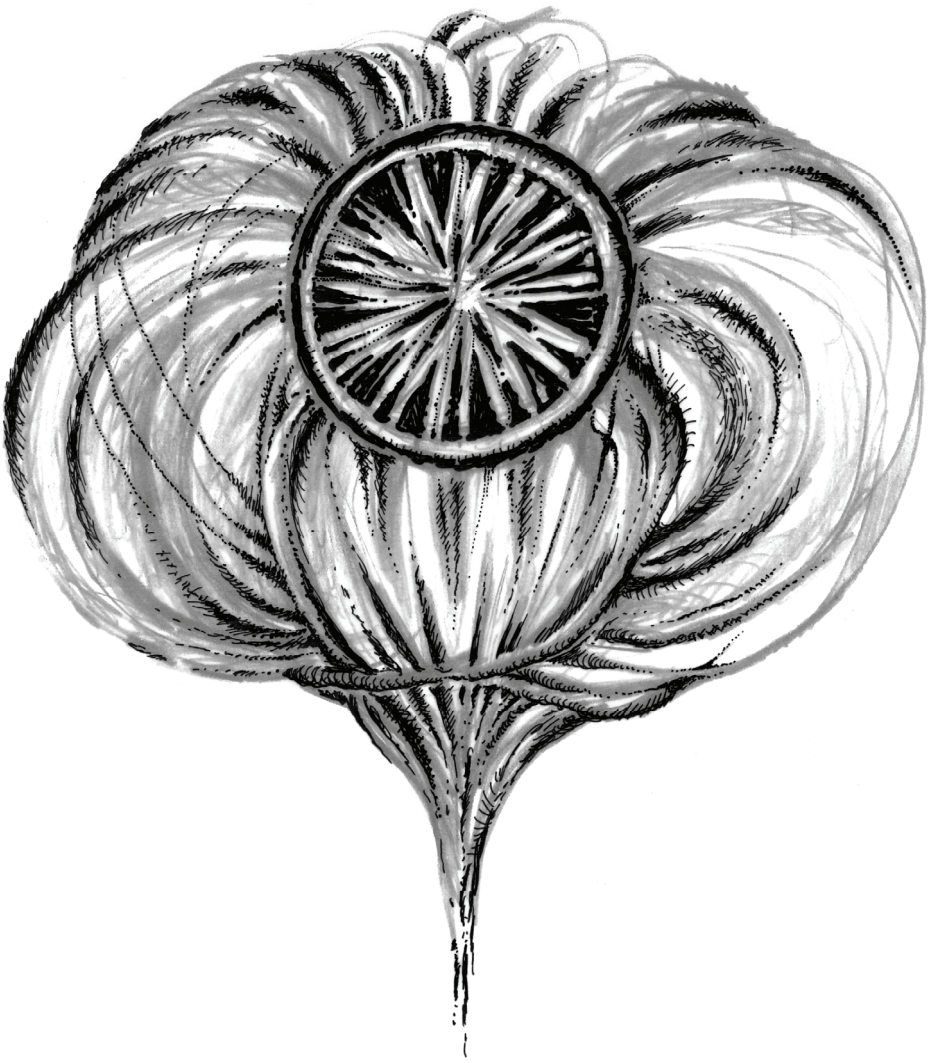
[66] Malcolm Ludvigsen: *General Relativity: A Geometric Approach* (2nd edition); Cambridge University Press 2021

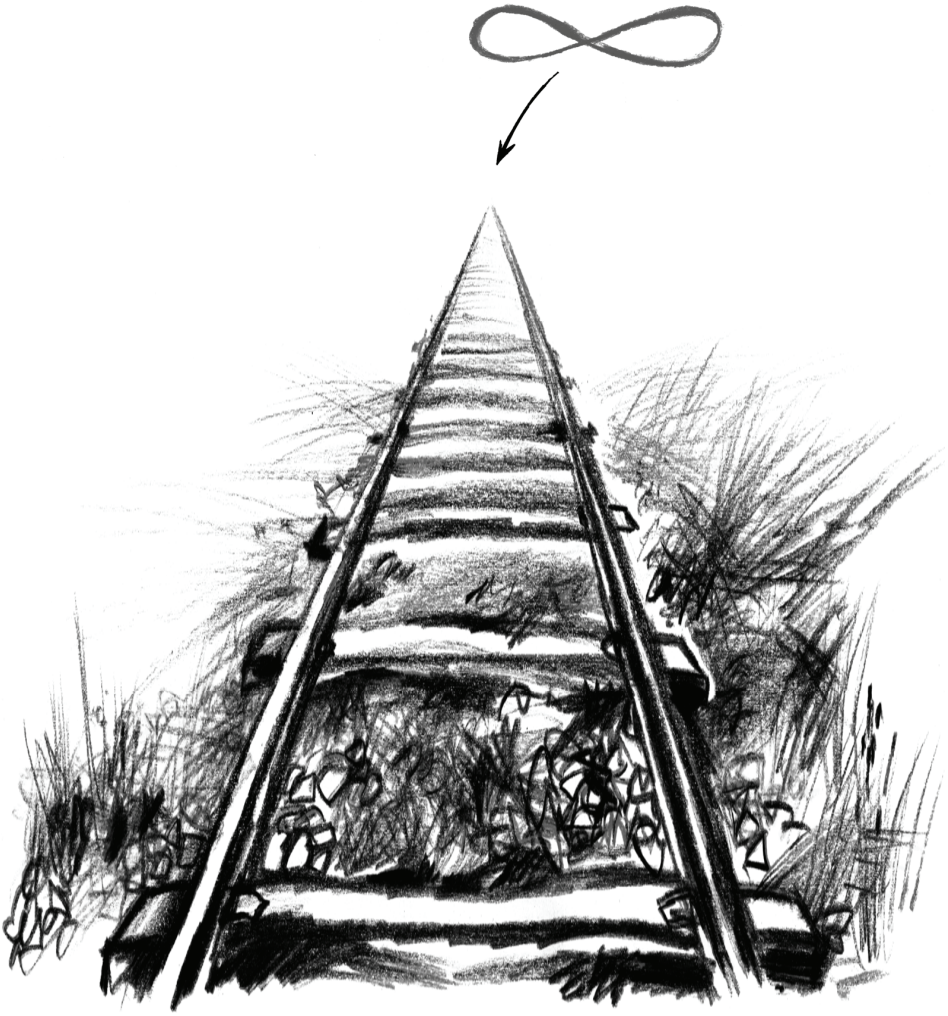
[67] Claudia de Rham: *The Beauty of Falling: A Life in Pursuit of Gravity*; Princeton University Press 2024

[68] Michael Tsamparlis: *Special Relativity: An Introduction with 200 Solved Problems*; Springer 2019

[69] Pankaj S. Joshi: *Spacetime, Geometry and Gravitation*; Springer 2019

Inside Other Volumes





Selected Topics in Theoretical Physics I

The first volume of “Selected Topics” serves as an introductory textbook for the entire series. It begins with the fundamental concepts of theoretical mechanics. The reader will become familiar with the elegance of describing mechanical and other phenomena using the Lagrangian and Hamiltonian equations, as well as Poisson brackets. The second part is dedicated to quantum theory. Relatively standard topics concerning the structure and interpretation of quantum theory are followed by problems related to the boundary between the quantum and classical worlds, the disproof of the existence of hidden variables, the EPR paradox, and Bell's inequalities. The final part of the first volume is devoted to the mathematical foundations necessary for studying physical phenomena. The reader may refer to this section of the textbook whenever they encounter mathematical tools required for reading the physics chapters.

THEORETICAL MECHANICS

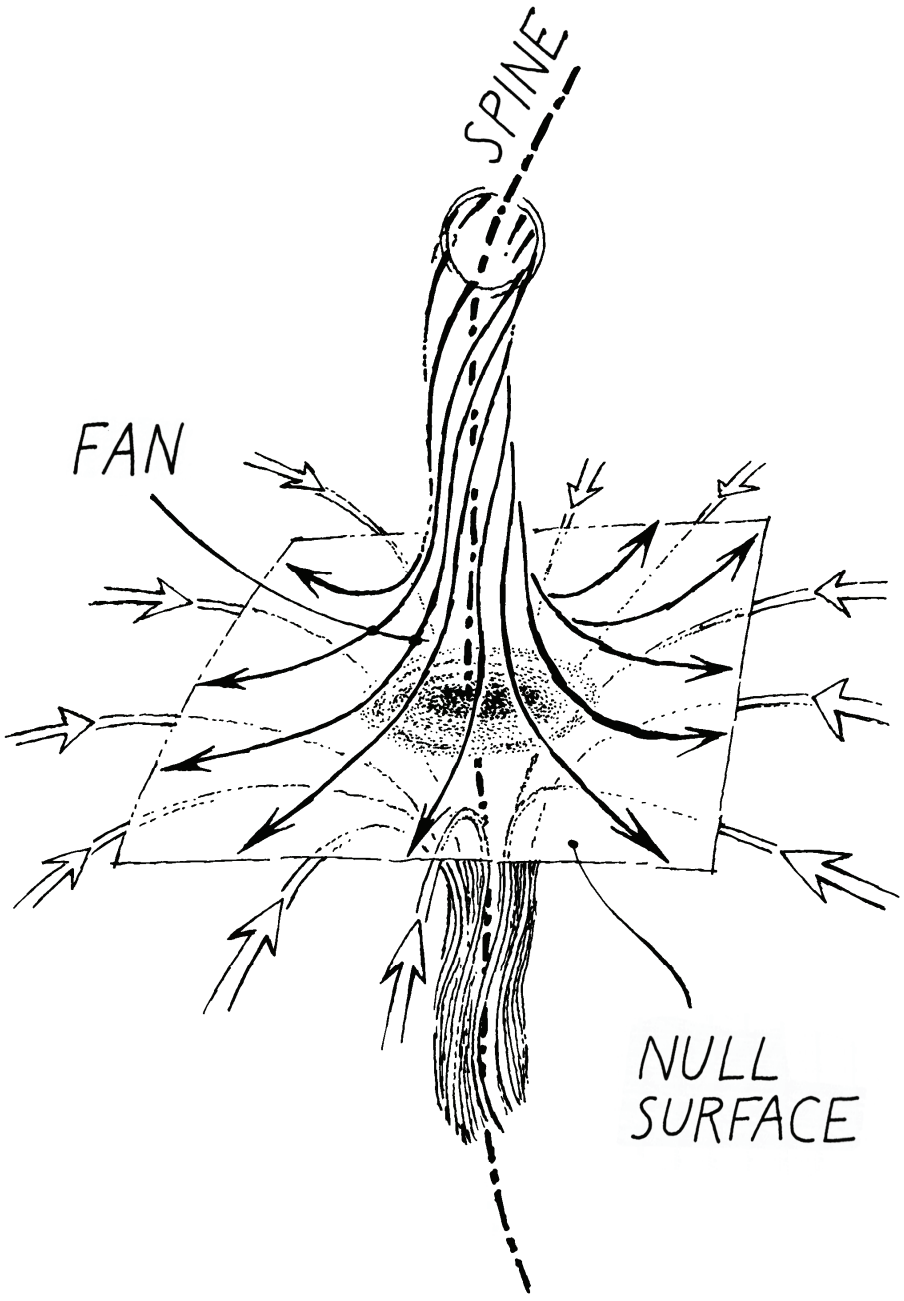
- Integral principles of mechanics
- Conservation laws in nature
- Hamilton's canonical equations
- Selected problems in theoretical mechanics
- Nonlinear dynamical systems
- Lagrangian equations for field problems

QUANTUM THEORY

- Micro-world and macro-world
- Fundamental principles of quantum theory
- Harmonic oscillator
- Simple one-dimensional systems
- Spherically symmetric potential
- Time evolution
- Relativistic quantum theory, spin
- Fermions and bosons
- Quantum theory and hidden variables

MATHEMATICS FOR PHYSICS

- Einstein summation convention
- Complex numbers and functions
- Vectors and tensors
- Dirac notation and operators in quantum theory
- From gradient to helicity
- Multidimensional integrals
- Some special functions
- Generalized functions
- Conic sections, trigonometry, and more


SPINE
FAN
NULL
SURFACE

Selected Topics in Theoretical Physics III

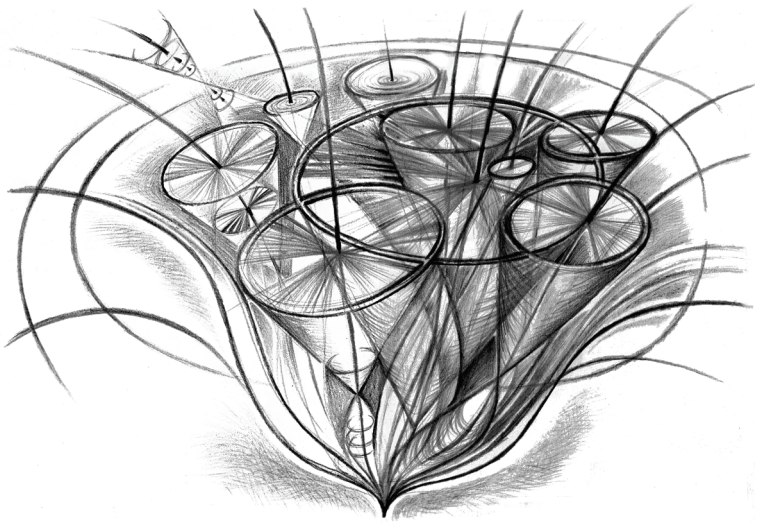
The third and final volume of "Selected Topics" is dedicated to plasma physics and numerical methods for solving physical problems. Plasma physics is a rapidly evolving scientific field that not only elucidates cosmic processes but also introduces revolutionary technologies into daily life, including novel therapeutic methods in medicine. This section expands upon the earlier textbook "Introduction to Plasma Theory," incorporating advanced non-linear topics – specifically solitons, turbulence, and the physics of runaway electrons, which pose a significant risk of damaging costly experimental equipment. The second part of this final volume introduces the fundamental principles of numerical simulations of physical processes. The rapid advancement of computing technology now enables fascinating simulations of phenomena occurring under extreme conditions that are often experimentally inaccessible.

PLASMA THEORY

- Charged Particle Motion
- Adiabatic Approximation
- Drifts
- Motion in Special Configurations
- Magnetohydrodynamics
- Helicity
- Fluid Dynamo
- Magnetic Reconnection
- Plasma Equilibrium
- Shock Waves
- Linear Waves in Plasma
- Electromagnetic Complex
- CMA Diagram
- Faraday Rotation
- Magnetoacoustic Waves
- Plasma Instabilities
- Statistical Description of Plasma
- Fokker-Planck Equation
- Nonlinear Phenomena: Turbulence, Solitons, Runaway Electrons

NUMERICAL SIMULATIONS

- Particle Motion Solvers
- Ordinary Differential Equations
- Partial Differential Equations
- Particle-in-Cell (PIC) Simulations
- Solving Polynomial Equations
- Monte Carlo Simulations
- Implementation of Probability Distributions
- Simulated Annealing
- Metropolis Method



Petr Kulhánek
Selected Topics in Theoretical Physics II
(1. Statistical Physics, 2. Elekctromagnetic Field, 3. Relativity)

No part of this publication may be published, reproduced, or distributed in any manner or by any means without the express written permission of the author and the AGA publisher.

Author: Prof. RNDr. Petr Kulhánek, Ph.D.
Graphics and Cover: Ing. arch. Ivan Havlíček
Format: 165×235 mm, 384 pages, 222 illustrations
Publisher: AGA (Aldebaran Group for Astrophysics)
Lecture Recordings: aldebaran.cz (Daniel Handl, FEE CTU)
Front Cover Image: Einstein cross Q2237+030, Gemini/GMOS
Back Cover Image: Einstein cross H1413+117, CFA/Harvard
Edition: Reprint of the 2nd edition, December 2024, Prague
Electronic Version Update: November 19, 2026
Printed by: TISK CENTRUM s.r.o.

ISBN 978-80-906638-3-1