University of Niš
Faculty of Electronic Engineering

Goran S. Ristić

PHYSICS

- Lectures -

Niš, 2026.

Contents

1 Physical Quantities and Units of Measure **1**
1.1 Physical Quantities . . . 1
1.2 International System of Units . . . 1
1.3 Definition of Base Units of Measure . . . 1
1.4 SI Decimal Units of Measure . . . 2
1.5 Dimensions of Physical Quantities . . . 2

2 Elements of Vector Calculus **4**
2.1 Scalar and Vector Quantities . . . 4
2.2 Addition and Subtraction of Vectors . . . 5
2.3 Multiplication of Vectors . . . 6
2.3.1 Scalar Product . . . 6
2.3.2 Vector Product . . . 7

3 Kinematics of a Material Point **8**
3.1 Subject of Classical Mechanics . . . 8
3.2 Rest and Movement . . . 8
3.3 Rectilinear Motion . . . 9
3.3.1 Average and Instantaneous Velocity . . . 9
3.3.2 Average and Instantaneous Acceleration . . . 10
3.3.3 Rectilinear Motion with Constant Acceleration . . . 11
3.4 Curvilinear Motion . . . 12
3.4.1 Circular Motion . . . 14
3.4.2 Projectile Motion . . . 15

4 Dynamics of a Material Point **17**
4.1 Newton's First Law of Motion . . . 17
4.2 Newton's Second Law . . . 17
4.3 Newton's Third Law . . . 18
4.4 Gravity . . . 18
4.4.1 Newton's Law of Gravitation . . . 18
4.4.2 Gravitational Field . . . 19
4.5 Body weight . . . 19
4.6 Friction Force . . . 20
4.7 Impulse . . . 21
4.8 Momentum and the Law of Conservation of Momentum . . . 21
4.9 Center of Mass . . . 22

5 Work and Energy **23**
5.1 Work . . . 23
5.2 Power . . . 24
5.3 Energy . . . 24
5.3.1 Kinetic Energy . . . 24
5.3.2 Potential Energy . . . 24
5.3.3 Law of Conservation of Energy . . . 26
5.4 Collisions . . . 27
5.4.1 Completely Elastic Collisions . . . 27
5.4.2 Completely Inelastic Collision . . . 27

6 Statics of a Rigid Body **29**
6.1 Condition of Equilibrium of a Material Point . . . 29
6.2 The Action of Forces on a Rigid Body . . . 30
6.3 Static Moment of a Force (Torque) . . . 30

7 Dynamics of Rotational Motion **31**
7.1 Moment of Inertia . . . 31
7.2 Kinetic energy, Work, and Power . . . 32
7.3 Angular Momentum and the Law of Conservation of Angular Momentum . . . 34

8 Elasticity **36**
8.1 Types of Deformation . . . 36
8.2 Stress and Relative Deformation . . . 36
8.3 Types of Elastic Deformations . . . 37
8.3.1 Stretch . . . 37
8.3.2 Shear . . . 38
8.3.3 Torsion . . . 39
8.3.4 Bending . . . 39
8.4 Internal Elastic Potential Energy of the Body . . . 39

9 Fluid Mechanics **41**
9.1 Pressure in the Fluid . . . 41
9.2 Pascal's Law . . . 41
9.3 Hydrostatic Pressure . . . 42
9.4 Buoyant Force and Archimedes' principle . . . 43
9.5 Characteristics of Flow . . . 44
9.6 Steady Flow. Flow Continuity Theorem . . . 44
9.7 Bernoulli's Principle . . . 45
9.8 Application of the Bernoulli's Equation . . . 46
9.8.1 Torricelli's Theorem . . . 46
9.9 Viscosity . . . 46

10 Heat and Temperature **48**
10.1 Temperature and Thermal Equilibrium . . . 48
10.2 Thermometers . . . 48
10.3 Temperature Scales . . . 49
10.4 Thermal Expansion of Solids and Liquids . . . 50
10.5 Thermal Stress of Materials . . . 50
10.6 Laws of Ideal Gases . . . 51
10.6.1 Boyle-Mariotte's Law . . . 51
10.6.2 Gay-Lussac's Law . . . 51
10.6.3 Charles's Law . . . 52
10.6.4 Avogadro's Laws . . . 53
10.6.5 Dalton's Laws . . . 53
10.7 Ideal Gas Law . . . 53
10.8 Quantity of Heat. Specific Heat Capacity . . . 54
10.8.1 Heat as a Form of Energy . . . 54
10.8.2 Specific Heat Capacity. Calorimetry . . . 54
10.9 Heat transfer . . . 55
10.9.1 Conduction . . . 55
10.9.2 Convection . . . 55
10.9.3 Radiation . . . 56

11 Fundamentals of Thermodynamics **57**
11.1 Basic Characteristics . . . 57
11.2 Internal Energy. Degree of Freedom . . . 57
11.3 Work of Gas During a Change in its Volume . . . 58
11.4 The First Law of Thermodynamics . . . 59
11.5 Specific Heat Capacities . . . 59
11.6 Work During a Change of Gas State . . . 60
11.7 Adiabatic process . . . 61
11.8 Cyclic Processes . . . 63
11.8.1 Reversible and Irreversible Processes . . . 63
11.8.2 Carnot Cycle Process . . . 63

11.9 The Second Law of Thermodynamics . . . 64

12 Oscillations **65**
12.1 Mechanical Oscillations . . . 65
12.1.1 Periodic Motion. Harmonic Oscillation . . . 65
12.1.2 Velocity and Acceleration in Harmonic Oscillation. The Harmonic Oscillation Equation . . . 65
12.1.3 Graphical Representation of Harmonic Oscillation . . . 66
12.1.4 Force in harmonic oscillatory motion. Simpler harmonic oscillators . . . 67
12.1.5 Energy in Harmonic Oscillatory Motion . . . 70
12.1.6 Damped Oscillations . . . 71
12.1.7 Forced Oscillations . . . 72

13 Wave Motion **73**
13.1 The Formation of Mechanical Waves . . . 73
13.2 Impulse Propagation Speed . . . 74
13.2.1 Sinusoidal Wave Motion . . . 74
13.3 Huygens' Principle . . . 75
13.4 Equation of a Progressive Wave . . . 76
13.5 Energy Transfer by Waves . . . 78
13.6 Wave Interference . . . 79
13.7 Standing Waves . . . 80
13.8 Reflection and Refraction of Waves . . . 81

14 Acoustics **83**
14.1 Fundamental Characteristics of Sound . . . 83
14.2 Sound Sources . . . 83
14.2.1 Taut Strings . . . 84
14.2.2 Rods . . . 84
14.2.3 Air Column . . . 85
14.3 Sound Intensity . . . 86
14.4 Resonance in Sound. Resonators . . . 86
14.5 Room Acoustics . . . 87
14.6 Doppler Effect . . . 87

15 Photometry **90**
15.1 Spectral Sensitivity of the Eye . . . 90
15.2 Photometric Quantities and Units . . . 91

16 Geometric Optics **93**
16.1 Laws of Geometric Optics . . . 93
16.2 Refraction of Light through a Prism . . . 93
16.3 Light Dispersion. Spectra . . . 94
16.4 Total Internal Reflection. Optical Fibers . . . 96
16.5 Spherical Lenses and Their Classification . . . 97
16.6 Thin Lens Equation . . . 98
16.7 Image Formation by Thin Lenses . . . 99
16.8 Optical Instruments . . . 100
16.8.1 Magnifying Glass . . . 100
16.8.2 Optical Microscope . . . 100

17 Wave (Physical) Optics **102**
17.1 Interference of Light . . . 102
17.1.1 Generation of Coherent Light. Interference Image . . . 102
17.2 Diffraction of Light . . . 104
17.2.1 Diffraction of Light at a Single Slit . . . 104
17.2.2 Diffraction of Light at Multiple Slits . . . 105
17.3 Polarization of Light . . . 106
17.3.1 Unpolarized (Natural) and Polarized Light . . . 106

18 Early Models of the Atom **108**
18.1 Thomson's and Rutherford's Models of the Atom . . . 108
18.2 Spectrum of Hydrogen Atoms . . . 109
18.3 Bohr's Postulates . . . 110
18.4 Franck-Hertz Experiment . . . 110

19 Fundamentals of Atomic and Molecular Physics **112**
19.1 Quantization of Angular Momentum . . . 112
19.2 Spatial Quantization . . . 112
19.3 The Stern-Gerlach Experiment . . . 114
19.4 Electron Spin . . . 114
19.5 Pauli Exclusion Principle . . . 115

20 Fundamentals of Nuclear Physics **117**
20.1 Basic Characteristics of the Atomic Nucleus . . . 117
20.2 Nuclear Binding Energy and Mass Defect . . . 118
20.3 Nuclear Forces. Nuclear Models . . . 118
20.4 Natural Radioactivity . . . 119
20.5 Radioactive Decay Law . . . 119
20.6 Activity of a Radioactive Source . . . 121
20.7 Radioactive Equilibrium . . . 121
20.8 Alpha and Beta Particles . . . 122
20.9 Gamma Radiation . . . 123
20.10 Electron Capture . . . 124
20.11 Acceleration of Charged Particles . . . 124
20.11.1 Linear Accelerators . . . 124
20.11.2 Cyclotron . . . 125
20.12 Nuclear Fission . . . 126
20.13 Nuclear Chain Reaction . . . 127
20.14 Nuclear Fusion . . . 127

Chapter 1

Physical Quantities and Units of Measure

1.1 Physical Quantities

The basis of all objective knowledge of nature is the experiment. The aim of an experiment is to determine the parameters from which the observed phenomenon depends. These parameters are called physical quantities. This introduction to the concept of physical quantities shows that they are not physical objects, states or processes, but can be understood as abstractions used to describe phenomena.

All physical quantities are divided into fundamental and derived quantities. Basic quantities are not defined in terms of expressions (equations), but rather by a description of the procedure for their measurement. They are chosen and declared by agreement at international conferences, following a certain amount of experience and knowledge about them. The term "fundamental" comes from the fact that all other physical quantities are derived from and defined in terms of them. Which quantities, and how many of them, are considered fundamental depends on the field of physics. For the field of mechanics, for example, three fundamental quantities are sufficient: length, time and mass, for electromagnetism four: length, time, mass and electric current strength, etc. The fundamental quantities are mutually independent, and are therefore also called independent quantities.

The number of fundamental quantities has changed over time. According to the decisions made to date by the General Conference of Weights and Measures (Conférence Générale des Poids et Mesures—CGPM), there are seven fundamental and independent physical quantities. These are: length, mass, time, electric current, thermodynamic temperature, luminous intensity, and amount of substance. It is considered that with these quantities, the entire field of physics and science in general can be covered.

Derived quantities are derived and defined from already known fundamental quantities using physical laws and expressions, or from already defined derived quantities. They depend on the fundamental quantities, and are also dependent on each other, so they are also called dependent quantities. Their number is not limited and is constantly changing with the development of physics and science.

1.2 International System of Units

To measure a quantity means to compare it with a pre-selected quantity of the same kind that is taken as the unit. The result of the measurement is a number showing how many times the selected unit is contained in the measured quantity.

The units of the seven basic and independent physical quantities are called **base units of measure**. These are: meter (for length), kilogram (for mass), second (for time), ampere (for electric current intensity), kelvin (for thermodynamic temperature), candela (for luminous intensity), and mole (for amount of substance). Units for derived quantities are derived from the base units and are called **derived units of measure**.

The system of units composed of these seven base units and the units derived from them constitutes the **International System of Units** (*Système International d'Unités*), abbreviated as **SI**. Supplementary SI units of measure include the radian (for plane angle) and steradian (for solid angle).

1.3 Definition of Base Units of Measure

Table 1.1 lists the base quantities and base units of measure in the SI.

- **Meter** is the length of the path traveled by light in a vacuum during a time interval of 1/299792458 of a second.
- **Kilogram** is the mass of the international prototype of the kilogram, kept at the International Bureau of Weights and Measures in Sèvres near Paris.
- **Second** is the duration of $9,192,631,770$ periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium 133 atom.

Table 1.1: Base Quantities and SI Base Units of Measure

Base Quantity (Name and Symbol)	Base Unit of Measure (Name and Symbol)
Length (l)	meter (m)
Mass (m)	kilogram (kg)
Time (t)	second (s)
Electric Current Intensity (I)	ampere (A)
Thermodynamic Temperature (T)	kelvin (K)
Luminous Intensity (I)	candela (cd)
Amount of Substance (N)	mole (mol)

- **Ampere** is the intensity of a constant electric current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed one meter apart in vacuum, produces a force between these conductors equal to $2 \cdot 10^{-7}$ newtons per meter of length.
- **Kelvin** is the thermodynamic temperature that is equal to 1/273.16 of the thermodynamic temperature of the triple point of water.
- **Candela** is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency $5.4 \cdot 10^{14}$ Hz and whose radiant intensity in that direction is 1/683 watt per steradian.
- **Mole** is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12.

1.4 SI Decimal Units of Measure

Decimal units are decimal multiples or sub-multiples of SI units, formed using internationally adopted prefixes (Table 1.2).

Table 1.2: Table 1.2: SI Decimal Units of Measure

Prefix Name	Prefix Symbol	Factor by which the unit is multiplied
atto	a	10^{-18}
femto	f	10^{-15}
pico	p	10^{-12}
nano	n	10^{-9}
micro	μ	10^{-6}
milli	m	10^{-3}
centi	c	10^{-2}
deci	d	10^{-1}
deka	da	10^{1}
hecto	h	10^{2}
kilo	k	10^{3}
mega	M	10^{6}
giga	G	10^{9}
tera	T	10^{12}
peta	P	10^{15}
exa	E	10^{18}

1.5 Dimensions of Physical Quantities

When speaking of dimensions, reference is made to the fundamental quantities and units by means of which a given quantity can be expressed. For example, the dimension of area is always the square of length and is denoted by $\left[L^2\right]$, and can be expressed in square metres, square centimetres, etc.; speed can be measured in units of m/s, cm/s, etc., but its dimension is always equal to the quotient of the dimension of length $[L]$ and the dimension of time $[T]$. As can be seen, the dimensions of physical quantities are denoted by capital letters in square brackets, for example the dimension of mass (m) with [M], length (l) with [L], time (t) with [T], etc.

Expressions that describe a physical quantity in different cases may be different, but the dimensions must remain the same. For example, although the formulas for the area of a triangle with base a and height h ($P = ah/2$) and a circle the radii r ($P{=}r^2\pi$) differ, the dimension of the area is the same in both cases and is equal to $\left[L^2\right]$.

When determining the dimensions of a derived quantity, the dimensions of the fundamental quantities are used. For example, the dimension of force as a derived quantity is equal to the product of the dimension of mass $[M]$ and the dimension of acceleration $\left[LT^{-2}\right]$, i.e. $\left[MLT^{-2}\right]$.

Proper use of dimensions can be a very effective aid in deriving an expression. This procedure is known as *dimensional analysis*. It can be used to check the correctness of an expression. In this case, two simple rules are used:

- *Only quantities of the same dimensions can be added together* (for example, mass and length cannot be added together).
- *The quantities on both sides of the equals sign must have the same dimensions.*

Let, for example, the expression for velocity be given as

$$v = v_o + \frac{1}{2}at^2,$$

where v is the velocity of the body after time t, v_o is the initial velocity and a is the acceleration. The dimensions in the above expression are

$$\left[LT^{-1}\right] = \left[LT^{-1}\right] + \left[LT^{-2}\right]\left[T^2\right] = \left[LT^{-1}\right] + [L]\,.$$

It is noted that the dimensions on the left and right-hand sides of the expression are not the same, which leads to the conclusion that an error was made in its derivation.

The matching of dimensions on either side of the equals sign does not mean that the correctness of the expression as a whole has been established. For example, the dimensionless multiplicative factor that is part of the expression may be incorrect. Therefore, a dimensional check can only indicate an expression's invalidity, but cannot be used to prove its validity.

Dimension analysis can also be used to quickly check the validity of a suspected expression. For example, is the expression for the period of oscillation of a mathematical pendulum $T = 2\pi\sqrt{l/g}$ or $T = 2\pi\sqrt{g/l}$, where l is the length of the pendulum, and g is the acceleration due to Earth's gravity? Dimensional analysis shows that the first expression,

$$[T] = \sqrt{[L]\,[L^{-1}T^2]} = \sqrt{[T^2]} = [T]\,,$$

is correct, and the second,

$$[T] \neq \sqrt{[LT^{-2}]\,[L^{-1}]} = \sqrt{[T^{-2}]} = \left[T^{-1}\right],$$

is incorrect. It should be noted that the factor 2π does not enter the final result.

The dimensions of any physical quantity in mechanics can be expressed using the formula

$$A = L^x M^y T^z,$$

where the exponents x, y and z can be whole numbers or fractions, positive or negative. If a dimension is missing from the expression, its exponent is said to be equal to zero.

Chapter 2

Elements of Vector Calculus

2.1 Scalar and Vector Quantities

Most physical quantities are either scalars or vectors. Scalar quantities are completely defined by their numerical value and the corresponding unit. Scalar quantities include: temperature, frequency, mass, density, work, energy, etc. For example, when it is stated that the duration of an event is $30\,s$, time as a scalar quantity is fully determined. Vector quantities are those that require direction and sense, in addition to a numerical value and a specific unit, to be fully defined. Vector quantities include velocity, acceleration, momentum, angular momentum, force, torque (moment of force), etc. For example, it is not sufficient to state that a force of $10\,N$ acts on a body; it is also necessary to know the direction and sense of its action. Besides scalar and vector quantities, tensor quantities also exist in physics. A tensor includes a vector as a special case.

The use of vectors in physics offers two major advantages. First, physical laws can be expressed independently of the choice of coordinate axes. Thus, vector notation serves as a "language" capable of conveying physical content without the need to introduce a coordinate system. The second advantage is that by using vector notation, many physical laws can be expressed in a simpler and more concise form compared to their representation within a specific coordinate system.

In physics, vectors are represented as directed line segments (Fig. 2.1) and are denoted by lowercase Latin letters with an arrow above ($\vec{a}$, $\vec{b}$, $\vec{c}$, etc.). Alternatively, vectors can be denoted using boldface letters (**a**, **b**, **c**, etc.), or by two uppercase letters representing the initial and terminal points of the vector (for instance, $\vec{AB}$ in Fig. 2.1). The line of action is the straight line on which the vector lies, and it defines the vector's direction. The sense of the vector is determined by the arrowhead at point B. Point A represents the initial point (origin), and point B represents the terminal point (tip) of the vector.

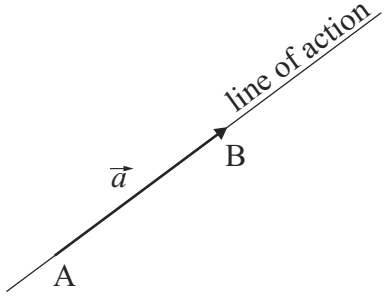

Figure 2.1: Vector notation

The magnitude of a vector is denoted by lowercase Latin letters without an arrow, for example a, or often as $|\vec{a}|$; it represents the numerical value of the length AB (Fig. 2.1).

A vector whose magnitude is equal to unity is called a unit vector. For example, the unit vector of vector $\vec{a}$ is denoted as $\hat{a}$, $ort\,\vec{a}$, or $\vec{a}_o$. It is defined as the ratio of the vector $\vec{a}$ to its magnitude:

$$\vec{a}_o = \frac{\vec{a}}{a}. \tag{2.1}$$

If $\vec{\imath}$, $\vec{\jmath}$, and $\vec{k}$ represent three mutually perpendicular (orthogonal) unit vectors that define a coordinate system (Fig. 2.2), then any vector in such a system can be expressed in the following way:

$$\vec{a} = a_x\vec{\imath} + a_y\vec{\jmath} + a_z\vec{k}, \tag{2.2}$$

where a_x, a_y, and a_z represent the components of vector $\vec{a}$ along the axes of the coordinate system. The component a_x can be expressed as

$$a_x = \vec{a} \cdot \vec{\imath},$$

since

$$\vec{a}\cdot\vec{\imath}=a_x\vec{\imath}\cdot\vec{\imath}+a_y\vec{\jmath}\cdot\vec{\imath}+a_z\vec{k}\cdot\vec{\imath}=a_x, \tag{2.3}$$

where it is assumed that

$$\vec{\jmath}\cdot\vec{\imath}=\vec{k}\cdot\vec{\imath}=0$$

and

$$\vec{\imath}\cdot\vec{\imath}=1.$$

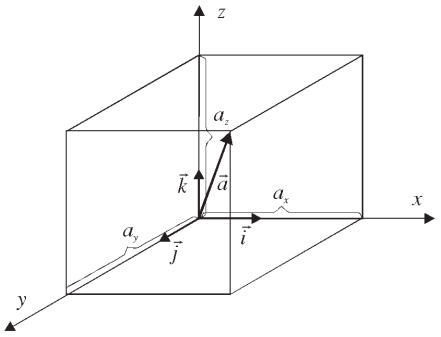

Figure 2.2: Vector components along the coordinate axes

The magnitude of vector $\vec{a}$ can be expressed in terms of its components a_x, a_y, and a_z as follows:

$$a=\sqrt{\vec{a}\cdot\vec{a}}=\sqrt{(a_x\vec{\imath}+a_y\vec{\jmath}+a_z\vec{k})\cdot(a_x\vec{\imath}+a_y\vec{\jmath}+a_z\vec{k})},$$

or

$$a=\sqrt{a_x^2+a_y^2+a_z^2}. \tag{2.4}$$

Vectors can be classified as free vectors or bound vectors. Free vectors can be moved parallel to themselves in space. Bound vectors are fixed to a specific point. A typical example is the weight vector of an object, which is bound to a point called the center of gravity.

Since vectors and scalars are different types of quantities, they cannot be added to or subtracted from one another; however, they can be multiplied and divided. Multiplying a scalar m by a vector $\vec{a}$ results in a vector $\vec{b}$, whose magnitude is m times greater than the magnitude of vector $\vec{a}$:

$$\vec{b}=m\vec{a}. \tag{2.5}$$

If $m<0$, the sense of vector $\vec{b}$ is reversed. If $m=0$, then the product of such a scalar and vector $\vec{a}$ is equal to zero (the zero vector):

$$0\vec{a}=0. \tag{2.6}$$

Dividing a vector $\vec{a}$ by a scalar m results in a vector $\vec{c}$, whose magnitude is m times smaller than the magnitude of vector $\vec{a}$:

$$\vec{c}=\frac{\vec{a}}{m}. \tag{2.7}$$

2.2 Addition and Subtraction of Vectors

The sum of two vectors $\vec{a}$ and $\vec{b}$ is defined by the geometric construction shown in Figure 2.3. This is the so-called parallelogram rule for vector addition.

By translating vector $\vec{b}$ parallel to itself until its initial point coincides with the terminal point of vector $\vec{a}$, the vector $\vec{c}$ is determined. This vector extends from the initial point of vector $\vec{a}$ to the terminal point of vector $\vec{b}$ and represents the sum of vectors $\vec{a}$ and $\vec{b}$.

When adding multiple vectors, the initial point of each vector is translated parallel to itself to the terminal point of the preceding vector. The resultant vector has its initial point at the start of the first vector and its terminal point at the end of the last vector (Fig. 2.4). The sum of a finite number of vectors is independent of the order in which they are added.

Three laws apply to vector addition:

- **commutative law**:

$$\vec{a}+\vec{b}=\vec{b}+\vec{a}, \tag{2.8}$$

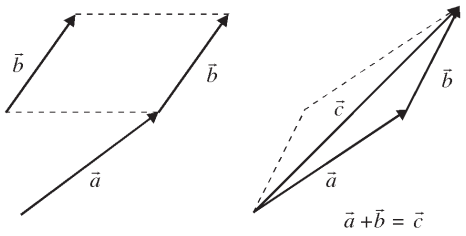

Figure 2.3: Addition of two vectors

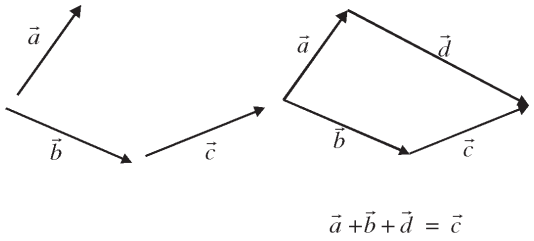

Figure 2.4: Addition of multiple vectors

- **distributive law**:

$$m(\vec{a}+\vec{b}) = m\vec{a} + m\vec{b}, \tag{2.9}$$

- **associative law**:

$$\vec{a} + (\vec{b}+\vec{c}) = (\vec{a}+\vec{b}) + \vec{c}. \tag{2.10}$$

The subtraction of vector $\vec{b}$ from vector $\vec{a}$ can be understood as a special case of addition. Specifically, the difference $\vec{e}=\vec{a}-\vec{b}$ can be written as $\vec{e}=\vec{a}+(-\vec{b})$, which means that vector $\vec{b}$ is multiplied by the scalar (-1), which only reverses its sense. The vector $(-\vec{b})$ is then added to vector $\vec{a}$ according to the rules of vector addition (Fig. 2.5).

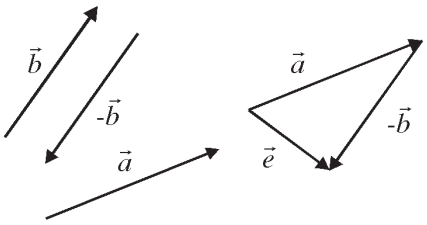

Figure 2.5: Subtraction of vectors

2.3 Multiplication of Vectors

The definition of certain complex physical quantities necessitates the definition of scalar and vector multiplication of vectors.

2.3.1 Scalar Product

Scalar multiplication of vectors $\vec{a}$ and $\vec{b}$ results in a scalar quantity c:

$$c = \vec{a} \cdot \vec{b}. \tag{2.11}$$

As seen from (2.11), the scalar product is symbolically denoted by a dot between the vectors being multiplied. The result of such multiplication is a number equal to the product of the magnitude of vector $\vec{b}$ and the projection of vector $\vec{a}$ onto the direction of vector $\vec{b}$:

$$c = b \cdot a\cos(\vec{a},\vec{b}) = ba\cos\theta. \tag{2.12}$$

The scalar product of two vectors $\vec{a}$ and $\vec{b}$ can be expressed in terms of their components as:

$$\vec{a} \cdot \vec{b} = a_x b_x + a_y b_y + a_z b_z. \tag{2.13}$$

If two vectors are mutually perpendicular ($\theta = 90°$, $\cos 90° = 0$), their scalar product is zero ($\vec{a} \cdot \vec{b} = 0$).

2.3.2 Vector Product

The vector product of vectors $\vec{a}$ and $\vec{b}$ results in a vector $\vec{c}$, which is normal to the plane formed by vectors $\vec{a}$ and $\vec{b}$ (Fig. 2.6):

$$\vec{c} = \vec{a} \times \vec{b}. \tag{2.14}$$

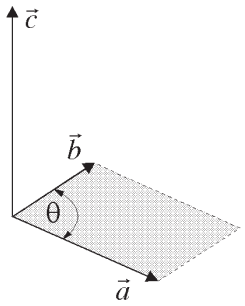

Figure 2.6: Vector product of vectors $\vec{a}$ and $\vec{b}$

The magnitude of this vector product

$$c = |\vec{a} \times \vec{b}| = ab \sin(\vec{a}, \vec{b}) = ab \sin\theta \tag{2.15}$$

is numerically equal to the area of the parallelogram formed by vectors $\vec{a}$ and $\vec{b}$. By convention, the direction (sense) of vector $\vec{c}$ is determined by the right-hand screw rule or the right-hand rule (Fig. 2.7).

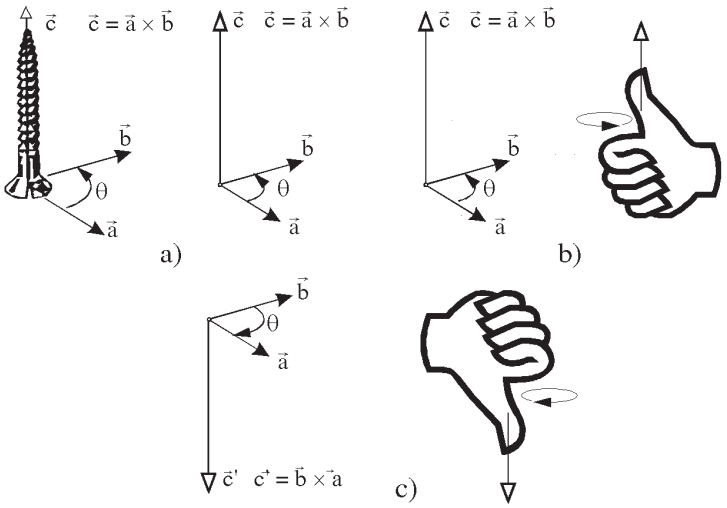

Figure 2.7: Determining the direction of vector $\vec{c}$ by the right-hand screw rule (a) and the right-hand rule (b and c)

According to the right-hand screw rule, the direction of vector $\vec{c}$ is determined by the direction of advancement of a right-handed screw when it is rotated through the angle θ (the angle between the directions of vectors $\vec{a}$ and $\vec{b}$).

The direction of vector $\vec{c}$ is determined by the right-hand rule in the following manner. First, the initial points of vectors $\vec{a}$ and $\vec{b}$ are brought to the same point, thereby defining the plane they form. Vector $\vec{c}$ is perpendicular to both vectors $\vec{a}$ and $\vec{b}$. If the curled fingers of the right hand indicate the direction of rotation from vector $\vec{a}$ toward vector $\vec{b}$ through the smaller of the two possible angles, then the thumb will point in the direction of vector $\vec{c}=\vec{a} \times \vec{b}$ (Fig. 2.7b and c). It follows from this that the vector product $\vec{b} \times \vec{a}$ has the opposite sign of the product $\vec{a} \times \vec{b}$:

$$-\vec{b} \times \vec{a} = \vec{a} \times \vec{b}. \tag{2.16}$$

Chapter 3

Kinematics of a Material Point

3.1 Subject of Classical Mechanics

Mechanics is the branch of physics that studies the laws of the motion of bodies, i.e. the change of a body's position over time. Such motions include, for example, the motion of planets, cars, projectiles, etc. Mechanics is divided into kinematics, dynamics and statics.

Kinematics studies the motion of a body without taking into account the causes that lead to that motion. It deals with quantities such as displacement, time, velocity and acceleration.

Dynamics studies the causes that lead to the motion of a body, i.e., it studies the motion of a body under the action of forces. In addition to the quantities used in kinematics, dynamics also operates with quantities such as force and mass.

Statics studies the equilibrium conditions of bodies acted upon by forces and moments of force (torques).

3.2 Rest and Movement

In mechanics, it is very important to know whether the observed body remains in the same place over time or moves relative to another body or an observer. If a body remains in the same place over time, it is said to be at rest or in a state of rest. If the body changes its position relative to another body or an observer during the observation, it is said to be in motion or in a state of motion. Therefore, to determine whether a body is moving or at rest, it is necessary to compare its position with the position of surrounding bodies that are assumed to be at rest. However, this assessment is highly unreliable. For example, a person in a moving train is at rest relative to the train, but is moving relative to the tracks. Thus, the person can be said to be at rest in one instance and in motion in another, depending on whether the motion is observed relative to the train or relative to the Earth. Furthermore, even if the train were at rest, one could say the person is at rest relative to the Earth, but it could not be claimed that they are at rest relative to the Sun, because the Earth, along with all bodies on it, is known to rotate about its axis and simultaneously revolve around the Sun. On the other hand, the Sun also moves through space, and since most likely larger or smaller parts of the universe perform similar motions, it can be concluded that in nature there is no body that is not moving and relative to which the motion of other bodies could be established in an absolute sense. Therefore, one can only speak of *relative rest* and *relative motion*.

Often, when studying motion, the dimensions of a body can be neglected, and the entire body can be replaced by a single point of finite mass equal to the actual mass of the body. Such a point is called a *material point* or *particle* and was introduced to simplify the description of the laws of motion.

It should be noted that in English literature, the term "particle" is used instead of the term "material point". However, we will use the term 'material point' throughout this text.

The position of a material point is determined relative to a body called the *reference body*. This body is in relative rest, and the coordinate system attached to it is called the *reference system* or *reference frame* for the position of the material point. The Cartesian rectangular coordinate system is most commonly used; however, for a more detailed analysis of the position and motion of a material point, it is more convenient in some cases to apply cylindrical or spherical coordinate systems.

During the motion of a material point P along a trajectory (Fig. 3.1), its position at any moment can be determined in the Cartesian coordinate system using the coordinates x, y, and z, or by the position vector (radius vector) $\vec{r}$, whose origin is at the coordinate origin O and whose endpoint is at point P.

When a material point moves, its coordinates change over time. In the general case, the motion of a material point is determined by three scalar equations:

$$x = x(t),\ y = y(t),\ z = z(t), \tag{3.1}$$

which are equivalent to the vector equation

$$\vec{r} = \vec{r}(t). \tag{3.2}$$

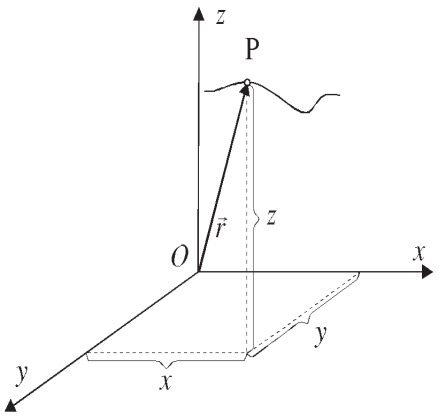

Figure 3.1: Determining the position of the material point

The magnitude of this position vector can be expressed as

$$r = \sqrt{x^2 + y^2 + z^2} \tag{3.3}$$

The number of independent coordinates that completely determine the position of a material point is called the *number of degrees of freedom*. If a material point moves in space, it possesses three degrees of freedom (coordinates x, y, and z); if it moves in a plane, it has two; and if the motion occurs along a straight line, the material point has one degree of freedom.

As can be seen from Figure 3.1, a material point describes a curve during its motion, which is called the *path* or *trajectory*. Accordingly, *the path or trajectory of a material point is understood as a sequence of successive points in space that the point occupies during its motion. The part of the trajectory that the material point traverses within a certain time interval is called the distance.*

3.3 Rectilinear Motion

3.3.1 Average and Instantaneous Velocity

In the previous section, it was emphasized that the position of a material point (or a body) in space can be determined using a rectangular coordinate system. When a material point moves in space, its projections move rectilinearly along the three axes. Therefore, its actual motion can be reconstructed from the motion of these projections. For this reason, it is of interest to consider the motion of a material point along a straight line, which in physics is known as *rectilinear motion*.

Let a material point move along the x-axis. At the moment t_1, it is at point M with coordinate x_1, and at the moment t_2, it is at point N with coordinate x_2 (Fig. 3.2). The moving of the material point in this case is defined by the vector $\vec{MN}$, whose magnitude is $x_2 - x_1 = \Delta x$, and is called *the displacement*. The average velocity $<v>$ of the material point's motion is defined as the quotient of the displacement Δx and the time interval $\Delta t = t_2 - t_1$

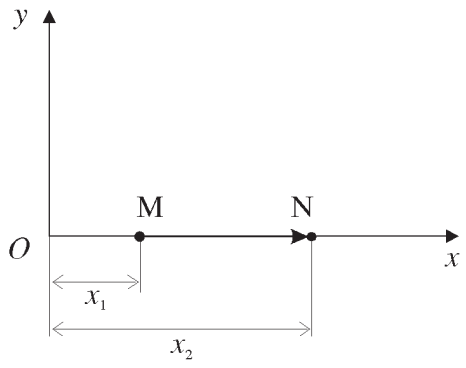

Figure 3.2: A material point moving along the x-axis

$$<v> = \frac{x_2 - x_1}{t_2 - t_1} = \frac{\Delta x}{\Delta t}. \tag{3.4}$$

The average velocity can also be determined from the dependence of the x-coordinate on time, t. Let this dependence be of the form shown in Figure 3.3. In that case, the numerical value of the average velocity can be determined from the slope of the chord M_1N_1 (from $\tan\alpha$):

$$<v>=\tan\alpha=\frac{\Delta x}{\Delta t}. \tag{3.5}$$

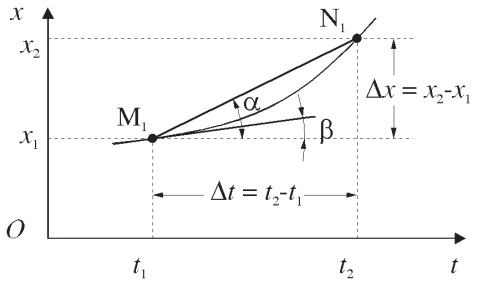

Figure 3.3: Dependence of the coordinate on time during the motion of a material point

A material point moving along a straight-line trajectory can cover equal or unequal distances in equal time intervals. When it covers equal distances, it is said to be in *uniform motion*. In that case, the ratio $\Delta x/\Delta t$ (the velocity) is constant over any time interval.

When a material point moving along a straight-line trajectory covers different distances in equal time intervals, it is said to be in *variable motion*. In this case, the velocity of the material point along the trajectory has a different value, called *the instantaneous velocity*.

For example, the instantaneous velocity of the material point shown in Fig. 3.2 is determined as follows. As has already been emphasised, the average velocity between points M and N is defined as the ratio of the displacement Δx and the time interval Δt.

Now, let point N approach point M, for which the average velocity is calculated for ever smaller displacements and ever smaller time intervals. The instantaneous velocity v at point M is defined as the limit of the average velocity as the point N approaches the point M. Therefore, instantaneous velocity represents the limit that the value of the average velocity approaches as the time interval tends to zero ($\Delta t \to 0$):

$$v=\lim_{\Delta t\to 0}\frac{\Delta x}{\Delta t}=\frac{dx}{dt}. \tag{3.6}$$

Therefore, the instantaneous velocity is equal to the differential quotient of the distance traveled and time (Newton's definition of the first derivative).

The concept of instantaneous velocity can also be understood with the help of Fig. 3.3. Namely, if point N_1 approaches point M_1, in the limit the chord will pass into the tangent at point M_1, whose slope ($\tan\beta$) gives the intensity of the instantaneous velocity. Therefore, the magnitude of the instantaneous velocity at any point on the graph of the coordinate as a function of time is equal to the slope of the tangent at that point.

3.3.2 Average and Instantaneous Acceleration

Experience shows that the velocity of a material point (or body), except in special cases, changes during its motion. The quantity that characterises this type of motion is called *acceleration*.

To derive an expression for the acceleration, consider the motion of a material point along the x-axis (Fig. 3.4), as well as the graph of the instantaneous velocity v as a function of time t (Fig. 3.5). The vector $\vec{v}_1$ (Fig. 3.4) represents the instantaneous velocity at point M, and $\vec{v}_2$ at point N.

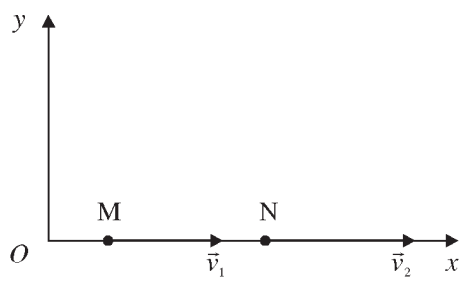

Figure 3.4: A material point moving along the x-axis

The magnitude of the average acceleration of a material point during its motion from point M to N is defined as the ratio of the change in velocity to the corresponding time:

$$<a>= \frac{v_2 - v_1}{t_2 - t_1} = \frac{\Delta v}{\Delta t}, \tag{3.7}$$

where t_1 and t_2 are the times corresponding to the velocities v_1 and v_2. The average acceleration is given by the slope of the chord M_1N_1 ($\tan\alpha$) (Fig. 3.5).

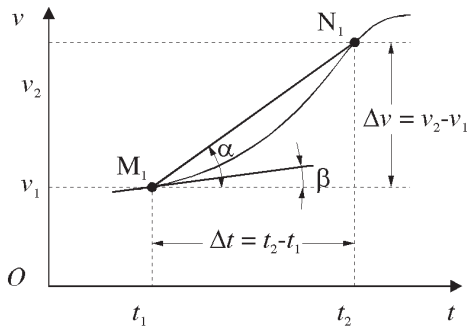

Figure 3.5: Dependence of the velocity on time during the motion of a material point

The instantaneous acceleration a of a material point, which represents the acceleration at a particular moment or at a particular point on its trajectory, is defined in a similar way to instantaneous velocity. Let point N approach point M (Fig. 3.4) while calculating the average acceleration for progressively shorter time intervals. The instantaneous acceleration a at point M will represent the limit that the average acceleration approaches as the time interval tends to zero:

$$a = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \frac{dv}{dt} = \frac{d^2x}{dt^2}. \tag{3.8}$$

From the graph of the instantaneous velocity as a function of time (Fig. 3.5) the scalar value (magnitude) of the instantaneous acceleration at the point M_1 is determined by the slope of the tangent ($\tan\beta$) at that point. Therefore, the instantaneous acceleration at any point of the velocity-versus-time graph is equal to the slope of the tangent to the graph at that point.

3.3.3 Rectilinear Motion with Constant Acceleration

The simplest form of accelerated motion is an rectilinear motion with a constant acceleration $a(t)=const.$ In this case, the velocity changes by equal amounts at equal time interval:

$$v = v_o + at. \tag{3.9}$$

where v_o is the initial velocity, a is the acceleration and t is the time (see Fig. 3.6).

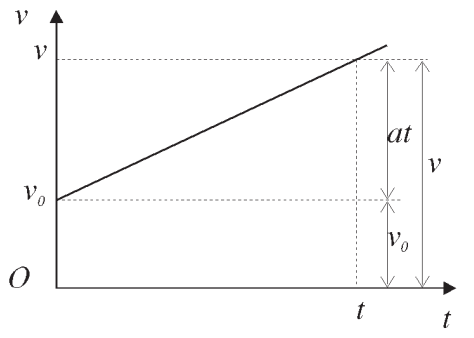

Figure 3.6: Velocity-time dependence for motion with constant acceleration

The special case of motion with constant acceleration is the case when it is equal to zero. In this case, the velocity is constant ($v=const.$), so the equation of motion takes the form:

$$x = vt. \tag{3.10}$$

3.4 Curvilinear Motion

So far, we have considered motion along a straight line — an rectilinear motion. Here we will consider the motion of a body along a curved trajectory that lies in a fixed plane, for example, a ball moving through the air, a shell fired from a gun, a body rotating around a fixed axis, the Moon or a satellite around the Earth, planets around the Sun, and so on.

Let a material point move in a plane, describing the trajectory shown in Figure 3.7.

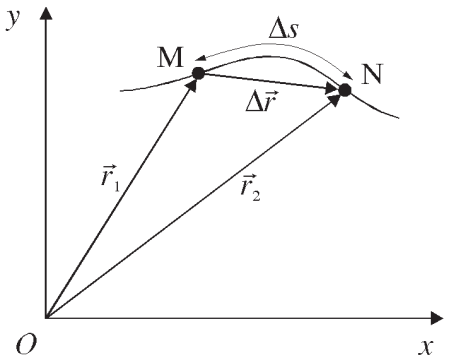

Figure 3.7: The trajectory of the material point in the xy plane

Let a material point be at point M at time t_1, and at point N at time t_2. The motion of a material point can be described using the coordinates x and y, or with the position vector. The vector $\vec{r}_1$ represents the position vector at time t_1, and $\vec{r}_2$ at time t_2. The displacement of the material point from M to N, over the time interval Δt, is given by the vector difference (so-called the displacement vector $\Delta\vec{r}$):

$$\Delta\vec{r} = \vec{r}_2 - \vec{r}_1, \tag{3.11}$$

and represents a displacement over a time interval $\Delta t = t_2 - t_1$. The vectors $\vec{r}_1$ and $\vec{r}_2$ can be expressed in terms of unit vectors, i.e.

$$\vec{r}_1 = x_1\vec{\imath} + y_1\vec{\jmath}, \quad \vec{r}_2 = x_2\vec{\imath} + y_2\vec{\jmath}, \tag{3.12}$$

where x_1, y_1 and x_2, y_2 are the coordinates of the points M and N, respectively.

Substituting (3.12) into (3.11) for the displaced vector, one obtains

$$\Delta\vec{r} = (x_2 - x_1)\vec{\imath} + (y_2 - y_1)\vec{\jmath}. \tag{3.13}$$

If the motion in space is observed, then $\Delta\vec{r}$ can be expressed in the following way:

$$\Delta\vec{r} = (x_2 - x_1)\vec{\imath} + (y_2 - y_1)\vec{\jmath} + (z_2 - z_1)\vec{k},$$

where z_1 and z_2 represent the positions of the material points along the z-axis at times t_1 and t_2, respectively. If the motion is along the x-axis, then $y_2 - y_1 = 0$ and $z_2 - z_1 = 0$, so the value of the displacement is $\Delta r = x_2 - x_1$, which coincides with the displacement in rectilinear motion.

The average velocity vector $<\vec{v}>$ for the time interval Δt is defined as

$$<\vec{v}> = \frac{\Delta\vec{r}}{\Delta t}. \tag{3.14}$$

Since the vector $<\vec{v}>$ represents the product of the vector $\Delta\vec{r}$ and the scalar $1/\Delta t$, its direction is the same as the direction of the displacement vector $\Delta\vec{r}$, and its magnitude is $\Delta r/\Delta t$.

Consider the case when $\Delta t \to 0$, so that the distance between M and N also tends to zero. Then the vector of the instantaneous velocity represents the limit of the average velocity:

$$\vec{v} = \lim_{\Delta t \to 0} \frac{\Delta\vec{r}}{\Delta t} = \frac{d\vec{r}}{dt}. \tag{3.15}$$

At any point on the trajectory, the velocity has the direction of the tangent, as shown in Figure 3.8. It can be noted that the average velocity, determined by the displacement vector $\Delta\vec{r}$, is not equal to the ratio of the distance travelled Δs and the time interval Δt (Fig. 3.7). In some cases, the directions of these speeds coincide (for example in the case of rectilinear motion), but in the general case they are different. However, in the limit, when $\Delta t \to 0 \Rightarrow \Delta r \to \Delta s$, so that the instantaneous velocities expressed in terms of Δr and Δs at any given moment coincide.

The instantaneous velocity of a material point in a plane can be expressed in terms of unit vectors

$$\vec{v} = \frac{d\vec{r}}{dt} = \frac{dx}{dt}\vec{\imath} + \frac{dy}{dt}\vec{\jmath} = v_x\vec{\imath} + v_y\vec{\jmath}.$$

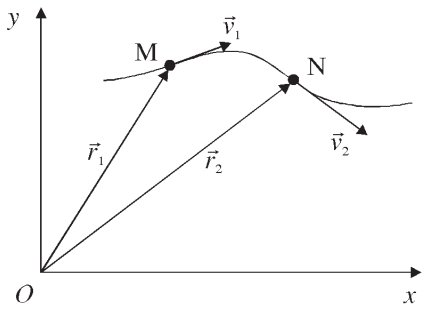

Figure 3.8: The velocity vectors $\vec{v}_1$ and $\vec{v}_2$ at times t_1 and t_2, respectively

The intensity of the velocity vector is

$$v = \sqrt{v_x^2 + v_y^2}, \tag{3.16}$$

where $v_x = dx/dt$ and $v_y = dy/dt$ represent the projections of the velocity onto the x and y axes, respectively.

The vector of the average acceleration, $<\vec{a}>$ during the time interval $\Delta t = t_2 - t_1$ is given by the expression

$$<\vec{a}> = \frac{\Delta\vec{v}}{\Delta t} = \frac{\vec{v_2} - \vec{v_1}}{t_2 - t_1}, \tag{3.17}$$

where $\Delta\vec{v}$ is the velocity change vector for the time interval Δt. It is observed that in many cases (for example, in the one shown in Figure 3.8) the vector $\vec{v}_2$ does not coincide in direction with $\vec{v}_1$. Therefore, $<\vec{a}>$ can have a direction that does not coincide with the vectors $\vec{v}_1$ and $\vec{v}_2$. Moreover, even when the velocities $\vec{v}_1$ and $\vec{v}_2$ are equal in magnitude, and differ in direction, their difference is not zero. From this, it follows that acceleration can arise both from a change in the magnitude of the velocity and from a change in its direction (or from a change in both).

The vector of the instantaneous acceleration is equal to the vector of the average acceleration when the time interval tends to zero

$$\vec{a} = \lim_{\Delta t \to 0} \frac{\Delta\vec{v}}{\Delta t} = \frac{d\vec{v}}{dt}. \tag{3.18}$$

The instantaneous acceleration can also be expressed in terms of unit vectors as

$$\vec{a} = \frac{dv_x}{dt}\vec{\imath} + \frac{dv_y}{dt}\vec{\jmath} = a_x\vec{\imath} + a_y\vec{\jmath},$$

while its intensity

$$a = \sqrt{a_x^2 + a_y^2}. \tag{3.19}$$

The total acceleration $\vec{a}$ is equal to the sum of the tangential $\vec{a}_\tau$ and normal $\vec{a}_n$ accelerations (Fig. 3.9):

$$\vec{a} = \vec{a}_\tau + \vec{a}_n. \tag{3.20}$$

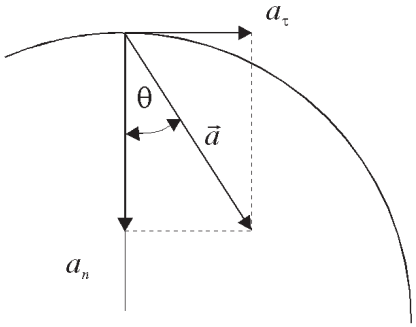

Figure 3.9: The tangential acceleration $\vec{a}_\tau$, the normal acceleration $\vec{a}_n$ and the total acceleration

The tangential acceleration is equal to the first derivative of the velocity, and its vector has the direction of the tangent and the sense of the material point's motion:

$$a_\tau = \frac{dv_\tau}{dt}. \tag{3.21}$$

The vector of the normal acceleration is in the direction of the radius and directed towards the centre of the circle. It is equal to:

$$a_n = \frac{v^2}{r}. \tag{3.22}$$

The value of the total acceleration at a given point on the trajectory can be determined from the values of the tangential and normal acceleration:

$$a=\sqrt{a_\tau^2+a_n^2}=\sqrt{\left(\frac{dv}{dt}\right)^2+\left(\frac{v^2}{r}\right)^2}. \tag{3.23}$$

3.4.1 Circular Motion

A special case of curvilinear motion is the circular motion (Fig. 3.10). If a material point travels a distance Δs over an interval of time Δt, then the ratio of the angular displacement $\Delta\theta$ to the time interval Δt is *average angular velocity* $<\omega>$:

$$<\omega>=\frac{\Delta\theta}{\Delta t}. \tag{3.24}$$

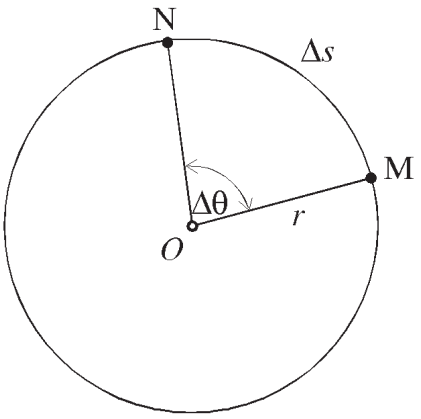

Figure 3.10: Circular motion

In the limit, as $\Delta t\to 0$, the angular velocity at a given moment (*instantaneous angular velocity*) ω is obtained as:

$$\omega=\lim_{\Delta t\to 0}\frac{\Delta\theta}{\Delta t}=\frac{d\theta}{dt}. \tag{3.25}$$

If ω_1 and ω_2 are the angular velocities at the times t_1 and t_2, then the ratio

$$<\alpha>=\frac{\omega_2-\omega_1}{t_2-t_1}=\frac{\Delta\omega}{\Delta t} \tag{3.26}$$

represents the average angular acceleration, and its limit

$$\alpha=\lim_{\Delta t\to 0}\frac{\Delta\omega}{\Delta t}=\frac{d\omega}{dt} \tag{3.27}$$

is the angular acceleration at a given moment (*instantaneous angular acceleration*).

The relationship between linear and angular quantities that characterize the motion of a material point around a circle can be expressed as follows. If s denotes the arc described by a material point moving along a circle of radius r, during which it has made an angular displacement θ, then

$$s=r\theta;\quad v=\frac{ds}{dt}=r\frac{d\theta}{dt}=r\omega. \tag{3.28}$$

Tangential and normal acceleration can be expressed as

$$a_\tau=\frac{dv}{dt}=r\frac{d\omega}{dt}=r\alpha, \tag{3.29}$$

$$a_n=\frac{v^2}{r}=r\omega^2. \tag{3.30}$$

For a constant angular acceleration ($\alpha=const.$), the expressions for the angular velocity and angular displacement have the form

$$\omega=\omega_o+\alpha t;\quad \theta=\omega_o t+\frac{1}{2}\alpha t^2, \tag{3.31}$$

where ω_o is the angular velocity at the initial moment $t=0$.

The angular velocity of rotation is represented by the vector $\vec{\omega}$, whose direction is along the axis of rotation, and its sense is determined by the right-hand rule.

3.4.2 Projectile Motion

Motion in which a body is given an initial velocity and then left to move under the influence of the gravitational force is called projectile motion. Examples of such motion include a thrown ball, a projectile fired from a cannon, a bomb dropped from an aircraft, a rocket after its fuel is exhausted, etc.

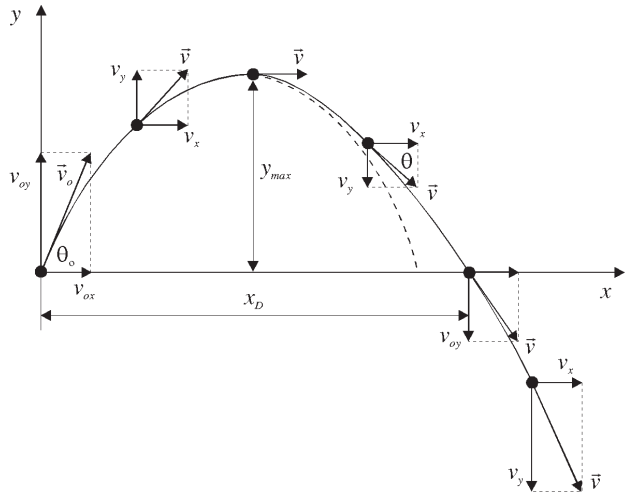

Figure 3.11: The trajectory of a body launched at an angle θ_o to the horizon with an initial velocity $\vec{v}_o$

Figure 3.11 shows the trajectory of the projectile, which has the shape of a parabola. The origin of the coordinate system is placed at the point where the body begins its free flight, and this moment is taken as the start of the observation, i.e., $t = 0$. The velocity at the origin is represented by the vector $\vec{v}_o$, which is called the *initial velocity vector*. The angle θ_o at which the body is launched is called the *angle of elevation*. The initial velocity v_o can be decomposed into a horizontal component $v_{ox} = v_o \cos\theta_o$ and a vertical component $v_{oy} = v_o \sin\theta_o$. Since the horizontal velocity component is constant, at any given moment it will have the value

$$v_x = v_{ox} = v_o \cos\theta_o. \tag{3.32}$$

Motion along the y-axis corresponds to uniformly decelerated motion; therefore, the vertical velocity component as a function of time is given by the expression:

$$v_y = v_{oy} - gt = v_o sin\theta_o - gt. \tag{3.33}$$

The resultant velocity of the projectile's motion is obtained by the vector addition of these components, and its magnitude is

$$v = \sqrt{v_x^2 + v_y^2}. \tag{3.34}$$

The coordinates of the projectile as a function of time are given by the following expressions:

$$x = v_o t \cos\theta_o, \tag{3.35}$$

$$y = v_o t \sin\theta_o - \frac{1}{2} g t^2. \tag{3.36}$$

The equation of the projectile's trajectory is obtained from expressions (3.35) and (3.36) by eliminating t:

$$y = x \tan\theta_o - \frac{g x^2}{2 v_o^2 \cos^2\theta_o}. \tag{3.37}$$

The quantities v_o, $\tan\theta_o$, $\cos\theta_o$, and g are constants, so the above equation takes the form

$$y = ax - bx^2 \tag{3.38}$$

and represents the equation of a parabola.

The most important parameters of a projectile's motion are the range, the angle of elevation, and the maximum height. The range of the projectile x_D is found from expression (3.37), for $x = x_D$ and $y = 0$:

$$x_D = \frac{v_o^2}{g} 2 \sin\theta_o \cos\theta_o = \frac{v_o^2}{g} \sin 2\theta_o. \tag{3.39}$$

For every value of initial velocity, there is a specific angle of elevation for which the projectile achieves its maximum range. This angle is obtained by differentiating expression (3.39) with respect to the angle and setting the resulting derivative to zero:

$$\frac{dx_D}{d\theta_o} = \frac{2v_o^2}{g}\cos 2\theta_o = 0 \tag{3.40}$$

From the above expression, it can be concluded that the maximum range is achieved when $\cos 2\theta_o = 0$, that is, for an angle of elevation of 45°. For any other range value, there are two complementary angles θ_o at which that range can be achieved.

The maximum height of the projectile $y = y_{max}$ is determined from the condition that at the highest point of the trajectory, $v_y = 0$. From expression (3.33), the time t required for the body to reach the highest point of its trajectory is

$$t = \frac{v_o \sin\theta_o}{g}. \tag{3.41}$$

By substituting expression (3.41) into (3.36), the maximum height is found to be

$$y_{max} = \frac{v_o^2 \sin^2\theta_o}{2g}. \tag{3.42}$$

Chapter 4

Dynamics of a Material Point

In kinematics, the motion of a body (material point) is considered on the basis of the concepts of velocity and acceleration without taking into account the causes that lead to this movement. A branch of mechanics that studies the relationship between the motion of a body and its the cause is called **dynamics**. The v velocities of the body, which are significantly lower than the speed of light c ($c = 3 \cdot 10^8 m/s$), will be considered here, which ignores relativistic effects.

From experience, it is known that a body cannot change its state of motion or rest, nor can it deform on its own, without the influence of other bodies. A physical quantity used as a measure of the mutual interaction of bodies is called *force*. Therefore, the definition of force is: *Force is the cause of a change in the state of motion of a body or its deformation.*

Every force acts with a certain magnitude and has a specific direction and sense. This means that force is a vector quantity. The point where the force is directly connected to a body or where the effect of the force is transmitted is called the point of application of the force.

Since the action of a force can lead to deformation or acceleration of a body, both of these phenomena can be used to measure it. In practice, it is considerably simpler to measure the deformation than the acceleration of the body.

4.1 Newton's First Law of Motion

Newton's First Law of Motion or the Law of Inertia is: *Every body tends to remain in a state of rest or uniform rectilinear motion, until the influence of external forces does not change its state* or *Every body remains in a state of rest or uniform rectilinear motion as long as no external forces act on it* or *A body left to itself does not change its velocity.*

The mathematical expression of this law is:

$$m\vec{v} = const., \qquad \vec{v} = const. \tag{4.1}$$

or

$$\vec{a} = 0. \tag{4.2}$$

Conditions (4.1) and (4.2) are satisfied only when no external forces act on the body, that is, when the vector sum of the forces acting is zero. That is why Newton's First Law can also be expressed mathematically in the following way:

$$\sum_{i=1}^{n} \vec{F}_i = 0. \tag{4.3}$$

4.2 Newton's Second Law

Newton's second law states that the acceleration of an object (body) is directly proportional to the net force acting on it and inversely proportional to its mass. This is commonly expressed by the equation $\vec{F} = m \cdot \vec{a}$, where $\vec{F}$ is the net force, m is the mass, and $\vec{a}$ is the acceleration. In essence, a greater force will cause greater acceleration, while a greater mass will result in less acceleration for the same force.

The figure 4.1 shows the relationship between the force acting on the body (on the material point) and the tangential and normal acceleration. In the figure, the force acting on the body is decomposed into two components: tangential $\vec{F}_\tau$, whose direction is along the tangent to the trajectory, and normal component $\vec{F}_n$, which has the direction as the radius, and the sense towards the center of the curve. Knowing that $a_\tau = dv/dt$ and $a_n = v^2/r$, as well as $v = r\omega$, it gives:

$$F_\tau = ma_\tau = m\frac{dv}{dt} \tag{4.4}$$

and

$$F_n = ma_n = \frac{mv^2}{r} = m\omega^2 r. \tag{4.5}$$

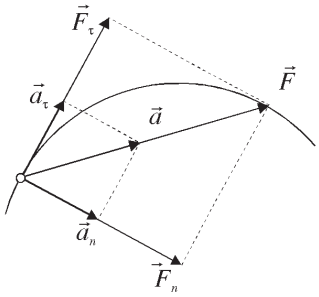

Figure 4.1: The action of force F on a body

4.3 Newton's Third Law

Newton's third law states that for every action, there is an equal and opposite reaction. This means that when two objects interact, they exert equal and opposite forces on each other simultaneously. For example, if object A pushes object B, object B pushes back on object A with the same force in the opposite direction.

Figure 4.2: The action of force F on a body

The figure 4.3 shows a proof of this law. Let two bodies connected by a dynamometer be on a smooth surface (a dynamometer is an instrument for measuring force). One of these two bodies is attached to a fixed support, and the other has another dynamometer attached to it. If the free dynamometer is pulled with the intention of detaching the body from the wall, it will be noticed that the pull on both dynamometers is equal, i.e., the dynamometer placed between the body will show the same force as the one being pulled. From this example, it can be concluded that the dynamometer connecting the bodies indicates the resisting force of one body against the action of the other. Thus, a pair of forces of equal magnitude but opposite direction occurs.

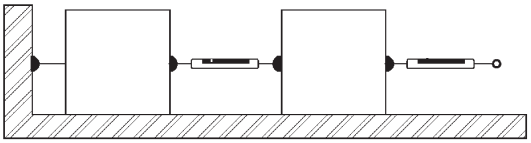

Figure 4.3: Bodies connected by dynamometers

Newton's third law, which shows the interaction between two bodies, can be expressed as: *Action is always opposite and equal to reaction; Mutual interactions of two bodies are always equal and directed in opposite directions.*

Mathematically, Newton's Third Law can be represented in the form:

$$\vec{F}_1 = -\vec{F}_2. \tag{4.6}$$

4.4 Gravity

The interaction between bodies caused by the existence of their masses is called *gravity*. The quantitative measure of gravity is the gravitational force, which is attractive in nature. Due to gravity, planets move in elliptical orbits around the Sun. The gravitational effect of the Earth influences the motion of objects (free fall, projectile, etc.). Finally, the weight of a body is also a result of the existence of Earth's gravity.

4.4.1 Newton's Law of Gravitation

All bodies in nature attract each other. The law according to which this attraction occurs was established by Newton in 1686. It states: *The force with which two material points attract each other is proportional to the product of*

their masses and inversely proportional to the square of the distance between them:

$$F = \gamma \frac{m_1 m_2}{r^2}, \tag{4.7}$$

where γ is the coefficient of proportionality and is called the gravitational constant.

The expression (4.7) can also be written in vector form. Namely, the force $\vec{F}_{12}$ with which a material point of mass m_2 acts on a material point of mass m_1 (Fig. 4.4) can be expressed in the following way:

$$\vec{F}_{12} = \gamma \frac{m_1 m_2}{r^2} \vec{r}_{o12}, \tag{4.8}$$

where $\vec{r}_{o12}$ is a unit vector that has the sense from m_1 to m_2.

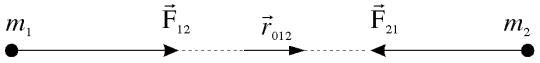

Figure 4.4: Gravitational force between two material points

According to Newton's third law, the force $\vec{F}_{21}$ with which a material point of mass m_1 acts on a material point of mass m_2 is of the same magnitude but opposite in direction to the force $\vec{F}_{12}$:

$$\vec{F}_{21} = -\vec{F}_{12} = -\gamma \frac{m_1 m_2}{r^2} \vec{r}_{o12}. \tag{4.9}$$

4.4.2 Gravitational Field

A field of forces, in a general sense, is a part of space in which the effect of these forces is felt. When the gravitational force exerted by one body on another is observed, it is said that the first body generates a gravitational field around itself, and this field exerts a force on the second body. However, the second body also generates a gravitational field around itself that affects the first body.

A quantitative measure used to describe a gravitational field is the gravitational field strength $\vec{G}$, which is determined by the ratio of the force to the mass on which the force acts:

$$\vec{G} = \frac{\vec{F}}{m}. \tag{4.10}$$

The mass m is called *a test mass* and is much smaller than the mass of the body whose field is being observed. This condition greatly simplifies the problem, because the field of the test mass can be neglected.

The unit of gravitational field is Nkg^{-1}.

As a specific example, the Earth's gravitational field will be considered. The expression for the strength of the Earth's gravitational field at a distance r from its center can be obtained if expression (4.7) for the gravitational force is substituted into expression (4.10):

$$G = \gamma \frac{M}{r^2}, \tag{4.11}$$

where M is the mass of the Earth. From expression (4.11), it can be seen that the Earth's gravitational field strength decreases with the square of the distance from its center.

The gravitational force according to Newton's second law is equal to $\vec{F} = m\vec{g}$, so from expression (4.10) it follows that the strength of the gravitational field is:

$$\vec{G} = \frac{m\vec{g}}{m} = \vec{g}. \tag{4.12}$$

Thus, the strength of the gravitational field is equal in intensity and direction to the acceleration due to Earth's gravity. For example, at a point where $g = 9,81\, ms^{-2}$, the strength of the gravitational field is $G = 9,81\, Nkg^{-1}$, and its direction is toward the center of the Earth.

4.5 Body weight

All objects that fall from a relatively small height from the Earth's surface have, if air resistance is neglected, the constant acceleration of g. This acceleration is known as *the gravitational acceleration* and depends on the height relative to the Earth's surface (see eqs. (4.11) and (4.12)). The force that causes this acceleration is called gravitational force, body weight, or gravity:

$$\vec{Q} = m\vec{g}. \tag{4.13}$$

The vectors $\vec{Q}$ and $\vec{g}$ have the same direction and sense, and they are oriented towards the center of the Earth.

4.6 Friction Force

Let there be a body on a flat surface on which the force $\vec{F}$ acts (traction force; Fig. 4.5). If this traction force is of low intensity, the body will be at rest, which means that the traction force is balanced with the frictional force. As the traction force increases, the frictional force also increases. The body will start moving when the traction force is greater than the frictional force. The maximum friction force that keeps an object at rest is called *the static frictional force.*

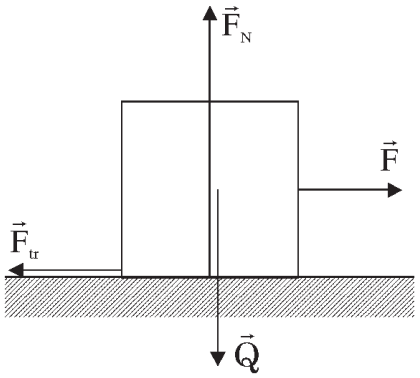

Figure 4.5: Friction force

The friction force that occurs when a body is moving is called *dynamic friction force.* The dependence of the friction force on the traction force is shown in the Fig. 4.6.

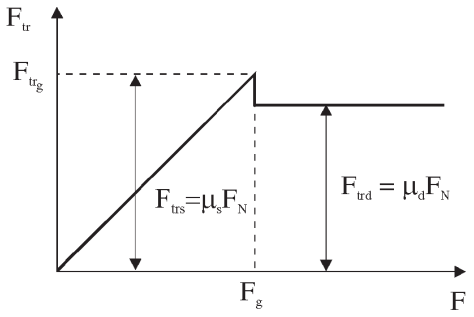

Figure 4.6: The dependence of the friction force on the traction force

It has been experimentally shown that the static friction force F_{trs} can be expressed as

$$F_{trs} = \mu_s F_N, \tag{4.14}$$

the force of dynamic friction F_{trd} as

$$F_{trd} = \mu_d F_N. \tag{4.15}$$

μ_s and μ_d represent the coefficients of static and dynamic friction, respectively. The static coefficient of friction is greater than the dynamic one. Both coefficients depend on the **type of materials, cleanliness, and roughness of the contact surfaces**.

The coefficient of friction can be determined using an inclined plane (Fig. 4.7). If a body is on a steep plane whose slope can be changed, the weight of the body can be decomposed into a component F along the steep plane and a normal component F_N. The angle increases until the body begins to slide. At that moment (the limiting angle θ_{max}), the component of the weight F is equal to the static friction force, i.e. $F = F_{trs}$. It can be seen in the figure 4.7.

$$F = Qsin\theta_{max} = F_{trs}; \quad F_N = Qcos\theta_{max}. \tag{4.16}$$

By combining equations (4.14) and (4.16) the coefficient of static friction is obtained:

$$\mu_s = tan\theta_{max}. \tag{4.17}$$

Accordingly, the coefficient of static friction is equal to the tangent of the angle θ_{max} at which a body begins to slide on an inclined plane.

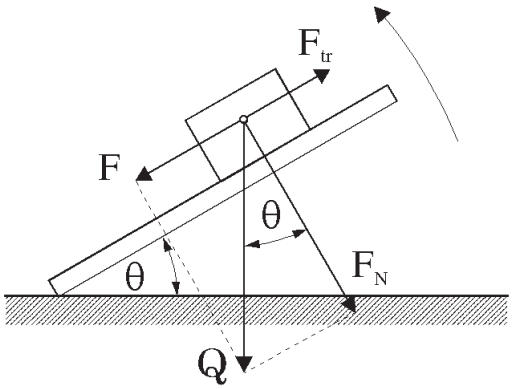

Figure 4.7: The dependence of the friction force on the traction force

4.7 Impulse

Impulse of force (impulse) is a physical quantity that quantifies the overall effect of a force acting on an object (body) over a period of time. Impulse $d\vec{p}$ is defined as the product of the force $\vec{F}$ applied to an object and the time interval dt over which the force acts:

$$d\vec{p} = \vec{F}dt. \tag{4.18}$$

It is a vector quantity, meaning it has both magnitude and direction, which is the same as the direction of the net force.

For a force that varies over time, impulse is calculated using the integral of the force with respect to time:

$$\vec{p} = \int_{t_1}^{t_2} \vec{F}dt. \tag{4.19}$$

When the force is constant, this expression takes the form

$$\vec{p} = \vec{F}(t_2 - t_1) = \vec{F}\Delta t, \tag{4.20}$$

where is $\Delta t = t_2 - t_1$.

4.8 Momentum and the Law of Conservation of Momentum

To describe the motion of a body of mass m, a vector quantity $\vec{K}$, called *momentum*, is used and is given by the expression:

$$\vec{K} = m\vec{v} = m\frac{d\vec{r}}{dt}, \tag{4.21}$$

where $\vec{v}$ is the velocity of the body. Momentum is a vector quantity of the same direction and sense as velocity whose intensity is m times greater (lower) than the intensity (magnitude) of velocity.

A body system is understood as a set of two or more bodies. Depending on the forces acting on them, systems can be closed (isolated) or open. In closed systems, only internal forces occur, i.e., the forces of mutual interaction of the bodies that make up that system. In nature, there is no ideal closed system, but in some cases, external forces (impulses of forces) are significantly smaller than internal ones, so they can be neglected. Therefore, to a certain degree of accuracy, such a system is considered closed. In an open system, in addition to internal forces, external forces also act on the bodies that make up that system.

Let two bodies (or material points) form a closed system. The mass of one body is m_1 and its velocity is $\vec{v}_1$, while the other body has a mass m_2 and a velocity $\vec{v}_2$. These two bodies interact with forces $\vec{F}_1$ and $\vec{F}_2$, of the same magnitude and direction, but opposite in sense. According to Newton's Third Law

$$\vec{F}_1 = -\vec{F}_2 \;\Rightarrow\; \vec{F}_1 + \vec{F}_2 = 0. \tag{4.22}$$

According to Newton's Second Law

$$\vec{F}_1 = m_1\vec{a}_1; \;\; \vec{F}_2 = m_2\vec{a}_2. \tag{4.23}$$

and by summing the forces

$$\vec{F}_1 + \vec{F}_2 = m_1\vec{a}_1 + m_2\vec{a}_2 = m_1\frac{d\vec{v}_1}{dt} + m_2\frac{d\vec{v}_2}{dt} = \frac{d(m_1\vec{v}_1)}{dt} + \frac{d(m_2\vec{v}_2)}{dt} = \frac{d}{dt}(m_1\vec{v}_1 + m_2\vec{v}_2) = (m_1\vec{v}_1 + m_2\vec{v}_2)'. \tag{4.24}$$

From the expression (4.22) and (4.24) follows

$$(m_1\vec{v}_1 + m_2\vec{v}_2)' = 0. \tag{4.25}$$

Since the first derivative of this expression is equal to zero, it means that the expression is constant (the first derivative of a constant is zero):

$$m_1\vec{v}_1 + m_2\vec{v}_2 = \sum_{i=1}^{2} m_i\vec{v}_i = const. \tag{4.26}$$

For a closed system of n bodies, the following applies

$$\sum_{i=1}^{n} m_i\vec{v}_i = const. \tag{4.27}$$

Expressions (4.26) and (4.27) represent the mathematical interpretation of the Law of Conservation of Momentum for a closed system. It can be expressed as follows: *The vector sum of the momenta of the individual bodies that make up a closed system remains constant during motion.*

4.9 Center of Mass

The center of mass of a system of bodies represents an imaginary point where the entire mass of the system is concentrated. In this way, the problem of the motion of a large number of bodies is reduced to the problem of the motion of a single material point. For a system consisting of n bodies, the masses m_1, m_2, m_3, ..., m_n (Fig. 4.8), the position vector of the center of mass can be defined:

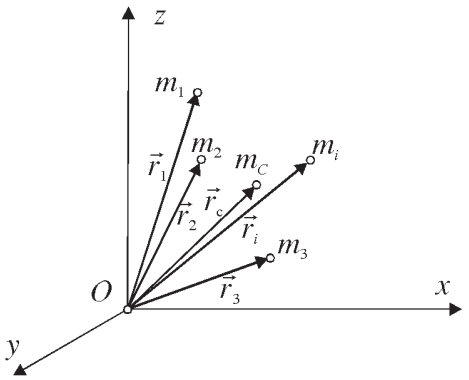

Figure 4.8: The center of mass of a system of bodies

$$\vec{r}_c = \frac{m_1\vec{r}_1 + m_2\vec{r}_2 + m_3\vec{r}_3 + \cdots + m_n\vec{r}_n}{m_1 + m_2 + m_3 + \cdots + m_n} \tag{4.28}$$

i.e.,

$$\vec{r}_c = \frac{\sum_{i=1}^{n} m_i\vec{r}_i}{\sum_{i=1}^{n} m_i} = \frac{1}{m}\sum_{i=1}^{n} m_i\vec{r}_i. \tag{4.29}$$

This expression for finding the center of mass of a system of bodies can be applied to finding the center of mass of a rigid body divided into a large number of small parts (elementary masses).

The velocity of the centar of mass is

$$\vec{v}_c = \frac{d\vec{r}_c}{dt} = \frac{1}{m}\sum_{i=1}^{n} m_i\frac{d\vec{r}_i}{dt} = \frac{1}{m}\sum_{i=1}^{n} m_i\vec{v}_i, \tag{4.30}$$

and momentum:

$$\vec{K}_c = m\vec{v}_c = \sum_{i=1}^{n} m_i\vec{v}_i. \tag{4.31}$$

Chapter 5

Work and Energy

5.1 Work

In everyday life, the word "work" is applied to activities that require any form of physical or mental exertion. In physics, the use of this term is restricted. Namely, it is well known that bodies can interact with each other through various forces (friction, gravity, elastic forces, electromagnetic forces, etc.), leading to their displacement. To characterize these phenomena, a quantity is used that is greater when the component of the force along the direction of displacement is larger, and when its point of application is moved over a greater distance. In physics, this quantity is known as *work*.

Let a force act on the body as shown in the Fig. 5.1. The most general case occurs when the force acting on a body changes in both direction and magnitude, while the motion takes place along a curvilinear path. To determine the work done by such a force along the path from s_1 to s_2 (Fig. 5.2), it is necessary to divide the path into elements ds, over which the force can be considered constant and the displacement rectilinear. In this case, the elementary work is

$$dA = \vec{F} \cdot d\vec{s} = F\, ds \cos\theta, \tag{5.1}$$

and the total work along the path from s_1 to s_2 is determined by the integral

$$A = \int_{s_1}^{s_2} \vec{F} \cdot d\vec{s} = \int_{s_1}^{s_2} F ds \cos\theta, \tag{5.2}$$

where $d\vec{s}$ is the displacement vector along the path (trajectory), and $\vec{F} \cdot d\vec{s}$ is the scalar product of the vectors $\vec{F}$ and $d\vec{s}$.

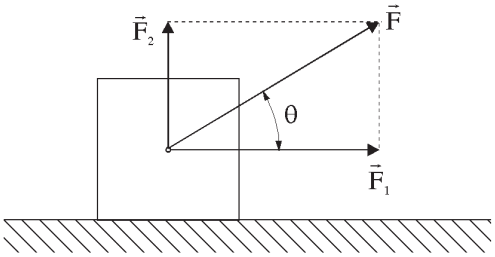

Figure 5.1: The action of force $\vec{F}$ on a body

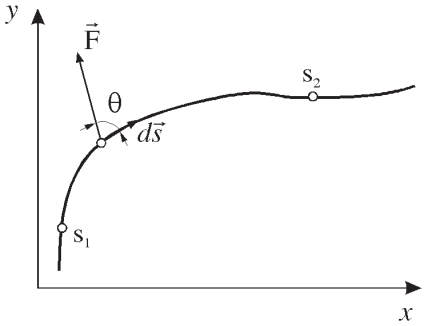

Figure 5.2: The case of a force with variable magnitude and direction

5.2 Power

In practice, besides knowing the work done by a force, it is often necessary to know the time during which that work was performed. Namely, from a technical perspective, a force is more useful if it performs a certain amount of work in a shorter time. To characterize the rate at which work is done, a physical quantity called *power* is used. Power is equal to the ratio of the work ΔA to the time interval Δt during which that work is performed:

$$P = \frac{\Delta A}{\Delta t}. \tag{5.3}$$

This expression holds for the case where the ratio $\Delta A/\Delta t$ is constant, that is, when power is a constant quantity. When power varies with time, the expression gives the average power within the interval Δt.

For the case of variable power (where different amounts of work are performed in equal time intervals), the concept of instantaneous power is introduced:

$$P = \lim_{\Delta t \to 0} \frac{\Delta A}{\Delta t} = \frac{dA}{dt}. \tag{5.4}$$

The unit of power is the Watt (W). $1\,W$ represents $1\,J$ u $1\,s$ ($W = J/s$).

5.3 Energy

The change in the energy of a body or a system of bodies ΔE is equal to the work A done by external forces acting on the body (system):

$$\Delta E = E_2 - E_1 = A. \tag{5.5}$$

This equality is called the *work-energy theorem.*

When a body (system) performs work, the work is positive, and when work is performed on the body (system), the work is negative.

There are several forms of energy, such as mechanical, thermal, electrical, chemical, nuclear, etc. Here, only mechanical energy will be discussed, which is divided into kinetic and potential energy.

5.3.1 Kinetic Energy

The kinetic energy of a body (material point) represents a measure of its mechanical motion and determines the work that must be performed to bring the body into motion. We will consider rectilinear motion.

Let the body have a velocity v_1 at the moment of observation when a force $\vec{F}$ begins to act in the direction of the body's motion. Under the action of the force F, its velocity increased to v_2, and the energy increased by the amount of work performed. The work done by the force F along the path from s_1, where the velocity is v_1, to s_2, where the velocity is v_2, is given by:

$$A = \int_{s_1}^{s_2} F\,ds. \tag{5.6}$$

As a result $F = m\,a = m\,dv/dt$, it follows

$$A = \int_{v_1}^{v_2} m\frac{dv}{dt}ds = \int_{v_1}^{v_2} mvdv = \frac{mv_2^2}{2} - \frac{mv_1^2}{2}. \tag{5.7}$$

where $mv_1^2/2$ is the kinetic energy of a body when it moves with velocity v_1, and $mv_2^2/2$ its kinetic energy when it moves with velocity v_2. Based on this, the kinetic energy in the general case is

$$E_k = \frac{mv^2}{2}. \tag{5.8}$$

In the case of a system of bodies (material points), whose masses are $m_1, m_2, m_3, \ldots, m_n$ and velocities are $v_1, v_2, v_3, \ldots, v_n$, the total kinetic energy of the system is

$$E_k = \sum_{i=1}^{n} \frac{m_i v_i^2}{2}. \tag{5.9}$$

5.3.2 Potential Energy

The body possesses energy not only when moving, but also when it is in a certain position in space. Such energy is called *potential energy.* A common example of potential energy is *gravitational potential energy*, E_p (the word "gravitational" is usually omitted).

A body of a certain mass raised to some height has gravitational potential energy due to its position, as it has the potential to do work when falling from that height.

Lifting a body of mass m vertically upwards is possible only if a force at least equal to its weight mg is applied (this holds for small heights, where $g = const.$). To raise the body to a height h from the Earth's surface (Fig. 5.3), work mgh must be done. If the body is allowed to fall from that height, it will do work mgh.

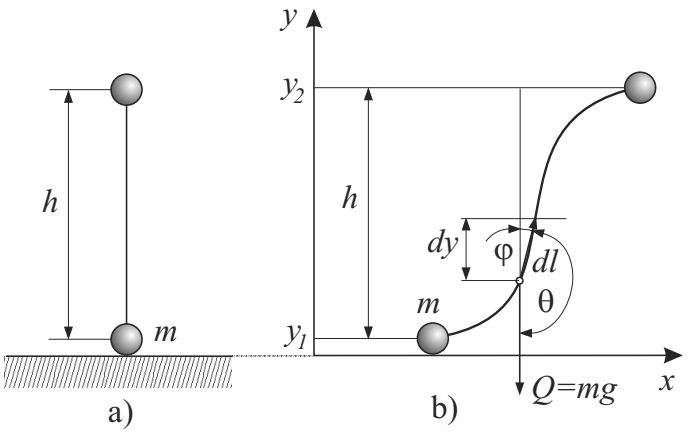

Figure 5.3: A body of mass m is lifted to a height h in two ways

Let a body move along an arbitrary path in the xy plane (Fig. 5.3b). It starts moving at a point with ordinate (height) y_1 and reaches height y_2, where $y_2 - y_1 = h$. The work done by the force when it moves by dl is

$$A = \int_{y_1}^{y_2} \vec{Q}\, d\vec{l} = \int_{y_1}^{y_2} mg\, dl\, \cos\theta. \tag{5.10}$$

From the figure, it is seen that $\theta = 180° - \varphi$, $\cos\theta = -\cos\varphi$, and $dl = dy/\cos\varphi$, giving

$$A = -\int_{y_1}^{y_2} mg\, dy = -mg(y_2 - y_1), \tag{5.11}$$

From expression (5.11), it can be seen that the work of a force depends on the height, but not on the shape of the trajectory. Such forces are called *conservative forces*. A force is conservative if the work done in moving a body between two points is independent of the trajectory taken. Conservative force are gravitational force, elastic spring force, electrostatic force, and so on.

In the case shown in Fig. 5.11b, $y_2 > y_1$, so the work of the weight force is negative because the body, i.e., the gravitational force of the body, does not perform work, but work is done by an external force acting against it. If $y_2 < y_1$ (the body falls downward), the case is the opposite (the gravitational force on the body does work), and the work is positive. If a body falls freely from a height $h = y_2 - y_1$, it can perform work on another body or system equal to $mgh = mg(y_2 - y_1)$.

In accordance with expressions (5.11) and theorem (5.5), the potential energy at a height h relative to the reference level is given by the expression

$$E_p = mgh. \tag{5.12}$$

Here, another form of potential energy will be considered, which is characteristic of elastic bodies and has many practical applications (the so-called *elastic potential energy*). The spring shown in the Fig. 5.4, whether compressed or stretched, possesses potential energy which, when released, can do work on the ball.

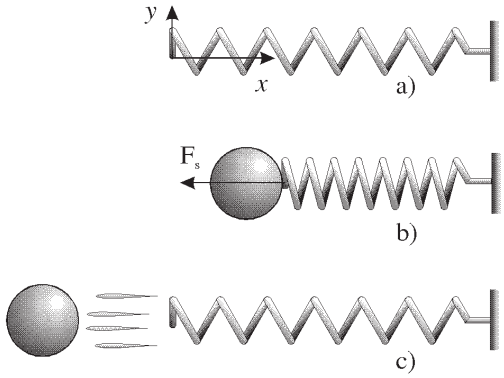

Figure 5.4: A spring can store elastic potential energy when it is stretched (a) or compressed (b); this energy can be used to perform work at the moment it is released (c)

Since the elastic force of the spring F_s is equal to

$$F_s = -kx, \tag{5.13}$$

an external force F of the same magnitude and direction, but opposite in sense ($F = kx$), must be applied to compress (or stretch) the spring.

Elastic potential energy is equal to the work done by the force F in compressing (stretching) the spring from the equilibrium position ($x = 0$) to a distance x:

$$E_p = A = \int_0^x F dx = \int_0^x kx dx = \frac{kx^2}{2}. \tag{5.14}$$

5.3.3 Law of Conservation of Energy

The law of conservation of energy originated from a series of experiments. To derive the law of conservation of mechanical energy, a system of bodies (material points) will be considered, whose masses are m_1, m_2, m_3,..., m_n, and velocities $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$,..., $\vec{v}_n$. Let internal forces $\vec{F}_1$, $\vec{F}_2$, $\vec{F}_3$,..., $\vec{F}_n$ and external forces $\vec{F}_{s1}$, $\vec{F}_{s2}$, $\vec{F}_{s3}$, ..., $\vec{F}_{sn}$ act on these bodies. From Newton's Second Law, it follows

$$\begin{aligned} m_1 \frac{d\vec{v}_1}{dt} &= \vec{F}_1 + \vec{F}_{s1}, \\ m_2 \frac{d\vec{v}_2}{dt} &= \vec{F}_2 + \vec{F}_{s2}, \\ m_3 \frac{d\vec{v}_3}{dt} &= \vec{F}_3 + \vec{F}_{s3}, \\ &\cdot \quad \cdot \quad \cdot \\ m_n \frac{d\vec{v}_n}{dt} &= \vec{F}_n + \vec{F}_{sn}. \end{aligned} \tag{5.15}$$

Let the bodies move for the same time interval dt by distances $d\vec{x}_1$, $d\vec{x}$, $d\vec{x}_n$. By multiplying each of these equations by the corresponding displacement $d\vec{x}_i$ and taking $d\vec{x}_i = \vec{v}_i dt$, we get

$$\begin{aligned} m_1(\vec{v}_1 d\vec{v}_1) - (\vec{F}_1 + \vec{F}_{s1})d\vec{x}_1 = 0, \\ m_2(\vec{v}_2 d\vec{v}_2) - (\vec{F}_2 + \vec{F}_{s2})d\vec{x}_2 = 0, \\ m_3(\vec{v}_3 d\vec{v}_3) - (\vec{F}_3 + \vec{F}_{s3})d\vec{x}_3 = 0, \\ \cdot \quad \cdot \quad \cdot \\ m_n(\vec{v}_n d\vec{v}_n) - (\vec{F}_n + \vec{F}_{sn})d\vec{x}_n = 0. \end{aligned} \tag{5.16}$$

Assuming that the system is closed:

$$\vec{F}_{s1} + \vec{F}_{s2} + \cdots + \vec{F}_{sn} = 0,$$

summing these equations gives:

$$\sum_{i=1}^{n} m_i \vec{v}_i d\vec{v}_i - \sum_{i=1}^{n} \vec{F}_i d\vec{x}_i = 0. \tag{5.17}$$

Term

$$\sum_{i=1}^{n} m_i \vec{v}_i d\vec{v}_i = \sum_{i=1}^{n} d(m_i v_i^2/2) = dE_k \tag{5.18}$$

represents an infinitesimal change in kinetic energy, and term $-\sum_{i=1}^{n} \vec{F}_i d\vec{x}_i = dE_p$ an infinitesimal change in the potential energy of the entire system. Therefore, expression (5.17) can be written in the form

$$dE_k + dEp = 0 \;\Rightarrow\; d(E_k + Ep) = 0. \tag{5.19}$$

Based on this, the total mechanical energy of the closed system is

$$E_k + E_p = const. \tag{5.20}$$

Equation (5.20) represents *the Law of Conservation of Mechanical Energy.*

The law of conservation of mechanical energy states: In a closed system of bodies between which only conservative forces act, mechanical energy is conserved, i.e., it does not change over time.

Regarding the law of conservation of mechanical energy, there is a theorem about the minimum of potential energy, which states: *The stable state of a closed system is the one in which the potential energy of the system is at a minimum.*

One of the simplest examples of the law of conservation of mechanical energy, which will be considered here, is the free fall of a body from rest ("the body is allowed to fall freely") under conditions where air resistance can be neglected. Let a body of mass m be at a height h above the Earth (Fig. 5.5). Since the body is at rest at that height, it possesses only potential energy, i.e., its total energy E is equal to the potential energy mgh. When the body is allowed to fall, its potential energy decreases while its kinetic energy increases during the motion. However, the total energy at any point along the trajectory remains the same.

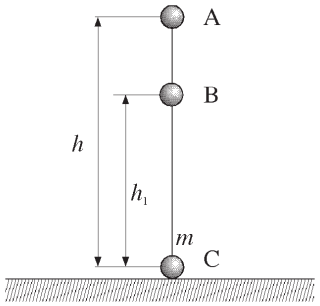

Figure 5.5: Example of the law of conservation of mechanical energy

5.4 Collisions

One example of the application of the law of conservation of momentum and energy is the collision of two bodies.

A collision can be *completely (perfectly) elastic*, *completely (perfectly) inelastic*, or *partially elastic*. A completely elastic collision is a collision of bodies in which their kinetic energy before and after the collision remains unchanged. This means that during a collision, the bodies can exchange part of their kinetic energy with each other, but the total kinetic energy of the system remains constant.

In the case of a completely inelastic collision, the bodies stick together after the collision and move together as one, or remain at rest. In this case, the total kinetic energy of the system is not conserved, because a part of it is converted into other forms of energy, such as heat or energy of deformation.

In all three cases, the law of conservation of momentum and the law of conservation of total mechanical energy apply.

5.4.1 Completely Elastic Collisions

Here, a completely elastic collision of two spheres (balls), whose centers of mass move along the same line, will be considered (Fig. 5.6). Such a collision is called a *completely elastic central collision*. The balls of mass m_1 before the collision has a velocity v_1, and the sphere of mass m_2 has a velocity v_2. The velocities of the balls after the collision are v_1 and v_2.

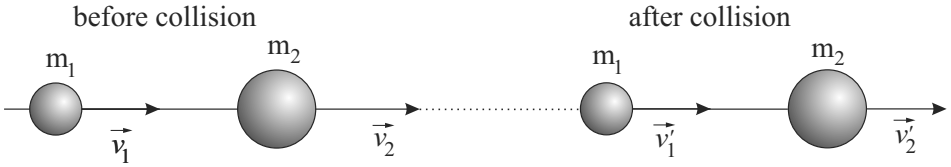

Figure 5.6: Completely elastic central collision of two balls

Equations that determine the motion of the balls before and after the collision are

- the law of conservation of momentum:

$$m_1 v_1 + m_2 v_2 = m_1 v_1' + m_2 v_2' \tag{5.21}$$

- the law of conservation of kinetic energy:

$$\frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 = \frac{1}{2} m_1 v_1'^2 + \frac{1}{2} m_2 v_2'^2. \tag{5.22}$$

5.4.2 Completely Inelastic Collision

A completely inelastic collision can be demonstrated with lead balls moving in the same direction (Fig. 5.7). After the collision, the balls continue to move together (they are stuck together).

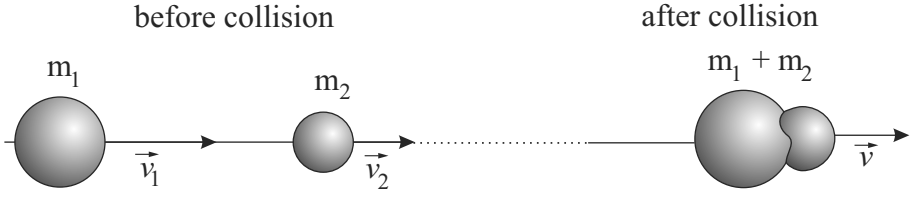

Figure 5.7: Completely inelastic central collision of two balls

If the masses of the balls are m_1 and m_2, and their velocity before collision v_1 and v_2, then the Law of Conservation of Momentum can be written in the form

$$m_1 v_1 + m_2 v_2 = (m_1 + m_2)v. \tag{5.23}$$

As emphasized, in a completely inelastic collision there is a loss of some kinetic energy due to its conversion into heat or other forms of energy such as the energy of deformation (the law of conservation of kinetic energy does not apply here). This loss can be determined from the difference in the kinetic energies of the bodies before and after the collision:

$$\Delta E_k = \left(\frac{m_1 v_1^2}{2} + \frac{m_2 v_2^2}{2}\right) - \frac{(m_1 + m_2)v^2}{2}. \tag{5.24}$$

Chapter 6

Statics of a Rigid Body

When a sufficiently strong force acts on a body, it accelerates or deforms. If deformation is neglected, when several forces act on a body simultaneously, it can move with acceleration, move uniformly, or remain at rest. If a body under the influence of forces moves in such a way that it accelerates, the forces are unbalanced. However, if a body moves uniformly or is at rest under the influence of forces, the forces are balanced. The branch of mechanics that studies the conditions of equilibrium of bodies under the action of forces is called statics.

Here, the action of forces on a material point and the conditions of its equilibrium will be considered first, followed by the action of forces on a rigid body. A rigid body is a body that does not deform under the action of forces. Otherwise, every real body, depending on the intensity of the forces, deforms more or less. However, when forces are of low intensity, solid bodies with high resistance experience only slight deformations. Such bodies can be considered as rigid bodies with a certain approximation.

6.1 Condition of Equilibrium of a Material Point

The simplest case of equilibrium is when two forces of the same magnitude and direction but opposite senses act on a material point.

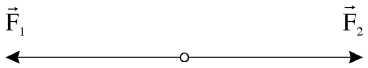

Figure 6.1: Equilibrium of a material point under the action of two forces

If multiple forces act on a material point, their resultant is equal to their vector sum, so the condition of equilibrium is:

$$\vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \cdots + \vec{F}_n = \sum_{i=1}^{n} \vec{F}_i = \vec{F}_R = 0. \tag{6.1}$$

For example, when three forces act on a material point, the forces $\vec{F}_1$ and $\vec{F}_2$ combine into a resultant $\vec{F}_R$, which must have the same magnitude and direction as the third force $\vec{F}_3$, but in the opposite sense (Fig. 6.2). A similar approach can be applied when a larger number of forces act on a material point.

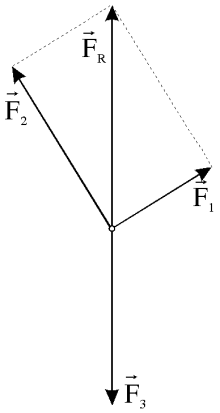

Figure 6.2: Equilibrium of a material point under the action of three forces

6.2 The Action of Forces on a Rigid Body

When two forces of the same line of action act on a rigid body at different points of application (Fig. 6.3), the condition for equilibrium is that the forces have the same magnitude and opposite sense (in that case, the vector sum of the forces is zero). However, if two forces of equal magnitude act on a body and their directions of action do not coincide (Fig. 6.4), the body will not move translationally, but will rotate..

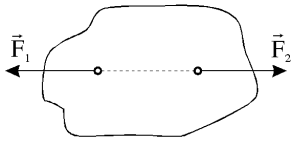

Figure 6.3: The action of forces whose lines coincide

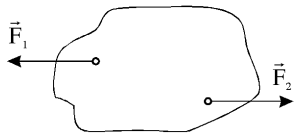

Figure 6.4: The action of forces whose directions do not coincide

6.3 Static Moment of a Force (Torque)

The quantity that characterizes the effect of a force on body rotation is a vector quantity called *the static moment of the force or torque* $\vec{M}$.

In order to derive the expression for the static moment of a force, the effect of the force on a body that can rotate around point O (Fig. 6.5) will be considered. When the direction of the force coincides with point O (Fig. 6.5a), the body will not perform rotational motion. If the line of action of force $\vec{F}$ does not pass through point O, the body will rotate.

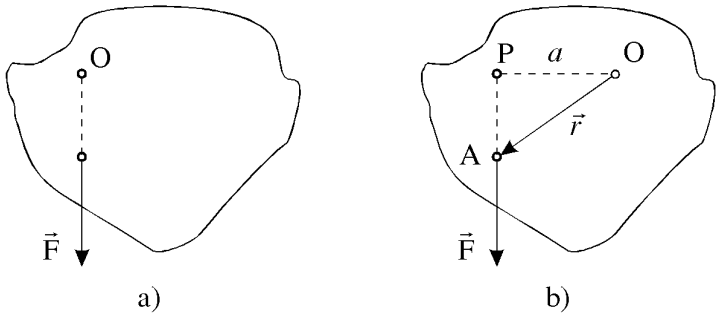

Figure 6.5: With an explanation of the static moment of force

The torque is defined as the vector product of the position vector of the point of application of the force $\vec{r}$ and the force $\vec{F}$ acting on the body:

$$\vec{M} = \vec{r} \times \vec{F}. \tag{6.2}$$

The torque is a vector quantity whose direction is perpendicular (normal) to the plane in which the force vector $\vec{F}$ and the position vector $\vec{r}$ lie, i.e., the direction of the vector $\vec{M}$ coincides with the axis of rotation. The axis of rotation stands perpendicular to the plane to which $\vec{r}$ and $\vec{F}$ belong. In Figure 6.5b, axis of rotation O is perpendicular to the drawing plane. The sense of the torque vector $\vec{M}$ is determined by the right-hand screw rule or the right-hand rule. The unit of the static moment of force is Nm.

The magnitude of the torque vector is

$$M = rF \sin(\vec{r}, \vec{F}) = Fa. \tag{6.3}$$

Chapter 7

Dynamics of Rotational Motion

7.1 Moment of Inertia

When a rigid body rotates around a fixed axis OO' under the action of force $\vec{F}_1$, all its points describe circles whose centers lie on this axis (Fig. 7.1). All points of the body at a given moment have the same angular velocity and angular acceleration.

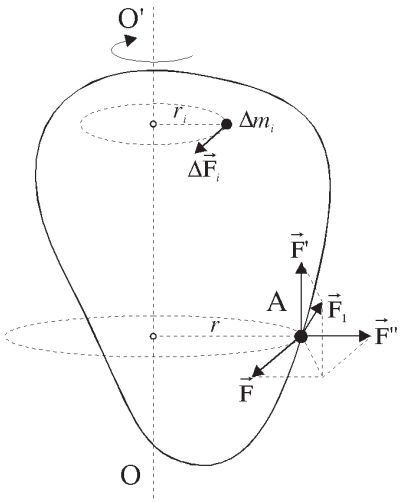

Figure 7.1: A body that rotates

The force $\vec{F}_1$ in Figure 7.1 can be decomposed into three components: $\vec{F}'$ parallel to the axis OO', $\vec{F}''$ perpendicular to this axis, and $\vec{F}$ perpendicular to both $\vec{F}'$ and $\vec{F}''$. It is obvious that the body performs rotational motion only under the action of the force component $\vec{F}$, which is tangentially directed to a circle of radius r. The effect of the force on rotation depends not only on its magnitude but also on the distance of the point of application from the axis of rotation, i.e., on the static moment of force (torque).

As already stated, the torque can be represented as the vector product of the position vector (radius vector) $\vec{r}$ of the point of application of the force and the force $\vec{F}$ acting on the body:

$$\vec{M} = \vec{r} \times \vec{F}. \tag{7.1}$$

As is well known, the magnitude of vector $\vec{M}$ can be calculated as the product of the magnitudes of vectors $\vec{r}$ and $\vec{F}$ and the sine of the angle θ between them ($\theta = \sphericalangle(\vec{r}, \vec{F})$): $|\vec{M}| = |\vec{r} \times \vec{F}| = rF \sin\theta$.

Since the tangent is perpendicular to the radius ($\theta = 90^o$, i.e. $\sin\theta = 1$), the magnitude of the torque $\vec{M}$ is

$$M = F\,r. \tag{7.2}$$

Let us divide the observed body into mass elements Δm_1, Δm_2, Δm_3,...,Δm_n. Since the force $\vec{F}$ causes the rotation of all particles, the mass element Δm_i will rotate with radius $\vec{r}$ under the action of the force element $\Delta\vec{F}_i$ (Fig. 7.1). According to Newton's Second Law, the magnitude of this force can be expressed as

$$\Delta F_i = \Delta m_i a_i, \tag{7.3}$$

where a_i is the tangential acceleration of the mass element Δm_i. When both sides of expression (7.3) are multiplied by r_i, we get

$$\Delta F_i r_i = \Delta m_i a_i r_i = \Delta m_i r_i^2 \alpha. \tag{7.4}$$

Since $\Delta F_i r_i = \Delta M_i$ is the moment of force that causes the rotation of an mass element Δm_i, considering that a mass element represents a material point whose moment of inertia is

$$\Delta I_i = \Delta m_i r_i^2, \tag{7.5}$$

expression (7.4) takes the form

$$\Delta M_i = \Delta I_i \alpha. \tag{7.6}$$

The static moment of the force (torque) M acting on all the mass elements that make up the body is given by the expression

$$M = \sum_{i=1}^{n} \Delta M_i = \alpha \sum_{i=1}^{n} \Delta I_i = \alpha I, \tag{7.7}$$

where

$$I = \sum_{i=1}^{n} \Delta I_i, \tag{7.8}$$

is *the moment of inertia of the body.*

This expression represents the fundamental law of rotational motion dynamics, or the Second Newton's Law for rotational motion.

In the case that the body has a regular geometric shape (ball, cylinder, rod) and is homogeneous (its density is the same in all its parts), the moment of inertia can be calculated using the volume integral:

$$I = \int r^2 dm. \tag{7.9}$$

7.2 Kinetic energy, Work, and Power

A rigid body rotating around a fixed axis OO' is observed (Fig. 7.2). Let the body be divided into small volumes with masses Δm_1, Δm_2, Δm_3,...,Δm_n (mass elements), at distances r_1, r_2, r_3,...,r_n from the axis of rotation.

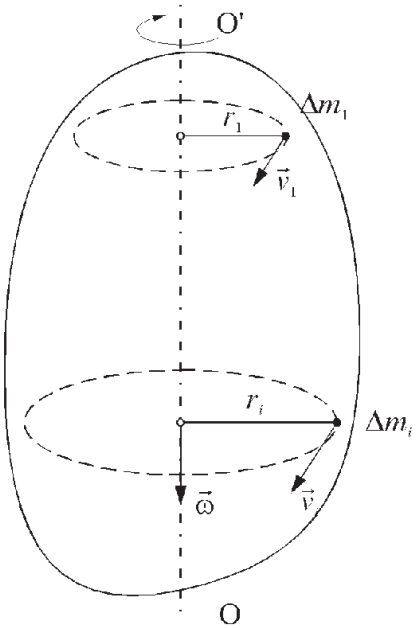

Figure 7.2: A body that rotates

During the rotation of a rigid body around a fixed axis, all mass elements describe circles of different radius r_i and have different peripheral (tangential) velocities v_i. However, since it concerns a rigid body, the angular velocity of all mass elements is the same:

$$\omega = \frac{v_1}{r_1} = \frac{v_2}{r_2} = \frac{v_3}{r_3} = \cdots = \frac{v_n}{r_n}. \tag{7.10}$$

The kinetic energy of a rigid body is equal to the sum of the kinetic energies of the mass elements:

$$E_k = \frac{\Delta m_1 v_1^2}{2} + \frac{\Delta m_2 v_2^2}{2} + \frac{\Delta m_3 v_3^2}{2} + \cdots + \frac{\Delta m_n v_n^2}{2} \tag{7.11}$$

and

$$E_k = \sum_{i=1}^{n} \frac{\Delta m_i v_i^2}{2}. \tag{7.12}$$

Using expression (7.10), the kinetic energy is

$$E_k = \sum_{i=1}^{n} \frac{\Delta m_i \omega^2 r_i^2}{2} = \frac{\omega^2}{2} \sum_{i=1}^{n} \Delta m_i r_i^2 = \frac{I\omega^2}{2}.$$

Accordingly, the kinetic energy of the rotational motion of a body is

$$E_k = \frac{I\omega^2}{2}. \tag{7.13}$$

It can be observed that the kinetic energy of a rigid body rotating around a fixed axis is given by an expression similar to the kinetic energy of a material point in translational motion (movement), where the moment of inertia I corresponds to the mass m, and the angular velocity ω corresponds to the peripheral (tangential) velocity v.

When a body simultaneously undergoes translational and rotational motion, its total kinetic energy is equal to the sum of its translational kinetic energy and rotational kinetic energy, i.e.

$$E_k = \frac{1}{2}mv^2 + \frac{1}{2}I\omega^2. \tag{7.14}$$

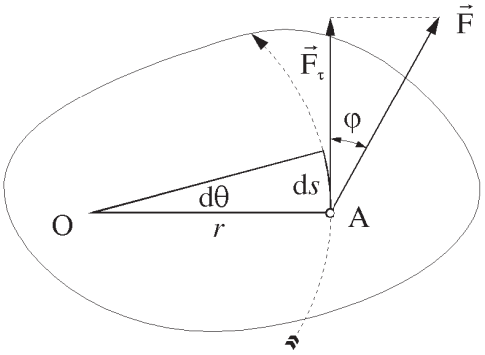

Figure 7.3: The work done by force $\vec{F}$ during an angular displacement $d\theta$

Let an external force $\vec{F}$ act on a rigid body at point A (Fig. 7.3). Under the influence of this force, the body rotates around a fixed axis that passes through point O and is perpendicular to the plane of the drawing. When the body is rotated by a small angle $d\theta$, point A moves by arc $ds = rd\theta$.

The elementary work dA done by the force $\vec{F}$ along the vector of arc, $\vec{ds}$, that has the direction of the tangent, is

$$dA = \vec{F} \cdot \vec{ds} = F\,ds\,\cos\sphericalangle(\vec{F}, \vec{ds}) = F\,ds\,\cos\varphi = F_\tau\,ds = F_\tau\,r\,d\theta, \tag{7.15}$$

where $\vec{F_\tau}$ is the tangential component of the force $\vec{F}$, which intensity is $F_\tau = F\cos\varphi$. The vectors $\vec{ds}$ and $\vec{F_\tau}$ have the same direction and course.

Since the magnitude of torque vector $\vec{M}$ is

$$|\vec{M}| = |\vec{r} \times \vec{F}| = rF\sin\sphericalangle(\vec{r}, \vec{F}) = rF\sin(\frac{\pi}{2} + \varphi) = rF\cos\varphi = rF_\tau, \tag{7.16}$$

then the expression (7.15) can be written in the form

$$dA = Md\theta. \tag{7.17}$$

For the final angular displacement, the work done is

$$A = \int_{\theta_1}^{\theta_2} Md\theta. \tag{7.18}$$

By substituting expression (7.7) into (7.18), the expression for work takes the form

$$A = \int_{\theta_1}^{\theta_2} \alpha I d\theta. \tag{7.19}$$

Since $\alpha = d\omega/dt$ and $d\theta = \omega dt$ (from $\omega = d\theta/dt$), the expression for work in rotational motion can be written as

$$A = \int_{\omega_1}^{\omega_2} I\omega d\omega = \frac{1}{2}I\omega_2^2 - \frac{1}{2}I\omega_1^2. \tag{7.20}$$

From this expression, it can be observed that the work done by the moment of a force is equal to the change in kinetic energy due to the rotational motion of a body, which is similar to the work-energy theorem for translational motion.

If expression (7.17) is divided by the time interval dt, one obtains

$$\frac{dA}{dt} = M\frac{d\theta}{dt}. \tag{7.21}$$

Since $dA/dt = P$ represents power, and $d\theta/dt = \omega$ angular velocity, the above expression takes the form

$$P = M\omega. \tag{7.22}$$

7.3 Angular Momentum and the Law of Conservation of Angular Momentum

First, the case of rotational motion of a material point of mass m, with a rotation radius r, will be considered (Fig. 7.4). Rotation is performed around a fixed axis that is perpendicular to the drawing plane and passes through point O.

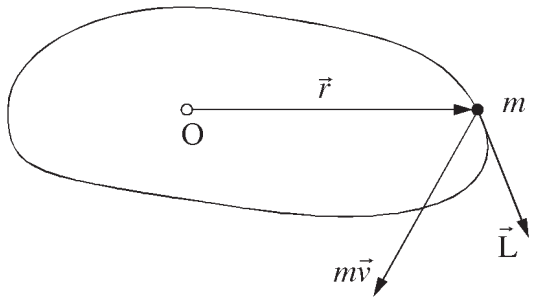

Figure 7.4: A material point during rotation

The angular momentum vector $\vec{L}$ is the vector product of the position vector $\vec{r}$ and the momentum vector $\vec{K}$ $(= m\vec{v})$:

$$\vec{L} = \vec{r} \times \vec{K} = \vec{r} \times m\vec{v}. \tag{7.23}$$

The magnitude of this vector is

$$L = rmv \sin \sphericalangle(\vec{r}, m\vec{v}). \tag{7.24}$$

Since the vectors $\vec{r}$ and $\vec{v}$ are perpendicular, expression (7.24) becomes:

$$L = mvr. \tag{7.25}$$

The angular momentum of a rigid body relative to a fixed axis can be obtained if it is divided into small volumes whose masses are Δm_1, Δm_2, Δm_3,. . . ,Δm_n (Fig. 7.2). Since during the rotation of a body, the peripheral (tangential) velocity $\vec{v_i}$ of a mass element (material point) Δm_i is always perpendicular to the radius $\vec{r_i}$, for the angular momentum L_i of a mass element Δm_i the expression (7.25) can be used:

$$L_i = \Delta m_i v_i r_i. \tag{7.26}$$

The angular momentum of a rigid body with respect to a fixed axis is equal to the sum of the angular momenta of all mass elements:

$$L = \sum_{i=1}^{n} \Delta m_i v_i r_i. \tag{7.27}$$

Since in rotational motion $v_i = \omega r_i$, the angular momentum is given by

$$L = \sum_{i=1}^{n} \Delta m_i r_i^2 \omega = \omega \sum_{i=1}^{n} \Delta m_i r_i^2, \tag{7.28}$$

i.e.

$$L = I\omega, \tag{7.29}$$

and in vector form

$$\vec{L} = I\vec{\omega}. \tag{7.30}$$

Therefore, the angular momentum of a rigid body with respect to a fixed axis of rotation is equal to the product of the body's moment of inertia about the same axis and its angular velocity.

Differentiating expression (7.30) with respect to time gives

$$\frac{d\vec{L}}{dt} = I\frac{d\vec{\omega}}{dt} = I\vec{\alpha} = \vec{M}. \tag{7.31}$$

For a closed system, the moment of external forces is equal to zero ($\vec{M}=0$), therefore

$$\frac{d\vec{L}}{dt} = \vec{L}' = 0,$$

where $\vec{L}'$ is the first derivative of the $\vec{L}$. Considering that the first derivative of the constant is equal to zero, it follows

$$\vec{L} = const.; \quad I\vec{\omega} = const. \tag{7.32}$$

Expression (7.32) represents the mathematical interpretation of *the law of conservation of angular momentum.* The angular momentum of a closed system is conserved, that is, it does not change over time. From expression (7.32), it can be observed that in the case of a constant moment of inertia, when there are no external forces, the angular velocity is also constant.

Chapter 8

Elasticity

8.1 Types of Deformation

In nature, there are no rigid bodies; all bodies, under the influence of forces, can not only change their velocity but they can also change their shape and dimensions, i.e., they can be deformed. Deformations can be completely elastic, partially elastic, or completely plastic. A completely elastic deformation is one in which the body returns to its original shape and dimensions once the forces stop acting. If a body permanently retains a deformed shape, this is plastic deformation. Real bodies exist somewhere between these two limiting cases, and it represents partially elastic deformation.

8.2 Stress and Relative Deformation

Let the surface S in Figure 8.1 represent an element of a body on the surface or inside, on which force $\vec{F}$ acts. The ratio of the intensity of this force to the surface S is called the stress $\sigma = F/S$ acting on the considered element. The

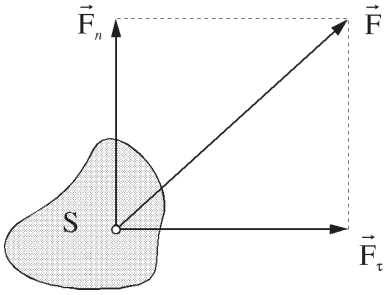

Figure 8.1: A force $\vec{F}$ acting on a small surface S, decomposed into a normal component $\vec{F}_n$ and a tangential component $\vec{F}_\tau$

force $\vec{F}$ can be decomposed into a component $\vec{F}_n$, normal to the surface element S, and a tangential component $\vec{F}_\tau$. The ratio of the normal component of the force to the area on which it acts is called the *normal stress* σ_n, and the ratio of the tangential component to the area is called *tangential stress* σ_τ or *shear stress* σ_τ

$$\sigma_n = \frac{F_n}{S}; \quad \sigma_\tau = \frac{F_\tau}{S}. \tag{8.1}$$

The unit of stress is *Pascal* ($Pa = Nm^{-2}$).

The deformation of a body is expressed as the ratio of the change in its dimensions caused by the action of a force to its original dimensions. This ratio is called *relative deformation* (δ). It has been experimentally shown that in elastic bodies, the stress σ is proportional to the relative deformation δ:

$$\sigma = E\,\delta. \tag{8.2}$$

The proportionality coefficient E is called *the modulus of elasticity*, and its value depends on the type of material. Expression (8.2) represents a mathematical interpretationi of *Hooke's law* for the elastic deformation of a body. This law applies up to a certain limit of material stress. This limit is called the proportionality limit and **lies before the elastic limit**. It has been experimentally shown that solid bodies have a complex dependence of deformation on time, which is not covered by Hooke's law. The deformation does not fully establish immediately at the onset of the applied force. Also, when the force ceases, the deformation does not disappear completely; instead, a portion of it remains and gradually decreases over time. The time dependence of elastic deformation is shown in Fig. 8.2. If at the moment t_o a force begins to act on a body, the initial elastic deformation AB establishes very quickly, and then under the action of the force, the deformation slowly increases over time (curve $BC'C$)). If at the moment t_1 the force stops acting, the deformation first quickly decreases by a value CD equal to AB, and then slowly decreases over time (curve DEE'). Thus, after the force stops acting, a certain deformation $A'D$ remains, which gradually decreases as the body returns to its original shape and dimensions. The deformation that the body retains for a certain time after the force stops acting is called *remanent* or *residual deformation*.

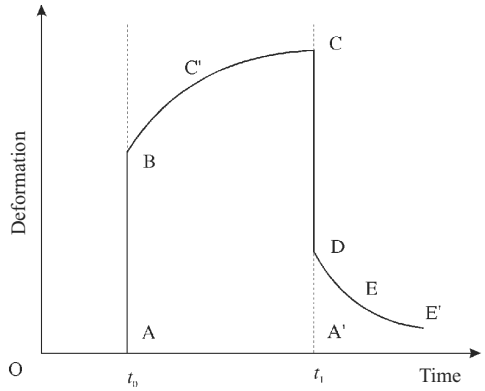

Figure 8.2: Change of elastic deformation over time (dependence of relative deformation during elastic deformations on time)

8.3 Types of Elastic Deformations

Elastic deformations of a body that occur under the action of forces can be categorized into two types: stretch (or compression) and shear. When a force acts perpendicular to the surface of a body, it changes the distances between the particles of the body (molecules, atoms, or ions), and this type of deformation is called stretching (or compression). When a force acts tangentially, it causes the layers of particles to shift relative to each other, and this type of deformation is called shearing.

8.3.1 Stretch

Let a rod (or wire) of length l and constant cross-sectional area S be subjected to a constant force $\vec{F}$, whose line of action coincides with the axis of the rod (Fig. 8.3). If the rod has a small cross-sectional area, it can be approximately assumed that only normal stress occurs, $\sigma = F/S$.

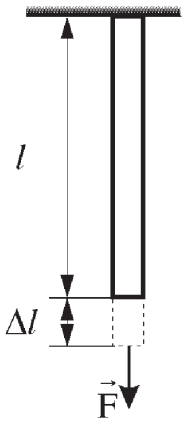

Figure 8.3: Stretch of a wire or rod under the action of force $\vec{F}$

Under the action of force F, the rod is stretched by Δl, so the relative deformation is $\delta = \Delta l/l$. Then, according to Hooke's law (8.2), the relationship between normal stress and relative deformation is given by the expression

$$\frac{F}{S} = E_y \frac{\Delta l}{l}, \tag{8.3}$$

where E_y is Young's modulus of elasticity.

The deformation of solid bodies follows Hooke's law only up to a certain limit. Strictly speaking, the dependence of relative deformation on stress is not linear, but deviations from linearity are not significant for small values of relative deformation. A graphical representation of the stretching process for a metal rod (or wire) is given in Figure 8.4. Linear dependence holds up to the proportional limit σ_p. After the force ceases, the body will return to its original shape and dimensions along line AO. With further increase in stress, the deformation is **still elastic**, but there is no linear relationship between normal stress and relative deformation (region AB). This behavior occurs up to the elastic limit σ_e. After the force stops, the body will also return to its original shape and dimensions along line BAO. If the stress continues to increase, residual deformation appears in the rod, so that after the force ceases, the reduction in deformation does not occur along the $CBAO$ line, but along CF. The stress at which residual deformation of approximately 2% appears is called the beginning of fluidity σ_f (corresponding to point C on the curve). In the CD

region, known as the fluidity region or plastic deformation region, deformation increases even though the stress does not. Materials for which the flow region is significant are called viscous, while those for which this region practically does not exist are brittle materials. With further stretching, the body undergoes complete plastic deformation (DE region). The maximum stress at which the body breaks is called the ultimate stress σ_k and corresponds to point E on the curve.

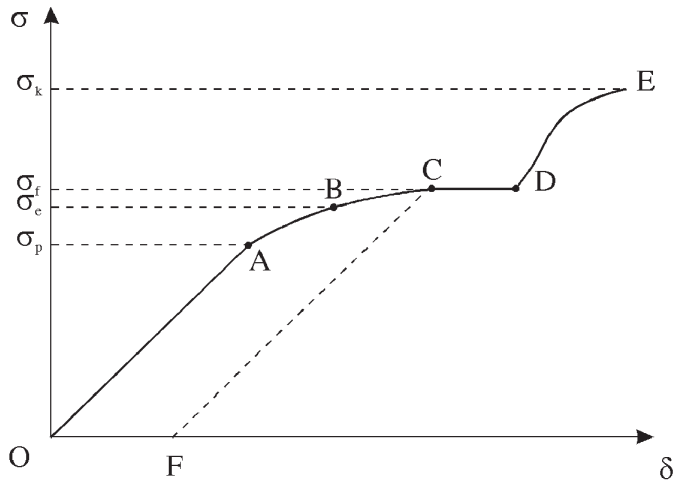

Figure 8.4: The relative deformation of solid bodies depending on stress

When a rod is subjected to a force causing stretching, its transverse (cross) dimensions decrease. Let there be a rod of length l, whose cross-section is a square with side a, subjected to a force that stretches it (Fig. 8.5). Due to the action of the force, the length of the rod becomes $l + \Delta l$, and the side $a - \Delta a$. The ratio of the relative transverse and longitudinal deformation

$$\mu = \frac{\Delta a/a}{\Delta l/l}, \tag{8.4}$$

is called the Poisson's ratio or the Poisson coefficient.

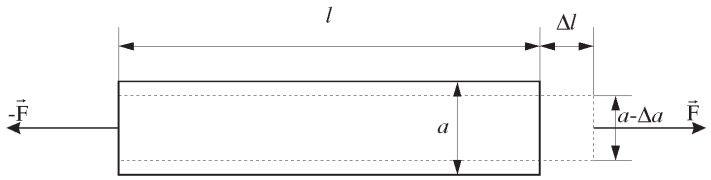

Figure 8.5: Change in the dimensions of the rod during stretch

This can be neglected if the longitudinal (lengthwise) dimensions of the body are much larger than the transverse (cross) dimensions.

8.3.2 Shear

A parallelepiped made of the given material, which is fastened to the base on the bottom side, will be observed (Fig. 8.6). A tangential force $\vec{F}$ acts on the upper surface of this parallelepiped, distributed across the entire surface. Since the parallelepiped is fastened to the base on the bottom side, a force of the same intensity but in the opposite direction (reaction force) will act on the bottom surface. Under the action of this force, the body will deform such that the upper layers of the material will shear relative to the lower layers.

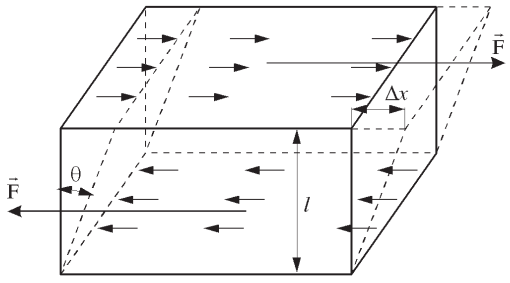

Figure 8.6: Change in body shape under shear

8.3.3 Torsion

Torsion (twisting or warping) is a special case of shear. The most common and also the simplest form of torsion occurs in a rod with a circular cross-section. Figure 8.7 shows the cross-section of such a rod of length l and radius R, when it is not under load (Fig. 8.7a) and when tangential forces causing torsion are applied (Fig. 8.7b).

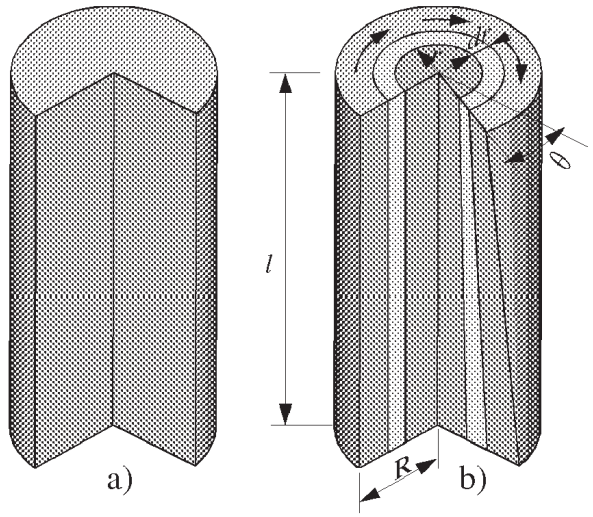

Figure 8.7: Torsion of a body

8.3.4 Bending

Besides torsion, an important type of deformation in engineering is bending. This type of deformation occurs in a rod that is fixed at one end, with a force acting on the other end, or in a rod that is fixed at both ends, with a force acting in the middle. In Figure 8.8a, a rod made of elastic material with a rectangular cross-section is shown, which is fixed at one end to a stationary support, while a force F acts on the other end. Due to the force, the rod will bend proportionally to the intensity of the force. During bending, the upper side of the rod becomes convex and longer, while the lower side becomes concave and shorter than its original length l. The case when the rod is fixed at both ends and the force acts at its middle is shown in Figure 8.8b.

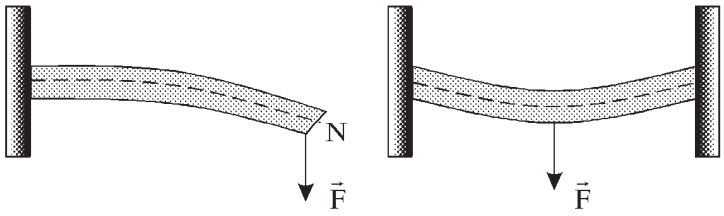

Figure 8.8: Bending of a rod fixed at one end (a) and at both ends (b)

8.4 Internal Elastic Potential Energy of the Body

During elastic deformation, there is a displacement of particles (atoms, molecules, or ions) within the body, between which attractive forces exist. This displacement requires overcoming this internal forces, i.e., to perform work that represents an increase in the internal elastic potential energy of a body.

Let an external force F act on a rod of length l and cross-sectional area S, which increases its length by Δl (Fig. 8.9).

A force F_1 acts on the rod at length $l + x$, which elongates (stretches) the rod by dx, performing work element:

$$dA = F_1 dx \tag{8.5}$$

From expressions (8.3) and (8.5):

$$dA = E_y S \frac{x}{l} dx. \tag{8.6}$$

The total work of the external force required to increase the length of the rod by Δl is equal to the internal elastic potential energy E_p and is obtained by integrating the above expression within the limits from $x=0$ to $x=\Delta l$:

$$E_p = A = \frac{E_y S}{l} \int_0^{\Delta l} x dx = \frac{1}{2} \frac{E_y S}{l} \Delta l^2 = \frac{1}{2} F \Delta l. \tag{8.7}$$

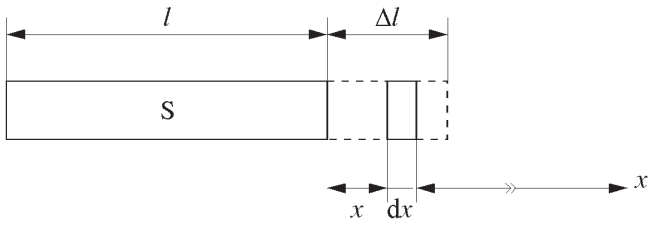

Figure 8.9: Work in rod deformation

Chapter 9

Fluid Mechanics

9.1 Pressure in the Fluid

A physical quantity equal to the ratio of **the normal force** ΔF acting on an element of the surface ΔS of a body in contact with a fluid is called pressure:

$$p = \frac{\Delta F}{\Delta S}. \tag{9.1}$$

If the pressure is constant at every point of the surface, expression (9.1) can be written in the form

$$p = \frac{F}{S}. \tag{9.2}$$

Pressure is *a scalar quantity*, and the unit is the Pascal (Pa).

It has been experimentally established that pressure in fluids is transmitted equally in all directions. There is *hydraulic (external) pressure* and *hydrostatic (internal) pressure*. External pressure occurs due to the action of an external force on a fluid contained in a closed container. Hydrostatic pressure arises from the weight of the fluid itself.

9.2 Pascal's Law

Unlike solid bodies, where external pressure is transmitted only in the direction of the applied force, in liquids it is transmitted equally in all directions. This can be demonstrated with the following experiment. A rubber ball is taken, which is attached to a thick-walled glass tube C containing a movable piston k (Fig. 9.1). The ball and the tube are filled with water. If a force is applied to the piston, the ball will maintain its shape, but its volume will increase. From this observation, it is concluded that pressure is transmitted equally in all directions.

This statement can also be verified using a container with some fluid (Fig. 9.2). As can be seen, the container contains two cylinders with movable pistons k_1 and k_2, whose cross-sectional areas are S_1 and S_2. In this case, the gravitational force will be neglected (there is no difference in the height of the container's centerline) and it will be assumed that *the fluid is incompressible*, i.e., that **under the action of a force, the volume of the fluid remains constant**. Let a constant force $\vec{F}_1$ act normally on the piston k_1 ($\vec{F}_1$ has the direction of the piston movement), the piston will move a certain distance x_1, displacing a fluid volume of $V_1 = S_1 x_1$ from the cylinder. Consequently, the fluid, i.e. force F_1 will apply a force F_2 on piston k_2 that will move a distance x_2, so the volume of fluid pressed into the second cylinder is $V_2 = S_2 x_2$. Since it is an incompressible fluid, these two volumes are equal ($V_1 = V_2$):

$$S_1 x_1 = S_2 x_2. \tag{9.3}$$

When the piston k_1 moves along the path x_1, the force F_1 does work (the force F_1 is constant and has the direction of the piston path):

$$A_1 = F_1 x_1 = p_1 S_1 x_1, \tag{9.4}$$

and the work done by the force F_2 in moving piston k_2 over the distance x_2 is (because the force F_1 is constant, and the force F_2, which is a result of force F_1, must also be constant):

$$A_2 = F_2 x_2 = p_2 S_2 x_2. \tag{9.5}$$

Since $A_1 = A_2$ (no friction), using expressions (9.3), (9.4) and (9.5), it follows

$$p_1 = p_2. \tag{9.6}$$

This expression represents the mathematical interpretation of Pascal's law, which can be formulated as follows: The pressure applied externally to any point of a completely enclosed fluid is transmitted equally in all directions within it.

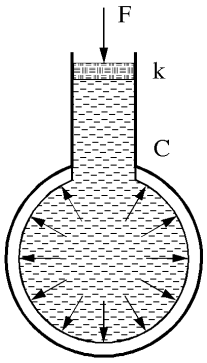

Figure 9.1: Change in the volume of a rubber ball due to the action of an external force

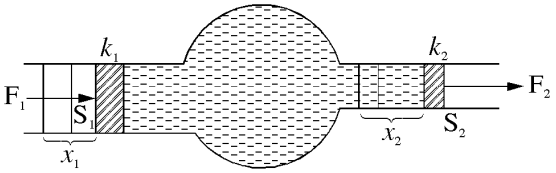

Figure 9.2: Proof of Pascal's Law

9.3 Hydrostatic Pressure

As already mentioned, hydrostatic pressure originates from the weight of the fluid itself. This kind of pressure occurs, for example, at the bottom of a vessel filled with liquid or in a column of air that presses on the surface. It is necessary to find an expression that gives the dependence of pressure on depth. Here, this will be done for a liquid ($\varrho = const.$). Let a liquid in a container (vessel) be exposed only to the effect of the gravitational force (Fig. 9.3). Imagine *a fluid element* in the shape of a thin cylinder with surface area S and thickness dy. Since pressure in liquids acts equally in all directions, the horizontal forces that the fluid exerts on the side surfaces of the observed element will be balanced with each other. Since the liquid is at rest, the vertical forces acting on this element must also be balanced. Let p be the pressure on the lower surface and $p + dp$ the pressure on the upper surface of the element. The upward force acting on the lower surface is pS, and the downward force acting on the upper surface is $(p + dp)S$. If the weight of the observed element is denoted by $dQ = \varrho g S dy$, then the balance of vertical forces can be expressed as follows:

$$(p + dp)S + \varrho g S dy = pS, \tag{9.7}$$

i.e.

$$dp = -\varrho g dy. \tag{9.8}$$

This expression describes the change in pressure with depth inside a fluid. The sign (–) indicates that the pressure

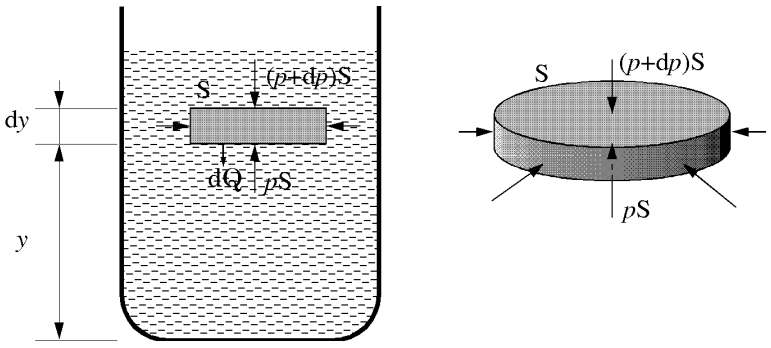

Figure 9.3: Forces acting on an element of fluid that is in equilibrium

decreases as the depth increases (which is not physically logical, but it is up to us which assumption we will introduce; actually, the pressure in a fluid increases with depth).

If inside the fluid at height y_1, relative to the bottom of the container, the pressure is p_1, and at height y_2, relative to the bottom of the container, the pressure is p_2 (Fig. 9.4), equation (9.8) can be written in the form (just integrate this equation)

$$\int_{p_1}^{p_2} dp = -\int_{y_1}^{y_2} \varrho\, g\, dy, \tag{9.9}$$

$$p_2 - p_1 = -\varrho g(y_2 - y_1). \tag{9.10}$$

The equation (9.10) shows that when $y_2 > y_1$, then $p_2 < p_1$, and when $y_2 < y_1$, then $p_2 > p_1$, which makes physical sense.

It is necessary to find the hydrostatic pressure p at a point A ($p_1 = p$) located at a height y_1 relative to the bottom of the container (Fig. 9.4). If the container is open, atmospheric pressure p_a acts on the surface of the liquid ($p_2 = p_a$), which is at a height y_2 relative to the bottom of the container, the eq. (9.10) takes the form

$$p = p_a + \varrho g h. \tag{9.11}$$

In the above expression, the term $\varrho g h$ defines the weight of the liquid column and is called *hydrostatic pressure.*

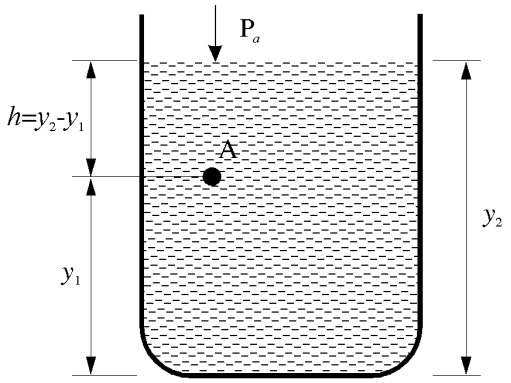

Figure 9.4: Pressure at height h

9.4 Buoyant Force and Archimedes' principle

On every body that is put (submerged) in a liquid (or gas), besides the force of gravity (acting downwards), there is another force acting upwards, which tends to push the body out of the liquid. This force is called *buoyant force.* Buoyant force occurs because the pressure in a liquid (or gas) increases with depth. As a result, the pressure force on the bottom surface of a submerged body (upward) is greater than the pressure force on the top surface (downward).

To simplify the problem, we will consider a body in the shape of a cylinder with height h and base area S, whose axis stands vertically relative to the surface of the liquid in which the cylinder is submerged (Fig. 9.5). A pressure $p_1 = \varrho g y_1$ acts on the upper surface of the cylinder, and the corresponding force is $F_1 = p_1 S = \varrho g y_1 S$. On the lower surface of the cylinder, a force $F_2 = p_2 S = \varrho g y_2 S$ acts. The difference between these two forces represents the buoyant force

$$F_p = F_2 - F_1 = \varrho g (y_2 - y_1) S = \varrho g h S = \varrho g V, \tag{9.12}$$

where $V = hS$ is the volume of the cylinder. Since ϱ is the density of the liquid (gas), the product $\varrho g V = mg$ is equal to the weight of the liquid that occupies a volume equal to the volume of the cylinder. Accordingly, the buoyant force acting on a cylinder is equal in magnitude to the weight of the liquid displaced by the cylinder. This applies to any shape of body. This phenomenon was first observed by Archimedes, and it was named Archimedes' principle. It states: *A buoyant force acts on every body submerged in a liquid, and this force is equal to the weight of the liquid displaced by that body.*

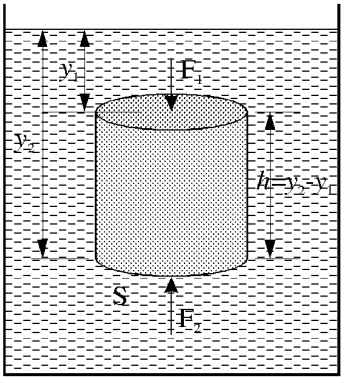

Figure 9.5: Buoyant Force

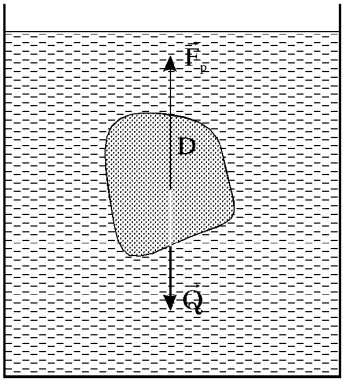

Figure 9.6: Archimedes' principle

9.5 Characteristics of Flow

The part of fluid mechanics that studies the laws of motion of liquids and gases is called *fluid dynamics*. There are two basic types of fluid flow: laminar and turbulent flow. If the flow is such that layers of fluid slide past each other, it is laminar flow. Characteristic of this type of motion is that each fluid particle moves along a smooth trajectory, with these trajectories not intersecting (Fig. 9.7a). When the fluid velocity exceeds a certain limit, which depends on a number of parameters, the flow becomes turbulent. This type of motion is characterized by the appearance of vortices (Fig. 9.7b), with significant internal friction occurring, known as viscosity. Viscosity also occurs in laminar flow, but it is much smaller than in turbulent flow.

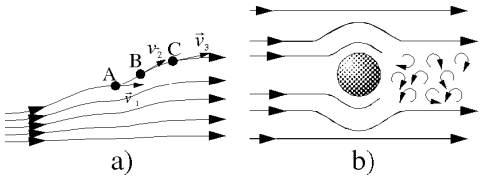

Figure 9.7: Laminar (a) and turbulent flow (b)

There are several important characteristics for laminar and turbulent flow. The first is that the fluid can be considered either compressible or incompressible. Although there is no material that is absolutely incompressible, the flow of many liquids is such that the change in their density is very small. Every fluid has internal friction (viscosity), whose value depends on the type of fluid. *Fluids whose viscosity and compressibility can be neglected are called ideal fluids.*

Fluid flow can be steady or unsteady. In steady flow (stationary flow), the velocity of the fluid at any point in space is constant over time, but this does not mean that it is the same at all points in space (Fig. 9.7a). If the velocity at a given point in space changes over time, the flow is unsteady. **We will consider an ideal fluid in a steady state (stationary regime)**.

9.6 Steady Flow. Flow Continuity Theorem

Let two cross-sections 1 and 2 be selected on a stream tube, whose cross-sectional areas are S_1 and S_2 (Fig. 9.8). The flow velocities of the fluid at these cross-sections are v_1 and v_2. Since this is an incompressible fluid, the density is the

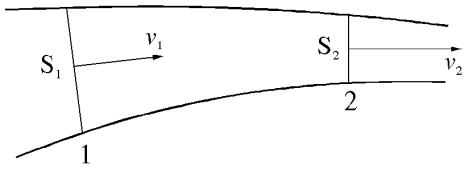

Figure 9.8: Stream tube in fluid

same at both cross-sections. The volume of fluid dV_1 passing through cross-section 1 during time dt is equal to the volume of a cylindrical element with base S_1 and height $dx_1 = v_1 dt$, i.e., $dV_1 = S_1 dx_1 = S_1 v_1 dt$. The mass of fluid that passes through this cross-section during the time dt is $dm_1 = \varrho dV_1 = \varrho S_1 v_1 dt$. The mass that passes through cross-section 2 in the same time is $dm_2 = \varrho dV_2 = \varrho S_2 v_2 dt$. Since we consider a steady flow of an incompressible fluid, the masses of fluid that pass through cross-sections 1 and 2 in the same time must be identical ($dm_1 = dm_2$):

$$\varrho S_1 v_1 dt = \varrho S_2 v_2 dt, \tag{9.13}$$

$$S_1 v_1 = S_2 v_2. \tag{9.14}$$

Since equality holds for any two sections of the stream tube, it can generally be written that

$$Sv = const., \tag{9.15}$$

i.e., that the product of the flow velocity and the cross-sectional area of the stream tube is a constant. Expression (9.15) is known as *the flow continuity theorem* and represents the volume of fluid that flows out per unit time (so-called volumetric flow rate).

9.7 Bernoulli's Principle

Let an ideal fluid flow in a stream tube that is inclined to the horizontal in a steady state (Fig. 9.9). At cross-section 1 of this tube, a constant force $\vec{F}_1$ acts on the fluid at height h_1 relative to the horizontal (reference level), giving the fluid velocity v_1 and the pressure p_1. The h_1 is the height between the reference level and the center of mass. Similarly, at cross-section 2, a constant force $\vec{F}_2$ acts on the fluid at height h_2, giving the fluid velocity v_2, and the pressure p_2. If $p_1 > p_2$, the fluid will move from left to right. For a time interval Δt, the same mass of fluid Δm passes through both cross-sections of the stream tube. The fluid masses Δm at the observed cross-sections 1 and 2 have kinetic energies

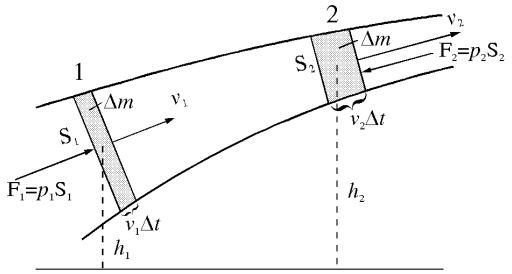

Figure 9.9: Fluid flow in a stream tube

$(\Delta m v_1^2)/2$ and $(\Delta m v_2^2)/2$ and potential energies $\Delta m g h_1$ and $\Delta m g h_2$. The change in total energy ΔE is equal to the difference in the total energies of the masses Δm that flow through these cross-sections during the time Δt

$$\Delta E = E_2 - E_1 = \frac{\Delta m v_2^2}{2} + \Delta m g h_2 - \left(\frac{\Delta m v_1^2}{2} + \Delta m g h_1 \right). \tag{9.16}$$

In accordance with the law of conservation of energy, the change in energy is equal to the work ΔA of external forces (pressure forces) used to move a mass Δm

$$\Delta E = \Delta A. \tag{9.17}$$

It is necessary to determine this work. The external constant force F_1 performs work ΔA_1 to move the displaced mass Δm along the path $v_1 \Delta t$. At the same time, the expelled mass Δm performs work ΔA_2 against the external constant force F_2 along the path $v_2 \Delta t$. Therefore, it can be written

$$\Delta A_1 = F_1 v_1 \Delta t; \quad \Delta A_2 = -F_2 v_2 \Delta t. \tag{9.18}$$

The total work is equal to the sum of these two works. From $F_1 = p_1 S_1$ and $F_2 = p_2 S_2$, it can be written:

$$\Delta A = \Delta A_1 + \Delta A_2 = p_1 S_1 v_1 \Delta t - p_2 S_2 v_2 \Delta t. \tag{9.19}$$

Since it concerns an incompressible fluid, it can be written as $S_1 v_1 \Delta t = S_2 v_2 \Delta t = \Delta V$, where ΔV is the volume of the fluid with mass Δm. Consequently, expression (9.19) takes the form

$$\Delta A = p_1 \Delta V - p_2 \Delta V. \tag{9.20}$$

Using expressions (9.16), (9.17) and (9.20) it is obtained

$$p_2 \Delta V + \frac{\Delta m v_2^2}{2} + \Delta m g h_2 = p_1 \Delta V + \frac{\Delta m v_1^2}{2} + \Delta m g h_1 \tag{9.21}$$

If equation (9.21) is divided by ΔV and it is taken that $\Delta m / \Delta V = \varrho$, where ϱ represents the density of the fluid, it finally results in

$$p_1 + \frac{\varrho v_1^2}{2} + \varrho g h_1 = p_2 + \frac{\varrho v_2^2}{2} + \varrho g h_2. \tag{9.22}$$

It is valid for any section of the stream tube:

$$p_u = p + \frac{\varrho v^2}{2} + \varrho g h = const., \tag{9.23}$$

which means that the total pressure in a moving fluid is constant at any cross-section.

The expression (9.23) is called *Bernoulli's equation* or *Bernoulli's law*. The first term represents external pressure, the second dynamic pressure, and the third hydrostatic pressure.

9.8 Application of the Bernoulli's Equation

9.8.1 Torricelli's Theorem

Figure 9.10 shows a container with a cross-sectional area S_1 containing a liquid of density ϱ. The container is open at the top, and it has a hole on the side with a cross-sectional area S_2 through which the liquid flows out at a speed of v_2. The S_1 is at a height h with respect to the center of hole S_2.

Atmospheric pressure p_a acts on both the top of the container and the hole. If $S_1 \gg S_2$, the velocity v_1 at which the liquid drains into the container will be considerably smaller than the outflow velocity v_2 ($v_1 \ll v_2$), so v_1 can be assumed to be approximately 0 ($v_1 \approx 0$).

The center of of hole S_2 is taken as the reference level, and Bernoulli's equation for this case can be written in the form

$$p_a + \varrho g h = p_a + \frac{\varrho v_2^2}{2}, \tag{9.24}$$

what gives

$$v_2 = \sqrt{2gh}. \tag{9.25}$$

This expression shows that the velocity (speed) of liquid flowing out of a wide container (container) is the same as the velocity of any body freely falling from a height h. This is *Torricelli's theorem*, which also applies to an opening at the bottom of the container.

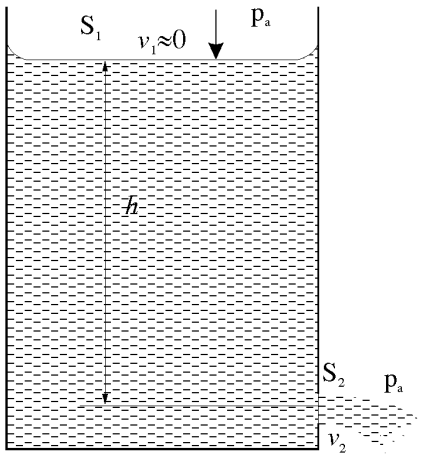

Figure 9.10: Leaking liquid through a small opening

9.9 Viscosity

As has already been emphasized, in a real fluid there is internal friction called viscosity. Viscosity can be imagined as the friction that occurs when layers of the medium in contact move relative to each other. In liquids, viscosity is caused by the cohesive forces acting between molecules, while in gases it results from collisions between atoms and molecules.

Viscosity depends on the type of liquid. For example, glycerin has a higher viscosity than water, lubricant grease for certain machine parts has a higher viscosity than motor oil, and so on. In general, liquids have a higher viscosity than gases.

Newton formulated a law according to which friction in fluids behaves analogously to friction in solid bodies in mechanics. Consider a fluid situated between two plates. One plate is stationary, while the other moves parallel to it at speed v, as shown in Fig. 9.11.

This fluid can be divided into thin layers. The layers adjacent to the plates experience adhesive forces between the fluid molecules and the molecules of the plate material. The layer of fluid next to the upper plate moves at the same speed as the plate, the layer below it moves at a lower speed, and so on, until the layer next to the stationary plate remains at rest. It is assumed that the layers slide over each other, i.e., that this is laminar flow, which can occur at relatively low velocities. If the upper plate moves uniformly under the action of force F. This means that the internal friction force F_{tr} is balanced by this force.

Newton established that the internal friction force depends on the type of fluid, the surface area of the contact layers S, and the velocity gradient (rate of change of velocity per unit layer thickness) dv/dx:

$$F_{tr} = F = \eta S \frac{dv}{dx}. \tag{9.26}$$

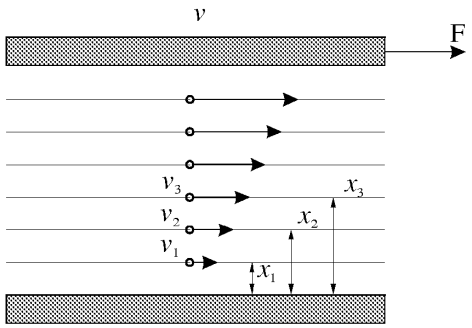

Figure 9.11: Laminar flow of a viscous fluid

The proportionality coefficient η depends on the type of fluid and is called *the dynamic viscosity coefficient.* The larger the viscosity coefficient, the more the fluid deviates from ideal behavior. The unit for the dynamic viscosity coefficient is $Pa\,s$.

Chapter 10

Heat and Temperature

10.1 Temperature and Thermal Equilibrium

The quantitative measure of a body's level of heat is called temperature. In this way, the concept of temperature defined here only reflects its qualitative nature. To determine the quantitative characteristics of temperature, it is necessary to introduce the concept of thermal equilibrium. This concept can be explained using the following example. If a piece of heated metal comes into contact with a piece of ice, the ice melts and the metal cools down. This process continues until the temperatures of both bodies equalize. When the bodies come into contact, heat is exchanged, causing the energy of the metal to decrease and the energy of the ice (during melting) to increase.

The term heat is used in two ways: as the thermal energy of a body and as the amount of heat exchanged (transfer of thermal energy between two systems), which is equivalent to work and represents a measure of energy change.

10.2 Thermometers

Instruments for measuring temperature are called thermometers. Some types of thermometers will be presented here. **A thermometer with liquid** consists of a bulb A with thin glass walls, from which a glass capillary B extends, sealed at the top and evacuated of air (Fig. 10.1). The liquid, such as mercury or colored alcohol, fills the bulb and partially the capillary. When the thermometer is placed in an environment with elevated temperature, the volume of the liquid increases, as well as the volume of bulb A and capillary B. Since the liquid expands more than the glass, its level in the capillary rises with the temperature increase. Thus, this thermometer uses the difference in expansion between the liquid and the glass.

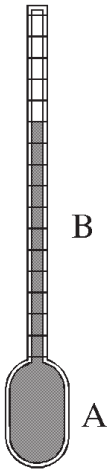

Figure 10.1: Thermometer with liquid

A resistance-based thermometer uses the property that the electrical resistance of a metal increases with temperature. As the resistance rises, the current through the metal decreases. In this way, temperature can be measured using an instrument that measures current. A schematic diagram of this thermometer is shown in Figure 10.2. The thermometer is connected in an electrical circuit with a current-measuring instrument.

A bimetal thermometer consists of a strip, one half of which is made of one metal and the other half of another metal (Fig. 10.3). Due to the different expansions of these two metals in the bimetal strip (they have different coefficients of linear expansion), the bimetal bends when the temperature increases. This thermometer has low accuracy and requires frequent calibration.

A thermocouple uses the phenomenon that in an electrical circuit of two different metals, when both junctions are not at the same temperature, a potential difference occurs that depends on the temperature difference of the junctions.

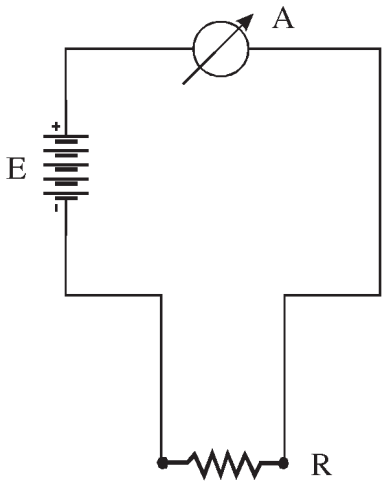

Figure 10.2: Thermometer based on measuring electrical resistance

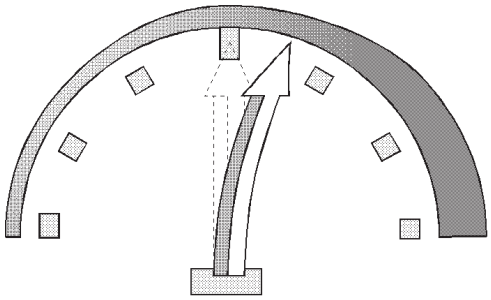

Figure 10.3: Thermometer with bimetal strip

A schematic representation of the thermocouple is shown in Figure 10.4. One end of the junction is kept at a constant temperature, while the other is at the place where the temperature needs to be measured. With this thermometer, the temperature can be measured very precisely.

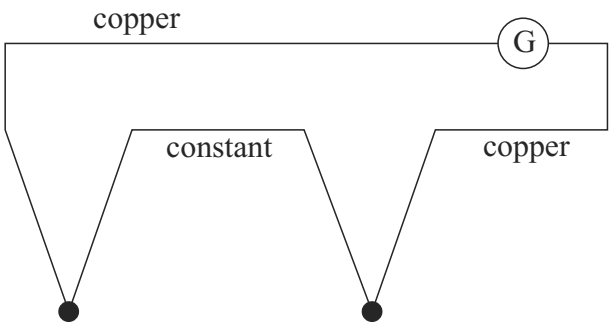

Figure 10.4: Thermocouple

10.3 Temperature Scales

When measuring temperature, it is necessary to choose a temperature scale. These scales are based on two fixed points that are easy to reproduce. Today, the most commonly used scales are the Celsius and Kelvin scales. According to Celsius, the first point is the melting point of ice, specifically the equilibrium temperature of a mixture of pure water and ice at normal atmospheric pressure. This temperature is assigned a value of zero (0 °C). The second point is the boiling temperature of pure water, i.e., the equilibrium temperature of water that boils together with its vapor at normal atmospheric pressure. This temperature is defined as 100 °C. This interval between the two fixed points is divided into 100 parts.

The Kelvin or thermodynamic scale has the same graduations as the Celsius scale, but its zero point (absolute zero) is approximately −273 °C. Therefore, when the temperature is expressed in Kelvins (absolute temperature), 273 must be added to the value read in Celsius degrees (for example, 20 °C is 20 + 273 = 293 K).

In some countries, the Fahrenheit scale is also used, in which the freezing point of water is at 32, and the boiling point of water is at 212. Thus, in the Fahrenheit scale, the interval between the freezing point of water and the boiling point of water is divided into 180 parts. One such part represents a Fahrenheit degree (°F).

10.4 Thermal Expansion of Solids and Liquids

Most bodies expand equally in all directions when heated. However, if the length of the body is significantly greater than the other two dimensions, as in a rod, such expansion is called *linear expansion*. Experimentally, it has been shown that the elongation of the rod Δl depends on its original length and the change in temperature in the following way

$$\Delta l = \alpha l_o \Delta t, \tag{10.1}$$

where l_o is the length of the rod at temperature t_o, $l = l_o + \Delta l$ is the length of the rod at temperature t ($\Delta t = t - t_o$) and α is the thermal coefficient of linear expansion. In the case when $t_o = 0\,^\circ$ C, the above expression takes the form

$$\Delta l = l - l_o = \alpha l_o t \Rightarrow l = l_o + \alpha l_o t = l_o(1 + \alpha t). \tag{10.2}$$

As a result of linear expansion, the volume of a body increases. Consider a body in the shape of a cube with side l_o at temperature t_o. Its volume at this temperature is $V_o = l_o^3$. After heating to temperature t, the side will have a length of $l_o(1 + \alpha t)$, and the body will have a volume

$$V = [l_o(1 + \alpha t)]^3 = l_o^3\,(1 + 3\alpha t + 3\alpha^2 t^2 + \alpha^3 t^3). \tag{10.3}$$

Since α for solid bodies has a small value (usually between 10^{-5} and $10^{-6}\,\mathrm{K}^{-1}$), the terms containing α^2 and α^3 can be neglected, so expression (10.3) takes the form

$$V \approx V_o(1 + \gamma t), \tag{10.4}$$

where $\gamma = 3\alpha$ is the thermal coefficient of volume expansion (volume expansion coefficient) of solid bodies. By substituting the body density $\varrho = m/V$ into expression (10.4), we get

$$\varrho = \frac{\varrho_o}{1 + \gamma t}, \tag{10.5}$$

where $\varrho_o = m/V_o$ is the density of the body at temperature $t = 0\,$°C, and ϱ is the density at temperature t. As can be seen, the density of the body decreases with temperature.

Expressions (10.4) and (10.5) also apply to fluids, but the volume expansion coefficient for liquids is higher than for solids and ranges from 10^{-3} to $10^{-4}\,\mathrm{K}^{-1}$. *Hollow bodies expand according to the same law as solid bodies (e.g., a glass in the shape of a cylinder and a solid cylinder in the shape of a glass)*.

An anomaly is observed in water as the temperature increases. Namely, when heated from 0 °C do 4 °C, its volume decreases. This means that the thermal coefficient of volume expansion in this range is negative. At temperatures above 4 °C, the volume of water increases when heated, as is the case with other substances. Since the volume of a given mass is smallest at a temperature of 4 °C, the density of water at this temperature is the highest. Figure 10.5 shows the dependence of the volume and density of water on temperature in the range from 0 °C to 10 °C.

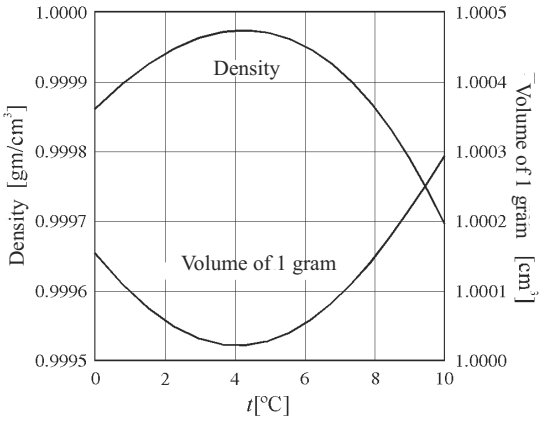

Figure 10.5: The density and volume of one gram of water in the temperature range from 0 °C to 10 °C

10.5 Thermal Stress of Materials

Experience has shown that if the thermal expansion of a body is restricted in any way, thermal stress occurs, meaning tensile or compressive stresses appear (so-called thermal stresses). These stresses can become very large, potentially causing the body to exceed its elastic limit or even leading to its fracture. To avoid such stresses, special measures are taken in technical structures. For instance, iron bridges are anchored to the foundation at only one end, while the other end rests on movable bearings.

Thermal stress in a rod that can expand or contract can be calculated in a relatively simple way. Let the ends of the rod of length l be fixed to prevent it from expanding or contracting. At temperature t_2, the rod is in an unstressed state. If the rod is cooled to temperature t_1, thermal stress will occur due to the inability of the rod to contract. If the rod could contract, the change in its length would be

$$\Delta l = \alpha l \Delta t, \tag{10.6}$$

but since that is not possible, a thermal stress force (stretching force) occurs. This force can be obtained from Hooke's law (expression (8.3)):

$$F = SE_y \frac{\Delta l}{l}. \tag{10.7}$$

By substituting Δl from (10.6) into (10.7), this force is obtained

$$F = SE_y \alpha \Delta t. \tag{10.8}$$

The F/S ratio represents the normal stress corresponding to thermal strain.

Here, only the case of thermal stress due to linear expansion is considered. Complex stresses occur in all directions within bodies. Various types of stresses can develop in materials as a result of cooling after different thermal production processes. The quality and durability of materials are significantly reduced when such stresses are present.

10.6 Laws of Ideal Gases

In studying gases, the ideal gas model is primarily used, so the laws to be discussed apply to this approximation. For an ideal gas, it is assumed that the volume of the molecules is negligible compared to the volume of the container holding the gas. Additionally, the effect of intermolecular forces is ignored, and collisions between gas molecules and the walls of the container are considered perfectly elastic. The behavior of ideal gases can be described using several laws. These are Boyle-Mariotte's, Gay-Lussac's, Charles's, Avogadro's, and Dalton's laws.

10.6.1 Boyle-Mariotte's Law

Boyle and Mariotte independently of each other established that the pressure of a gas changes inversely with its volume, if the temperature of a certain amount of gas is kept constant. Mathematically, this law can be expressed as

$$pV = const. \qquad (T = const., m = const.) \tag{10.9}$$

Expression (10.9) in the $p - V$ diagram represents an equilateral hyperbola called *an isotherm*. In Figure 10.6, a family of isotherms for three different temperatures is shown.

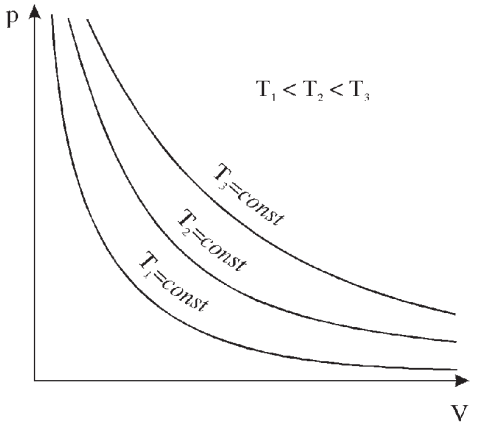

Figure 10.6: Isotherms of an ideal gas

10.6.2 Gay-Lussac's Law

It is known that gases expand much more than solids when heated. Here, the thermal expansion of a given amount of gas at constant pressure is considered. In this case, the volume of the gas V changes linearly with temperature t according to Gay-Lussac's law:

$$V = V_o + V_o \gamma t = V_o(1 + \gamma t), \qquad (p = const.) \tag{10.10}$$

where V_o is the volume of the gas at 0 °C, V is the volume at t °C, and γ is the thermal coefficient of volume expansion (volume expansion coefficient) of the gas, which for all gases is approximately $1/273\,\mathrm{K}^{-1}$. The process of gas expansion

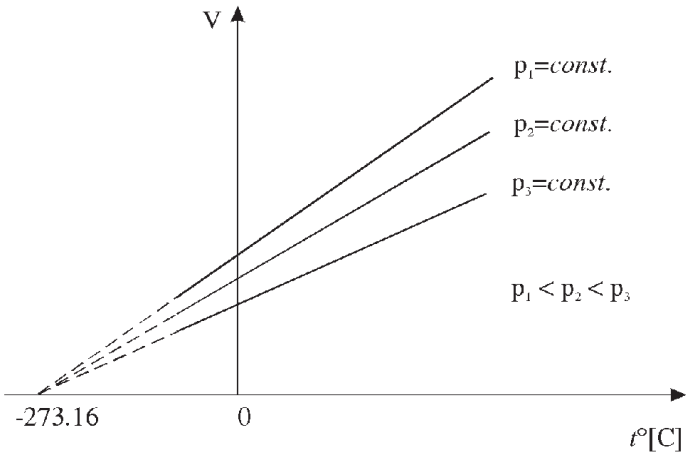

Figure 10.7: Isobars of an ideal gas

at constant pressure is called isobaric. In the $V-t$ diagram, this process is represented by a straight line called **an isobar** (Fig. 10.7).

From expression (10.10) it follows that the isobars intersect the t-axis at the point $t \approx -273\,℃$. If the origin of the coordinate system is moved to this point, we arrive at the Kelvin or thermodynamic scale, where the relationship between temperature expressed in Kelvins and in degrees Celsius is given by the relation

$$T(K) = t\,(℃) + 273. \tag{10.11}$$

It follows that $0\,\mathrm{K} = -273\,^{\circ}\mathrm{C}$.

10.6.3 Charles's Law

Gas can also be heated in such a way that its volume is kept constant. In this case, the gas pressure changes linearly with temperature according to Charles's law:

$$p = p_o + p_o\gamma_1 t = p_o(1+\gamma_1 t), \qquad (V = const.) \tag{10.12}$$

where p_o is the gas pressure at $0\,℃$, p is the pressure at $t\,℃$, and γ_1 is the pressure change coefficient, which for ideal gases is equal to the volume expansion coefficient ($\gamma_1 = \gamma \approx 1/273\,\mathrm{K}^{-1}$). In the $p-t$ diagram, this process is represented by a straight line called *an isochore* (Fig. 10.8). From expression (10.12) it can also be seen that the isochore intersect

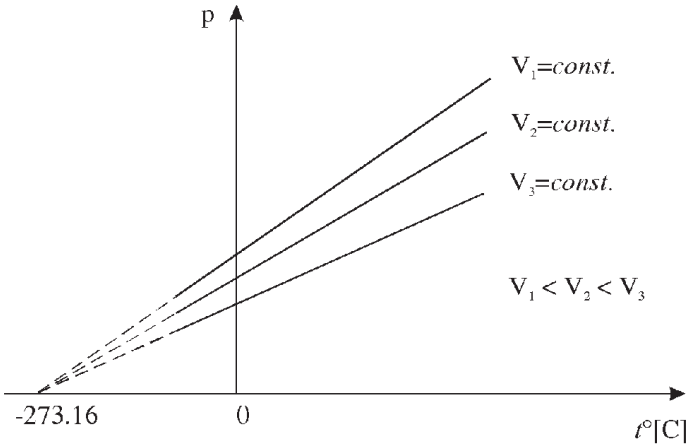

Figure 10.8: Isohore of an ideal gas

the t-axis at the point $t \approx -273\,℃$.

From expression (10.10) it follows that at a temperature of $0\,\mathrm{K}$

$$V = V_o\left(1 + \frac{-273}{273}\right) = 0,$$

i.e., that at a temperature of 0K the volume of a gas would become zero, which does not correspond to reality. This shows that in the area of low temperatures, the laws of ideal gases are not applicable. At low temperatures, a material cannot exist in a gaseous state but instead transitions to a liquid or, most often, solid state.

Expression (10.10) can be represented in a simpler form using thermodynamic temperature

$$V = V_o(1+\gamma t) = V_o\left(1+\frac{1}{273}t\right) = V_o\frac{273+t}{273} = V_o\frac{T}{T_o}, \tag{10.13}$$

where $T_o = 273\,\mathrm{K}$ corresponds to $0\,^\circ\mathrm{C}$. The above expression can be written in the form

$$\frac{V}{V_o} = \frac{T}{T_o}, \qquad (p = const.) \tag{10.14}$$

At constant pressure, the volume of a gas is proportional to its thermodynamic temperature (definition of Gay-Lussac's law).

By the same transformation of expression (10.12), we get

$$\frac{p}{p_o} = \frac{T}{T_o}, \qquad (V = const.) \tag{10.15}$$

At constant volume, the pressure of a gas is proportional to the thermodynamic temperature (definition of Charles's law).

10.6.4 Avogadro's Laws

Based on experiments with various gases, Avogadro came to the following law: Equal volumes of ideal gases, at the same temperature and pressure, contain the same number of molecules (atoms).

One mole of any gas at the same temperature and pressure occupies the same volume V_m (molar volume). In a mole of any gas, there is the same number of molecules (atoms). This number is called Avogadro's number N_a.

10.6.5 Dalton's Laws

Dalton's law can be stated as follows: In a mixture of gases that do not chemically react with each other, each gas exerts its own pressure independently, as if the other gases were not present.

Assuming there is a gas in a container (vessel) with pressure p_1, and if a second gas is added which, if it were alone in the same container, would have pressure p_2, the total pressure in the container will be

$$p = p_1 + p_2, \tag{10.16}$$

or, in the case of multiple gases:

$$p = p_1 + p_2 + p_3 + \cdots + p_n. \tag{10.17}$$

10.7 Ideal Gas Law

Let a certain mass of gas at temperature T_1 and pressure p_1 occupy a volume V_1. This mass of gas, in some other arbitrary state, is characterized by the parameters p_2, V_2, T_2 (Fig. 10.9). Let the transition from state 1 to state 2 take place in the form of two processes: isothermal (isotherm $1-1'$) and isochoric (isochore $1'-2$).

In accordance with Boyle–Mariotte's law (10.9) and Charles's law (10.12), it can be written as

$$p_1V_1 = p_1'V_2 \tag{10.18}$$

and

$$\frac{p_1'}{p_2} = \frac{T_1}{T_2}. \tag{10.19}$$

By eliminating p_1' from expressions (10.18) and (10.19), we get

$$\frac{p_1V_1}{T_1} = \frac{p_2V_2}{T_2}. \tag{10.20}$$

Since states 1 and 2 were chosen arbitrarily, it can be concluded that for a given mass of gas, the value of pV/T remains constant, i.e.

$$\frac{pV}{T} = B = const. \tag{10.21}$$

The constant B depends on the type of gas and is called the gas constant.

According to Avogadro's law, at the same pressure and temperature, one mole of any gas occupies the same volume V_m (molar volume), so the constant B is the same for all gases. Now, the expression (10.21) can be written for one mole as:

$$pV_m = RT, \tag{10.22}$$

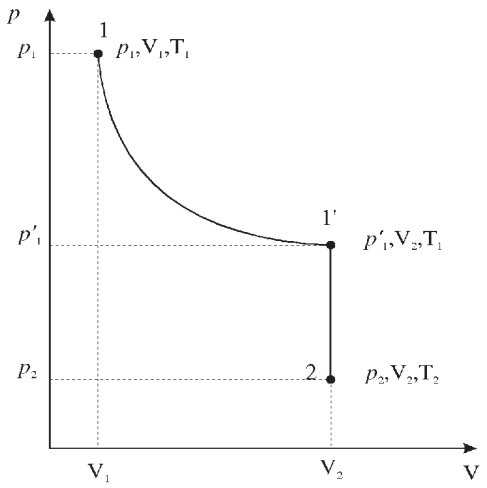

Figure 10.9: An isothermal process and an isochoric process

where R is called *the universal gas constant.*

If at a given pressure p and temperature T one mole of gas, which has a mass M (the so-called molar mass), occupies a volume V_m, then under the same conditions n moles of gas with mass m will occupy a volume V. Obviously, a proportion can be written: $V/m = V_m/M$, which gives $V_m = VM/m$. When V_m is substituted into equation (10.22):

$$pV = \frac{m}{M}RT. \tag{10.23}$$

It is clear that the number of moles of a gas n is equal the ratio of the mass of the gas m to its molar mass M ($n = m/M$), giving:

$$pV = nRT. \tag{10.24}$$

Expressions (10.23) and (10.24) are known as *the Equation of State for an Ideal Gas* or *the Ideal Gas Law.*

10.8 Quantity of Heat. Specific Heat Capacity

10.8.1 Heat as a Form of Energy

When considering the Law of Conservation of Mechanical Energy, it is emphasized that the total mechanical energy in a closed system remains constant (there are no external forces, such as friction, for example). That is, a part of kinetic energy can be converted into potential energy and vice versa, but their sum remains unchanged at any given moment. Here, this law will be applied to thermal phenomena. For this purpose, a relatively simple example will be considered. Suppose a body, given a certain initial velocity, slides over a surface. Its velocity (speed) will gradually decrease due to the friction between the body and the surface. At the same time, it is observed that both the body and the surface heat up. If the body moves over a horizontal surface, its gravitational potential energy does not change, but there is a decrease in kinetic energy, as a result of heating.

It is said that the lost part of the kinetic energy of a body is transferred to the body and the surroundings, which is accompanied by a change (increase) in their temperature. Therefore, it can be concluded that heat is just one form of energy. By increasing the temperature, the internal energy of the body increases, which represents the sum of the kinetic and potential energy of the molecules in a system in thermodynamic equilibrium. The temperature of a body, therefore, can increase at the expense of a decrease in the body's kinetic energy. In this process, the increase in thermal energy is equal to the loss of kinetic and/or potential energy.

Body temperature, and consequently internal energy, can also increase at the expense of cooling another body. For example, the air temperature in a closed room rises at the expense of the stove's temperature decreasing. In this process, thermal energy is transferred from one body to another, without the movement of the bodies themselves.

10.8.2 Specific Heat Capacity. Calorimetry

The heat capacity of a body C_t is defined as the ratio of the amount of heat Q received to the increase in temperature Δt

$$C_t = \frac{Q}{\Delta t}. \tag{10.25}$$

The heat capacity depends on the mass of the body and the type of material, as well as on the temperature of the body ($C_t = f(t)$). The C_t corresponds to the amount of heat a body receives to warm up by Δt, and represents the amount of heat per one degree.

In order to obtain a quantity that characterizes a specific material and does not depend on the mass of the body, the specific heat capacity c (specific heat) is defined. It represents the ratio of the heat capacity to the mass of the body:

$$c = \frac{C_t}{m} = \frac{Q}{m\Delta t}. \tag{10.26}$$

Since the heat capacity is a function of temperature, the specific heat capacity is also function of temperature ($c_t = f(t)$), and the expression (10.26) should be presented in differential form

$$c = \frac{dQ}{mdt}. \tag{10.27}$$

The c is the amount of heat required to raise the temperature of 1 kg of a material by 1℃ (or 1 K).

From the above expression, the amount of heat supplied to or removed from a body can be calculated

$$Q = m\int_{t_1}^{t_2} cdt, \tag{10.28}$$

where c is a function of temperature.

The specific heat capacity is often defined per one mole, the so-called molar specific heat capacity C. It is the amount of heat required to raise the temperature of one mole of a material by 1℃ or 1 K. It can be expressed as

$$C = cM, \tag{10.29}$$

where M is the molar mass of the material (the mass of one mole).

10.9 Heat transfer

Heat can be transferred in three ways: *conduction*, *convection*, and *radiation*. In practice, this transfer is often carried out by a combination of these three methods, depending on the properties of the system through which the heat is transferred.

10.9.1 Conduction

If two parts of a solid body are at different temperatures, heat is then transferred through that body by conduction from the place of higher temperature to the place of lower temperature. This form of heat transfer is mainly determined by atoms, which transmit their vibrations to neighboring atoms. In metals, heat transfer is carried out by free electrons.

The amount of heat transferred by conduction is given by the so-called Fourier's law

$$\frac{dQ}{d\tau} = -\lambda S\frac{dt}{dx}, \tag{10.30}$$

where dQ is the amount of heat transferred during time $d\tau$ through the surface area S that is the normal to the direction of heat propagation, $dQ/d\tau$ is the instantaneous rate of heat transfer, λ is the thermal conductivity coefficient, and dt/dx is the temperature gradient. The sign (–) indicates that heat is transferred in the direction of decreasing temperature.

In the steady-state (stationary, time-independent) regime of heat conduction, the temperature is constant at every point, so neither the gradient dt/dx nor the heat transfer rate depends on time, which gives $dt/dx = \Delta t/\Delta x$ and $dQ/d\tau = Q/\tau$.

If heat conduction through a plate of thickness d is observed in a steady-state regime (Fig. 10.10), expression (10.30) can be written in the form

$$\frac{Q}{\tau} = \lambda S\frac{\Delta t}{d}, \tag{10.31}$$

where $\Delta t = t_2 - t_1$ is the temperature difference on the opposite surfaces S of the wall of the plate.

According to the value of the thermal conductivity coefficient, materials can be divided into thermal conductors and thermal insulators. As with the division according to electrical conductivity, there is no sharp boundary between these two groups. Although, as a rule, materials that are good electrical conductors also conduct heat well, there are exceptions. Thus, for example, aluminum oxide Al_2O_3, a good electrical insulator, conducts heat relatively well.

10.9.2 Convection

Heat transfer by convection is characteristic of fluids. This can be observed in the example of a solid body surrounded by a fluid. When the temperature of the body is different from the temperature of the fluid, heat exchange occurs between them. If the temperature of the body is higher than the temperature of the surrounding fluid, the fluid heats up. This leads to a decrease in the density of the fluid, and it rises. In the opposite case, the fluid cools down, and due

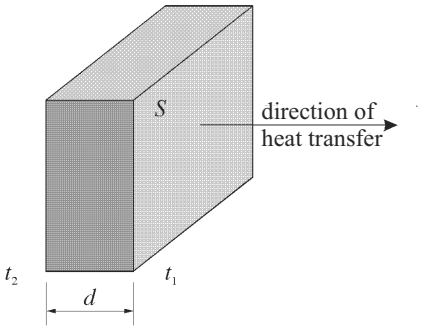

Figure 10.10: Heat transfer by conduction

to an increase in its density, it flows downward near the body. In both cases, it is the heat propagation thanks to the movement of the fluid, or *convection*.

The amount of heat transferred by convection between a solid body and the surrounding fluid is given by Newton's law:

$$q_c = \frac{dQ}{d\tau} = hS(t_t - t_f), \tag{10.32}$$

where q_c is the heat transfer rate, h is the heat transfer coefficient from the surface of the body to the fluid, S is the surface area of the body, t_t is the temperature of the body surface, and t_f is the temperature of the fluid mass to which the heat is transferred. The value of the heat transfer coefficient largely depends on the body shape, temperature, etc., and is usually determined experimentally.

10.9.3 Radiation

Heat transfer by radiation occurs through electromagnetic waves emitted by heated bodies. This means that heat (thermal waves) is part of the electromagnetic spectrum. More precisely, thermal radiation falls within the infrared radiation. The wavelength of emitted radiation depends on the temperature of the body. *As the temperature of the body increases, the wavelength of the emitted radiation decreases.*

A heated body emits (radiates) energy, which another body absorbs and converts into heat. Heat transfer from a body to the surroundings by conduction and convection depends on the temperature difference between the body and its environment. Radiation of a body depends only on its temperature. This means that above a certain temperature, radiation becomes the dominant process compared to conduction and convection. The dependence of emitted thermal energy on the temperature of the body is given by the Stefan–Boltzmann law:

$$E = e\sigma T^4, \tag{10.33}$$

where E is the emitted power, i.e., the thermal energy emitted from a unit area of a body per unit of time, σ is the proportionality constant (Stefan-Boltzmann constant), and e is a constant known as the emissivity of a body, which, depending on the structure of its surface, can have values between 0 and 1. The emissivity of all bodies is generally less than one, and it is lower the more polished the surface of the body is. An absolutely black body should have an emissivity equal to one.

Chapter 11

Fundamentals of Thermodynamics

11.1 Basic Characteristics

Thermodynamics is a branch of physics that studies phenomena related to the conversion of thermal energy into other forms of energy (mechanical, electrical, chemical, etc.), and vice versa. Thermodynamics also studies the conditions under which thermal energy is converted into mechanical work.

The conversion of one form of energy into another is always associated with some thermodynamic system. One of the simplest thermodynamic systems is an ideal gas in a cylinder with a movable piston. A state of gas system in which pressure and temperature are the same at all points in the gas is **equilibrium state**, and the ideal gas equation of state can be applied.

In thermodynamics, the properties of macroscopic systems are studied without taking their microstructure into account. Therefore, *macroscopic quantities* (pressure, volume, temperature, internal energy, etc.) are used, which determine the state of a thermodynamic system.

A thermodynamic process is understood as a change in the state of a system. Cyclic processes, in which the system returns to its initial state after a series of changes (a series of of states), play a special role. Heat engines precisely repeat a specific thermodynamic cyclic process, converting thermal energy into mechanical work.

Thermodynamic processes can be **reversible** and **irreversible**. A process is reversible if the system passes through a series of *equilibrium states*; otherwise, it is said to be irreversible. An ideal reversible process does not exist in nature, and real thermodynamic processes are always more or less irreversible. However, theoretically, it is only possible to consider reversible processes. From the results obtained in this way, certain conclusions about the behavior of real processes can be drawn.

11.2 Internal Energy. Degree of Freedom

Internal energy is an important characteristic of a thermodynamic system. It represents the sum of the energy of chaotic (thermal) motion of the system's microparticles (molecules, atoms, electrons, nuclei, etc.), which represents their kinetic energy, and the potential energy of their mutual interactions, as a result of intermolecular forces.

In an ideal gas, the effect of intermolecular forces is neglected (potential energy of their mutual interactions is neglected), so its internal energy is equal only to the kinetic energy of all the molecules. In any state, a system has a certain internal energy.

If two thermodynamic systems (for example, systems with an ideal gas) are at different temperatures, when they come into contact, the molecules of the higher-temperature system will transfer energy to the molecules of the lower-temperature system through collisions until the temperatures equalize. Heat will, therefore, flow from the hotter body to the cooler body, and in the process, the internal energy of the hotter body will decrease, while that of the cooler one will increase.

The number of degrees of freedom of a system represents the number of independent quantities (coordinates) that determine the position of the system in space. For example, the position of a material point is determined by three coordinates. Based on the above assumption, atoms have three degrees of freedom for translational motion.

Atoms are considered as material points (Fig. 11.1a), where the energy of their rotational motion can be neglected ($r \to 0,\ I = mr^2 \to 0,\ E_k = I\omega^2/2 \to 0$). In classical mechanics, diatomic gas molecules are considered in the first approximation as a system of two material points (two atoms), which are rigidly connected by a bond that cannot be deformed (for example, two spheres connected by a rod). This system, in addition to three degrees of freedom of translational motion, also has two degrees of freedom of rotational motion (Fig. 11.1b). Rotation around the third axis (the axis passing through both atoms) is unlikely due to the small moment of inertia of the molecule. Triatomic molecules have six degrees of freedom—three due to translational motion and three due to rotational motion (Fig. 11.1c).

It is natural that strong bonds between atoms do not exist. Therefore, for real molecules, it is necessary to also take into account the degrees of freedom of oscillatory motion (atoms in a molecule oscillate around their equilibrium

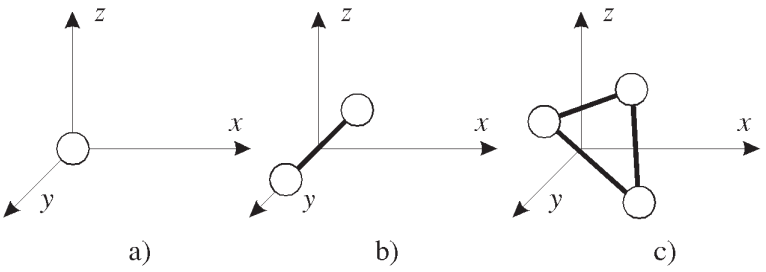

Figure 11.1: With an explanation of the number of degrees of freedom

positions). This means that in this case, intermolecular forces, as well as the potential energy of interactions between molecules, cannot be neglected (it will be explained in more detail in the section 11.5).

Regardless of the total number of degrees of freedom of a molecule, three degrees of freedom are always associated with translational motion. These degrees of freedom are equivalent, and each of them has energy $<E_1>$ equal to 1/3 of the average translational energy $<E>$ $(= 3kT/2)$ of the molecules:

$$<E_1>= \frac{<E>}{3} = \frac{3kT/2}{3} = \frac{kT}{2}. \tag{11.1}$$

From Boltzmann's law of the uniform distribution of energy per degree of freedom, which is given in classical statistical physics, it follows that for statistical systems in thermodynamic equilibrium, each degree of translational and rotational freedom on average corresponds to kinetic energy of $kT/2$, and each degree of oscillatory motion corresponds to kT. The degree of freedom of oscillatory motion has twice the energy because, in addition to kinetic energy (as in the case of translational and rotational motion), there is also potential energy between molecules, and the average value of kinetic and potential energy being equal. Therefore, the average energy of molecules is

$$<E>= \frac{i}{2}kT, \tag{11.2}$$

where i is the total number of degrees of freedom of translational (i_T), rotational (i_R), and double degrees of freedom of molecular oscillatory motion $2i_O$:

$$i = i_T + i_R + 2i_O. \tag{11.3}$$

The internal energy of one mole of gas U_m is equal to the sum of the average energies of N_a molecules (N_a is Avogadro's number):

$$U_m =<E> N_a = \frac{i}{2}kTN_a = \frac{i}{2}RT. \tag{11.4}$$

For n moles of gas, the internal energy is

$$U = nU_m = n\frac{i}{2}RT = \frac{m}{M}\frac{i}{2}RT. \tag{11.5}$$

11.3 Work of Gas During a Change in its Volume

The simplest thermodynamic system will be observed, an ideal gas in a cylinder with a movable piston (Fig. 11.2). It is always assumed that the piston moves along the cylinder without friction. If an amount of heat Q is supplied to the gas, it will be used to increase its internal energy and to increase its volume (gas performs work).

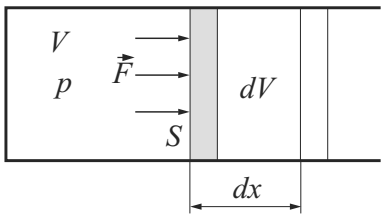

Figure 11.2: A cylinder with a movable piston containing an ideal gas

If the piston is fixed, the supplied amount of heat will be used only to increase the internal energy of the gas ΔU. However, if the piston is not fixed, supplying heat causes the gas to expand, the piston moves, and work is performed. If, during the expansion of the gas, the piston moves a small distance dx, the infinitesimally small work done is

$$dA = Fdx = pSdx = pdV, \tag{11.6}$$

where S is the surface area of the piston, and $Sdx=dV$ is the infinitesimally small volume of the gas. Total work done during the change in the gas volume from V_1 to V_2 can be found by integrating the above expression

$$A = \int_{V_1}^{V_2} pdV. \tag{11.7}$$

The work done during a certain process can be represented graphically in a $p-V$ diagram. For example, the change in gas pressure during its expansion is shown by a curve in Figure 11.3. When the volume increases by dV, the work done is equal to pdV, i.e., it is determined by the shaded area in the figure. The total work done by the gas during expansion from volume V_1 to volume V_2 is determined by the area bounded by the horizontal axis, the curve $p=f(V)$, and the lines V_1 and V_2.

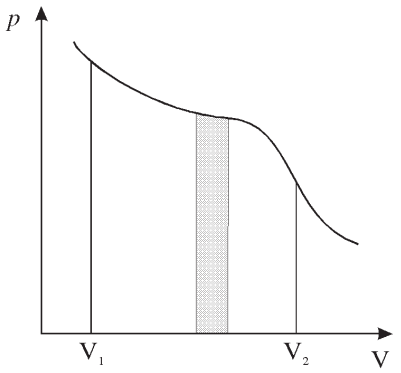

Figure 11.3: Change in gas pressure during its expansion

Graphically, only equilibrium processes can be represented, i.e., processes in which the change of the gas system's state occurs slowly enough. All real processes are non-equilibrium, but in a large number of cases this non-equilibrium can be neglected (the slower the process occurs, the closer it is to equilibrium). Furthermore, only *equilibrium processes* will be considered.

11.4 The First Law of Thermodynamics

When an amount of heat (thermal energy) Q is supplied to a system (gas), a part of it is used to increase its internal energy ΔU, and the rest is used for the work A that the system performs on the surrounding bodies. Therefore, according to the law of conservation of energy, it follows

$$Q = \Delta U + A. \tag{11.8}$$

Expression (11.8) represents the mathematical interpretation of the first law of thermodynamics. For small changes in the state of the system, this expression takes the form

$$dQ = dU + dA. \tag{11.9}$$

11.5 Specific Heat Capacities

As has already been said, the specific heat capacity of a material represents the amount of heat needed to raise the temperature of $1\,kg$ of the material by one Kelvin or one degree Celsius:

$$c = \frac{dQ}{mdT}. \tag{11.10}$$

The molar specific heat capacity of a material represents the amount of heat required to raise the temperature of one mole of the material by one degree:

$$C_M = \frac{dQ}{ndT}, \tag{11.11}$$

where $n=m/M$ is the number of moles.

The relationship between C_M and c is

$$C_M = c\,M, \tag{11.12}$$

where M represents the molar mass of the gas (the mass of one mole of the gas).

There is the specific heat capacity of a gas at constant volume (c_V) when, during the heating of a gas, its volume is kept constant ($V = const.$) and the specific heat capacity of a gas at constant pressure (c_p), when its pressure is kept constant ($p = const.$), during its heating. There is also the molar specific heat capacity of a gas at constant volume ($C_V = c_V M$) and the molar specific heat capacity of a gas at constant pressure ($C_p = c_p M$).

Using equations (11.9) and (11.11), the first law of thermodynamics (11.9) for one mole of gas takes the form

$$C_M dT = dU_m + pdV_m. \tag{11.13}$$

If one mole of a gas is heated at constant volume, then $C_M = C_V$ and the work done by external forces is zero ($pdV_m = 0$), so the heat added to the gas is used solely to increase its internal energy

$$C_V = \frac{dU_m}{dT}, \tag{11.14}$$

i.e., the C_V is equal to the change (increase) in the internal energy of one mole of the gas when its temperature increases by one degree. Namely, the eq. (11.14) represents the internal energy per degree.

According to equation (11.4): $U_m = iRT/2 \Rightarrow dU_m = iRdT/2$, the following holds

$$C_V = \frac{i}{2}R. \tag{11.15}$$

If one mole of the gas is heated at constant pressure ($p = const. \Rightarrow C_M = C_p$), expression (11.13) can be written in the form

$$C_p = \frac{dU_m}{dT} + \frac{pdV_m}{dT} = C_V + \frac{RdT}{dT} = C_V + R = \frac{i+2}{2}R, \tag{11.16}$$

where for pdV_m, the the Equation of State for an Ideal Gas (10.24) is used: $pV = nRT \Rightarrow pV_m = RT \Rightarrow pdV_m = RdT$.

When considering thermodynamic processes, it is important to know the relationship of specific heat capacities for each gas:

$$\kappa = \frac{C_p}{C_V} = \frac{i+2}{i}. \tag{11.17}$$

From expressions (11.15) and (11.16), it follows that the molar specific heat capacities depend only on the number of degrees of freedom i, and not on the temperature T. Such behavior has been established by molecular-kinetic theory for *monatomic gases* over a wide temperature range.

A diatomic molecule has three translational degrees of freedom, two rotational degrees of freedom, and one degree of freedom for vibrational motion ($i = i_T + i_R + 2i_O = 3+2+2\cdot 1 = 7$). According to the law of energy distribution among degrees of freedom (eq. (11.2)), for a diatomic gas at room temperature, it should hold that: $C_V = 7R/2$. However, for diatomic gases, experiments have shown that the number of degrees of freedom i depends on the temperature.

The figure 11.4 shows the experimental results of the dependence of the molar specific heat capacity C_V on temperature T for the hydrogen molecule (C_V is an easily measurable physical quantity). From the figure, it can be seen that C_V depends on temperature: at low temperatures (up to $50\,K$), $C_V = 3R/2$ ($i = i_T = 3$); at room temperature, $C_V = 5R/2$ (instead of the calculated $7R/2$; $i = i_T + i_R = 5$); and at very high temperatures, $C_V = 7R/2$. This can be explained by assuming that at low temperatures only translational motion exists, at room temperature translational and rotational motion, and at high temperatures also vibrational motion of the molecules. When studying the specific

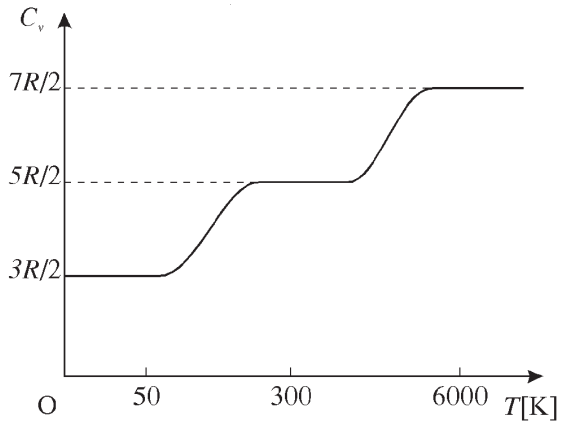

Figure 11.4: Specific heat capacity C_V for hydrogen molecule as a function of temperature

heat capacity, the quantized energy of the rotational and vibrational motion of molecules must be considered (*not any value* of rotational and vibrational energy is possible, but only certain discrete sets of their values). If the energy of thermal motion is insufficient, for example, to excite vibrations, then these vibrations do not affect the specific heat capacity (that degree of freedom is neglected in the law of equal energy distribution). This explains why the specific heat capacity of a diatomic gas (hydrogen molecules) at room temperature is $5R/2$ instead of $7R/2$. Similarly, the decrease in specific heat capacity at low temperatures (the "freezing" of the rotational degree of freedom) and the increase at high temperatures (the "excitation" of the vibrational degree of freedom) can be explained.

11.6 Work During a Change of Gas State

Using the first law of thermodynamics, the work can be calculated for some processes in which the state of an ideal gas changes.

a) **Isochoric process** ($V = const.$). The diagram of isochoric process in the $p - V$ diagram represents a line parallel to the ordinate (Fig. 11.5). Here, $1 - 2$ represents the process of isochoric heating, and $1 - 3$ of isochoric cooling. In an isochoric process, no work is done, because $dV = 0$, so $dA = pdV = 0$. The first law of thermodynamics for this case can be written in the form:

$$dQ = dU. \tag{11.18}$$

Thus, in an isochoric process, the supplied amount of heat is used to increase the internal energy of the system.

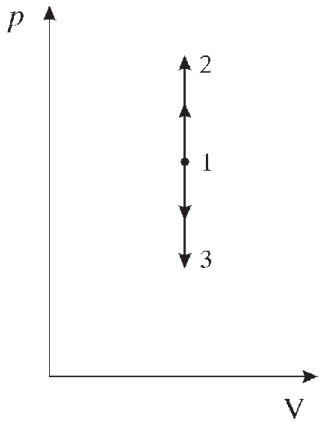

Figure 11.5: Isochoric process

Isobaric process ($p = const.$). The diagram of this process (isobar) in a $p - V$ diagram is represented by a line $1 - 2$, parallel to the abscissa axis (Fig. 11.6). The work of the gas (eq. (11.7)) during its isobaric expansion from volume V_1 to V_2 is given by the expression

$$A = \int_{V_1}^{V_2} pdV = p(V_2 - V_1), \tag{11.19}$$

and is proportional to the area of the shaded rectangle in Figure 11.6.

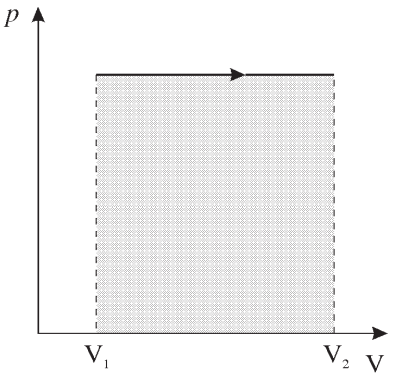

Figure 11.6: Isobaric process

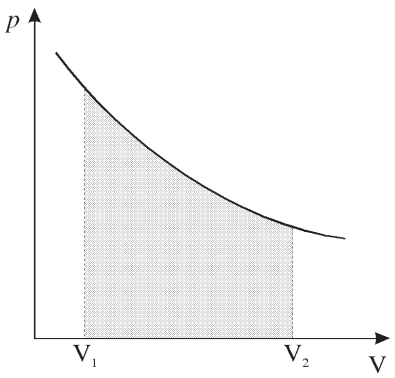

Figure 11.7: Isothermal process

Isothermal process ($T = const.$). The diagram of this process (isotherm) in the $p - V$ diagram, as previously noted, represents a hyperbola (Fig. 11.7). Thus, the gas behaves according to Boyle's law. By using expressions (11.7) and (10.23), an expression for the work done by a gas during an isothermal process can be obtained:

$$A = \int_{V_1}^{V_2} pdV = \int_{V_1}^{V_2} \frac{m}{M} RT \frac{dV}{V} = \frac{m}{M} RT \ln \frac{V_2}{V_1} = \frac{m}{M} RT \ln \frac{p_1}{p_2}, \tag{11.20}$$

which is equal to the shaded area under the isotherm in the $p - V$ diagram.

11.7 Adiabatic process

An adiabatic process refers to a change in the state of a system (gas) without any heat exchange with its surroundings. Such processes do not exist in nature, but processes that occur rapidly approximate them in their characteristics.

From the first law of thermodynamics (11.9), it follows for an adiabatic process

$$dA = -dU \;\Rightarrow\; pdV = -dU, \tag{11.21}$$

i.e., a work is done at the expense of the system's internal energy. Accordingly, an adiabatic process is the opposite of an isothermal one, in which work is performed at the expense of heat supplied from outside.

Using equations (11.5) and (11.15), for an arbitrary mass of gas:

$$U = \frac{m}{M}\frac{i}{2}RT \;\Rightarrow\; dU = \frac{m}{M}\frac{i}{2}RdT = \frac{m}{M}C_V dT. \tag{11.22}$$

Using eqs. (11.21) and (11.22):

$$pdV = -\frac{m}{M}C_V dT. \tag{11.23}$$

By differentiating the equation of state for an ideal gas: $pV = mRT/M$ with respect to temperature T, one obtains

$$pdV + Vdp = \frac{m}{M}RdT. \tag{11.24}$$

Dividing equation (11.24) by equation (11.23) gives

$$\frac{pdV + Vdp}{pdV} = \frac{\frac{m}{M}RdT}{-\frac{m}{M}C_V dT} = -\frac{R}{C_V} = -\frac{C_p - C_V}{C_V}. \tag{11.25}$$

where eq. (11.16) is used: $C_p = C_V + R$.

By separating the variables and taking $\kappa = C_p/C_V$:

$$\frac{dp}{p} = -\kappa\frac{dV}{V}. \tag{11.26}$$

By integrating this equation from p_1 to p_2, which corresponds to volumes V_1 and V_2, we arrive at the final expression

$$\frac{p_2}{p_1} = \left(\frac{V_1}{V_2}\right)^{\kappa},$$

i.e.

$$p_1 V_1^{\kappa} = p_2 V_2^{\kappa}. \tag{11.27}$$

Since states 1 and 2 were chosen arbitrarily, in the general case the above expression can be written in the form

$$pV^{\kappa} = const.. \tag{11.28}$$

To switch to the variables T and V or T and p, the equation of state for an ideal gas (10.23) can also be used, whereby the pressure or volume is eliminated. This way, two more important expressions can be obtained:

$$TV^{\kappa-1} = const. \tag{11.29}$$

$$T^{\kappa}p^{1-\kappa} = const. \tag{11.30}$$

Expressions (11.28) to (11.30) represent the equations of an adiabatic process. In these equations, the dimensionless quantity (11.17)

$$\kappa = \frac{C_p}{C_V} = \frac{c_p}{c_V} = \frac{i+2}{i} \tag{11.31}$$

is called the Poisson coefficient. For monoatomic gases (H, He, etc.) $i = 3$, so $\kappa = 1,67$, for diatomic gases (H_2, N_2, O_2, etc.), up to temperatures of several thousand Kelvin (Fig. 11.4), $i = 5$ ($\kappa = 1,4$). The values for κ calculated using expression (11.31) agree well with experimental results.

In Figure 11.8, a comparison between an isotherm ($pV = const.$) and an adiabatic curve ($pV^{\kappa} = const.$) is shown. Since $\kappa > 1$, the adiabatic curve is steeper than the isotherm, where $\kappa = 1$. This is explained by the fact that during adiabatic compression, the gas pressure increases not only due to the reduction in its volume (as in an isothermal process) but also due to the rise in temperature.

The work in an adiabatic process is (no need to know by heart)

$$A = -\frac{nR}{\kappa - 1}(T_2 - T_1), \tag{11.32}$$

where T_1 and T_2 are the temperatures corresponding to the initial and final volumes V_1 and V_2.

The work done during adiabatic expansion (equal to the shaded area in Figure (11.8)) is smaller than that of isothermal expansion. This is explained by the fact that during adiabatic expansion the gas cools, while in isothermal expansion the temperature is maintained constant by the heat supplied from outside.

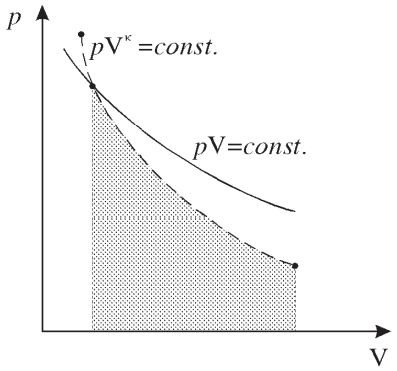

Figure 11.8: Isotherm and adiabatic processes

11.8 Cyclic Processes

11.8.1 Reversible and Irreversible Processes

As highlighted at the beginning of this chapter, thermodynamic processes can be reversible (returnable) and irreversible (non-returnable). A process is reversible if the system passes through a series of equilibrium states. In that case, the process can also occur in the opposite direction, returning through the same equilibrium states to the initial one. Each of these equilibrium states can be represented by a point in the $p-V$ diagram, so the entire process will be represented by a curve $p=f(V,T)$. This can be illustrated with a simple example. If the temperature of an ideal gas is increased by a small amount dT at constant volume V, its pressure will increase by dp, and the gas will transition from one state of thermodynamic equilibrium to another. Before the process, all points of the system had identical thermodynamic coordinates (p, V , T), and after the process, all points of the system again have identical thermodynamic coordinates ($p+dp$, V, $T+dT$). The gas can be returned to its original state simply by reducing the temperature by dT, which will lower the pressure by dp.

A reversible process must proceed slowly enough so that the system is in equilibrium at every point. Such a process takes the system from the initial state, through a series of equilibrium states, to the final state. For example, if the volume of a gas is reduced through a series of infinitesimally small pressure changes, allowing equilibrium to be established after each change, this constitutes a reversible process because it can be reversed through the same equilibrium states while maintaining defined thermodynamic coordinates at every moment.

If the compression occurs suddenly, the gas will go from the initial state to the final state through a series of non-equilibrium states, meaning the process will be irreversible. Such a process cannot be represented by a curve on a $p-V$ diagram, nor can it be reversed so that the system and the surroundings return to the same initial state.

11.8.2 Carnot Cycle Process

In the steam engines it has been observed that only a portion of the thermal energy is converted into mechanical work. The ratio of the utilized to supplied energy, η, is called *the efficiency* (degree of utilization).

Carnot, while studying the operation of heat engines, theoretically analyzed a cyclic process in which a system with an ideal gas returns to its initial state after two isothermal and two adiabatic processes, thereby partially converting the supplied amount of heat into mechanical work.

Figure 11.9 shows the Carnot cycle in a $p-V$ diagram. Isothermal expansion and compression are represented by

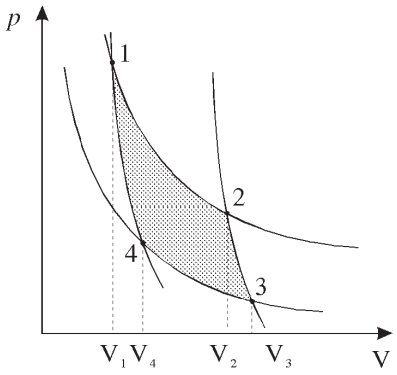

Figure 11.9: Carnot cycle

curves $1-2$ and $3-4$, and adiabatic expansion and compression by curves $2-3$ and $4-1$, respectively. In an isothermal process ($T=const.$), the internal energy does not change ($U=const.$), and the amount of heat Q_1 received by the gas

is equal to the work it performs when transitioning from state 1 to state 2:

$$A_{12} = \frac{m}{M} RT_1 \ln \frac{V_2}{V_1} = Q_1. \tag{11.33}$$

During the adiabatic expansion $2-3$, there is no heat exchange with the surroundings, so the work done during the expansion of gas A_{23} is performed at the expense of the change in internal energy, i.e.

$$A_{23} = -\frac{m}{M} C_V (T_2 - T_1). \tag{11.34}$$

The Q_2 heat released during isothermal compression is equal to the work done on the system

$$A_{34} = \frac{m}{M} RT_2 \ln \frac{V_4}{V_3} = -Q_2. \tag{11.35}$$

In this case, work and heat are negative quantities. The work during adiabatic compression is

$$A_{41} = -\frac{m}{M} C_V (T_1 - T_2) = -A_{23}. \tag{11.36}$$

The total work performed in a cyclic process

$$A = A_{12} + A_{23} + A_{34} + A_{41} = Q_1 + A_{23} - Q_2 - A_{23} = Q_1 - Q_2 \tag{11.37}$$

is equal to the shaded area in Figure 11.9.

The efficiency is:

$$\eta = \frac{A}{Q_1} = \frac{Q_1 - Q_2}{Q_1}. \tag{11.38}$$

11.9 The Second Law of Thermodynamics

The first law of thermodynamics provides a quantitative relationship between the amount of heat, work, and the change in the internal energy of a system, but it does not determine the direction of processes in nature. According to this principle, the transfer of energy as heat from a body at higher temperature to a body at lower temperature, and vice versa, is possible.

The direction of processes related to the change of energy in nature is determined by the second law of thermodynamics. This principle arises from the analysis of the work of heat engines (Fig. 11.10). Heat engines operate by taking an amount of heat Q_1 from a higher temperature reservoir T_1 (the heater) during the cycle and delivering an amount of heat Q_2 to a lower temperature reservoir T_2 (the cooler). The work performed during the cycle is $A = Q_1 - Q_2$.

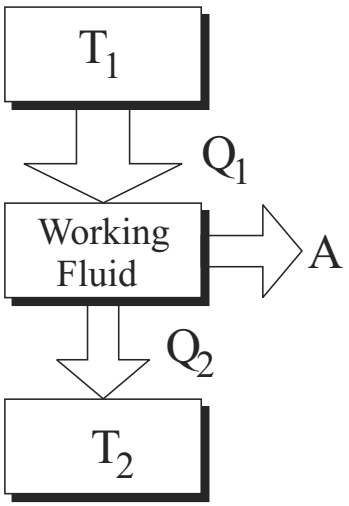

Figure 11.10: Heat engine

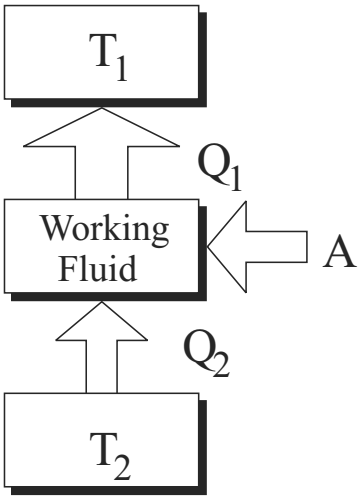

Figure 11.11: Cooling machine

For the efficiency of a heat engine to be equal to one, it is necessary that $Q_2 = 0$, i.e., that the heat engine has only one heat reservoir. Carnot showed that this is not possible and that the operation of a heat engine requires at least two heat reservoirs at different temperatures.

The process opposite to the operation of heat engines is used in cooling machines, whose working principle is shown in Figure 11.11. During the cycle, an amount of heat Q_2 is taken from a reservoir at a lower temperature T_2, and an amount of heat Q_1 is delivered to a reservoir at a higher temperature T_1. For a cyclic process, $Q = A$, specifically $Q = A = Q_2 - Q_1 < 0$. From this, $Q_2 - Q_1 = -A$ or $Q_1 = Q_2 + A$, i.e., the amount of heat Q_1 is greater than Q_2 by the amount of work done on the system. Therefore, without performing work, it is not possible to remove heat from a colder body and transfer it to a hotter body.

Chapter 12

Oscillations

12.1 Mechanical Oscillations

12.1.1 Periodic Motion. Harmonic Oscillation

Periodic motion refers to the movement of a body or a material point when it returns to its original position after a certain time interval. Simpler periodic motions include the movement of a body along a path that forms a closed curve of any shape. One example of such motion is the rotation of planets along elliptical orbits.

If periodic motion occurs along a path that is not closed, alternately to one side or the other from the equilibrium position, it is referred to as *oscillatory motion*. An example of simpler oscillatory motion is the movement of a body of mass m, suspended from a spring or thread, around the equilibrium position.

A system that performs oscillatory motion is called *an oscillator*. If the oscillator performs oscillations without the influence of external forces, it is said to undergo free (natural) oscillations. An oscillator can also undergo forced oscillations due to external forces. Every real oscillator always experiences friction, so it undergoes damped (attenuated) oscillations.

Oscillatory motion is characterized by the following physical quantities: the oscillation period T, elongation (displacement) x, and amplitude A. The oscillation period T represents the time it takes for a body to complete one full oscillation. It is the time that passes between two identical physical states of the system. The reciprocal of the oscillation period is called the frequency (ν) and represents the number of oscillations per unit of time. Elongation represents the distance of the oscillator from the equilibrium position at any given moment. Amplitude represents the maximum value of the elongation. Thus, the range of the body's motion is $2A$.

The mathematical interpretation of oscillatory motion is quite complicated. Therefore, further presentations of this issue will be limited to the analysis of a special type of oscillatory motion, which is called harmonic oscillatory motion. This is a type of motion in which the elongation changes over time according to a sine or cosine law, i.e.

$$x = A\sin(\omega t + \phi); \qquad x = A\cos(\omega t + \phi), \tag{12.1}$$

where $\omega = 2\pi\nu = 2\pi/T$ is the angular frequency. The argument of the sine or cosine $(\omega t + \phi)$ represents the phase of the oscillation at time t, and ϕ is the initial phase. The initial phase determines the value of the displacement at the moment of observation ($t = 0$). For sinusoidal oscillations at $t = 0$, $x = A\sin\phi$. If the observation of the oscillatory motion is made at the moment $t = 0$ (when $x = 0$), then $\phi = 0$.

Expressions (12.1) describe harmonic oscillatory motion, which occurs along either a straight or curved trajectory. In that case, to determine the position of the oscillating body, it is sufficient to know the value of the elongation x. If the body oscillates around the equilibrium position along a straight trajectory, it is referred to as simple harmonic motion.

12.1.2 Velocity and Acceleration in Harmonic Oscillation. The Harmonic Oscillation Equation

The oscillator shown in Figure 12.1 consists of an elastic spring from which a body of mass m is suspended. Let the body be displaced from the equilibrium position (point O) by a distance x, with the initial phase $\phi = 0$. Since the body performs harmonic oscillations, as will be shown later in the text, the extension can be expressed in the form

$$x = A\sin\omega t, \tag{12.2}$$

where A are ω are constants. The velocity of the body is the derivative of the elongation with respect to time:

$$v = \frac{dx}{dt} = A\omega\cos\omega t = v_o\cos\omega t = v_o\sin(\omega t + \frac{\pi}{2}), \tag{12.3}$$

where $v_o = A\omega$ is the maximum value of velocity (velocity amplitude).

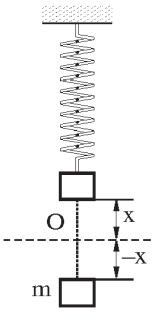

Figure 12.1: Oscillation of a body of mass m suspended on an elastic spring

From expression (12.3), it can be seen that the velocity of the body, as well as the elongation, varies over time according to a sinusoidal law with frequency ω, but the phase difference between these two quantities is $\pi/2$. This means that at the moment when the elongation is zero, the body moves at maximum velocity.

Since the velocity of a body in harmonic oscillation changes continuously over time, this motion is accelerated, and the acceleration changes over time according to a sinusoidal law

$$a = \frac{dv}{dt} = -A\omega^2 \sin \omega t = a_o \sin(\omega t + \pi), \tag{12.4}$$

where $a_o = A\omega^2$ is the maximum value of acceleration (acceleration amplitude).

The phase difference between acceleration and elongation is equal to π, and the phase difference between velocity and acceleration is $\pi/2$. By substituting $A \sin \omega t$ with x in expression (12.4):

$$a = -\omega^2 x, \tag{12.5}$$

i.e., in harmonic oscillation of a body, the acceleration is proportional to the elongation and has the opposite sign. In Figure 12.2, the change of elongation, velocity, and acceleration over time is shown. As can be seen from the figure, a change in phase of 2π corresponds to the oscillation period. For example, if at moment t the phase of oscillation of any of the three mentioned quantities was $\omega t + \phi$, then after time T it will be equal to $\omega(t + T) + \phi$, so the difference of those phases is equal to

$$[\omega(t + T) + \phi] - (\omega t + \phi) = 2\pi.$$

Expression (12.5) can also be written in the following form:

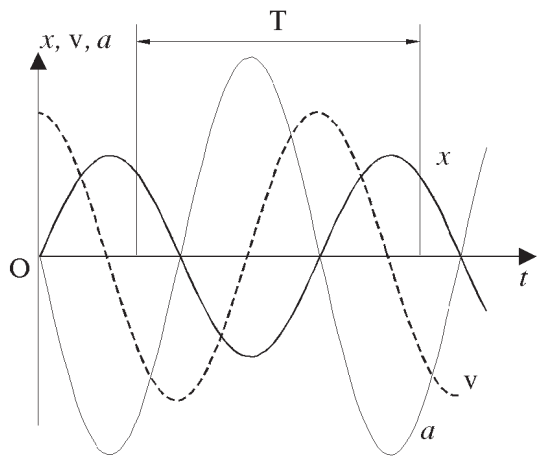

Figure 12.2: Dependence of elongation, velocity, and acceleration on time

$$\frac{d^2x}{dt^2} + \omega^2 x = 0. \tag{12.6}$$

This is a differential equation that determines the time dependence of the distance of a body from the equilibrium position. It can be shown that, when x changes with time according to expressions (12.1), it satisfies equation (12.6).

12.1.3 Graphical Representation of Harmonic Oscillation

Since expressions for harmonic oscillation include sine or cosine functions, there must be a connection between this type of motion and circular motion. That is, if a material point moves along a circular trajectory with uniform velocity v, its projections onto the coordinate axes will undergo harmonic oscillations (Fig. 12.3).

$$x = r \cos \theta, \qquad y = r \sin \theta.$$

Since in uniform circular motion

$$\theta = \frac{v}{r} t = \omega t,$$

that means

$$x = r\cos\omega t, \qquad y = r\sin\omega t.$$

The projection of a material point moving along a circular trajectory with uniform motion performs harmonic oscillation. The angular frequency of this oscillation is given by the angular velocity $\omega = v/r$. A graphical representation of the oscillation over time represents a curve whose period is $T = 2\pi/\omega$ (Fig.12.3).

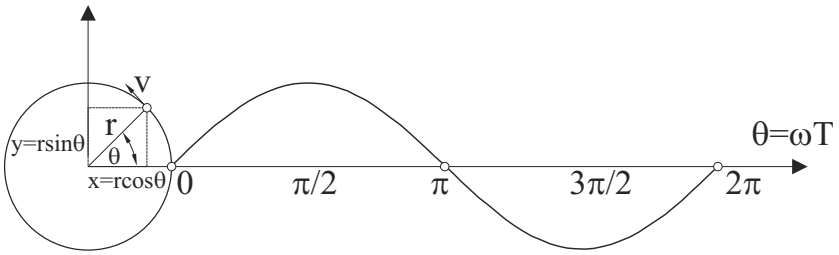

Figure 12.3: Graphical representation of harmonic oscillation

12.1.4 Force in harmonic oscillatory motion. Simpler harmonic oscillators

As previously noted, oscillatory motion belongs to accelerated motions, because it occurs under the influence of some force. If a body of mass m performs harmonic oscillation (Fig. 12.1), then, according to Newton's second law, a force F must act on it:

$$F = ma = -m\omega^2 x = -kx, \tag{12.7}$$

where $k = m\omega^2$ is the coefficient, numerically equal to the force at unit elongation (the so-called directional force). The direction of the force coincides with the direction of acceleration, and the acceleration vector, according to expression (12.5), will be directed toward the equilibrium position. Therefore, for a body to perform simple harmonic oscillation, it must be acted upon by a force that is always directed toward the equilibrium position and whose magnitude is proportional to the elongation. This force is called *the restoring force.*

From expression (12.7), the expression for the oscillation period of the oscillator shown in Fig. 12.1 can be obtained. Since $k = m\omega^2$, and $\omega = 2\pi/T$, the oscillation period is

$$T = 2\pi\sqrt{\frac{m}{k}}. \tag{12.8}$$

So, the period of oscillation depends on the mass of the body and the directional force, which for the oscillator shown in Figure 12.1 is also called the spring stiffness.

Here, several examples of simpler oscillators will be considered.

a) *A body of mass m suspended on an elastic spring*

In this case, the restoring force acting on a body of mass m is represented by the force of an elastic spring caused by its deformation (Fig. 12.1). This force, for small deformations of the spring, is proportional to the elongation of the oscillating body (expression 12.7). If only the restoring force F acts on the oscillator, with k remaining constant, the oscillation will be harmonic.

This can be verified using the system shown in Fig. 12.4. When a body of mass m, on which a pencil is placed, oscillates perpendicular to the direction of the uniform motion of a paper strip, the pencil will draw a sinusoidal line.

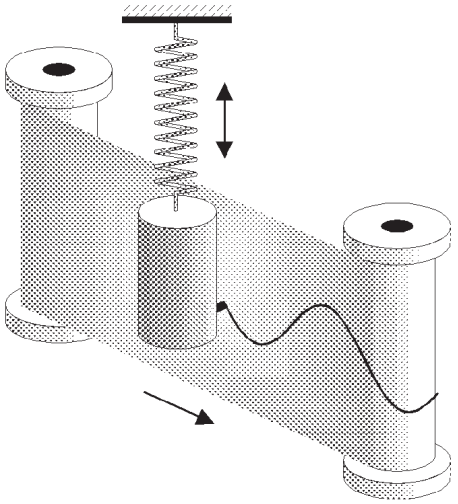

Figure 12.4: Demonstration of simple harmonic motion

b) *A system that performs torsional oscillations (torsion pendulum)*

In torsional oscillation (Fig. 12.5), a restoring torque acts on the body, preventing its deviation from the equilibrium position, and when the body is displaced from the equilibrium position, it provides acceleration toward the equilibrium position. In this case, the restoring torque appears due to the deformation of the wire when the circular plate suspended from it is rotated by a certain angle in the horizontal plane. For a small angle of deviation from the equilibrium position, the restoring torque is proportional to that angle.

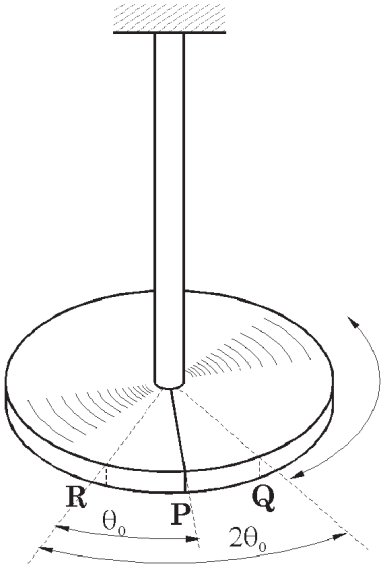

Figure 12.5: Torsion pendulum

If the torsional oscillation is harmonic:

$$\theta = \theta_o \sin \omega t, \tag{12.9}$$

then the angular velocity $d\theta/dt$ and the angular acceleration $d\omega/dt = d^2\theta/dt^2$ also change according to the sinusoidal law:

$$\frac{d\theta}{dt} = \theta_o \omega \cos \omega t; \quad \frac{d^2\theta}{dt^2} = -\theta_o \omega^2 \sin \omega t = -\omega^2 \theta. \tag{12.10}$$

The restoring torque (moment of restoring force) M is equal to the product of the body's moment of inertia I and the angular acceleration

$$M = I \frac{d^2\theta}{dt^2} = -I\omega^2 \theta = -c\theta, \tag{12.11}$$

where $c = I\omega^2$ is a torsion constant and can be determined if the value of the restoring moment M is known, by measuring the angle of torsion:

$$c = \frac{M}{\theta}.$$

The angular frequency and the oscillation period are:

$$\omega = \sqrt{\frac{c}{I}}; \quad T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{I}{c}}. \tag{12.12}$$

c) *Physical pendulum*

A physical pendulum is a rigid body of any shape that can oscillate around a horizontal axis under the influence of gravitational force (Fig. 12.6). This means that the restoring torque represents the torque due to the action of gravitational force and has a direction opposite to the displacement x, i.e.

$$M = -Qx = -mgl_o \sin \theta, \tag{12.13}$$

where l_o is the distance from the pivot point to the center of mass of the body. For a small displacement from the equilibrium position, it can be assumed that $\sin\theta \approx \theta$, so the restoring moment is

$$M = -mgl_o\theta = -c\theta. \tag{12.14}$$

By comparing expression (12.14) with (12.11), it is obtained that $I\omega^2 = mgl_o$, so the angular frequency and oscillation period are given by the expressions

$$\omega = \sqrt{\frac{mgl_o}{I}}; \quad T = 2\pi\sqrt{\frac{I}{mgl_o}}. \tag{12.15}$$

d) *Mathematical pendulum*

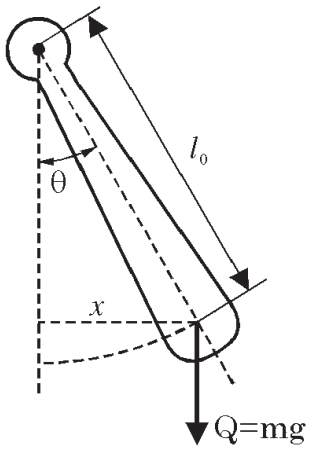

Figure 12.6: Physical pendulum

A mathematical pendulum is a small-bodied object (for example, a small sphere) suspended on a light, inextensible string of length l, which can oscillate around the equilibrium position under the influence of gravitational force (Fig. 12.7). The small sphere moves along an arc s. The restoring force $\vec{F}$ is a component of force $\vec{Q}$ and is directed along the tangent of the path along which the body moves.

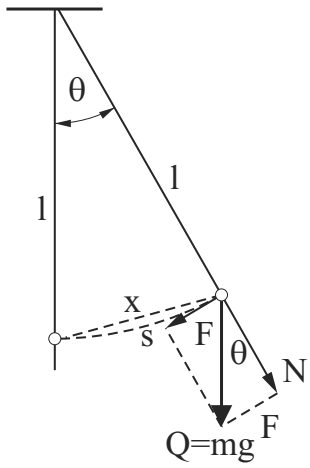

Figure 12.7: Mathematical pendulum

The expression for the force F, based on Fig. 12.7, can be written in the form:

$$F = Q \sin\theta. \tag{12.16}$$

For a small displacement from the equilibrium position ($\theta < 10^o$), it can be assumed that $\sin\theta \approx \theta$, so the force F is

$$F \approx Q\theta \approx mg\theta. \tag{12.17}$$

In circular motion, the following applies

$$s = l\theta. \tag{12.18}$$

For small oscillations, the path s that the ball travels from the equilibrium position to the amplitude position is approximately equal to the elongation x ($s \approx x$):

$$x \approx l\theta. \tag{12.19}$$

Combining equations (12.17) and (12.19) gives:

$$F = \frac{mg}{l}x. \tag{12.20}$$

Now we get an equation

$$F = -kx, \tag{12.21}$$

where $k = mg/l = const.$

The expression (12.21) for the restitution force is obtained, which applies to harmonic oscillatory motion. Since this force is always directed towards the equilibrium position and has a direction opposite to the elongation x, a minus sign is used.

By equating expression (12.20) with (12.7):

$$\frac{mg}{l} = m\omega^2; \quad T = 2\pi\sqrt{\frac{l}{g}}. \tag{12.22}$$

The oscillation period does not depend on the mass of the body that is oscillating, but on the length of the pendulum and the acceleration due to Earth's gravity.

12.1.5 Energy in Harmonic Oscillatory Motion

Oscillatory motion is accompanied by the alternating conversion of an oscillator's kinetic energy into potential energy and vice versa. If the oscillator performs free oscillations, the sum of kinetic and potential energy (total energy) at any distance from the equilibrium position and at any time will remain constant.

The kinetic energy of a body of mass m performing harmonic oscillation is equal to

$$E_k = \frac{mv^2}{2} = \frac{1}{2}m(\frac{dx}{dt})^2. \tag{12.23}$$

Using expression (12.3) for the velocity v, the kinetic energy can be expressed in the form:

$$E_k = \frac{1}{2}m(A\omega\cos\omega t)^2 = \frac{1}{2}mA^2\omega^2(\cos\omega t)^2 = \frac{kA^2}{2}\cos^2\omega t. \tag{12.24}$$

If the oscillation occurs under the action of a spring, the potential energy of a body undergoing harmonic oscillatory motion is obtained from the expression for the potential energy of the spring's elastic deformation:

$$E_p = \frac{kx^2}{2}. \tag{12.25}$$

By using expression (12.2) for elongation, the final expression for the potential energy of a body performing harmonic oscillation is obtained

$$E_p = \frac{kA^2}{2}\sin^2\omega t. \tag{12.26}$$

The total energy E of a body performing harmonic oscillation is equal to the sum of its kinetic and potential energy

$$E = E_k + E_p = \frac{kA^2}{2}(\cos^2\omega t + \sin^2\omega t) = \frac{kA^2}{2} = \frac{mA^2\omega^2}{2} = E_0 = const. \tag{12.27}$$

The dependence of kinetic, potential, and total energy on time is shown in Figure 12.8.

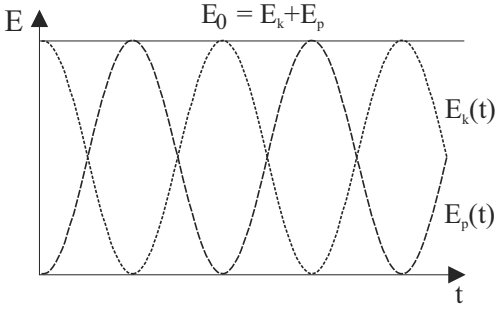

Figure 12.8: Kinetic, potential, and total energy as a function of time

Equation (12.27) shows that the sum of kinetic and potential energy in harmonic oscillation is indeed constant. The significance of equation (12.27) can be seen in Figure 12.9, which shows the oscillator's energy E as a function of elongation x. The total energy of the oscillator is given in the figure as the line $E_o = const.$, and each point on the parabola divides this energy into potential (below) and kinetic (above) the curve components. The oscillation amplitudes are marked as $\pm A$. Harmonic oscillation is energetically analogous to the motion of a frictionless ball in a gravitational field along a parabolic path (a well) of height E_o, where it remains within such a well. Thus, for a body undergoing harmonic oscillation, it can be said that it is within a potential well. The concept of a potential well is frequently used in quantum mechanics.

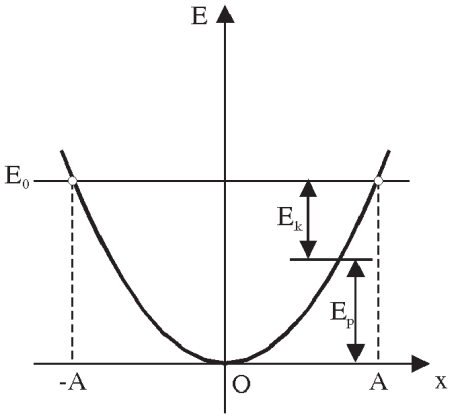

Figure 12.9: Kinetic and potential energy in simple harmonic motion

12.1.6 Damped Oscillations

In the previous consideration of oscillations, the force of friction, which always occurs when an oscillator moves in a medium, was not taken into account. The action of this force significantly changes the nature of the motion, i.e., the oscillation becomes *damped (attenuated)*. Such a situation can occur if a body of mass m, suspended on a spring, is in a liquid (Fig. 12.10). In this case, during oscillation, in addition to the restoring force, the frictional force F_{tr} of the liquid will also act. Then the equation of motion will take the following form:

$$m\frac{d^2x}{dt^2} = -kx + F_{tr}. \tag{12.28}$$

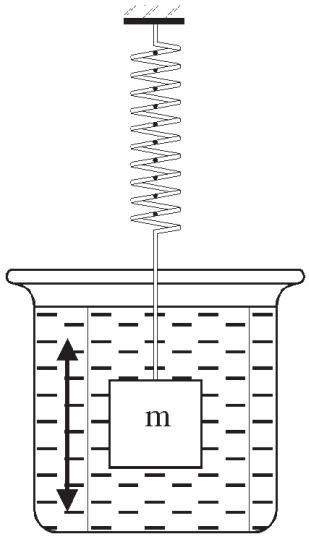

Figure 12.10: Damped oscillations under the action of the fluid's viscosity force

To solve this differential equation, it is necessary to know the parameters on which the friction force depends. It is usually assumed that, at small amplitudes and oscillation frequencies, the friction force F_{tr} is proportional to the velocity of the body's movement and acts in the opposite direction to this velocity:

$$F_{tr} = -\mu v = -\mu\frac{dx}{dt}, \tag{12.29}$$

where μ is the coefficient of friction and depends on the nature of the medium in which the oscillations are carried out. By substituting expression (12.29) into (12.28) and dividing by m:

$$\frac{d^2x}{dt^2} + 2\beta\frac{dx}{dt} + \omega^2 x = 0, \tag{12.30}$$

where $2\beta = \mu/m$ and $\omega^2 = k/m$. Here, β is *the damping coefficient*, and ω is the angular frequency of free (undamped) oscillations. When $\beta = 0$ the oscillator does not transfer energy to the medium during oscillation.

The solution of equation (12.30) depends on the difference:

$$\omega^2 - \beta^2 = \omega_1^2, \tag{12.31}$$

where ω is the angular frequency of free (undamped) oscillations and ω_1 is the angular frequency of damped oscillations.

When $\omega^2 - \beta^2 > 0$, the angular frequency of the damped oscillations ω_1 represents a real quantity, and the solution of equation (12.30), in the case of small damping, has the form:

$$x = A_o \exp\,(-\beta t)\cos\,(\omega_1 t + \phi), \tag{12.32}$$

The amplitude of damped oscillations is

$$A(t) = A_o \exp\,(-\beta t), \tag{12.33}$$

whereA_o is the amplitude at the start of observing the oscillation ($t = 0$).

The graphical representation of function (12.32) is given in Figure 12.11 with a solid line, while the amplitude reduction (expression 12.33) is shown on the same figure with dashed lines. The time interval τ for which the amplitude decreases by a factor of $e(=2.72)$ from its initial value ($A(\tau) = A_o/e$) is called *the relaxation time* ($\tau = 1/\beta$).

The period of damped oscillations depends on the damping coefficient and is determined by the expression:

$$T_1 = \frac{2\pi}{\omega_1} = \frac{2\pi}{\sqrt{\omega^2 - \beta^2}}. \tag{12.34}$$

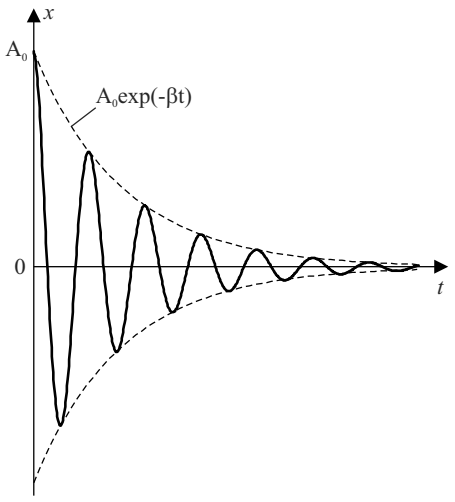

Figure 12.11: Dependence of the elongation and amplitude of damped oscillations on time

From expression (12.34) it can be seen that for small damping ($\beta^2 \ll \omega^2$) the period of damped oscillations is approximately equal to the period of undamped oscillations ($T \approx 2\pi/\omega$). For strong damping ($\beta^2 > \omega^2$), the oscillation period is an imaginary quantity.

If $A(t)$ and $A(t+T)$ are the amplitudes of two consecutive oscillations, separated by a time interval T, then the ratio $A(t)/A(t+T)$ is called the *decrement of damping*, and its natural logarithm is the logarithmic decrement of damping:

$$\lambda = \ln \frac{A(t)}{A(t+T)} = \ln \frac{A_o \exp(-\beta t)}{A_o \exp(-\beta(t+T))} = \ln \frac{\exp(-\beta t)}{\exp(-\beta t)\exp(-\beta T)} = \beta T = \frac{T}{\tau} = \frac{1}{N}. \tag{12.35}$$

In expression (12.35), N represents the number of oscillations that occur until the amplitude decreases by a factor of e from its initial value.

In the case of strong damping ($\beta^2 > \omega^2$), the oscillation period is an imaginary quantity. Such motion is called *aperiodic*.

In Figure 12.12, some forms of oscillatory motion are shown. It should be emphasized that some measuring instruments with a moving coil (analog voltmeters, ammeters, etc.) operate in the critical damping mode. Critical damping refers to such motion of an oscillator where, after being displaced from its equilibrium position, it returns to that same position in the shortest possible time and stops there.

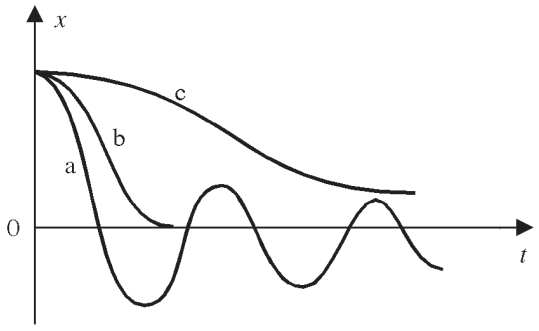

Figure 12.12: Elongation graph as a function of time: a) damped oscillations, b) critical damping, c) aperiodic case

12.1.7 Forced Oscillations

For a real oscillator to perform undamped oscillations, it is necessary for the lost energy to be continuously restored. The restoration of energy in mechanical oscillations is most often carried out by an external force (driving force) that varies according to the following law

$$F = F_o \cos \omega' t, \tag{12.36}$$

where F_o and ω' are the amplitude and the angular frequency of the oscillating driving force. The differential equation of motion of the oscillator for this case takes the form:

$$m\frac{d^2x}{dt^2} = -kx - \mu\frac{dx}{dt} + F_o \cos \omega' t. \tag{12.37}$$

By dividing this equation by m and introducing the substitutions $\mu/m = 2\beta$ and $F_o/m = f_o$, a nonhomogeneous differential equation is obtained

$$\frac{d^2x}{dt^2} + 2\beta\frac{dx}{dt} + \omega^2 x = f_o \cos \omega' t. \tag{12.38}$$

The solution and discussion of this differential equation is beyond the level of this physics course.

Chapter 13

Wave Motion

13.1 The Formation of Mechanical Waves

If a single particle (element) of an elastic medium in any aggregate state is displaced from its equilibrium position, the elastic equilibrium at that location will be disturbed. Over the period of time during which the elastic forces re-establish equilibrium, the impulse of the displaced particle is transferred to its immediate neighbors, and subsequently to all more distant particles of the medium, which are also displaced from their equilibrium positions. In this way, the initially induced impulse can propagate through the specified medium.

If the mentioned particle of the medium is given only a single, short-lived impulse, it is displaced from its equilibrium position, resulting in a local material compression that propagates through the medium. The elastic equilibrium will be re-established and the particles will return to their equilibrium position first at the point of application, and then the equilibrium—i.e., the state of particle rest—will be successively established throughout the entire medium. In this case, all particles of the medium have executed only the motion corresponding to half of one oscillation. The propagation of such an impulse through the medium, followed by the successive re-establishment of the particle equilibrium position, occurs at a specific speed c which depends on the properties of the medium. Such a phenomenon is called a wave pulse. The speed of wave c should not be mistaken for the speed of particle motion v, since $c \gg v$.

If the particles of the elastic medium continuously oscillate under the action of some external periodic force of a sinusoidal form, then the impulse (disturbance) and the areas of compression and rarefaction of the medium continuously propagate through space, also at speed c. Thus, the stationary propagation of a disturbance through an elastic medium is called a wave or wave motion, and the particles of the medium from which the wave originates are called the wave source.

The transfer of impulse can be observed in the following examples. If an oscillatory motion is induced at one end of a taut rope, perpendicular to its direction, it will be transferred along the rope at a specific speed. This phenomenon can also be observed in a row of identical balls arranged in a line (Fig. 13.1). If a new ball of the same mass strikes this row of balls head-on at a specific speed, after a short time the ball at the far end of the row will separate and move at the same speed as the incident ball, while the rest will remain at rest. Thus, the incident ball imparted a specific impulse to the medium (the row of balls), and this impulse propagated through the medium at a specific speed.

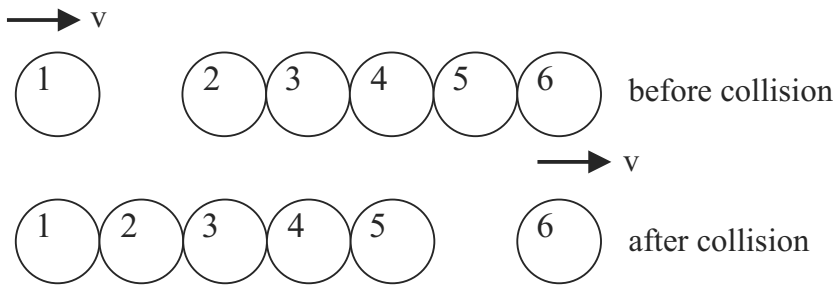

Figure 13.1: Impulse transfer in a row of balls

The mechanism of impulse propagation within an elastic medium can be explained using a simple model. It is known that molecules within a medium are bound by intermolecular forces, which can be represented by a multitude of identical spheres (balls) connected by elastic springs. If one sphere in such a system is displaced from its equilibrium position, the elastic springs connecting it to the other spheres will be compressed or stretched. This will cause the displacement of surrounding spheres, which is then transferred further through the system. The speed c and direction of the impulse propagation do not depend on the speed v of the individual spheres.

Wave motion is closely related to simple harmonic oscillatory motion. When a wave travels through an elastic medium, every particle of that medium oscillates around its equilibrium position, such that the oscillations of a large number of particles must be considered instead of just one. The speed of wave propagation c must be distinguished from the speed of the particles v of that medium. Specifically, during wave propagation, the particles merely oscillate

around the equilibrium position, while the phase of oscillation, or rather the energy from the wave source, is transferred through the medium.

When the propagation of an impulse occurs in only one direction, such a wave is called a *linear wave*. When impulses propagate along a surface, it is a *surface wave*, and if they propagate in all directions, it is a *spatial wave* (or volume wave). If the medium is homogeneous, the impulses propagate in all directions at the same speed, and such a wave is called a *spherical wave*.

The surface that the wave simultaneously reaches from a given source constitutes the wavefront. The wavefront can also be defined as the surface on which all particles of the medium oscillate in the same phase. The shape of the wavefront depends on the shape of the wave source and the medium through which the wave propagates. In the case of a point source, the shape of the wavefront will be a sphere. If the wave source is located at infinity, it can be considered a plane wavefront (or plane wave).

For the simpler visualization of spherical and surface waves, the concept of a ray has been introduced. A ray is understood to be a narrow beam of waves along a single direction of propagation and is represented by a straight line. In the case of spherical waves, rays represent the radii r of the spheres, which are normal (perpendicular) to the wavefront.

13.2 Impulse Propagation Speed

In a homogeneous and isotropic elastic medium, impulses (the wave) propagate in all directions at the same speed. A homogeneous medium means that its density is constant, and an isotropic medium means that its physical properties are the same in all directions. An elastic medium does not absorb energy, i.e., when moving through it there is no loss of energy, meaning it does not transform into other forms of energy such as heat.

If the direction of impulse propagation is perpendicular to the direction of particle oscillation, then it is a *transverse wave*, and if the direction of impulse propagation is the same as the direction of particle oscillation, the wave is *longitudinal*.

The propagation speed c of a transverse impulse in a tensioned wire of length l is:

$$c = \sqrt{\frac{F}{\mu}}, \tag{13.1}$$

where F is the tension force of the wire, and μ is the linear density (mass per unit length: $\mu = m/l$). From this expression, it can be seen that the speed of propagation of transverse pulses (transverse wave) c does not depend on the speed v of particle movement in the medium, but only on the tension force F of the string and the linear density μ.

The speed of propagation of longitudinal impulses (waves) in solid bodies is:

$$c = \sqrt{\frac{E_y}{\varrho}}, \tag{13.2}$$

where E_y is Young's modulus of elasticity and ϱ is the density. For fluids, in this equation, instead of E_y, the bulk modulus of elasticity E_V appears.

The speed of propagation of longitudinal pulses in gases is given by the expression

$$c = \sqrt{\frac{p\kappa}{\varrho_t}}, \tag{13.3}$$

where p is the gas pressure, κ is the so-calles adiabatic constant that represents the ratio of specific heat capacities at constant pressure (c_p) and constant volume (c_V), and ϱ_t is the gas density at temperature t^oC. With an increase in temperature, the speed of impulses in a gas, unlike in a liquid, increases significantly. For $p = const.$, $\varrho_t = \varrho_o/(1+t_o/273)$, where ϱ_o is the gas density at a temperature of 0^oC. By substituting the expression for ϱ_t into (13.3):

$$c = \sqrt{\frac{p\kappa}{\varrho_o}(1 + \frac{t^oC}{273})} = c_o\sqrt{1 + \frac{t^oC}{273}},$$

where $c_o = p\kappa/\varrho_o$ is the impulse propagation speed at temperature $0^o\,C$.

The speed of propagation of longitudinal pulses c, as seen from expressions (13.2) and (13.3), depends on the properties of the material. It is important to note that *transverse impulses (waves)* can only propagate **in solids** (because they have a significant shear modulus), while *longitudinal pulses* can propagate in bodies of **any state of matter**.

13.2.1 Sinusoidal Wave Motion

If an oscillator (wave source) that performs simple harmonic oscillations transmits them through an elastic medium in the form of sinusoidal pulses, then the individual particles of the medium that the pulses reach begin to oscillate themselves according to the sinusoidal law by which the wave source also oscillates. Such a wave, in which the oscillation of the particles follows the sinusoidal law, is called a sinusoidal wave.

As already mentioned, impulses can propagate transversely and longitudinally, and accordingly, transverse and longitudinal sinusoidal waves can be formed.

Figure 13.2 shows a diagram of the propagation of transverse waves. The five rows in the figure represent the arrangement of particles (which could be, for example, a line of atoms in a solid) at specific moments. The numbers 1, 2, 3, etc., denote particles that are positioned at distances of $cT/4$ (c - the speed of impulse propagation, and T - the period of particle oscillation), i.e., at distances that the wave travels in a quarter of the motion period. The first row shows the positions of the particles at time $t = 0$, when all particles occupy their equilibrium positions and when an upward vertical impulse (marked with an arrow in the figure) is applied to the first particle, which will then set the subsequent particles in the sequence into motion. After $t = T/4$, particle 1 reaches its highest position (the amplitude of its oscillation), while at the same time particle 2 receives an upward vertical impulse. After $t = T/2$, particle 1 passes through the equilibrium position moving downward, particle 2 reaches its highest position, and particle 3 begins to move upward from the equilibrium position, as it has been struck by an impulse, and so on. By the time $t = T$, particle 1 completes one full oscillation, returning to the position it had at the start of the motion. In that time, the wave travels a distance cT, reaching particle 5.

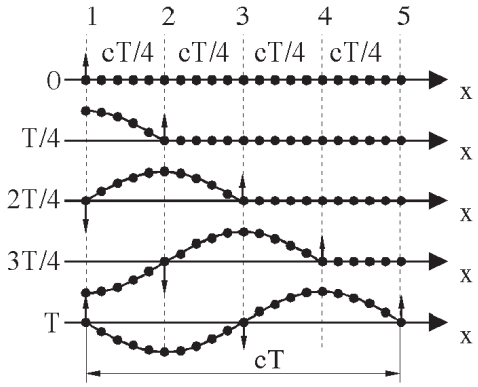

Figure 13.2: Diagram of the propagation of transverse waves

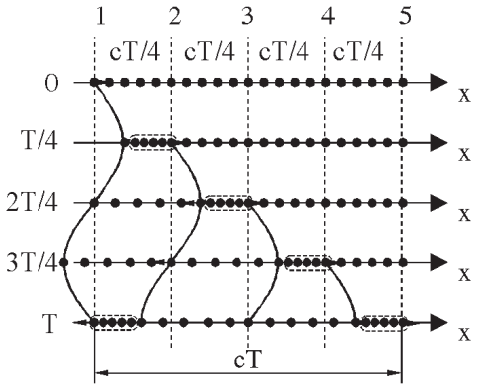

Figure 13.3: Diagram of the propagation of longitudinal waves

Figure 13.3 shows a diagram of the propagation of longitudinal waves. All considerations for transverse waves can also be applied to longitudinal waves (left to the reader to work out). The only difference is that in longitudinal waves, the particles of the medium oscillate in the direction of the wave propagation. The figure shows that during the propagation of longitudinal waves, the medium undergoes alternating compressions (areas circled in the figure) and rarefactions. The wave propagation process is accompanied by the movement of these compression and rarefaction regions through the medium.

During time T from the beginning of oscillation of a point source (particle 1 in Fig. 13.2), during which it completed a full oscillation, the impulse traveled a distance λ. This distance is called *the wavelength*

$$\lambda = cT. \tag{13.4}$$

From Figures 13.2 and 13.3, it can be seen that the wavelength is the minimum distance between two consecutive particles along the direction of wave propagation, which oscillate with a phase difference of 2π. By replacing T with $1/\nu$ in expression (13.4), where ν is the frequency of the source oscillation, the speed of wave propagation is obtained:

$$c = \lambda\nu$$

. Thus, the speed of wave propagation is equal to the product of the wavelength and the frequency of particle oscillation.

13.3 Huygens' Principle

The propagation of waves is often treated as the formation of a new wavefront at a given moment from the wavefront formed at the previous moment. This can be explained using *Huygens' principle*. Let a wavefront move through a homogeneous medium and at some moment occupy position 1 (Fig. 13.4). It is necessary to determine its position after a time interval Δt. According to Huygens' principle, every point of the medium reached by the wave becomes a source of secondary waves. This means that new spherical waves spread from every point reached by the wave. To define the secondary waves, spheres of radius $\Delta x = c\Delta t$ should be drawn around the points of the initial wavefront 1.

According to Huygens' principle, secondary waves cancel each other out in all directions except in the direction of wave propagation. In other words, a new wavefront is formed along the outer envelope of the secondary waves. In this way, the position of the wavefront is determined. Based on Huygens' principle, the phenomena of reflection, refraction, and diffraction of waves can be explained.

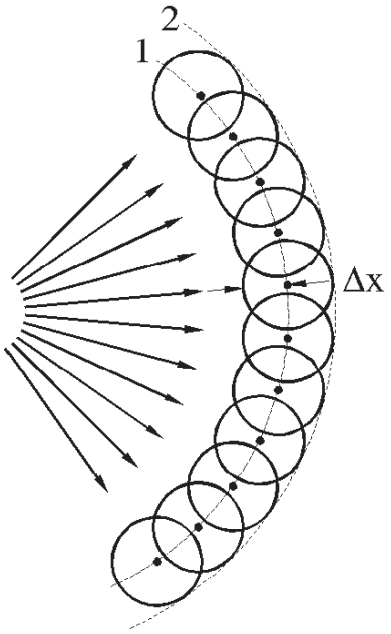

Figure 13.4: The way of wavefront propagation according to Huygens' principle

13.4 Equation of a Progressive Wave

A progressive wave is defined as a wave that propagates in the same direction and sense, ideally, throughout the entire elastic medium. When a wave propagates through an elastic medium, there is no energy loss, i.e., no conversion of wave energy into other forms of energy, primarily into heat.

To derive the equation of a progressive wave, whose parameters depend on spatial coordinates and time, a plane sinusoidal wave will be considered. Let the wave propagate along the x-axis. Then the wavefront will be perpendicular to the x-axis, so all points on the wavefront will oscillate in the same phase. Such oscillations can, regardless of the type of wave formed, be represented by the well-known sinusoidal law for simple harmonic motion (12.1), where the displacement depends only on x and t, i.e., $y = y(x, t)$.

The oscillation of particles lying in the plane $x = 0$ (Fig. 13.5) can be described by the equation:

$$y(0, t) = A \sin \omega t. \tag{13.5}$$

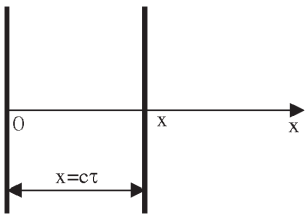

Figure 13.5: Position of the wave front at the beginning of observation ($t = 0$) and after some time τ

It is necessary to find the equation of the oscillation of particles in a plane corresponding to an arbitrary value of x. For the wave to travel the distance from the plane $x = 0$ to the plane at distance x, a time $\tau = x/c$ must elapse. This means that the oscillation of particles in the plane at x will lag behind the oscillation of particles in the plane $x = 0$ by the time τ, due to the finite speed of wave propagation, so the equation of oscillation in the plane at x has the form:

$$y(x, t) = A \sin \omega(t - \tau) = A \sin \omega(t - \frac{x}{c}), \tag{13.6}$$

where A is the amplitude of the wave and ω is the angular frequency. Equation (13.6) is called the equation of a progressive wave. This holds for the case when no energy is transferred to the elastic medium during the propagation of the wave.

Since $\omega = 2\pi/T$ and $c = \lambda/T$, equation (13.6) can be written in the following form:

$$y = A \sin 2\pi(\frac{t}{T} - \frac{x}{\lambda}) = A \sin(\omega t - kx). \tag{13.7}$$

If the initial phase is ϕ, then equation (13.7) takes the following form:

$$y = A \sin(\omega t - kx + \phi), \tag{13.8}$$

where

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{c} \tag{13.9}$$

is the wavenumber.

Since equation (13.7) involves two independent variables (x and t), the phenomenon of wave motion can be studied either by observing the elongation over time at a single point in space, or by examining the distribution of the oscillation phase in space at a specific moment. A graphical representation of equation (13.7) is shown in Figure 13.6.

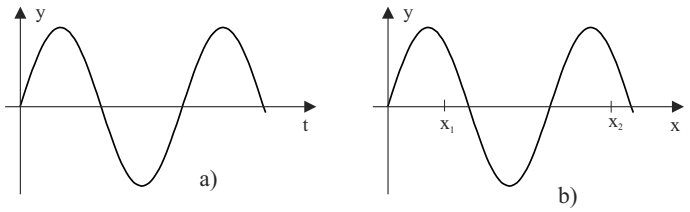

Figure 13.6: Graphical representation of equation (13.7)
a) elongation of a certain point in space as a function of time
b) elongation as a function of coordinate x at time t.

Let the phase difference at points x_1 and x_2 (Fig. 13.6b)

$$\Delta\phi = 2\pi(\frac{t}{T} - \frac{x_1}{\lambda}) - 2\pi(\frac{t}{T} - \frac{x_2}{\lambda}) = \frac{2\pi}{\lambda}(x_2 - x_1).$$

If $\Delta\phi = n2\pi$, the particles in the medium will oscillate in phase. The distance between the particles of the medium is

$$x_2 - x_1 = n\lambda, \tag{13.10}$$

where $n = 1, 2, 3,....$ For n =1, the distance $x_2 - x_1$ is equal to the wavelength λ, i.e.

$$x_2 - x_1 = \lambda = \frac{c}{\nu}.$$

This is a common expression for wave motion.

The particles of the medium will oscillate in opposite phases if there is a phase difference

$$\Delta\phi = (2n + 1)\pi, \tag{13.11}$$

where $n = 0, 1, 2,...$ that is, if the particles are at distances

$$x_2 - x_1 = (2n + 1)\frac{\lambda}{2}. \tag{13.12}$$

By an analogous procedure as for plane waves, it can be shown that the motion of a spherical sinusoidal wave in an elastic medium (progressive spherical sinusoidal wave) can be represented by an equation

$$y = \frac{A_o}{r}\sin(\omega t - kx), \tag{13.13}$$

where A_o is the wave amplitude, which is constant in this case, and r is the distance from the wave center to the considered point of the medium. Thus, in the case of a spherical wave, even in media that do not absorb energy (elastic media), the oscillation amplitude does not remain constant, but decreases with distance from the source according to the law $1/r$. Equation (13.13) applies only for values of r that are much larger than the dimensions of the source (the wave source can be considered a point source).

From equation (13.9), it follows:

$$c = \frac{\omega}{k}. \tag{13.14}$$

That is, the phase speed of sinusoidal waves depends on their frequency. This phenomenon is called *wave dispersion*, and the medium in which dispersion occurs is called a dispersive medium.

Equation (13.7) applies in the case when the amplitude does not depend on the coordinate x, i.e. the amplitude is constant. This means that the wave does not transfer its energy to the medium through which it propagates (elastic medium). If the wave transfers its energy to the medium while propagating, wave intensity decreases (wave attenuation occurs). Namely, the energy of the waves transforms into other forms of energy, such as heat (non-elastic medium). Experimentally, it has been shown that the amplitude of such a wave decreases according to the law $A = A_o\exp(-\beta x)$, so the equation of plane waves can be written in the form:

$$y = A_o\exp(-\beta x)\sin(\omega t - kx), \tag{13.15}$$

and the equation for spherical waves

$$y = \frac{A_o}{r}\exp(-\beta x)\sin(\omega t - kx), \tag{13.16}$$

where A_o is the amplitude of the particle in the plane $x = 0$ (Fig. 13.5).

13.5 Energy Transfer by Waves

The process of wave propagation in any elastic medium involves the transfer of energy by the particles of the elastic medium. Let S be the area of a part of the plane wavefront at time t (Fig. 13.7). After a time Δt, this front will move a distance $\Delta x = c\Delta t$, during which all the particles of the medium within the volume $\Delta V = S\Delta x$ will execute oscillatory motion (c is the speed of wave propagation). Namely, during the time Δt, the medium of volume ΔV will receive energy E through the surface S.

Let w be the energy of oscillation of the medium's particles per unit volume (w is calculated as the ratio of the energy E of all particles in the ΔV to the ΔV: $w = E/\Delta V$). It will be assumed that the volume $\Delta V = S\Delta x$ is very small, so that within this volume the energy w is uniform.

Therefore, over the time interval Δt, energy E is transferred through the surface S:

$$\Phi = \frac{E}{\Delta t} = \frac{w\Delta V}{\Delta t} = \frac{wS\Delta x}{\Delta t} = wSc. \tag{13.17}$$

The quantity Φ is called *the energy flux of the wave* and is expressed in watts.

The wave intensity or *wave strength* I (strength of the energy flux) represents the energy transferred per unit of time through a unit area normal to the direction of wave propagation:

$$I = \frac{E}{\Delta t S} = \frac{\Phi}{S} = wc. \tag{13.18}$$

The intensity of a wave is expressed in W/m^2. Since the velocity of wave propagation c is a vector quantity, the intensity of the wave I is also a vector quantity and is called the Umov vector.

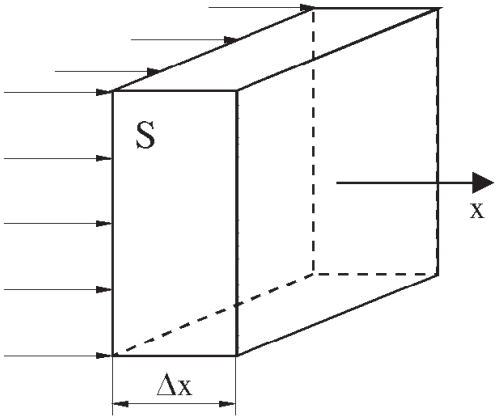

Figure 13.7: Propagation of a part of a plane wavefront

Let a point source emit energy E_1 through an elastic medium every time interval Δt. This energy will be evenly distributed over the spherical surface $S = 4\pi r^2$ of the wavefront. Therefore, $I = E_1/(4\pi r^2 \Delta t)$, i.e., $I \sim 1/r^2$. This expression, as well as expression (13.18), applies only to media where there is no conversion of oscillatory motion energy into other forms of energy (for example, into thermal energy). It shows that for spherical waves originating from a point wave source, the wave intensity decreases with the square of the distance (r^2) from the source.

It is necessary to find the relationship between the intensity of the wave I, the maximum speed reached by a particle during oscillation (speed amplitude) v_o, and the wave propagation speed c. Therefore, the volume of an elastic medium whose dimensions are smaller than the wavelength λ of the wave will be considered, so that within this volume ΔV it can be assumed that all particles oscillate with the same frequency and amplitude.

Let the mass of a single particle be m, and their number per unit volume be $n = N/\Delta V$, where N is the total number of particles in volume ΔV. Then $w = E/\Delta V = NE_o/\Delta V = nE_o$, where E_o is the energy of a particle.

Since the total oscillatory energy of a single particle according to expression (12.27) is equal to $E_0 = mA^2\omega^2/2$, then, the total oscillatory energy of all particles in a unit volume is:

$$w = n\frac{mA^2\omega^2}{2} = \frac{\varrho v_o^2}{2}, \tag{13.19}$$

where $\varrho = Nm/\Delta V = nm$ is the density of the medium (Nm is the mass of all particles in volume ΔV), A is the oscillation amplitude, ω is the angular frequency of particle oscillation, and $v_o = A\omega$ is amplitude of particle oscillation speed (eq. 12.3).

By substituting expression (13.19) into (13.18) for the wave intensity, the following expression is obtained:

$$I = \frac{\varrho v_o^2}{2}c. \tag{13.20}$$

As already shown, the intensity of a spherical wave in an elastic medium decreases with the square of the distance r from the wave source ($I \sim 1/r^2$). Based on equations (13.18) and (13.19), it follows: $I \sim w \sim A^2 \sim 1/r^2$. Accordingly,

the amplitude of particle oscillation A decreases inversely with that distance r ($A \sim 1/r$), what was used in expression (13.13).

When a wave propagates in a non-elastic medium, its energy is absorbed, which causes the conversion of mechanical energy into heat. Let the plane waves propagate along the x-axis. As the wave passes through a layer of thickness dx, the intensity of the wave I decreases by dI ($dI \sim Idx$):

$$dI = -\mu I dx, \tag{13.21}$$

where μ is called *the coefficient of absorption* of wave energy in a given medium. The minus sign is included because the intensity decreases with distance.

If intensity is I_o at the point $x = 0$ and intensity is $I(x)$ at the point x, then integrating the previous equation gives

$$I(x) = I_o \exp(-\mu x). \tag{13.22}$$

Using relation $I \sim A^2$, it can be concluded that in a non-elastic medium the amplitude decreases according to the same (exponential) law:

$$A_{ox} = A_o \exp(-\beta x), \tag{13.23}$$

where β is *the damping coefficient* of the oscillation of the particles of the medium. This equation was used in expressions (13.15) and (13.16).

The following expression applies to the change in pressure of the medium during wave propagation:

$$\Delta p = \Delta p_o \exp(-\beta x). \tag{13.24}$$

13.6 Wave Interference

If the tips of two sticks simultaneously touch the calm surface of the water, waves in the form of concentric circles will spread independently from those points (Fig. 13.8).

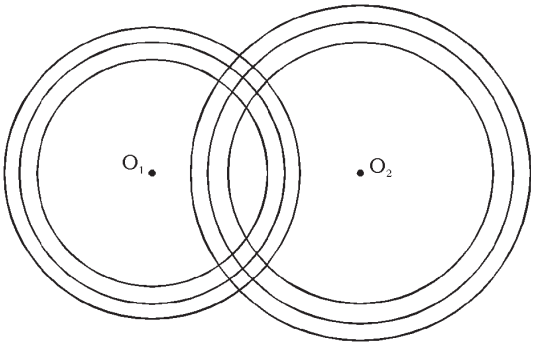

Figure 13.8: Wave propagation from two sources

Although waves originating from two points of contact (wave sources) overlap at certain points, they will continue to propagate independently as if they had not met. In the region of wave overlap, the oscillations of the particles superpose on one another, resulting in interference or wave superposition, and as a result, in some locations, the particles of the medium may oscillate with greater amplitude, while in others with smaller amplitude. To find the resulting displacement of a particle affected by waves, it is necessary to determine its displacement caused by each wave individually and then combine them scalarly if these displacements occur in the same direction, or vectorially if they occur in different directions. Most often, the interference of coherent waves is observed. Coherent waves are understood to be those waves whose frequencies are identical and that have a constant phase difference. Sources of coherent waves can be most easily produced if they are connected in such a way that they oscillate together.

For example, two interconnected rods, attached at one end to a tuning fork, with their free ends submerged in water (Fig. 13.9). Let the resulting oscillation occur at point M. The distance from point M to source O_1 is x_1, and to source O_2 is x_2. We will consider the case when the initial phases are zero. The wave equations at point M, originating from sources O_1 and O_2, according to expression (13.7), are

$$y_1 = A_1 \sin 2\pi(\frac{t}{T} - \frac{x_1}{\lambda}), \tag{13.25}$$

$$y_2 = A_2 \sin 2\pi(\frac{t}{T} - \frac{x_2}{\lambda}), \tag{13.26}$$

where A_1 and A_2 are the amplitudes of the waves at point M. If the difference $x_1 - x_2$ is very small, it can be assumed that the wave amplitudes are equal, i.e., $A_1 \approx A_2 \approx A$.

The resulting elongation of the particle at point M will be equal to the sum of the elongations individually produced by these waves, i.e.

$$y = y_1 + y_2 = 2A \cos 2\pi(\frac{x_1 - x_2}{2\lambda}) \sin 2\pi(\frac{t}{T} - \frac{x_1 + x_2}{2\lambda}),$$

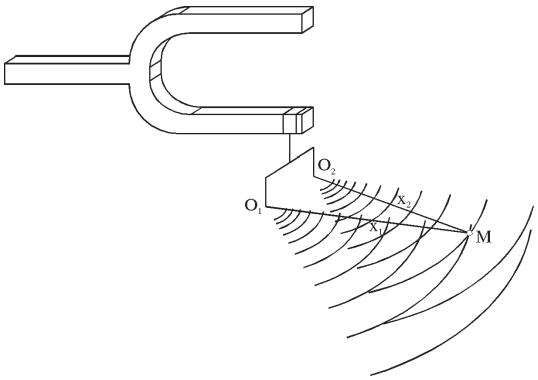

Figure 13.9: Interference of two coherent waves

or

$$y = B \sin 2\pi(\frac{t}{T} - \frac{x_1 + x_2}{2\lambda}) \tag{13.27}$$

where is

$$B = 2A \cos 2\pi(\frac{x_1 - x_2}{2\lambda}) \tag{13.28}$$

the amplitude of the resulting oscillation. Here, the trigonometric transformation for the sum of the sines of two angles, e.g. α and β, is applied: $sin(\alpha) + sin(\beta) = 2sin((\alpha+\beta)/2)cos((\alpha-\beta)/2)$, where $\alpha = 2\pi(\frac{t}{T} - \frac{x_1}{\lambda})$ and $\beta = 2\pi(\frac{t}{T} - \frac{x_2}{\lambda})$.

From expression (13.27) it can be seen that point M oscillates harmonically with the same period as the wave sources O_1 and O_2. The amplitude of the resulting oscillation B will depend on the difference in distance of point M from the wave sources O_1 and O_2, which is called *the path difference*. If this path difference $x_1 - x_2$ is equal to an integer number of wavelengths

$$x_1 - x_2 = n\lambda, \tag{13.29}$$

where $n = 0, 1, 2,...$, then

$$B = 2A. \tag{13.30}$$

Thus, the amplitude of the resulting oscillation doubles compared to the amplitude of an individual wave. It is said that the waves are amplified to the maximum.

If the path difference is equal to an odd number of half wavelengths

$$x_1 - x_2 = (2n+1)\frac{\lambda}{2}, \tag{13.31}$$

where $n = 0, 1, 2,...$, then

$$B = 0, \tag{13.32}$$

i.e. if the amplitude of oscillation of a particle affected by these waves is zero, it is said that the waves cancel each other (are damped).

In the case where the amplitudes A_1 and A_2 are different at points that satisfy condition (13.29), the resulting amplitude will also be equal to their sum $A_1 + A_2$. At points that satisfy condition (13.31), complete wave cancellation will not occur; instead, the amplitude of the resulting particle oscillation will reach the minimum possible value, equal to $|A_1 - A_2|$. At other intermediate points, depending on the path difference of the waves, the resulting amplitude may have different values, continuously varying within the range $|A_1 - A_2| \leq B \leq A_1 + A_2$. In a similar manner, the interference of a larger number of coherent waves originating from different sources could also be considered.

13.7 Standing Waves

A special case of wave interference is the standing wave. A standing wave is formed when two progressive waves of equal frequency and amplitude meet, moving in opposite directions. Let two waves meet along the x-axis in elastic medium. Since the medium is elastic, waves do not transfer energy to it, and the waves retain the same frequency and amplitude. The initial coordinate is usually chosen at the point where both waves have initial phases equal to zero. Then the corresponding equations for the wave propagating along the positive direction of the x-axis and the wave propagating in the opposite direction will have the following form:

$$\begin{aligned} y_1 &= A \sin(\omega t - kx), \\ y_2 &= A \sin(\omega t + kx). \end{aligned} \tag{13.33}$$

The equation of the resulting oscillation (resulting wave) is

$$y = y_1 + y_2 = 2A \cos kx \sin \omega t = 2A \cos(\frac{2\pi}{\lambda}x) \sin \omega t. \tag{13.34}$$

From equation (13.34), it can be seen that at every point of a standing wave, oscillation of the same frequency occurs with an amplitude $B_s = |2A\cos(2\pi x/\lambda)|$, which depends on the coordinate x of the considered point.

At points in the medium where it applies

$$\frac{2\pi}{\lambda}x = \pm n\pi, \tag{13.35}$$

where $n = 0, 1, 2,...$, the amplitude of the standing wave reaches its maximum $B_s = 2A$.

At points

$$\frac{2\pi}{\lambda}x = \pm(2n+1)\frac{\pi}{2}, \tag{13.36}$$

where $n = 0, 1, 2,...$, the amplitude of the standing wave reaches its minimum $B_s = 0$.

Points at which the amplitudes of standing waves are maximum ($B_s = 2A$) are called *antinodes* of the standing wave, and points at which the amplitude of the standing wave is zero ($B_s = 0$) are called *nodes* of the standing wave.

From expressions (13.35) and (13.36), the corresponding coordinates of the antinodes x_t and nodes x_c of the standing wave can be obtained

$$x_t = \frac{n\lambda}{2}, \tag{13.37}$$

$$x_c = (2n+1)\frac{\lambda}{4}. \tag{13.38}$$

From expressions (13.37) i (13.38), it can be seen that the distance between two adjacent antinodes or nodes is equal to $\lambda/2$, and the distance between an antinode and a node is $\lambda/4$.

Unlike progressive waves, where all particles oscillate with the same amplitudes but with a phase that lags (in progressive waves, the phase of oscillation depends on the x-coordinate of the observed point), all points of a standing wave between two nodes oscillate with different amplitudes but with the same phase (the argument of the sine in expression (13.34) does not depend on x). On the other hand, at the node, the amplitude B_s changes sign, with the phase of oscillation differing by π. Therefore, particles located on opposite sides of the node will oscillate in opposite phases.

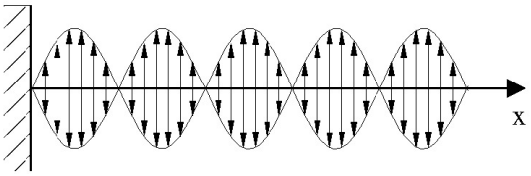

Figure 13.10: Standing waves

13.8 Reflection and Refraction of Waves

Wave reflection occurs when a wave strikes a boundary and is reflected back into its original medium. This phenomenon obeys the law of reflection (Fig. 13.11), which states that the angle of incidence α is equal to the angle of reflection α':

$$\alpha = \alpha', \tag{13.39}$$

both measured against a perpendicular ("normal") line. This law applies if the wave is reflected from flat surfaces whose *roughness is significantly smaller than the wavelength of the wave.*

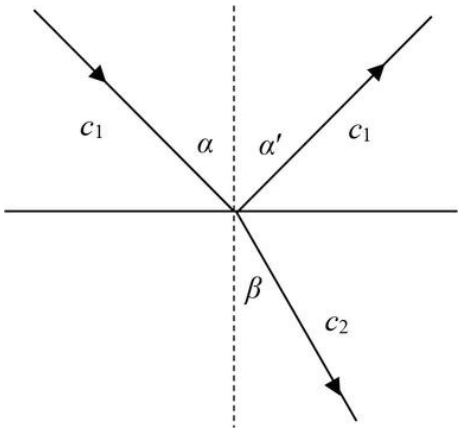

Figure 13.11: Reflection and refraction of waves

The reflected wave maintains the same frequency, wavelength, and speed as the incident wave, although its direction of travel changes. Examples include echoes from sound waves and the image formed by a mirror from light waves. The

figure 13.11 shows the incident and reflected waves (beams) in the first medium. The speed of the wave in this medium is c_1.

Wave refraction is the bending of a wave as it passes from one medium to another, caused by a change in the wave's speed. When a wave slows down, it bends towards the "normal" (a line perpendicular to the boundary), and when it speed up, it bends away from it. The figure 13.11 shows a case where the wave bends toward the normal, which means it slows down in the second medium.

This effect is responsible for phenomena like light bending through a lens. The frequency of the wave remains constant during refraction, but its wavelength changes.

Figure 13.11 shows the refraction of a wave coming from a medium with a wave speed of c_1 and at an angle α with respect to the normal (angle of incidence) hitting the interface between two media. The wave refracts into a medium with a wave speed of c_2 at an angle β with respect to the normal (angle of refraction).

The law of refraction states:

$$\frac{\sin\alpha}{\sin\beta} = \frac{c_1}{c_2}. \tag{13.40}$$

Chapter 14

Acoustics

14.1 Fundamental Characteristics of Sound

Acoustics is the science of sound. Mechanical oscillations that propagate in an elastic medium (solid, liquid, or gaseous) and can be perceived by the sense of hearing are called *sound waves*. The human ear detects sound waves with frequencies approximately from $20\,Hz$ to $20\,kHz$. Waves with frequencies below $20\,Hz$ are called *infrasound*. Infrasound occurs during various vibrations, earthquakes, and tremors caused by traffic, etc. Sound with frequencies above $20\,kHz$ is called *ultrasound*. Ultrasonic waves can be produced using electrical and magnetic methods. Some animals can produce and detect ultrasound. For example, dolphins and whales can detect ultrasound up to $30\,kHz$. Bats emit ultrasound, which helps them navigate in completely dark spaces. These animals, while flying, emit ultrasonic signals at specific time intervals lasting about $0,01\,s$, with frequencies between $50\,Hz$ to $80\,kHz$, and then listen for the echo from the nearest reflecting surfaces.

Based on the time it takes for the echo to return and the direction of that reflected sound, a bat determines the direction and distance of obstacles, allowing it to fly freely without the risk of colliding with walls or other objects.

Ultrasound with a frequency in the GHz range or higher is called hypersound. The upper limit of ultrasound frequency (in crystals $10^{12} - 10^{13}\,Hz$) corresponds to a frequency whose wavelength is on the order of the interatomic spacing in the crystal. The same limitation in air corresponds to the order of the mean free path of molecules.

The sensitivity of the human ear depends on the frequency of sound. It is highest for sound frequencies from $1,5$ to $3\,kHz$.

As already mentioned, only longitudinal sound waves can propagate in liquids and gases, whereas both longitudinal and transverse sound waves can propagate in solid bodies. In a vacuum, sound waves cannot propagate because particles of the medium are necessary to transmit mechanical waves.

Different materials transmit sound differently, i.e., they have different acoustic conductivity. For example, soft and porous materials conduct sound poorly, so they are used as sound insulators.

The sound perceived by the sense of hearing is divided into tone and noise. A tone, in a certain sense, has regular oscillations, that is, a specific frequency. A pure or simple tone is a sound whose source performs simple harmonic oscillations. The basic characteristic of a pure tone is its frequency.

The sound produced by musical instruments is a complex tone. A complex tone can be decomposed into a series of simple tones. The lowest frequency ν in that series characterizes *the fundamental tone*, while the others are *higher harmonics*. The fundamental frequency determines the pitch of the tone. Tones with frequencies 2ν, 3ν, 4ν,...,$n\nu$ represent higher harmonics. *The timbre of a tone* depends on which higher harmonics are present in the frequency spectrum of the tone and with what intensity. Depending on the timbre, human voices, as well as the tones of different musical instruments, can be distinguished.

The sound intensity (I) depends on the amplitude of oscillation and represents the physical (objective) intensity of the sound. Since the ear's sensitivity to different frequencies varies, there is also a subjective (physiological) intensity of the sound (loudness) L.

Noise represents a complex wave with varying frequencies and amplitudes. If a tone contains more frequencies that are not its harmonics, it is closer to noise. These types of sounds occur during the vibration of machines, rustling of leaves, creaking, etc.

A tone has a line spectrum. Such a spectrum is produced by various musical instruments. Noise has a continuous spectrum. An example of line and continuous spectra is shown in Figure 14.1.

14.2 Sound Sources

A sound source is defined as any mechanical oscillator that can properly oscillate within the range of sound frequencies. In engineering, the most commonly used sound sources are stretched strings, rods, air columns, membranes, and plates.

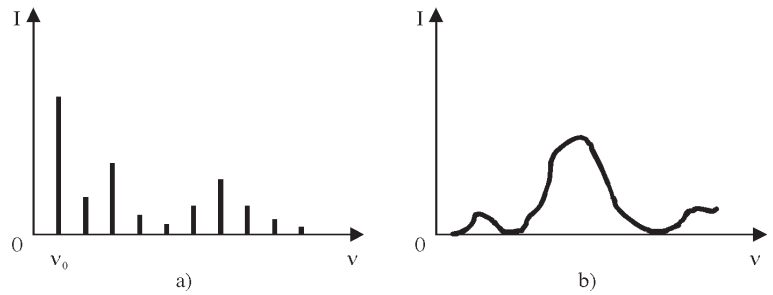

Figure 14.1: Spectrum of a) tones, b) noise

14.2.1 Taut Strings

Figure 14.2 shows a wire that is fixed at both ends and stretched by a force F between two immovable supports. If a transverse deformation is applied to the middle of the wire, a transverse standing wave will form due to interference. A standing wave will form if the length of the wire is

$$l = n\frac{\lambda}{2}, \tag{14.1}$$

where $n = 1, 2, 3,...$, and λ is the wavelength of the transverse wave.

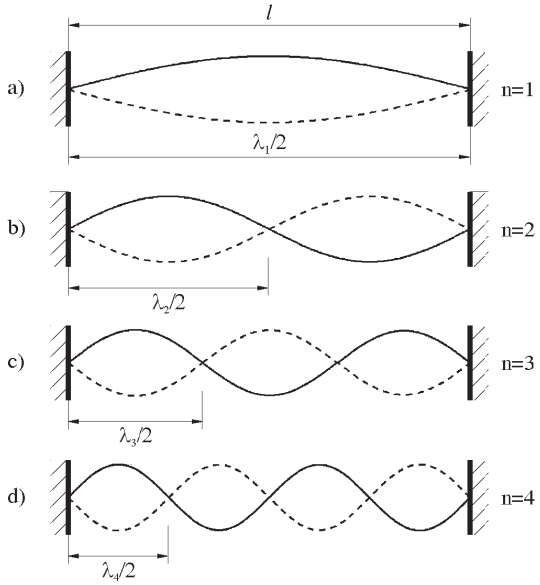

Figure 14.2: Oscillation of a taut string (wire), fixed at both ends

Figure 14.2 shows standing waves for $n = 1, 2, 3$, and 4. A string oscillating in one of the ways shown in Figure 14.2 produces a simple tone. If the oscillation is complex, for example by superposing these four types of oscillations from the figure, a complex tone is obtained, in which the pitch is determined by the frequency of the fundamental tone (produced by the oscillation in Fig. 14.1.a), and the other superimposed oscillations represent higher harmonics that determine the timbre of the tone.

Since the frequency $\nu = c/\lambda$, where $c = \sqrt{F/\mu}$ is the wave propagation speed, expression (14.1) can be written in the form:

$$\nu = \frac{c}{\lambda} = \frac{n}{2l}c = \frac{n}{2l}\sqrt{\frac{F}{\mu}}. \tag{14.2}$$

14.2.2 Rods

Both transverse and longitudinal oscillations can occur in rods. Figure 14.3 shows the transverse oscillations of a rod. If a rod is fixed at one end and its free end is displaced from the equilibrium position, it will start to oscillate, and this disturbance will propagate in the form of a wave towards the fixed end of the rod, from where it will reflect and return to the free end. As a result, a standing wave will form, with its antinode located at the free end of the rod.

A standing transverse wave is formed such that the length of the rod is $l = \lambda/4$ for the case in Figure 14.3.a, and $l = 3\lambda/4$ for the case in Figure 14.3.b. In general, it can be written

$$l = (2n + 1)\frac{\lambda}{4}, \tag{14.3}$$

$$\nu = (2n+1)\frac{c}{4l}, \tag{14.4}$$

where $n = 0, 1, 2, 3, ...$

If the rod is fixed at two points (Fig. 14.3.c), nodes of the standing wave occur at those points.

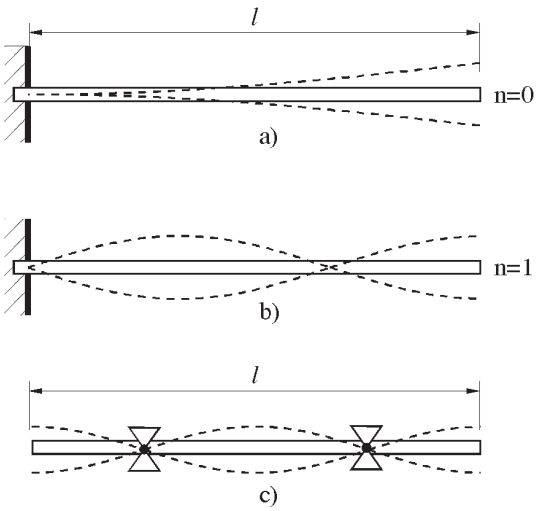

Figure 14.3: Transverse oscillations of a rod
a) fundamental frequency when the rod is fixed at one end
b) first harmonic when the rod is fixed at one end
c) fundamental frequency when the rod is fixed at two points

14.2.3 Air Column

In a pipe closed at one end (Fig. 14.4.a), the wave is reflected from the closed end and a standing wave is formed with a node at the closed end and an antinode at the open end. This case corresponds to the oscillation of a rod fixed at one end, so the same expressions (14.3) and (14.4) will apply here as well

$$l = (2n+1)\frac{\lambda}{4},\ \nu = (2n+1)\frac{c}{4l}, \tag{14.5}$$

where $n = 0, 1, 2, 3, ...$, and c is the speed of propagation of longitudinal waves in the air.

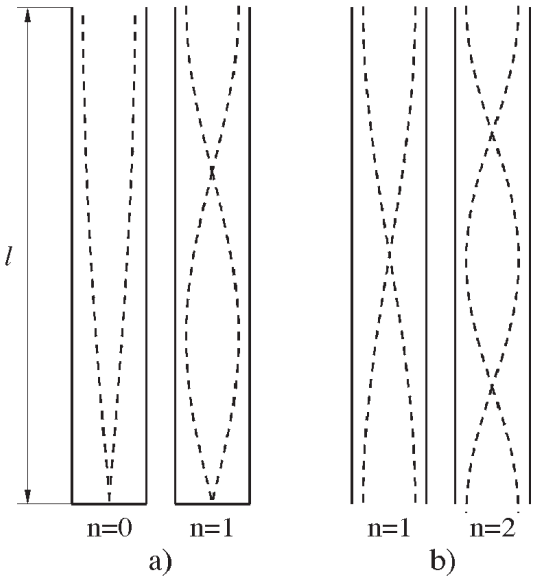

Figure 14.4: Standing waves in air columns
a) fundamental frequency and first higher harmonic in an air column closed at one end
b) fundamental frequency and first higher harmonic in an air column open at both ends

In an air column open at both ends (Fig. 14.4.b), the oscillations are such that the antinodes of the standing wave are always located at the open ends ($l = \lambda/2$ for the first case in Figure 14.4.b, and $l = 2\lambda/2$ for the second case in Figure 14.4.b):

$$l = n\frac{\lambda}{2},\ \nu = \frac{nc}{2l}, \tag{14.6}$$

where $n = 1, 2, 3, ...$

14.3 Sound Intensity

The intensity or loudness of sound I (the energy flux density of sound waves) represents the average amount of energy that sound waves transfer per unit of time through a unit area perpendicular to the direction of wave propagation. The intensity of sound waves, as with other mechanical waves, is given by equation (13.18).

The propagation of longitudinal sound waves in an elastic medium causes volumetric deformation of that medium. As a result, the pressure changes at every point in the medium. The sound intensity can be expressed through the pressure amplitude of the medium, i.e.

$$I = \frac{1}{2}\frac{p_m^2}{c\varrho}, \tag{14.7}$$

where p_m is the maximum value of the pressure (pressure amplitude), ϱ is the density of the medium in its undisturbed state, and c is the speed of sound propagation in that medium.

The human ear is limited in perceiving sound not only by frequencies but also by its intensity. The ear can detect sounds whose intensity exceeds a certain minimum value, known as the threshold of hearing. The hearing threshold depends on frequency and varies between different individuals. The ear is most sensitive at frequencies from 1500 Hz to 3000 Hz, when the threshold of hearing is around $10^{-12}\,W/m^2$.

A sound that is too loud causes a feeling of pressure and pain in the ear. The intensity of sound at which ear pain occurs is called the threshold of pain. It, like the threshold of hearing, depends on the frequency of sound waves. Between the threshold of hearing and the threshold of pain lies the region of audibility (Fig. 14.5).

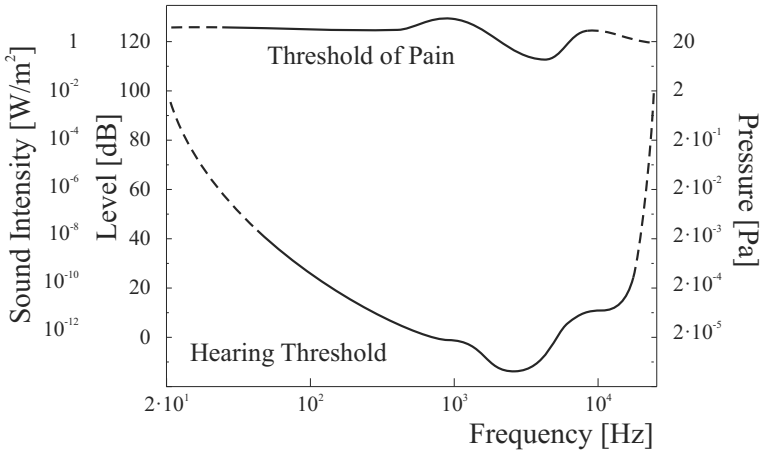

Figure 14.5: Hearing range

Expression (13.18) applies to the objective value of sound intensity that can be measured. However, the human ear perceives the gradation of sound loudness according to a logarithmic law (subjective loudness). This phenomenon was studied by Weber and Fechner, who formulated the law known as the Weber-Fechner psychophysical law. According to this law, the sense of hearing perceives the gradation of sound loudness approximately as the logarithm of the sound intensity expressed in W/m^2. Based on this logarithmic law, a scale of sound intensity has been established, called *the sound level* or *loudness* L. According to the Weber-Fechner law, the sound level is given by the following expression:

$$L = k \log \frac{I}{I_o}, \tag{14.8}$$

where $I_o = 10^{-12}\,Wm^{-2}$ is the sound intensity at the threshold of hearing, and I is the measured value of the sound intensity. For $k = 1$, the sound level is measured in bels (B), and for $k = 10$, in decibels (dB). In Figure 14.5, in addition to the sound intensity, the sound level values are also given.

14.4 Resonance in Sound. Resonators

Resonance is a phenomenon that occurs in forced oscillations. Since sound sources are mechanical oscillators, forced oscillations can also be established in them. Any sound source can be brought into resonance by a periodic force, usually provided by another sound source. This can be demonstrated using a simple device schematically shown in Figure 14.6. A tube open at both ends, is submerged in a container of water. The length of the air column l in the tube can be changed by raising or lowering it.

When a tuning fork is set into vibration and brought near the upper opening of a tube, forced vibrations will occur in the air column under the influence of these oscillations. The natural frequency of the air column depends on its length. By raising or lowering the tube, the natural frequency of the air column changes. When the length l is adjusted such

that the air column's natural frequency equals the frequency of the tuning fork, resonance will occur. This produces intense oscillation of the air column, which can be observed by the increased intensity of the tone. When a tuning fork

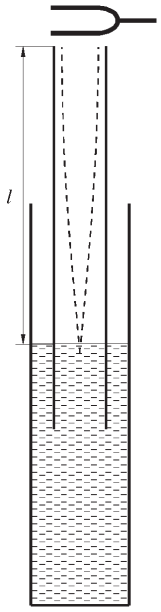

Figure 14.6: With an explanation of resonance in an air column

oscillates, the energy of the sound waves is small due to its small surface area. However, when resonance occurs, in addition to the tuning fork, the air column also emits energy, and because of the more intense emission of sound energy, the tone is heard more loudly.

14.5 Room Acoustics

When a sound source oscillates in an enclosed room, the listener will not hear only the sound coming directly from the source, but also the sound reflected from the reflective surfaces in the room, i.e., the walls, ceiling, floor, and all objects present in the room. The intensity of the reflected sound depends on the shape, size, and arrangement of the reflective surfaces, as well as the materials they are made of. If the sound source continuously emits waves that the walls and other surfaces poorly absorb, standing waves can form in certain parts of the room. As a result, in some areas of the room the sound will be heard more loudly, and in others more weakly.

Different materials absorb sound waves to different extents. This property of a material is characterized by the absorption coefficient A, which is given by the expression

$$A = \frac{P_a}{P_u}, \tag{14.9}$$

where P_a is the sound power absorbed by a given surface, and P_u is the total sound power incident on that surface.

If materials absorb less, the reflection of sound waves will be more intense, and vice versa. From the initially emitted sound signal in a room, part of it will be absorbed, and part will be reflected off reflective surfaces. These reflected secondary waves will then be absorbed and reflected again, and so on, until the sound intensity drops to zero.

Thus, sound continues to persist in a room even after the emission has stopped. This phenomenon is known as *reverberation*, and the time during which the sound is still audible after the emission has ceased is called *the reverberation time*. For better definition and easier measurement of reverberation time in engineering, it has been established that it is calculated from the moment the sound source stops until the intensity of the sound drops to one millionth of its original value, i.e., by $60\,dB$.

Reverberation has a significant impact on the acoustics of a room. If reflective surfaces absorb sound poorly, the reverberation time will be long, and the room will have an echo that is unpleasant to the ear. This can be corrected by placing good sound absorbers on the reflective surfaces, reducing the reverberation time.

If the material's absorption coefficient is high, a short reverberation time will be achieved. A room in which this is specifically enhanced through the choice of materials and surface configuration is called an anechoic chamber. In it, the reverberation time can be 100 times or more shorter than in ordinary rooms. It is used in acoustics to test the characteristics of loudspeakers, speaker boxes, sources of unwanted noise from various devices, etc.

14.6 Doppler Effect

The Doppler effect is the change in the frequency of oscillation of a sound source that is detected by a receiver (listener) during their relative motion. For example, it is known that the pitch of a vehicle's siren is higher when it approaches the listener and lower when it moves away from them, compared to the pitch heard when both the sound source and the

listener are stationary. Therefore, the motion of the sound source relative to the receiver changes the frequency of the received oscillations. This change can also be observed when the sound source is stationary and the receiver is moving.

When considering the Doppler effect, it is usually assumed that the sound source and the receiver move along the same straight line. Let v_i and v_p be the velocities of the source and the receiver, respectively, which are positive if the sound source or the receiver is approaching, and negative if they are receding. The oscillation frequency of the source is ν_o. Depending on the motion of the sound source and receiver, the following cases can be observed:

a) The receiver and the source are stationary with respect to the medium. Then $v_i = v_p = 0$. If c is the speed of sound waves in the considered medium, their wavelength is

$$\lambda = cT = \frac{c}{\nu_o}. \tag{14.10}$$

Propagating in the medium, the waves reach the receiver, which detects the frequency

$$\nu = \frac{c}{\lambda} = \frac{c}{cT} = \frac{1}{T} = \nu_o. \tag{14.11}$$

From expression (14.11), it can be seen that the frequency of the sound registered by the receiver ν is equal to the frequency of the sound source ν_o.

b) The receiver approaches a stationary sound source. Then $v_p > 0$, and $v_i = 0$. In this case, the velocity of wave propagation relative to the receiver is equal to $c + v_p$. Since the wavelength does not change in this case (Fig. 14.7), the frequency ν registered by the receiver is given by

$$\nu = \frac{c + v_p}{\lambda} = \frac{c + v_p}{cT} = \frac{c + v_p}{c}\nu_o. \tag{14.12}$$

It follows from expression (14.12) that the frequency of oscillation registered by the receiver is $(c + v_p)/c$ times higher than the frequency of oscillation of the source.

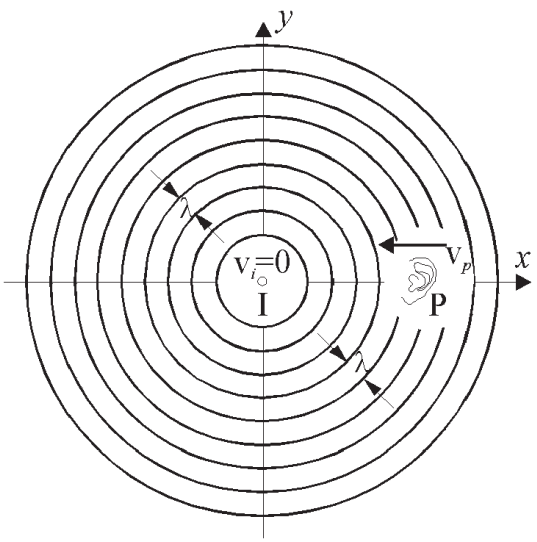

Figure 14.7: Wave fronts detected by a receiver in motion

c) The source is approaching a receiver that is at rest. Then $v_p = 0$, and $v_i > 0$. The propagation speed of the wave depends only on the properties of the medium. Therefore, during the time equal to the oscillation period of the sound source, the wave will propagate a distance cT (equal to the wavelength λ), regardless of whether the source is at rest or moving. During that time, the sound source moves a distance in the direction of the wave $v_i T$ (Fig. 14.8). Consequently, the wavelength of the wave in the direction of the source's motion shortens and is equal to $\lambda' = \lambda - v_i T = (c - v_i)T$, and the frequency of the wave recorded by the receiver

$$\nu = \frac{c}{\lambda'} = \frac{c}{(c - v_i)T} = \frac{c}{c - v_i}\nu_o, \tag{14.13}$$

i.e., the frequency ν registered by the receiver increases by a factor of $c/(c - v_i)$. In the cases b) and c), if $v_p < 0$ and $v_i < 0$, the sign in expressions (14.12) and (14.12) will be opposite.

d) The source and the receiver are moving relative to each other. Using expressions (14.12) and (14.13), a general expression can be written for the oscillation frequency detected by the receiver

$$\nu = \frac{c \pm v_p}{c \mp v_i}\nu_o. \tag{14.14}$$

The plus sign in the numerator of expression (14.14) corresponds to the receiver approaching the source, while the minus sign corresponds to the receiver moving away from the source. The minus sign in the denominator of the same expression corresponds to the sound source approaching the receiver, while the plus sign corresponds to the source moving away.

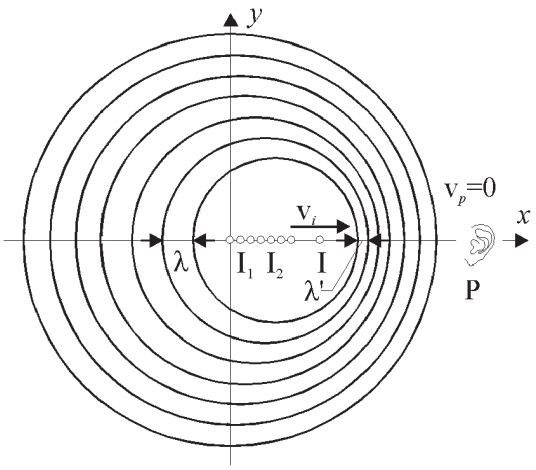

Figure 14.8: Wave fronts detected by a source in motion

It can be shown that in the case of uniform motion of the source and receiver with arbitrary directions of the velocity vectors $\vec{v_p}$ and $\vec{v_i}$, expression (14.14) takes the following form:

$$\nu = \frac{1 \pm (v_p/c)\cos\theta_2}{1 \mp (v_i/c)\cos\theta_1}\nu_o, \tag{14.15}$$

where θ_1 and θ_2 are the angles formed by the vectors $\vec{v_p}$ and $\vec{v_i}$ with the line connecting the source and the receiver.

Chapter 15

Photometry

15.1 Spectral Sensitivity of the Eye

Photometry, in a narrower sense, is a branch of optics that deals with measuring the energy of electromagnetic waves in the visible part of the spectrum, as well as the quantities associated with it. The measurement of energy across all regions of this spectrum is handled by a special branch of physics called radiometry. Standard photometry focuses on measurements in the wavelength range from $100\,nm$ to $1\,mm$, while the visible wavelength range is from 380 to $780\,nm$.

Photometry deals with the measurement of quantities related to point light sources. Since such sources do not exist, a point source is considered to be a measurement in which the detector is at least 10 times farther away than the linear length of the source.

The human eye is not equally sensitive to all wavelengths of electromagnetic radiation in the visible part of the spectrum. Therefore, two types of photometric measurements must be performed – **energetic** and **visual (photometric)**. For example, the human eye perceives a much stronger stimulus at a wavelength of $\lambda = 555\,nm$ than at other wavelengths, for the same intensity of the light source. Radiation outside the visible spectrum generally does not cause a visual sensation, although it can be very harmful to the eye.

To assess the degree of light sensation produced on the eye by rays of different wavelengths, the concept of the relative spectral sensitivity of the human eye to monochromatic light (V_λ) was introduced, and the function $V_\lambda = f(\lambda)$ was measured across the entire visible part of the spectrum.

The value of V_λ for a given wavelength λ is obtained as follows. The eye is exposed to a luminous flux at the wavelength $\lambda_{max} = 555\,nm$ nm from a source with power $P_{\lambda max}$. Then the power of the source P_λ is measured, which under the same conditions produces the same visual sensation in the eye at wavelength λ as the sensation from the source with power $P_{\lambda max}$, which can be determined very precisely using a suitable visual photometer with both sources acting simultaneously. The value of V_λ represents the ratio of these two radiant powers:

$$V_\lambda = \frac{P_{\lambda max}}{P_\lambda}. \tag{15.1}$$

Figure 15.1 shows two curves. Curve a was recorded as described under daylight conditions (above $10\,lx$). Curve b was recorded under dim night light (below $0.1\,lx$). In photometry, curve a is used. Curve a has a maximum at $555\,nm$, while curve b has a maximum at a wavelength of $505\,nm$.

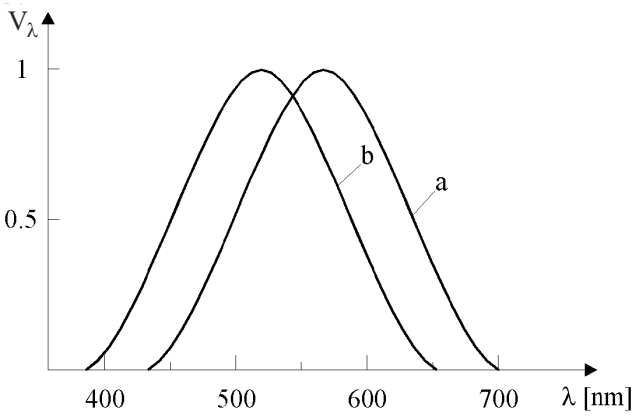

Figure 15.1: Relative spectral sensitivity of the average human eye to monochromatic light (V_λ) as a function of wavelength (λ) a) sensitivity above $10\,lx$ (daytime vision) b) illumination below $0,1\,lx$ (night vision)

It should be noted that these curves were recorded multiple times, the average values corresponding to the human eye were determined and then adopted as the international standard, thus completely removing any traces of subjectivity in photometry.

Due to the mentioned phenomena in photometry, in addition to the energy units, visual (lighting) units are also introduced, which are based on the sensitivity of the human eye, i.e., on the function $V_\lambda = f(\lambda)$.

15.2 Photometric Quantities and Units

Photometric quantities are divided into two main groups. The first group includes quantities that relate to the emission of light energy from a source (emitter). These are: luminous flux, luminous intensity, illuminance or emissive power or emittance, and the brightness of a light source or luminance.

The second group includes quantities related to the illuminated surface (detector). These include the following quantities: illuminance, luminous emittance, brightness of the illuminated surface (luminance), luminous exposure, and reflective power.

One of the most important quantities in photometry is the amount of light (light energy) Q. The energy quantity for the amount of light energy is $J = Ws$, and the visual one is $lm\,s$.

Luminous flux (radiant flux). Let a wavefront spread from a point light source O (Fig. 15.2). At some distance from the source, there is a surface corresponding to the solid angle $d\Omega$. The luminous flux $d\Phi$ in a given direction is defined as the amount of light energy transmitted through the given surface per unit time, i.e., within the given solid angle $d\Omega$. It can be mathematically expressed as follows:

$$d\Phi = \frac{dQ}{dt}, \tag{15.2}$$

where dQ is the amount of light (light energy) in the solid angle $d\Omega$.

Luminous flux can also be defined as the power of radiation through a surface dS. If the medium does not absorb light energy, then the same flux passes through any cross-section of the cone shown in Figure 15.2. The total luminous flux Φ emitted from the source O in all directions is obtained by integrating the elemental fluxes in individual directions

$$\Phi = \int_V d\Phi. \tag{15.3}$$

The energy unit for luminous flux is the watt (W), and visual unit is *lumen* (lm).

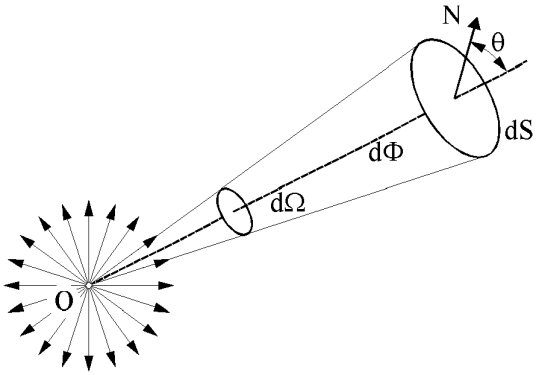

Figure 15.2: With the definition of luminous flux

Luminous (light) intensity I represents the energy emitted per unit of time per unit of solid angle. Light intensity is the ratio between the luminous flux that a source emits in a given direction into an element of solid angle and the element of solid angle:

$$I = \frac{d\Phi}{d\Omega}. \tag{15.4}$$

If the luminous flux is the same in all directions (an isotropic light source), i.e., if a uniform luminous flux Φ emitted by a light source into the solid angle Ω is considered, the above expression can be represented in a simpler form:

$$I = \frac{\Phi}{\Omega} = \frac{\Phi}{4\pi}. \tag{15.5}$$

Light sources are most often not isotropic (emission differs in various directions). Therefore, the average light intensity is defined as the ratio of the total luminous flux Φ to the solid angle Ω

$$I_{sr} = \frac{\phi}{\Omega}. \tag{15.6}$$

The energy unit for light intensity is W/sr, and the visual unit is candela ($cd = lm/sr$).

Illuminance is a measure of the intensity of light falling on a given surface. It is defined as the ratio of the luminous flux falling on the given surface to the area of that surface itself (Fig. 15.3):

$$E = \frac{d\Phi}{dS}. \tag{15.7}$$

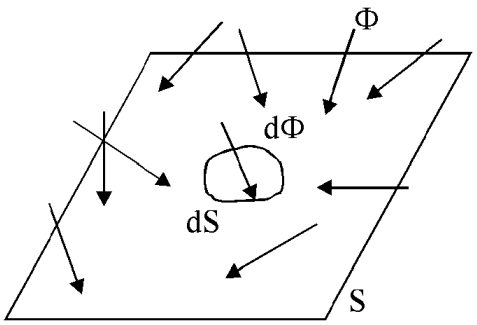

Figure 15.3: The surface on which the light falls, whose flux is Φ

The illuminance of a surface represents the density of the light flux on it. If the luminous flux is not uniform, then the average illuminance is determined as the ratio of the total flux to the area on which it falls, i.e.,

$$E_{sr} = \frac{\Phi}{S}. \tag{15.8}$$

Illuminance can be expressed in terms of light intensity, using equation (15.4), from which it follows that $d\Phi = Id\Omega$. From Figure 15.4, it can be seen that the solid angle is equal to

$$d\Omega = \frac{dS_n}{r^2} = \frac{dS\,\cos\theta}{r^2}, \tag{15.9}$$

and

$$d\Phi = Id\Omega = \frac{IdS\,\cos\theta}{r^2}. \tag{15.10}$$

By substituting this expression into (15.7), the final expression for the illumination of the surface dS, which is at a

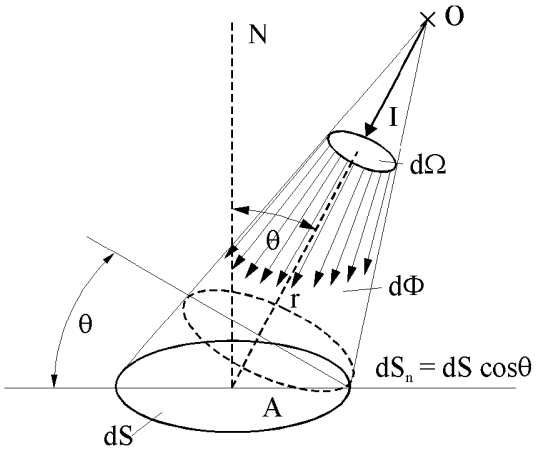

Figure 15.4: With the definition of the illumination of a given surface

distance r from the light source, is obtained

$$E = \frac{I}{r^2}\cos\theta, \tag{15.11}$$

where θ is the angle between the normal to the surface being illuminated and the direction OA in Figure 15.4. Expression (15.11) represents Lambert's law, which shows that the illumination of the surface at point A is proportional to the cosine of the angle θ and inversely proportional to the square of the distance between the observed point and the light source. If light falls normally on the surface, then $\theta = 0$, and therefore

$$E = \frac{I}{r^2}. \tag{15.12}$$

The energy unit for illuminance is W/m^2, and the visual unit is *lux* ($lx = lm/m^2$).

If the illumination at a certain point comes from multiple light sources, then the total illumination is equal to the algebraic sum of the illuminations coming from those sources:

$$E = E_1 + E_2 + E_3 + ... + E_n = \sum_{i=1}^{n} E_i. \tag{15.13}$$

Chapter 16

Geometric Optics

16.1 Laws of Geometric Optics

To solve a number of problems such as the formation of a light beam, the formation of images, and similar issues, it is not necessary to use the wave theory of light. In the first approximation, it can be assumed that light in homogeneous and isotropic media travels along certain lines called light rays. They represent the directions of light energy propagation. Geometrical methods are applied to solve problems of this type, and the branch of optics that deals with this is called *geometric optics (ray optics)*.

Geometric optics represents the limiting case when the wavelength of the light used approaches zero ($\lambda \to 0$). In this case, phenomena such as interference, diffraction, and polarization can be neglected. The basis of geometric optics consists of four laws. The first law is *the law of rectilinear propagation of light*. It applies to homogeneous media and is clearly observed in the formation of shadows by opaque objects.

The second law is *the law of independent propagation of light rays*. It can be formulated as follows: two light rays interact at the point of their intersection, and then continue to propagate independently.

The third law, or *the law of reflection of light rays from interfaces (boundary surfaces)*, states: a) *the angle of incidence of the light ray is equal to the angle of reflection* (eq. 13.39), b) *the incident and reflected rays lie in the same plane with the normal to the boundary surface at the point of incidence*. This law applies in cases when rays are reflected from smooth and optically polished surfaces. These are surfaces on which the roughness is significantly smaller than the wavelength of the light used. If the surface is rough, light rays are reflected diffusely, in all directions, i.e., they are scattered, and then *this law doesn't apply*.

The fourth law, or *the law of refraction of light at a flat boundary surface (interface)* between two stationary, homogeneous, and isotropic media, states: a) *The ratio of the sine of the angle of incidence α to the sine of the angle of refraction β is equal to the ratio of the speed of light in the incident medium v_1 to the speed of light in the refractive medium v_2:* $\sin\alpha/\sin\beta = v_1/v_2$ (eq. 13.40) , b) *the incident and refracted rays lie in the same plane with the normal to the boundary surface passing through the point of refraction*. If the speed of light in the medium from which the ray comes is greater than the speed in the medium into which the ray enters, the ray bends toward the normal, and vice versa.

In optics, *the refractive index of a medium n* is introduced. It is a dimensionless value that describes how light (or other radiation) propagates through a specific medium. It is one of the most fundamental properties of a transparent optical material. The refractive index is defined as the ratio of the speed of light in a vacuum (c) to the speed of light in the given medium (v):

$$n = \frac{c}{v}. \tag{16.1}$$

By using the refractive index, the law of refraction of light (Snell's law) can be written in the form:

$$\frac{v_1}{v_2} = \frac{\sin\alpha}{\sin\beta} \;\Rightarrow\; n_1 \sin\alpha = n_2 \sin\beta, \tag{16.2}$$

where $n_1 = v_1/c$ is the refractive index of the first (incident) medium, and $n_2 = c/v_2$ of the second (refractive) medium. The α and β are the angle of incidence and the angle of refraction, respectively, relative to the normal at the interface between media.

16.2 Refraction of Light through a Prism

Figure 16.1 shows a triangular prism made of optical material with polished sides AC and AB (optical prism), with an apex angle θ (prism angle). The refractive index of the prism material is n. The prism is in air, which refractive index n_o is approximately 1 (as in a vacuum: $n_o \approx 1$). A monochromatic light beam falls on the boundary surface AC, at point O, at an angle α_1 and is refracted towards the normal at an angle β_1. At the second boundary surface AB, at point O', it falls at an angle β_2 and is refracted away from the normal at an angle α_2. After double refraction, the beam deviates by an angle δ, which is called *the angle of deviation*.

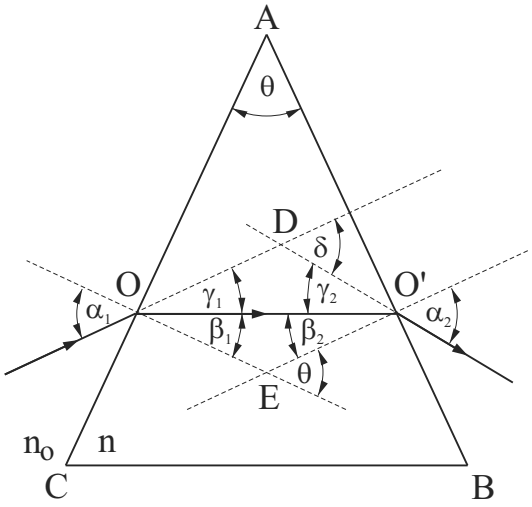

Figure 16.1: Refraction of a light beam through a prism

From the triangle $OO'D$, it follows:

$$180^o = (180^o - \delta) + \gamma_1 + \gamma_2, \tag{16.3}$$

$$180^o = (180^o - \delta) + (\alpha_1 - \beta_1) + (\alpha_2 - \beta_2), \tag{16.4}$$

$$\delta = (\alpha_1 + \alpha_2) - (\beta_1 + \beta_2). \tag{16.5}$$

From the triangle $OO'E$, it follows:

$$180^o = (180^o - \theta) + \beta_1 + \beta_2. \tag{16.6}$$

$$\theta = \beta_1 + \beta_2, \tag{16.7}$$

From (16.5) and (16.7):

$$\delta = (\alpha_1 + \alpha_2) - \theta. \tag{16.8}$$

The minimum deviation δ_{min} occurs when $\alpha_1 = \alpha_2$, i.e., when the light after refraction at the boundary surface AC moves parallel to the base of the prism (minimum deviation condition). In this case, the angle of deviation is

$$\delta_{min} = 2\alpha_1 - \theta, \tag{16.9}$$

$$\alpha_1 = \frac{\delta_{min} + \theta}{2}. \tag{16.10}$$

The law of refraction for point O can be written as

$$n_o \sin\alpha_1 = n \sin\beta_1, \tag{16.11}$$

and for point O':

$$n \sin\beta_2 = n_o \sin\alpha_2. \tag{16.12}$$

If α, θ and n are small, and if the minimum deviation condition of the light ray is fulfilled, then the sines of the incident and refracted angles are approximately equal to the angles themselves, so the law of refraction of light on the two observed surfaces can be written in the form ($n_o \approx 1$):

$$\alpha_1 \approx n\,\beta_1, \quad \alpha_2 \approx n\,\beta_2 \tag{16.13}$$

$$\alpha_1 + \alpha_2 = n\,(\beta_1 + \beta_2) = n\,\theta. \tag{16.14}$$

By substituting the values for $\alpha_1 + \alpha_2$ from expression (16.14) into (16.8), the deviation angle δ of the light ray is obtained

$$\delta = \theta(n - 1). \tag{16.15}$$

As can be seen, the deviation angle is a function of the prism angle and the refractive index of the prism material.

16.3 Light Dispersion. Spectra

When considering the refraction of light through a prism, it was assumed that light contains rays of only one wavelength, i.e., that it is *monochromatic*. However, light is in most cases considered as a collection of a larger number of rays of different wavelengths, which is why it is called *polychromatic light*. The speed of light in a vacuum is the same regardless of wavelength.

However, when light propagates through different transparent materials, its speed, and consequently the refractive index, are functions of wavelength ($v = \lambda/\nu$, but, as previously mentioned, the frequency ν does not change during

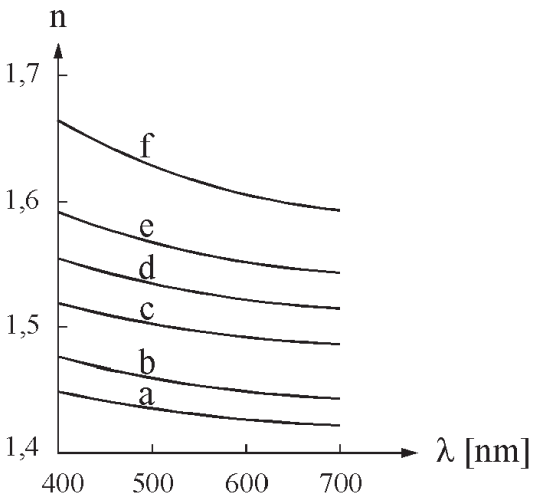

Figure 16.2: Refractive index as a function of wavelength for some optical materials

refraction). Therefore, when polychromatic light passes through such materials, it is dispersed according to wavelength. This phenomenon is called *dispersion*.

Figure 16.2 shows the dependence of the refractive index on the wavelength of light for several optical materials. The figure shows that as the wavelength increases, the refractive index decreases. This decrease depends on the type of optical material.

Let a beam of white polychromatic light fall on an optical prism (Fig. 16.3). Red light has the longest wavelength in the spectrum, while violet has the shortest. The greatest light propagation speed v corresponds to the longest wavelength λ because $v = \lambda/\nu$ (the ν is unchanged during refraction).

This means red light will travel the fastest, and thus have the lowest refractive index ($n = c/v$), bending the least (the largest angle β in expression (16.2)). Namely, a larger angle of refraction β means less bending of light when passing into another medium. Violet light bends the most, while the other colors (indigo, blue, green, yellow, orange) are in between.

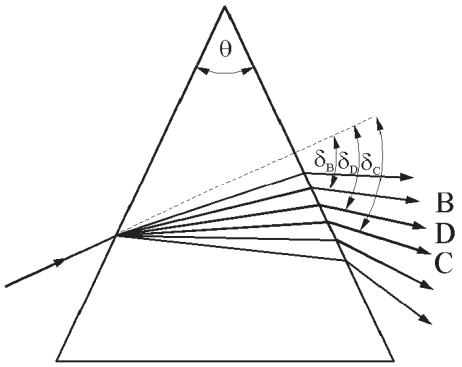

Figure 16.3: Dispersion of polychromatic light using an optical prism

When decomposing polychromatic light into a spectrum, two quantities are characteristic: *the measure of dispersion* and *the dispersive power*. The measure of dispersion, or just dispersion, represents the difference between two angles of deviation. Let δ_B be the angle of deviation of light B, and δ_C the angle of deviation of light C. Then the difference $\delta_C - \delta_B$ represents the measure of dispersion. If the incidence angle of the polychromatic light, the prism's refractive index, and the prism angle are small, and the condition of approximately minimum deviation of the light ray is met, then expression (16.15) can be applied for the angles of deviation:

$$\begin{aligned} \delta_B &= (n_B - 1)\theta, \\ \delta_C &= (n_C - 1)\theta. \end{aligned} \tag{16.16}$$

The following expression is obtained for the measure of dispersion:

$$\delta_C - \delta_B = (n_C - n_B)\theta. \tag{16.17}$$

Dispersion power represents the ratio of the dispersion measure $\delta_C - \delta_B$ to the deviation angle of the central ray δ_D in the spectrum

$$\omega = \frac{\delta_C - \delta_B}{\delta_D} = \frac{n_C - n_B}{n_D - 1}. \tag{16.18}$$

The shape of the spectrum depends on the nature of the light source. Three basic types of spectra are distinguished: *continuous*, *line*, and *band*.

In a continuous (uninterrupted) spectrum, all colors (wavelengths) are present, with the transition from one color to another being gradual. Continuous spectra arise as a result of the mutual interaction of a large number of molecules

and atoms in the radiating body. The main role in emission is played by the chaotic, oscillatory, and rotational motion of these particles due to increased temperature.

The line spectrum consists of a series of discrete lines that represent specific wavelengths. The width of these lines is not infinitely small, but includes a certain range of wavelengths. This is the case for gases at low pressures. Emission occurs due to the transition of bound electrons from higher to lower energy levels (orbits) in the atom.

The band spectrum consists of a large number of lines arranged in specific groups. When observing such a spectrum with low-resolution spectroscopic instruments, they appear as bands separated by dark areas. Each spectral band corresponds to a specific range of wavelengths. Band spectra are emitted by individual (free) molecules. The emission is caused by electronic transitions as well as by vibrational and rotational movements of atoms within the molecule.

The study of spectra plays a very important role in explaining the processes that occur in atoms and molecules, as well as in understanding the structure of materials.

16.4 Total Internal Reflection. Optical Fibers

Total reflection (complete reflection) is the phenomenon of light being reflected entirely from the boundary surface (interface) that separates two media with different refractive indices. It occurs when light rays come from an optically denser medium (with higher refractive index) toward the boundary surface. An example of total reflection is shown in Figure 16.4.

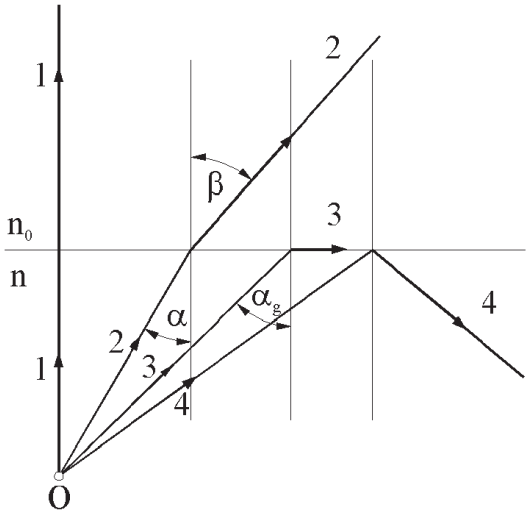

Figure 16.4: With an explanation of total reflection

Light rays from a point light source O, which is located in a medium with a refractive index n, strike the boundary surface that separates this medium from a medium with a refractive index n_o ($n > n_o$). When passing from one medium to another, the rays are refracted according to the law:

$$n \sin\alpha = n_o \sin\beta, \tag{16.19}$$

where $\beta > \alpha$, since $n > n_o$.

By changing the angle of incidence α, the angle of refraction can vary from 0^o to 90^o. The angle of incidence α_g, at which the angle of refraction is 90^o, is called *the limiting* or *critical angle* of total internal reflection. If the angle of incidence is greater than α_g, the light is reflected back into the same medium. This is referred to as *total internal reflection.*

The boundary angle of total reflection can be determined from expression (16.19), assuming that $\beta = 90^o$

$$\sin\alpha_g = \frac{n_o}{n}. \tag{16.20}$$

Total internal reflection in optical prisms is used in optical instruments. Thanks to this phenomenon, it is possible to bend a light beam by 90^o (Fig. 16.5.a), 180^o (Fig. 16.5.b), and other angles.

Optical fibers or *light guides* operate on the principle of total internal reflection. An optical fiber is a dielectric thread with a circular cross-section, consisting of two layers: an inner, transparent layer called *the core*, and an outer layer called *the fiber cladding* (Fig. 16.6). The refractive index of the core n_1 is slightly higher than the refractive index of the cladding n_2. Light rays entering the fiber core at a small angle relative to the zz' axis will encounter a less optically dense medium at the core-cladding interface, at an angle greater than the critical angle, which will result in total internal reflection. In this way, the amount of light that entered the fiber under these conditions will propagate along the fiber without noticeable losses. The fundamental limitation in transmitting information through such an optical fiber arises from the spreading of short light pulses due to the different propagation times of light along the fiber axis and at certain angles relative to the axis. From this figure, it can be seen that two light beams entering this optical fiber at the same time will not exit it at the same time.

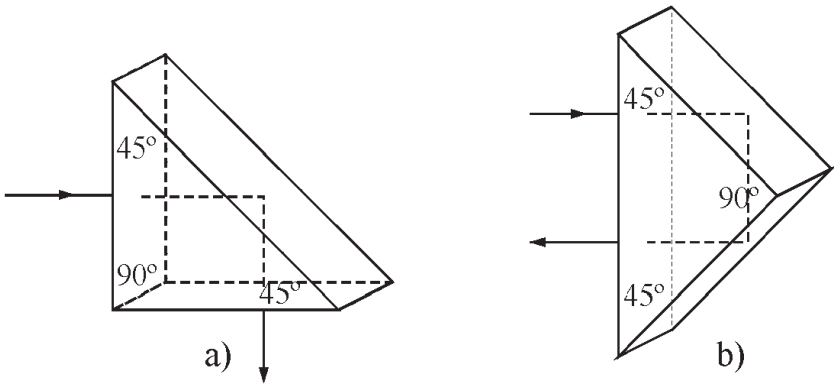

Figure 16.5: Total internal reflection of light using a prism

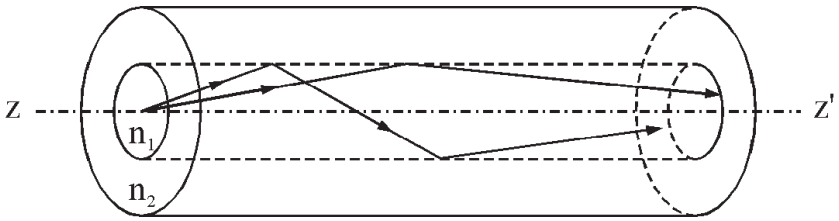

Figure 16.6: Light propagation in a fiber with a constant refractive index core

The transmission characteristics of an optical fiber can be significantly improved by using fibers whose core refractive index continuously decreases along the radial coordinate. These are *graded-index optical fibers*. Since the refractive index of the core changes in a graded-index fiber, rays in the core propagate along curved paths. A typical appearance of two groups of data (rays entering in the plane of the optical axis and those that miss it) is shown in Figure 16.7. It is also characteristic for them that the path of the ray increases with increasing angle of incidence. Since it propagates through an optically less dense medium, the speed of that ray is greater than the speed of the ray propagating near the optical axis of the fiber. In this way, the longer path length is compensated, so that the beams exit the optical fiber at the same time. Information is transmitted using optical fibers based on light modulation (most commonly by changing

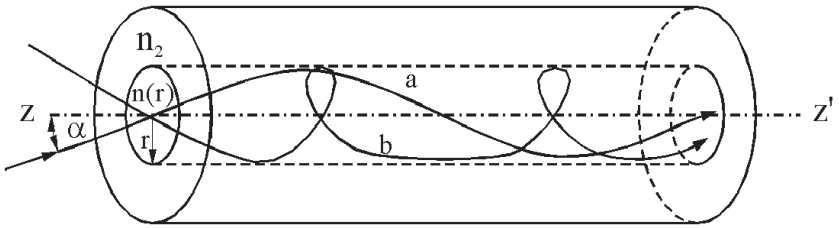

Figure 16.7: Propagation of rays in a fiber with a graded-index change in the core
a) rays that lie in the plane of the optical axis
b) rays that are off-axis relative to the optical axis

the intensity). Properly manufactured fibers can carry information over much greater distances than wired systems. Moreover, a significantly larger number of pulses (information) can be transmitted through these fibers per unit of time than through conventional electrical signal conductors.

16.5 Spherical Lenses and Their Classification

A spherical lens is most commonly understood as a body of transparent optical material, bounded by two optically polished spherical surfaces, which act as the refractive surfaces for light rays. If lenses have only two refractive surfaces (usually both spherical or one flat and the other spherical), they are called *simple lenses*.

Figure 16.8 shows the types of simple lenses and their basic elements: the line zz' that passes through the centers of curvature is called *the optical axis* of the lens; R_1 and R_2 are the radii of curvature of the lens surfaces. The boundary surfaces of lenses can be convex, concave, or flat (planar), and accordingly, lenses can be biconvex, plano-convex, and concave-convex, or biconcave, plano-concave, and convex-concave.

Lenses can be divided into *thin* and *thick*. A lens is thin if its thickness is significantly smaller than the radii of the spherical surfaces that form it. For such a lens, the deviation of a light ray as it passes through the center can be neglected. If this condition is not met, the lens is called thick. Lenses in which refracted rays intersect the optical axis are called *converging lenses*, while lenses where the refracted rays diverge from the optical axis are called *diverging*

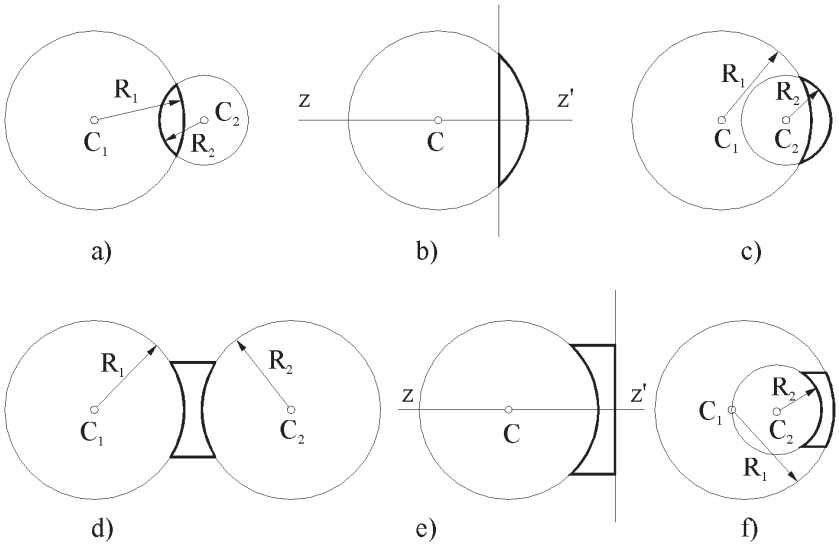

Figure 16.8: Types of lenses: a) biconvex, b) plano-convex, c) concave-convex, d) biconcave, e) plano-concave, and f) convex-concave

lenses. This applies *only* when the refractive index of the lens material is greater than the refractive index of the surrounding medium.

16.6 Thin Lens Equation

One of the forms of the equation for a thin lens is:

$$\frac{1}{p}+\frac{1}{l}=(\frac{n}{n_o}-1)(\frac{1}{R_1}-\frac{1}{R_2}), \tag{16.21}$$

where p is the distance of the object from the optical center of the lens, l is the distance of the image from the optical center of the lens, n is the refractive index of the material from which the lens is made, n_o is the refractive index of the medium in which the lens is placed, R_1 and R_2 are the radii of curvature of the first and second lens surfaces.

The focal length f of a thin lens represents the distance from the lens to the image when the object is at infinity. Namely, if $p \to \infty$ then $f \to l$, so equation (16.21) takes the following form:

$$\frac{1}{f}=(\frac{n}{n_o}-1)(\frac{1}{R_1}-\frac{1}{R_2}). \tag{16.22}$$

Based on this, equation (16.21) can be written in a simpler form:

$$\frac{1}{p}+\frac{1}{l}=\frac{1}{f}. \tag{16.23}$$

Table 16.1 presents the conventionally adopted signs for thin lenses.

Table 16.1: Conventionally adopted signs for thin lenses

p is + if the object is in front of the lens
p is – if the object is behind the lens
l is + if the image is formed behind the lens
l is – if the image is formed in front of the lens
R_1, R_2 is + if the center of the curve is behind the lens
R_1, R_2 is – if the center of the curve is in front of the lens

The focal point (F) is the specific point on the optical axis where light rays parallel to the optical axis converge (meet) after passing through a convex lens or the point from which they appear to diverge after passing through a concave lens. This is illustrated in Fig. 16.9a for a biconvex (positive f) and in Fig. 16.9b for biconcave (negative f) lens. A thin lens is characterized by two focal points (two foci), formed at the intersection of refracted rays, on the left and right sides of the lens.

Lenses can be classified as converging or diverging based on the properties of the focal points. If the focal points are real, i.e., if parallel rays converge at a real point after passing through the lens (Fig. 16.9.a), the lenses are converging or positive. If the focal points are imaginary, i.e., if parallel rays diverge after refraction (Fig. 16.9.b), the lenses are diverging or negative.

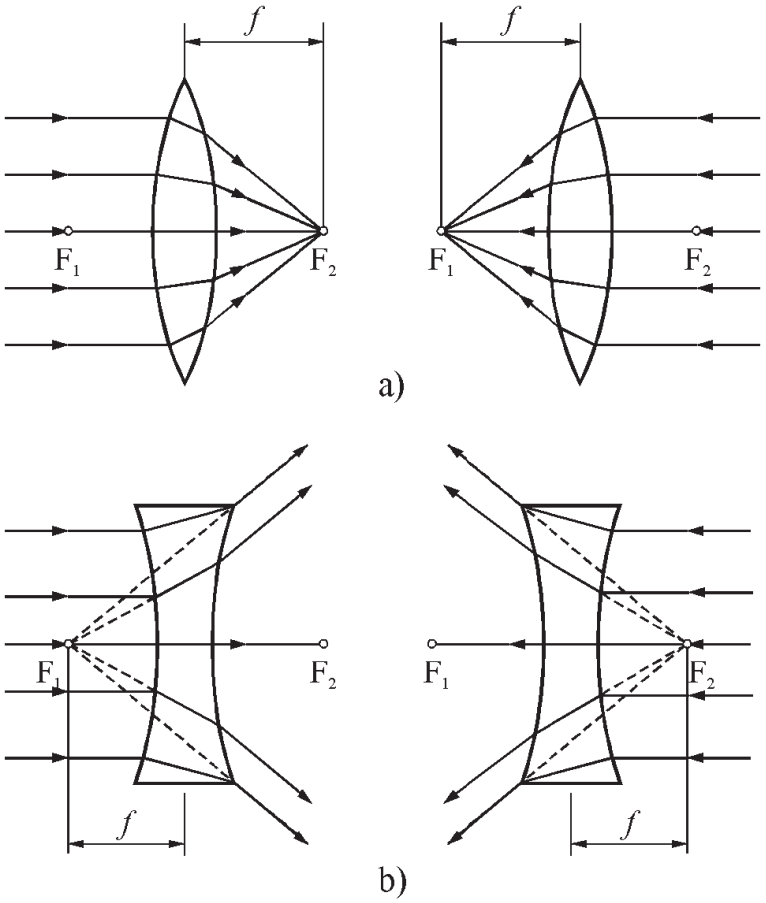

Figure 16.9: Focal lengths (f) and focal points (F) for convex lens (a) and concave lens (b)

As a measure of the refractive power (the light bending strength) of a lens, a quantity called *optical power (strength)* ω is used, expressed in the following way:

$$\omega = \frac{n}{f}, \tag{16.24}$$

where n is the refractive index of the medium in which the lens is located. For a lens in air ($n \approx 1$), the optical power is

$$\omega = \frac{1}{f}. \tag{16.25}$$

Optical power is expressed in diopters (D) when the focal length is expressed in meters. Converging lenses are indicated by a positive value of optical power, while diverging lenses are indicated by a negative value.

16.7 Image Formation by Thin Lenses

Figure 16.10 shows the scheme for image formation of an object of finite dimensions using a thin biconvex (converging) lens. The object AB is located in a plane perpendicular to the optical axis zz', at a distance p from the lens optical center. The image $A'B'$ is obtained in a plane also perpendicular to the optical axis zz', and is located at a distance l from the lens. The distances p and l are related by equation (16.23). To obtain the image, for example of point B, the path of the characteristic rays is traced. The first ray (BC) travels parallel to the optical axis and after refraction passes through the focal point (focus). The second ray (BF) passes through the focal point F and after refraction through the lens travels parallel to the optical axis. The third ray (BT), passing through the otpical center T of the lens, does not refract (bend), since for the thin lens, the displacement of light rays inside the lens can be neglected. After exiting the lens, the rays intersect at point B', which represents the image of point B. Using a similar procedure, the other points of the object AB can also be located. Their set represents the image $A'B'$. The image formation using a biconcave (diverging) lens is shown in Fig. 16.11.

Whether the image will be magnified or reduced compared to the object depends on the distances between the object and the lens. The ratio of the linear dimensions of the image to the object is called *the linear* or *transverse magnification* U of a thin lens:

$$U = \frac{A'B'}{AB} = \frac{l}{p}. \tag{16.26}$$

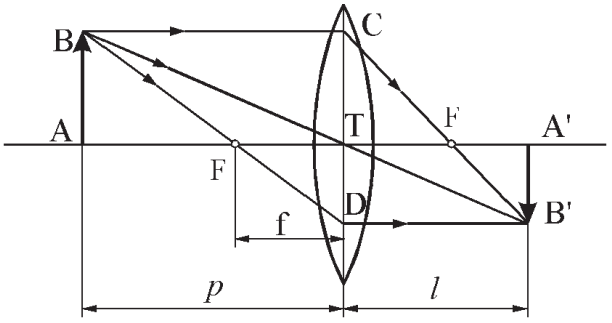

Figure 16.10: Forming an image with a thin converging lens

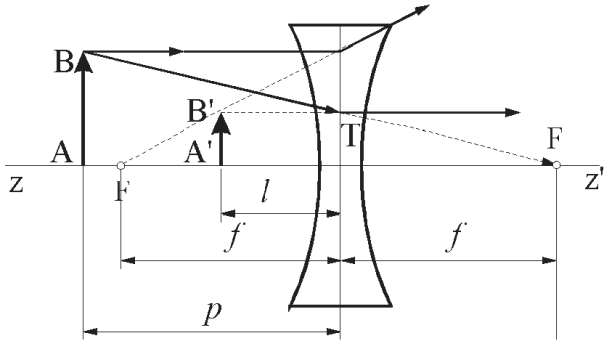

Figure 16.11: Forming an image with a thin diverging lens

16.8 Optical Instruments

16.8.1 Magnifying Glass

A magnifying glass (magnifier) is the simplest optical instrument, most often consisting of a single convex lens, and is used for observing small objects at close distances.

It is known that the size of the image on the retina of the eye depends on the angle at which the object is viewed. To make the image as large as possible, and thus the viewing angle of the image, it is necessary to bring the object as close to the eye as possible. However, the eye cannot sharply focus on objects at distances shorter than *the least distance of distinct vision (the near point).* Therefore, if the object is at the near point, the eye sees it under the maximum angle (Fig. 16.12.a). In practice, the near point is usually taken as 25 cm (Fig. 16.12.a). To increase the angle at which an

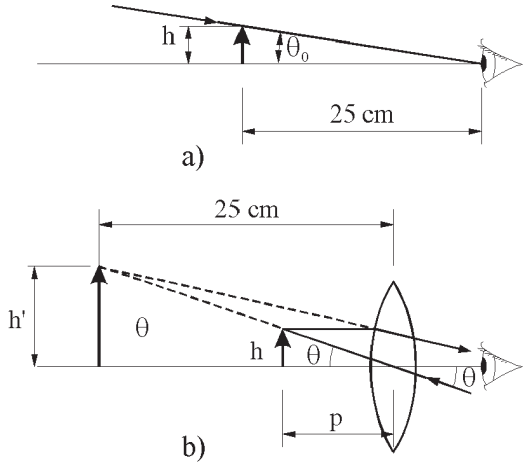

Figure 16.12: Observation of objects
a) without a magnifier,
b) with a magnifier

object is seen, it is necessary to place a converging lens in front of the eye with the object located at point O, which is very close to the lens's focus (Fig. 16.12.b). In this case, the lens forms a virtual, upright, and magnified image, as shown in the figure. Thus, the lens increases the angle at which the object is clearly seen. The angular magnification W is defined as the ratio of the angle under which the object is seen with a magnifying glass (angle θ in Fig. 16.12.b) to the angle under which it is seen without the magnifying glass (angle θ_o in Fig. 16.12.a):

$$W = \frac{\theta}{\theta_o}. \tag{16.27}$$

16.8.2 Optical Microscope

To observe very small objects at close distances, significant magnification is required, which cannot be achieved with a magnifying glass. For this purpose, a more complex optical instrument is used, called a microscope.

A simplified diagram of an optical microscope is shown in Fig. 16.13. It consists of an objective lens with a very short focal length f_{ob} ($f_{ob} < 1\,cm$) and an ocular lens (eyepiece or magnifier) with a focal length f_{ok} of a few centimeters.

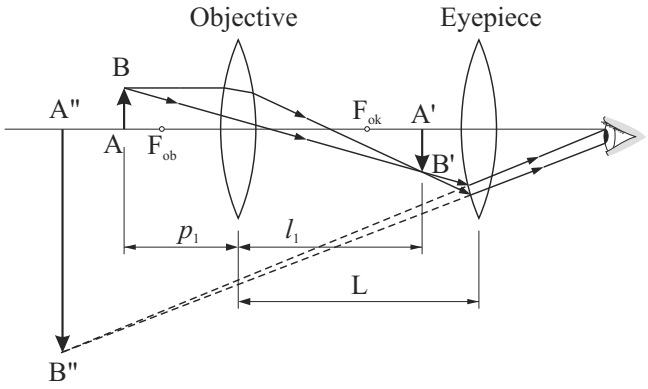

Figure 16.13: Optical microscope

These two lenses are separated by a distance L, which is much greater than both focal lengths f_{ob} and f_{ok}. The object AB is placed near the focal point of the objective. A real and inverted image $A'B'$ is formed by the objective near the focal point of the eyepiece. The eyepiece then forms a virtual image $A''B''$ from $A'B'$, which is also inverted relative to the object.

Chapter 17

Wave (Physical) Optics

17.1 Interference of Light

17.1.1 Generation of Coherent Light. Interference Image

Only the interference of mechanical waves has been considered. It is natural to assume that this phenomenon is also characteristic of light waves, which belong to the group of electromagnetic waves (question for the reader: which waves are part of the electromagnetic spectrum?) Experiments have shown that interference effects can also appear in light waves if those waves are coherent, i.e., if they have a constant phase difference at the same frequencies, approximately the same intensities, and a defined orientation of the polarization planes.

It has been shown that two bright (luminous) bodies can never be used as coherent sources of light. The light emitted by a luminous source (for example, a filament lamp) consists of a set of electromagnetic waves emitted by individual particles (atoms and molecules) of the body. The emission conditions of these particles change rapidly and continuously. For light to be coherent, the waves emitted by all the particles of one source must maintain a constant phase difference relative to the waves emitted by particles of another source over a longer period of time. Such a phenomenon is impossible, except in lasers. For this reason, certain methods are used to obtain coherent light sources. One of these methods is the "splitting" of light emitted from a single light source (Jung's method). This splitting can be performed, for example, using a screen with two small openings (Fig. 17.1). According to Huygens' principle, secondary light sources O_1 and O_2 are formed at the openings in the screen z from the primary light source O. In that case, every change in the phase of the wave emitted by source O is also followed by the waves emitted by the secondary sources O_1 and O_2. Therefore, for the waves emitted by sources O_1 and O_2, the phase difference remains constant at all times.

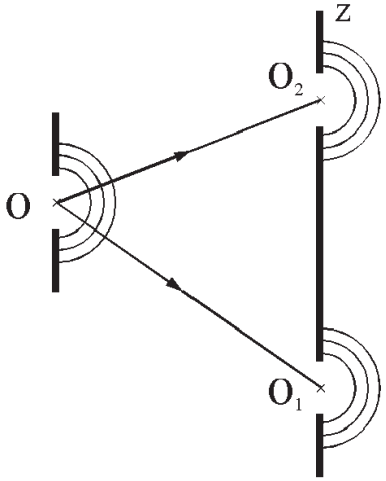

Figure 17.1: Generation of coherent light waves using Jung's method

The second method for obtaining coherent light is the Fresnel method, which is based on the reflection of light from two flat mirrors joined at an angle slightly less than 180^o (Fig. 17.1). Such an optical system is called 'Fresnel mirrors.' The points O_1 and O_2 of the source O serve as coherent light sources.

The result of interference at point A will depend on the path difference of the waves from sources O_1 and O_2. If *the path difference* of the rays $\Delta l = O_1A - O_2A$ is equal to an integer number of wavelengths of light

$$\Delta l = n\lambda; \quad n = 0, 1, 2, ..., \tag{17.1}$$

then, it will be a maximum light intensity or an interference maximum at point A.

If the path difference is equal to an odd number of half wavelengths:

$$\Delta l = (2n + 1)\frac{\lambda}{2}; \quad n = 0, 1, 2, ..., \tag{17.2}$$

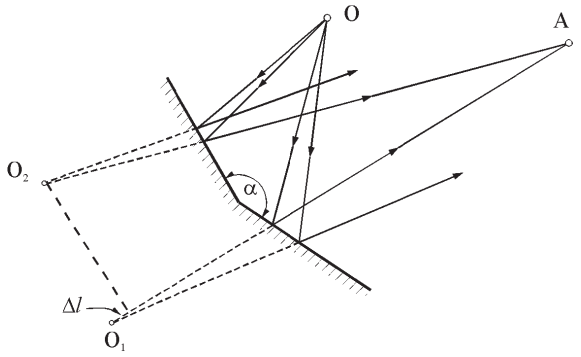

Figure 17.2: Obtaining coherent light waves using the Fresnel method

an interference minimum will occur. If the amplitudes of the waves are equal, complete cancellation of light will occur. For different wave amplitudes, at point A complete cancellation will not occur; instead, there will be a reduction in light intensity.

Let an interference pattern (image) be formed on the screen z by two coherent light sources of monochromatic light O_1 and O_2. The distance between these sources is d, and the perpendicular (normal) distance from the sources to the screen is L, where $d \ll L$ (Fig. 17.3). It is necessary to determine the distance x between points D and A where the interference maxima (minima) occur.

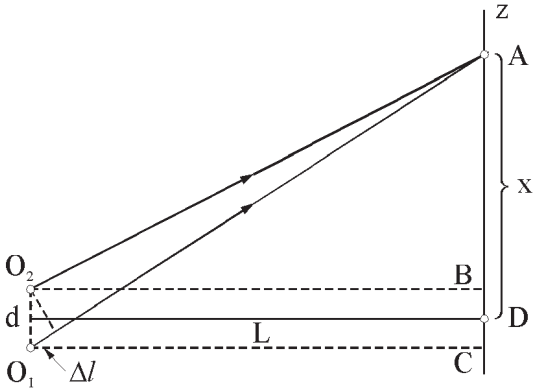

Figure 17.3: With an explanation of the interference pattern

From triangles ACO_1 and ABO_2:

$$(AO_1)^2 = L^2 + (x + \frac{d}{2})^2 \tag{17.3}$$
$$(AO_2)^2 = L^2 + (x - \frac{d}{2})^2$$

$$(AO_1)^2 - (AO_2)^2 = 2xd. \tag{17.4}$$

Expression (17.4) can be written in the form

$$(AO_1 - AO_2)(AO_1 + AO_2) = 2xd. \tag{17.5}$$

Since it is

$$AO_1 - AO_2 = \Delta l; \quad AO_1 + AO_2 \approx 2L$$

substitution into expression (17.5) for the distance of the interference maximum or minimum is obtained

$$x = L\frac{\Delta l}{d}. \tag{17.6}$$

Using expressions (17.1) and (17.2), an expression is obtained for finding the distance between interference maxima of light

$$x = n\frac{\lambda L}{d}, \tag{17.7}$$

and interference minima of light

$$x = (2n+1)\frac{\lambda L}{2d}, \tag{17.8}$$

from the point D.

17.2 Diffraction of Light

Diffraction is a phenomenon characteristic of both mechanical and light waves. In physics, narrow slits or holes are most commonly used to demonstrate diffraction. Here, the diffraction of light through one and two parallel slits will be examined in more detail. The slits are most often rectangular, with a width much smaller than the length and on the order of the wavelength of light.

17.2.1 Diffraction of Light at a Single Slit

Let a plane wave front of monochromatic light fall on a screen B with a narrow rectangular slit (aperture) of width a (Fig. 17.4). Most of the rays from the beam pass through the slit without deviation (Fig. 17.4.a). After refraction through the lens S, the rays converge into a narrow band (stripe) on the screen z, which is perpendicular to the drawing plane and lies in the focal plane of the lens. The projection of this stripe onto the drawing plane is symbolically represented by the point O. From the standpoint of geometric optics, this stripe represents the image of the slit. The maximally illuminated stripe passes through point O because the rays along the slit are coherent, and the path difference from the slit to point O is zero.

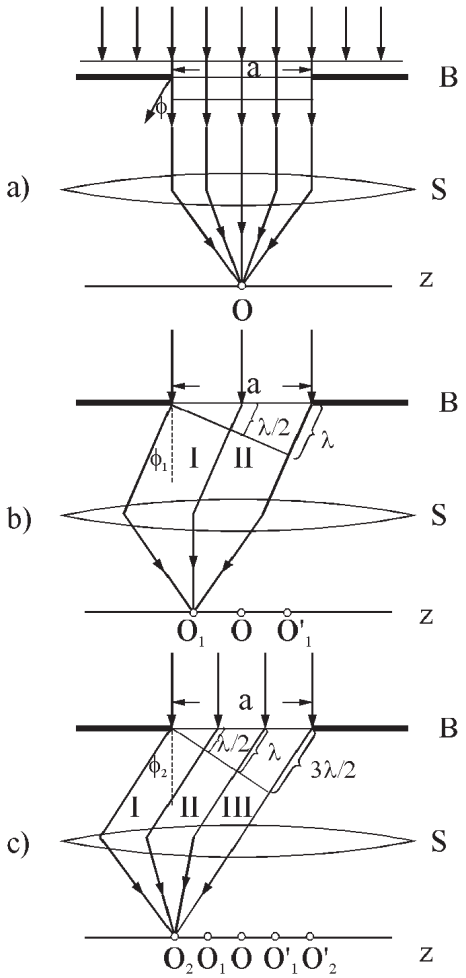

Figure 17.4: Diffraction of light at a single slit

Due to diffraction, a portion of the light in the beam that encounters a slit is diffracted at various angles relative to the original direction (the angle ϕ is called *the diffraction angle*). Let the light beam be diffracted by the slit at an angle $\phi = \phi_1$ such that the path difference Δl between the extreme rays of the beam is equal to the wavelength of light, i.e., $\Delta l = \lambda$ (Fig. 17.4.b).

A light beam can be divided into two zones (Fresnel zones), so that the path difference between the ray from the first zone and the corresponding ray from the second zone is equal to $\lambda/2$. These rays, previously collected by the lens at O_1, cancel each other out after interference. Thus, a dark band (diffraction minimum) appears at point O_1. A diffraction minimum also occurs at point O'_1, which is at the same distance as minimum O_1 but on the opposite side of maximum O (the rays forming this minimum in figure 17.4.b are not shown).

In the diffraction beam, there are also rays that bend at an angle $\phi = \phi_2$, whose path difference is $\Delta l = 3\lambda/2$ (Fig. 17.4.c). The entire beam can then be divided into three Fresnel zones (I, II, and III), each with a path difference of $\lambda/2$. Two adjacent zones cancel each other out, leaving the third, which will produce a diffraction maximum at point O_2. The same maximum will appear at point O'_2. The illuminance of the O_2 and O'_2 maxima will be significantly lower than the maximum of O.

By analogy reasoning, it can be shown that behind the maxima O_2 and O'_2 there appear minima originating from the light in the beam, which after diffraction for some angle $\phi = \phi_3$ can be divided into four Fresnel zones ($\Delta l = 4\lambda/2$),

and so on.

In general, it can be said that light beams diffracting at angles corresponding to an odd number of Fresnel zones produce diffraction maxima on the screen, while light beams diffracting at angles corresponding to an even number of Fresnel zones produce diffraction minima. The illumination of the maxima decreases as the diffraction angle that produces these maxima increases.

In this way, the diffraction pattern obtained from a single narrow slit consists of a series of alternating dark and bright fringes symmetrically arranged on both sides of the central maximum O. This diffraction pattern is shown in Fig. 17.5. The same figure includes a graph of the intensity distribution of the maxima formed in diffraction, where x is the distance from the center of the image.

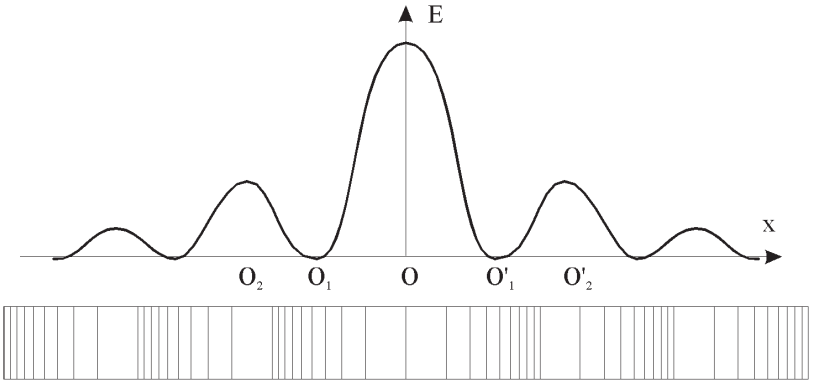

Figure 17.5: Curve of the dependence of illuminance on distance in diffraction through a single slit

As can be seen, diffraction maxima occur at a path difference $\Delta l = (2n+1)\lambda/2$, and diffraction minima at a path difference $\Delta l = n\lambda$, where $n = 1, 2, 3, ...$

From Figure 17.4, it can be seen that $\Delta l = a \sin\phi$, where a is the slit width. Then the diffraction maxima can be written in the following form:

$$\sin\phi = (2n+1)\frac{\lambda}{2a}, \tag{17.9}$$

and the diffraction minima:

$$\sin\phi = \frac{n\lambda}{a}. \tag{17.10}$$

17.2.2 Diffraction of Light at Multiple Slits

Let a beam of monochromatic light fall normally on screen B with two narrow parallel slits separated by a distance d (Fig. 17.6). These slits act as coherent sources of light.

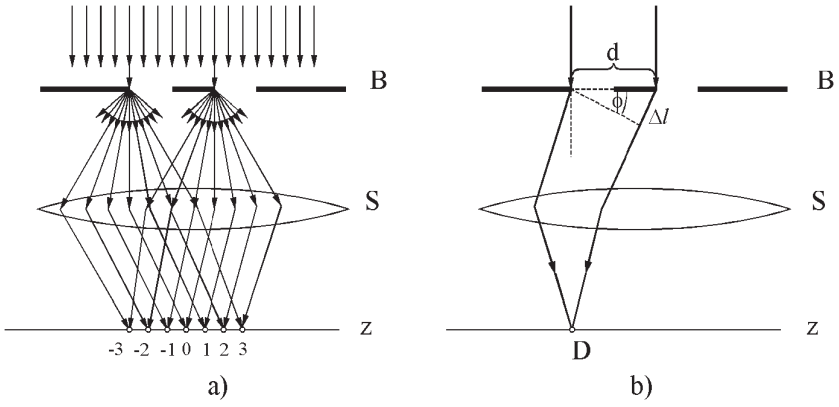

Figure 17.6: Diffraction at two slits

If a collecting lens S is placed behind a mask B with slits, then a diffraction pattern (image) appears on the mask Z, which is located in the focal plane of the lens, as a result of two processes: the diffraction of light at each individual slit and the interference of light from both slits.

The rays that fall, for example, on the left edge of both slits will be observed. Due to diffraction, the light will spread in all directions after passing through the slit (Fig. 17.6.a). Here, the deviation of the rays at an angle ϕ will be considered (Fig. 17.6.b), where their path difference is

$$\Delta l = d \sin\phi. \tag{17.11}$$

An interference maximum will appear on the screen when the path difference is equal to an integer number of wavelengths

$$d \sin \phi = n\lambda. \tag{17.12}$$

An interference minimum will appear on the screen when the path difference is equal to an odd number of half wavelengths

$$d \sin \phi = (2n+1)\frac{\lambda}{2}. \tag{17.13}$$

The number of maxima n can be obtained by expression (17.12):

$$\sin \phi = n\lambda/d. \tag{17.14}$$

Since $\sin \phi \leq 1$, it follows

$$n\lambda/d \leq 1 \Rightarrow n \leq \frac{d}{\lambda}. \tag{17.15}$$

The illuminance of the maxima is unequal. The central maximum ($n = 0$) has the highest illuminance, the first subsequent maxima ($n \pm 1$) are less bright, the next ones even less, and so on. Figure 17.7 shows a graph of the illuminance distribution of the maxima obtained by diffraction through two slits.

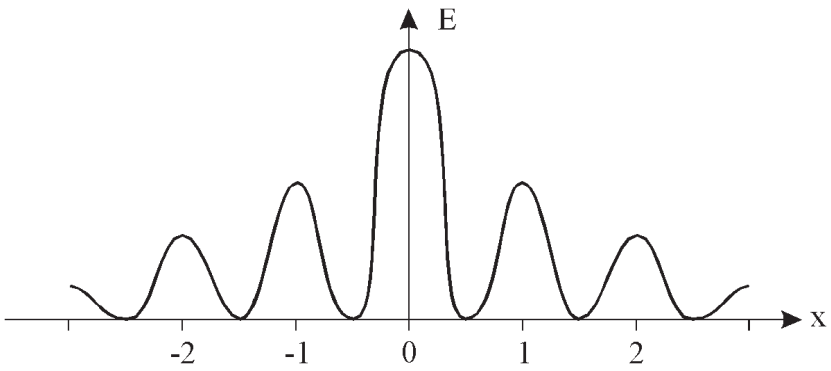

Figure 17.7: The Illuminance distribution curve for diffraction through two slits

When light diffracts through multiple narrow slits arranged parallel at small distances from each other, a diffraction pattern of a similar shape is obtained. In this case, the maxima are sharper and narrower than those formed by two slits. The distance between adjacent maxima and their sharpness increase with the number of slits. The positions of these maxima are also determined from expression (17.12).

A series of parallel narrow slits, separated by equal opaque surfaces, forms a *diffraction grating*. The distance d between two slits (Fig. 17.6.b) is called *the grating constant* or *the grating spacing*.

17.3 Polarization of Light

17.3.1 Unpolarized (Natural) and Polarized Light

Interference and diffraction of light are phenomena that revealed the wave nature of light. However, they did not provide an answer to the question of whether light waves are transverse or longitudinal. The phenomenon of polarization shows that light is a transverse electromagnetic wave.

It has been experimentally established that the light emitted by atoms consists of transverse waves with mutually perpendicular electric field $\vec{E}$ and magnetic field $\vec{H}$ vectors, which are perpendicular to the direction of light propagation. Therefore, to fully describe the polarization state of a light beam, it is sufficient to know the orientation of just one of the vectors. These considerations are often made with respect to the electric field vector $\vec{E}$, because when light interacts with materials, the electric field of the wave, which acts on the atoms of the material, plays the primary role. It is accepted that the vector $\vec{E}$ is called the light vector and is used to describe phenomena in optics. The plane in which the vector $\vec{E}$ oscillates is called *the plane of oscillation of the light wave*, which is simultaneously *the polarization plane of light waves*.

A light wave emitted by a single atom as a light source in one emission event has the property that its $\vec{E}$ vector oscillates in a strictly defined plane, while its $\vec{H}$ vector oscillates in a plane perpendicular to the plane of oscillation of the $\vec{E}$ vector. Such a wave is called *linearly polarized*. In this case, the electric vector varies in intensity in only one direction, i.e., in the plane of polarization.

Linearly polarized light is schematically represented in the manner shown in Fig. 17.8.a. The direction of the beam is perpendicular to the plane of the figure, while the arrows represent the amplitudes of the oscillation of the $\vec{E}$ vector.

In nature, light that originates from a single atom is very rare. Every real light source consists of a large number of atoms that emit numerous linearly polarized waves whose planes of oscillation are equally oriented in all directions

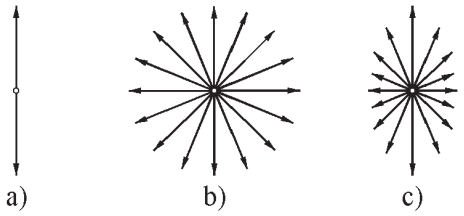

Figure 17.8: Linearly polarized (a), unpolarized (b), and partially polarized light (c)

(Fig. 17.8.b). Such light is referred to as unpolarized or natural. There are also cases in which the amplitudes of the $\vec{E}$ vectors of light waves vary in different planes of oscillation. Such light is called partially polarized (Fig. 17.8.c).

For natural light to be polarized, it is necessary to create conditions under which the oscillation of the light vector $\vec{E}$ is possible only in one plane. Such conditions can be achieved if light passes through media that are anisotropic with respect to electric oscillations ((Fig. 17.9)). Certain crystals possess this property. Experiments have shown that many crystals, both natural and artificial, can polarize natural light as it passes through them.

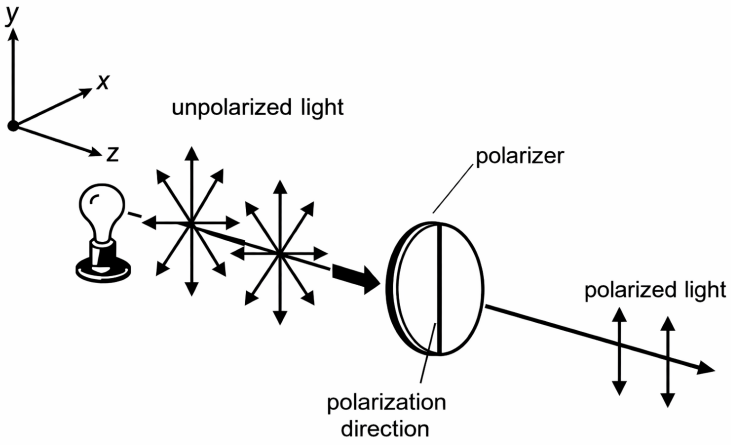

Figure 17.9: Unpolarized and polarized light

Chapter 18

Early Models of the Atom

18.1 Thomson's and Rutherford's Models of the Atom

The first model of the atom was proposed by J. J. Thomson. This model was called the Thomson or plum-pudding model. According to this model, the atom is a sphere with a diameter of about $10^{-10}\,m$, which is positively charged. Isolated negative charges - electrons are embedded in this sphere and oscillate around their equilibrium positions. The total negative charge is equal to the positive charge of the sphere.

Although Thomson's model proved unable to describe the actual behaviour of atoms, it played a positive role in science. Using this model, the phenomenon of electromagnetic radiation emitted by atoms as oscillators was explained, and the study of the linear harmonic oscillator from the perspective of quantum physics was intensified. With this model, the nature of the dielectric constant was investigated, and it also predicted the existence of isotopes. This model also made it possible to study atoms with multiple electrons.

Experiments by Rutherford and his collaborators played a significantly greater role in explaining the structure of the atom. At that time, radioactivity had already been discovered, so they observed the scattering of alpha particles as they passed through foils of various materials. An alpha particle represents the nucleus of helium ${}_2^4He$, which is doubly positively charged (α^{++}).

The diagram of the equipmen used by Rutherford and his collaborators to observe the scattering of alpha particles is shown in Fig. 18.1. A metal foil F (made of gold, platinum, copper, or silver) with a thickness of $1\,\mu m$ is located inside chamber K, in the shape of a sphere, from which the air has been evacuated. Inside the chamber, there is a container B with a radioactive source I. A narrow beam of alpha particles (marked as α in the figure) falls on the metal foil F, whose atoms scatter them. After passing through this foil, the alpha particles deflect in different directions. Some of them hit screen S made of fluorescent material (for example, a screen coated with zinc sulfide).

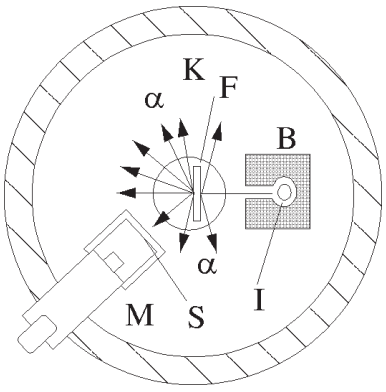

Figure 18.1: Equipmen for observing the scattering of alpha particles

When alpha particles strike such fluorescent material, scintillations (flashes) occur. Thanks to these scintillations, the scattering of alpha particles can be tracked. The scintillations were observed using microscope M, which, together with the fluorescent screen S, could be rotated around an axis passing through the center of this sphere. In this way, it was possible to record alpha particles scattered at different angles, i.e., the number of alpha particles as a function of the deflection angle.

They noted that as most alpha particles, passed through these foils, they were deflected at very small angles (2^o do 3^o). However, a smaller number of alpha particles were deflected at much larger angles.

Seeking an explanation for the scattering of alpha particles at larger angles, Rutherford assumed that such a deviation was caused by the interaction (collision) of alpha particles, and the positive part of the atom, as shown in Fig. 18.2.

Deflection at small angles is the result of alpha particles passing near that positive charge (influence of Coulomb force). From these experiments, it was concluded that the entire positive charge of the atom and almost all of its mass are concentrated in a very small volume with a diameter on the order of $10^{-14} - 10^{-13}\,m$, which is called *the atomic nucleus*, because very few alpha particles are scattered at an angle of 180^o degrees. It is important to note that the

influence of the electrons orbiting the nucleus on the motion of the alpha particle can be neglected, as the mass of the electron is much smaller than the mass of the alpha particle.

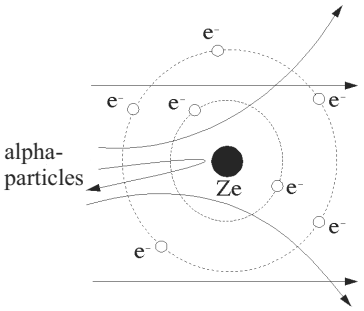

Figure 18.2: Interaction of alpha particles with the atomic nucleus

Based on theoretical and experimental considerations, Rutherford proposed an atomic model called the planetary (nuclear) model of the atom. According to this model, electrons orbit around a positively charged nucleus, whose charge is Ze, where Z is the number of electrons (and protons), and e is the electron's charge. Since the atom is neutral, the charge of the nucleus is equal to the sum of the charges of the electrons that make up the atom. Rutherford assumed that electrons move around the nucleus in circular orbits with radius r. In this case, the Coulomb force is equal to the centrifugal force, i.e.

$$\frac{Ze^2}{4\pi\epsilon_o r^2} = \frac{mv^2}{r}, \tag{18.1}$$

where v is the velocity of the electron in an orbit of radius r, $\epsilon_o = 8,85 \cdot 10^{-12}\, Fm^{-1}$ is the vacuum permittivity.

In equation (18.1), the orbit radius and the speed at which the electron moves along the orbit can take arbitrary values, i.e., their values can change continuously. The system's energy E_{max} is greatest when the electron is on the orbit farthest from the nucleus. When an electron moves from one orbit to another, any amount of energy less than or equal to E_{max} can be emitted or absorbed. In other words, according to this model, the atom's energy spectrum is continuous.

From equation (18.1) it follows that for an orbit radius $r \approx 10^{-10}\, m$, the electron's velocity is $v \approx 10^6\, ms^{-1}$, and the acceleration $v^2/r \approx 10^{22}\, ms^{-2}$. According to the laws of classical electrodynamics, due to the accelerated motion of charged body (the electron), electromagnetic radiation is emitted, causing the electron to continuously lose energy. As a result, the electron would get closer and closer to the nucleus and eventually fall into it. Such an atom would be unstable, which does not correspond to reality, because the atom behaves like a stable particle.

The attempt to formulate a model within classical physics did not yield the desired results. Namely, Thomson's atomic model was replaced by Rutherford's model, which proved to be electrodynamically unstable and could not account for certain experimental results. This means that certain phenomena in the microscopic world could not be explained based on the laws of classical physics, and this atomic model met a similar fate as Thomson's model. It was replaced by a new one - the Bohr model, which was based on a quantitatively new theory called *the quantum theory of the atom.*

18.2 Spectrum of Hydrogen Atoms

By studying the atomic spectra in rarefied gases (emission spectra of atoms), it has been shown that each gas has its characteristic line spectrum, consisting of a large number of lines or groups of lines. Figure 18.3 shows some lines of the hydrogen atom spectrum.

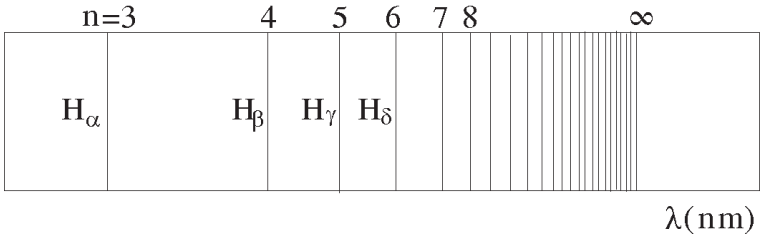

Figure 18.3: Spectrum of hydrogen atoms

Balmer was the first to determine that the wavelengths of the spectral lines of the hydrogen atom can be expressed using an empirical formula

$$\frac{1}{\lambda} = R_H\left(\frac{1}{2^2} - \frac{1}{n^2}\right), \tag{18.2}$$

where $n = 3, 4, 5, ...$, $R_H = 10970900\, m^{-1}$ is the Rydberg constant, and $c = 3 \cdot 10^8\, ms^{-1}$ is the speed of light in a vacuum.

From expression (18.2), it can be seen that different spectral lines correspond to different values of n. These lines form a group or series called the Balmer series. As the value of n increases, the spectral lines get progressively closer to each other, as shown in Fig. 18.3. The series limit is reached for $n \to \infty$.

Expression (18.2) is often expressed in terms of frequency ν

$$\nu = R'_H(\frac{1}{2^2} - \frac{1}{n^2}), \tag{18.3}$$

where $n = 3, 4, 5,...$, $R'_H = R_H \cdot c = 3,28985.10^{15}\, s^{-1}$ and is also called the Rydberg constant.

Research of the hydrogen spectrum has shown that there are other series of spectral lines (the Lyman series in in the ultraviolet region of the electromagnetic spectrum, Paschen series, Brecket series, and so on).

All series of the hydrogen spectrum can be expressed as follows:

$$\frac{1}{\lambda} = R_H(\frac{1}{m^2} - \frac{1}{n^2}) \tag{18.4}$$

$$\frac{c}{\lambda} = \nu = R'_H(\frac{1}{m^2} - \frac{1}{n^2}), \tag{18.5}$$

where m and n are integers. For a given series, $n = m + 1$, $m + 2$, $m + 3$, and so on. For the Lyman series, $m = 1$, for the Balmer series, $m = 2$, for the Paschen series, $m = 3$, and so on.

The line (discrete) structure of atomic emission and absorption spectra shows that an atom cannot emit (or absorb) just any value of energy, but only a precisely defined amount called a quantum. Thus, an atom can exist only in certain (discrete) states, which is in contrast to the Rutherford model (this is the second major drawback of this model). When the atom transitions from one state to another, it emits (or absorbs) a quantum equal to the difference in energy between the initial and final states.

18.3 Bohr's Postulates

The first attempt to establish a new (quantum) theory of the atom was made by Bohr. He tried to connect the empirical laws of the line spectrum, Rutherford's atomic model, and the quantum nature of emission and absorption of light.

At the core of his theory, Bohr proposed three postulates. The first postulate (the postulate of stationary orbits) states: *An electron in an atom cannot move along just any orbit, but only along orbits with precisely defined radii.* On these orbits, which are called stationary or discrete, the electron's angular momentum satisfies the condition:

$$L = mv_n r_n = n\hbar, \tag{18.6}$$

where m is the mass of the electron, v_n its speed in the n-th orbit of radius r_n, n is an integer ($n = 1, 2, 3,...$) which is called *the principal quantum number*, and $\hbar = h/(2\pi)$ where $h = 6,64.10^{-34}\, Js$ is Planck's constant, and $\hbar$ is the reduced Planck constant.

The second postulate of Bohr: *In the movement of electrons along stationary orbits, energy is neither emitted nor absorbed.*

The third postulate (the frequency rule) states: *The transition of an electron from one orbit to another is governed by the emission or absorption of a quantum of energy.* Thus, when an electron moves from one stationary orbit to another, the atom emits or absorbs a quantum of energy $h\nu$, which is equal to the difference in the energies of those stationary states

$$h\nu = E_2 - E_1 = \Delta E, \tag{18.7}$$

where E_2 and E_1 are the energies of the stationary states of the atom before and after emission (absorbtion). If the atom absorbs a quantum of energy $h\nu$, the electron moves from an orbit closer to the nucleus (energy E_1) to an orbit further from it (energy E_2), and the opposite leads to emission.

From expression (18.7), it can be observed that the frequency of electromagnetic radiation emitted by an atom is not determined by the frequency of the electron's orbit in the atom, but by the the energy difference of the atom's stationary states.

Bohr's postulates can be expressed in a somewhat more general form as follows: *an atom is stable only in states corresponding to discrete energy values ($E_1, E_2, E_3,...$). The transition of an atom from one stationary state to another is accompanied by the emission or absorption of a quantum of energy $h\nu$.*

18.4 Franck-Hertz Experiment

The hypothesis of discrete energy states of atoms, on which Bohr formulated his postulates, was confirmed by the Franck-Hertz experiment. The experimental setup is shown in Fig. 18.4, and it consists of a glass tube B from which the air is evacuated and which is filled with mercury in a gaseous state. A voltage U_1, whose value can be adjusted, is applied between the cathode K and the anode A that is in the form of a grid. The anode is at a potential $U_2 \approx 0.5\, V$ higher than the collector C. The cathode is heated and electrons are emitted (released) from the cathode through thermo-electronic emission. They are accelerated in the space between the cathode and anode by the electric field

created by the applied voltage U_1. It is known that a potential difference of 1 V in vacuum gives an electron kinetic energy of 1 eV.

If the kinetic energy of the electrons upon reaching the anode is greater than 0.5 eV, they will overcome the retarding voltage U_2 and reach the collector C, at which point the galvanometer G will register a current i.

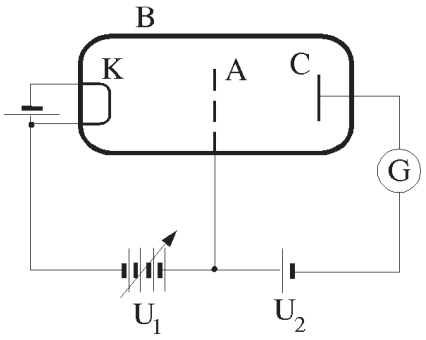

Figure 18.4: Experimental setup of the Franck-Hertz experiment

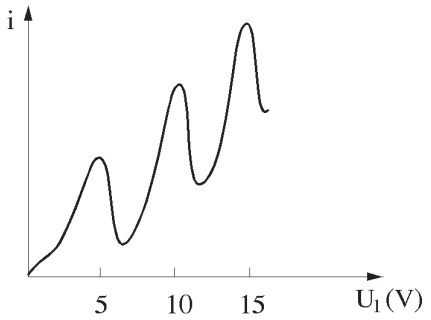

Figure 18.5: Current-voltage characteristic

The graph of the dependence of the collector current i on the voltage U_1 between the cathode and anode is shown in Fig. 18.5. As the voltage U_1 increases, the current initially rises, then drops sharply at $U_1 \approx 4.9$ V. With further increase of the voltage U_1, the current begins to rise again and also drops sharply at $U_1 \approx 2 \cdot 4,9$ V. The next drop in current is at $U_1 \approx 3 \cdot 4,9$ V, and so on. Therefore, the distance between neighboring maxima on the graph corresponds to a change in voltage U_1 of approximately $4,9$ V. This means that in the field created by a voltage $U_1 \approx 4,9$ V, the electron acquires kinetic energy of $4,9$ eV ($1\,\text{eV} = 1,6 \cdot 10^{-19}$ J).

This shape of the graph can be explained by the fact that atoms, when colliding with electrons (moving toward the anode), can receive (absorb) only a certain (discrete) amount of energy, approximately equal to $4,9$ eV.

When the energy of electrons is less than $4,9$ eV ($U_1 < 4,9$ V), the collision of electrons with mercury atoms is elastic, meaning the electron does not change the magnitude of its velocity, and thus its kinetic energy. In this case, the internal energy of the mercury atom also does not change. Therefore, the electron reaches the collector C, as its energy is sufficient to overcome the retarding voltage U_2. As the voltage U_1 increases, the collector current i also increases.

When the voltage U_1 reaches approximately $4,9$ V, an electron traveling from the cathode to the anode acquires an energy of $4,9$ eV. At that moment, the collision of the electron with mercury atoms is inelastic. In this case, the electrons transfer all their energy to the mercury atoms, which then become excited. Having expended all their energy, the electrons do not reach the collector C because they cannot overcome the retarding voltage U_2. Only those electrons that have not undergone an inelastic collision with mercury atoms reach the collector. Consequently, the current i decreases sharply.

When the voltage U_1 reaches a value at which some electrons, having undergone inelastic collisions with mercury atoms, acquire the energy between the cathode and anode necessary to overcome the retarding voltage U_2, the current starts to rise again. When the voltage U_1 reaches $2 \cdot 4,9$ V, some electrons after the first inelastic collision have an energy of $4,9$ eV, sufficient for a second inelastic collision with mercury atoms. In that case, the electrons transfer this energy and are stopped at the anode. In this way, the second sudden drop in current is explained (see the graph in Fig. 18.5). A similar explanation applies to the other sudden drops in current (for $U_1 \approx 3 \cdot 4,9$ V, etc.).

The results of the Franck-Hertz experiment show that electrons can transfer only specific amounts of energy to mercury atoms. The value of $4,9$ eV is the smallest possible amount of energy (quantum of energy) that a mercury atom in its ground state can absorb. In this way, Bohr's hypothesis about the existence of energy levels in the atom has been fully experimentally confirmed.

A mercury atom, upon receiving energy of $4,9$ eV from an electron during a collision, transitions to an excited state and, upon returning to the ground state, emits, according to Bohr's third postulate (expression (18.7)), a quantum of energy with frequency $\nu = \Delta E/h$. The wavelength of the emitted radiation is

$$\lambda = \frac{hc}{\Delta E} \approx 253,7\,nm. \tag{18.8}$$

If this theory is correct, the mercury atoms should be the source of ultraviolet rays when bombarded with electrons of 4.9 eV energy. Radiation of this wavelength was indeed observed in the Franck-Hertz experiment. Using this experiment, discrete energy levels of absorption and emission were also found in other atoms, such as sodium, helium, etc. The only difference is in the wavelengths of the emitted radiation.

Chapter 19

Fundamentals of Atomic and Molecular Physics

19.1 Quantization of Angular Momentum

The angular momentum of an electron in an atom in classical mechanics is represented by the vector product (cross product) of the electron's position vector and momentum

$$\vec{L} = \vec{r} \times \vec{p}. \tag{19.1}$$

Quantum-mechanical determination of angular momentum is a complex procedure. Therefore, here only the final equation for determining its quantum value is given

$$L = \hbar\sqrt{\ell(\ell+1)}, \tag{19.2}$$

where $\hbar$ is the reduced Planck constant, and ℓ is an integer called *the orbital quantum number*, which can take values $0, 1, 2, ..., (n-1)$. This means that for a given value of the principal quantum number n (the given energy level of an electron in an atom), the following quantum states (energy sublevels) of the electron in the atom are possible:

$n = 1$	$\ell = 0$	($1s$ state)
$n = 2$	$\ell = 0$	($2s$ state)
	$\ell = 1$	($2p$ state)
$n = 3$	$\ell = 0$	($3s$ state)
	$\ell = 1$	($3p$ state)
	$\ell = 2$	($3d$ state), an so on.

From equation (19.2), it follows that an electron in an atom in an s-state, i.e., in a state with $\ell = 0$, has orbital angular momentum equal to zero ($L = 0$). From the point of view of Bohr's theory, such a state would correspond to an electron's path passing through the nucleus. Quantum mechanics allows for the existence of electron states in which the orbital angular momentum is not associated with motion along an orbit. For example, s-electron states have a spherically symmetric probability distribution, so their most probable orbits are circular, and their orbital momentum is zero. Electron states labeled p, d, f, etc., have an elliptical shape whose eccentricity (elongation) increases with increasing ℓ, allowing these electrons to penetrate into the interior of the atom, where electrons of other levels reside, thus undergoing additional interactions. As a result, their energies change.

An orbital is not a trajectory in the classical sense of the term. Namely, an orbital is a region of space around the nucleus of an atom where there is a high probability of finding an electron. It is often thought of as an electron cloud. It represents a three-dimensional volume where an electron is likely to be located 90 % to 95 % of the time. While Bohr's model describes orbits as fixed paths, modern quantum mechanics uses orbitals to describe the fuzzy regions of electron density. In Fig. 19.1, the s, p_x, p_y and p_z orbitals are shown. Each orbital can contain at most 2 electrons.

19.2 Spatial Quantization

It is known from electrodynamics that an electron moving around the nucleus, in addition to orbital angular momentum, also has an orbital magnetic moment. The vectors of orbital angular momentum $\vec{L}$ and orbital magnetic moment $\vec{p}_m$ have the same direction but opposite sense. The relationship between these two vectors is as follows:

$$\vec{p}_m = -\frac{e}{2m}\vec{L} = -g_\ell \vec{L}, \tag{19.3}$$

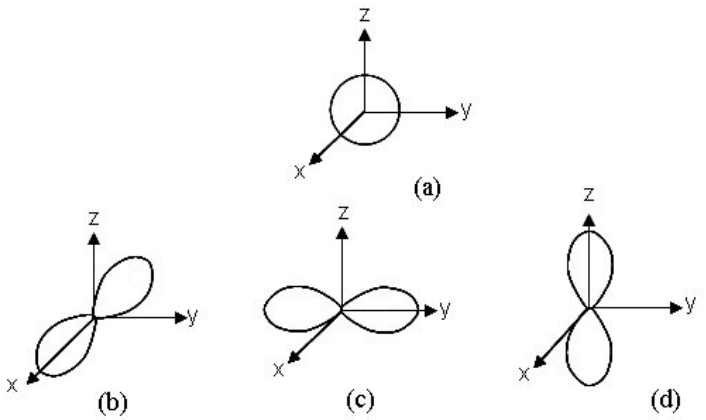

Figure 19.1: Shapes of the s, p_x, p_y and p_z orbitals

where m is the mass of the electron, and $g_\ell = e/2m$ is the orbital gyromagnetic ratio. An analogous relationship exists between the orbital angular momentum and the orbital magnetic moment for the atom as a whole. The relationship between the vectors $\vec{L}$ and $\vec{p}_m$ also holds in Bohr's theory.

In quantum mechanics, it is not possible to determine the orientation of the vectors $\vec{L}$ and $\vec{p}_m$ relative to the plane of the electron orbit. To determine the orientation of these vectors, a direction in space must be chosen, and the position of the vector $\vec{L}$ can be fixed by the angle this vector forms with that direction. To determine this direction, one either chooses the direction of the external magnetic field or the direction of the internal magnetic field produced by all electrons, except the one under consideration, and by the nucleus of the atom.

In classical physics, the vector $\vec{L}$ (or $\vec{p}_m$) can have arbitrary orientations relative to the direction of the magnetic field. The theory of paramagnetism is based on this assumption.

According to Bohr's theory, the arbitrary orientation of the vector $\vec{L}$ in space implies that the plane of the electron's orbit in space can be arbitrarily oriented with respect to the external magnetic field. It has been shown that such assumptions are incorrect, so a principle called the principle of spatial quantization has been adopted. This principle states that *the electron's angular momentum vector $\vec{L}$ can have only such orientations in space for which the projection L_z of the vector $\vec{L}$ onto the z-axis of the external magnetic field has quantized values.*

In quantum mechanics, it has been proven that the projections L_z of the vector $\vec{L}$ onto the direction z of the magnetic field can only take integer multiples of $\hbar$:

$$L_z = m_\ell \hbar, \tag{19.4}$$

where m_ℓ can take values $0, \pm 1, \pm 2, ..., \pm \ell$ (a total of $2\ell + 1$ possible values) and is called the magnetic quantum number.

Figure 19.2 shows the possible orientations of the vector $\vec{L}$ with respect to the external magnetic field $\vec{H}$ for $\ell = 1$ (p-state) and $\ell = 2$ (d-state).

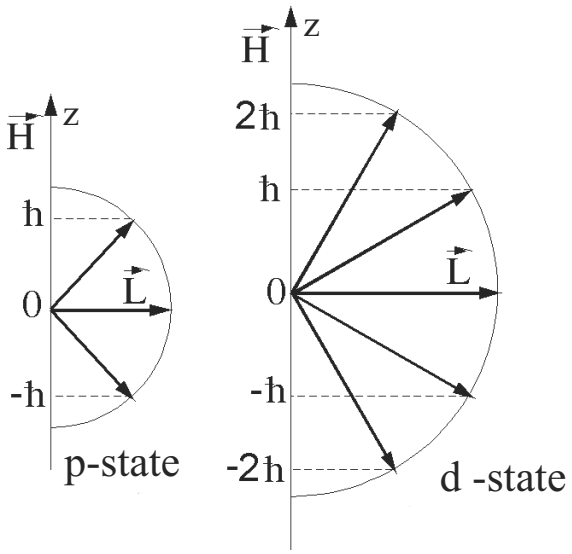

Figure 19.2: Possible orientations of vector $\vec{L}$ relative to the direction of the external magnetic field $\vec{H}$

19.3 The Stern-Gerlach Experiment

Stern and Gerlach performed experiments to confirm spatial quantization. The experiments were performed using an equipment consisting of a tube C in which a vacuum of $10^{-3}\,Pa$ was established (Fig. 19.3). Inside the tube, there was a piece of silver that could be heated, two diaphragms B_1 and B_2 with narrow slits, the poles of an electromagnet N and S, and a glass plate A that served as a screen. The inhomogeneous magnetic field was produced using a strong SN electromagnet with specially shaped poles (one pole shaped like a gutter, the other like a blade extending into the gutter).

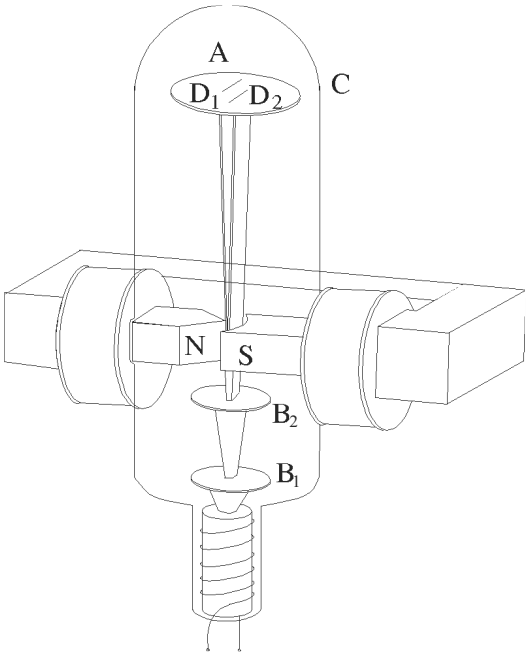

Figure 19.3: Equipment for testing the Stern–Gerlach experiment

When silver is heated, it vaporizes and its atoms escape in different directions. From these emitted atoms, diaphragms B_1 and B_2 allow only a narrow beam to pass. The diaphragms are adjusted so that atoms that have passed through the first diaphragm B_1 also pass through the second diaphragm B_2. This results in a beam of atoms in which all move in the same direction.

In the absence of a magnetic field, atoms move in a straight line and fall onto the plate so that an exact image of the slit on the diaphragms through which they passed is obtained on plate A. When the field is established, classical physics would predict that these atoms scatter and their traces appear all around the center of plate A. There should be a certain continuity here, on both sides from the center position up to a certain maximum value. However, experiments have shown that the atom beam splits into two parts, so instead of a single image of the slit, two images, D_1 and D_2, are obtained, which are symmetrically shifted relative to the original direction of the beam.

The results of the Stern–Gerlach experiment confirmed spatial quantization, i.e., that the orientation of the angular momentum of silver atoms in space is a quantized quantity. Specifically, some of these atoms have a parallel orientation, while others have an antiparallel orientation of angular momentum relative to the lines of the magnetic field.

Substituting expression (19.2) into (19.3) gives the magnetic moment of the electron

$$p_m = \frac{e}{2m}L = \frac{e\hbar}{2m}\sqrt{\ell(\ell+1)} = \mu_B\sqrt{\ell(\ell+1)}. \tag{19.5}$$

The quantity $\mu_B = e\hbar/(2m) = 9,27.10^{-24}\,JT^{-1}$ is called the Bohr magneton.

19.4 Electron Spin

As already emphasized, the Stern-Gerlach experiment confirmed spatial quantization. The results of this experiment clearly showed that the magnetic moment measured in this experiment acquires two orientations. If spatial quantization did not exist, the orientation of the atom's magnetic moment vector $\vec{p}_m$ would be arbitrary.

Stern and Gerlach initially conducted experiments with elements of the first group of the periodic table. The atoms of these elements have one valence electron (one electron in the outer orbit). In the ground state of the atom, this electron is in an s-state, i.e., the value of the orbital quantum number ℓ is zero. Therefore, in a state with $\ell = 0$, the orbital angular momentum L is also zero, as follows from expression (19.2). The results of their experiment with such atoms should not have shown spatial quantization of the orbital angular momentum. However, experiments did show the existence of spatial quantization.

This fact led to serious difficulties in interpreting the Stern-Gerlach experiments. As a result, there were significant challenges in explaining the results of that experiment. These difficulties were resolved in 1925 when Uhlenbeck and

Goudsmit proposed that the electron rotates around its own axis. Since the electron is charged, this motion is associated with the existence of the electron's own magnetic field. This means that there exists an intrinsic angular momentum of the electron, $\vec{L}_s$, and an intrinsic magnetic moment of the electron, $\vec{p}_{ms}$. Such electron rotation is known as *electron spin*. Often, both of these quantities ($\vec{L}_s$ and $\vec{p}_{ms}$) are also called the electron spin.

The results of the Stern–Gerlach experiment, performed with atoms of the first group of the periodic table, received a simple explanation after the introduction of electron spin. In these experiments, spatial quantization of intrinsic angular momentum $\vec{L}_s$ appears. Similar to orbital angular momentum vector $\vec{L}$, the intrinsic angular momentum vector (spin vector) $\vec{L}_s$ and its projection $\vec{L}_{sz}$ along the z-axis, which coincides with the direction of the magnetic field, must be quantized.

In quantum mechanics, it has been shown that the magnitude of spin vector $\vec{L}_s$ is quantized according to the rule

$$L_s = \hbar\sqrt{s(s+1)}. \tag{19.6}$$

Equation (19.6) is analogous to equation (19.2), but instead of the quantum number ℓ, the quantum number s is introduced, which is called *the spin quantum number*.

The electron spin vector $\vec{L}_s$ must be spatially quantized as vector $\vec{L}$, but it has $(2s+1)$ orientations. It means that the projection vector $\vec{L}_{sz}$ of spin vector $\vec{L}_s$ onto the direction of the magnetic field is also quantized.

Since the results of the Stern-Gerlach experiment show only two spin orientations, we have $2s+1=2$, i.e., $s=1/2$. The spin quantum number s has only one value and differs in this regard from the principal quantum number (n), the orbital quantum number (ℓ), and the magnetic quantum number (m_ℓ). From equation (19.6), the numerical value of the electron spin can be determined

$$L_s = \hbar\sqrt{\frac{1}{2}(\frac{1}{2}+1)} = \frac{\sqrt{3}}{2}\hbar. \tag{19.7}$$

By analogy with the spatial quantization of the orbital angular momentum vector $\vec{L}$ (eq. (19.4)), an equation can be written for the spatial quantization of spin vector $\vec{L}_s$, i.e. for the magitude of projection vector $\vec{L_{sz}}$ of the vector $\vec{L_s}$ onto the direction of the magnetic field:

$$L_{sz} = m_s\hbar, \tag{19.8}$$

where m_s is the magnetic spin quantum number, which can take two values $m_s = +1/2$ and $m_s = -1/2$, in accordance with the fact that the projection of spin onto the direction of the magnetic field can have two values:

$$L_{sz} = \pm\frac{1}{2}\hbar. \tag{19.9}$$

It is often said that the projection of electron spin equals $\pm\hbar/2$ and can be oriented either along the magnetic field or opposite to it.

19.5 Pauli Exclusion Principle

From the previous exposition, it follows that the motion of electrons in an atom is characterized by four quantum numbers: the principal quantum number n, the orbital quantum number ℓ, the magnetic quantum number m_ℓ, and the magnetic spin quantum number m_s. Their values according are:

principal quantum number:	$n = 1, 2, 3, ...,$
orbital quantum number:	$\ell = 0, 1, 2, ..., (n-1),$
magnetic quantum number:	$m_\ell = 0 \pm 1, \pm 2, ..., \pm\ell$ ($2\ell+1$ values),
magnetic spin quantum number:	$m_s = +1/2, -1/2.$

If an atom contains multiple electrons, they interact with each other with forces that lead to changes in their motion. However, it has been shown that, with a good approximation, the motion of these electrons can be described using the mentioned four quantum numbers. From experimental results obtained through optical spectroscopy methods, Pauli formulated a quantum-mechanical principle known as the Pauli principle. In simplified form, this principle states: *in one atom, no two electrons can have the same values for all four quantum numbers.*

Electrons in an atom with the same principal quantum number n form an energy shell. Depending on the value of n, there are the K-level for $n=1$, L-level for $n=2$, M-level for $n=3$, N-level for $n=4$, O-level for $n=5$, and so on. Table 19.1 shows the maximum number of electrons N that can be found in individual energy shell, which is calculated using the formula: $N = 2n^2$.

At each energy level, electrons are arranged in sublevels (subshells). A sublevel corresponds to a specific value of the orbital quantum number ℓ, as shown in Table 19.1.

Figure 19.4 shows the quantum numbers of individual energy sublevels, and Fig. 19.5 shows the distribution of electrons in a sodium atom $_{11}$Na by energy sublevels (electron configuration). The sodium atom $_{11}$Na has 11 electrons. The upward arrow indicates that the electron has a magnetic spin quantum number of $+1/2$, and the downward arrow -1/2.

Table 19.1: Maximum number of electrons N in individual energy level n

	ℓ:	0	1	2	3	4	
	Sublevel:	s	p	d	f	g	
n	Level						N
1	K	2					2
2	L	2	6				8
3	M	2	6	10			18
4	N	2	6	10	14		32
5	O	2	6	10	14	18	50

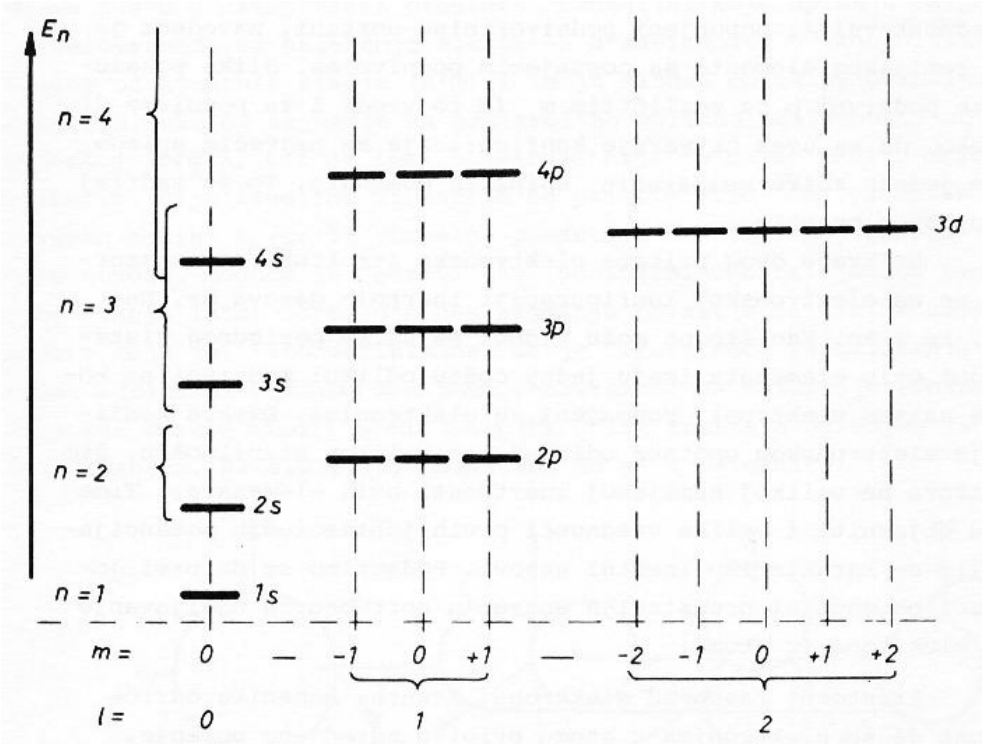

Figure 19.4: Quantum numbers of energy sublevels

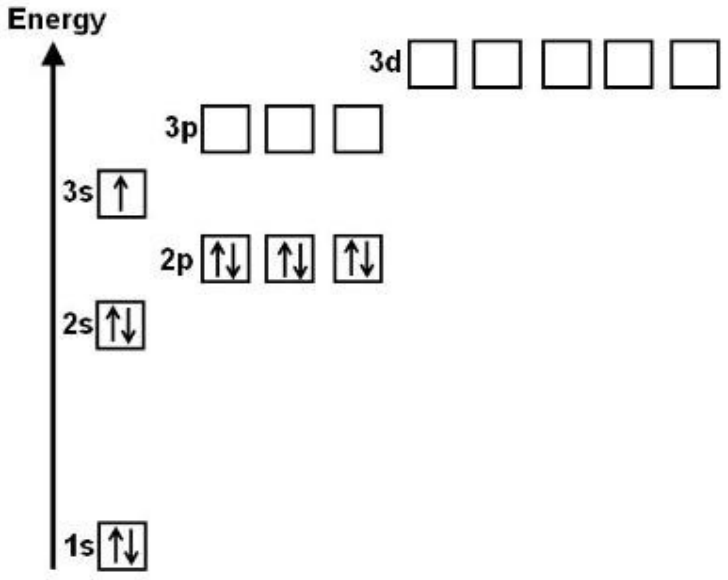

Figure 19.5: Electron configuration of sodium atom $_{11}$Na

Chapter 20

Fundamentals of Nuclear Physics

20.1 Basic Characteristics of the Atomic Nucleus

By observing the scattering of fast alpha particles as they pass through metal foils, Rutherford concluded that the atom consists of a positively charged nucleus around which electrons orbit.

The nucleus consists of positively charged particles – protons and neutral particles – neutrons, which are collectively called nucleons. The number of protons in the nucleus, Z, represents the charge number or, when referring to the corresponding atom, the atomic number. The charge of the nucleus can be expressed as follows:

$$q_j = Ze^+, \tag{20.1}$$

where e^+ is the positive charge numerically equal to the charge of an electron (elementary charge).

The number of nucleons in a nucleus, A, is called the mass number or nuclear number. If N denotes the number of neutrons in the nucleus, then the mass number can be expressed as follows:

$$A = Z + N. \tag{20.2}$$

The mass of the atom is approximately equal to the mass of the nucleus, because the total mass of the electrons, compared to the mass of the nucleus, can be neglected. It is usually expressed in atomic mass units (u). $1\,u$ is defined as $1/12$ of the mass of the carbon isotope with a mass number of 12. Therefore, $1\,u = (1,66043 \pm 0,00031).10^{-27}\,kg$.

The masses of a proton m_p and a neutron m_n are:

$m_p = (1,00727663 \pm 0,00000008)\,u = (938,550 \pm 0,005)\,MeV/c^2$,

$m_n = (1,0086654 \pm 0,0000004)\,u = (939,550 \pm 0,005)\,MeV/c^2$.

The nucleus is symbolically represented as ${}^A_Z X$, where the letter X denotes an element, the subscript Z denotes the atomic number, and the superscript A denotes the mass number.

The first assumptions about the dimensions of the atomic nucleus were obtained from Rutherford's experiments. Later studies involving the scattering of fast neutrons and electrons on different nuclei showed that the nucleus does not have a constant density ϱ, but its density gradually decreases with the distance r from the center (Fig. 20.1). The nucleus radius r_o is the distance from the center of the nucleus at which the density decreases to half of its value at the center of the nucleus.

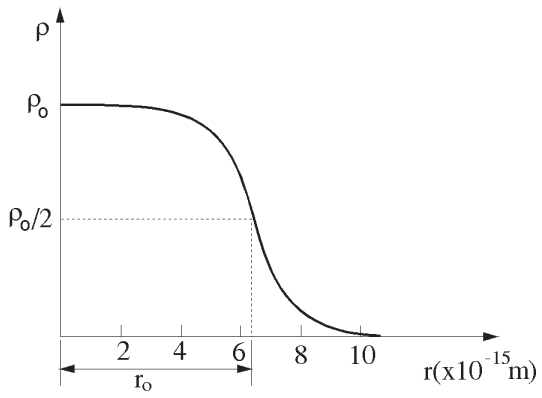

Figure 20.1: Nucleus density as a function of the distance from its center

Although not all nuclei are spherical in shape, the dimensions of nuclei are determined using their radii. Only the most stable nuclei, known as magic nuclei, have a spherical shape in the ground state, while the remaining nuclei are deformed and most often have a paraboloid shape.

Nuclei with the same atomic number Z but different mass numbers A are called *isotopes*. Most chemical elements have several isotopes. Thus, for example, hydrogen has three isotopes: ${}_1^1\mathrm{H}$ ($Z = 1$, $N = 0$), ${}_1^2\mathrm{H}$ ($Z = 1$, $N = 1$) , which is called deuterium, and ${}_1^3\mathrm{H}$ ($Z = 1$, $N = 2$), which is called tritium. Deuterium and tritium are denoted by the symbols D and T. Ordinary hydrogen and deuterium are stable, whereas tritium is radioactive. Oxygen also has three isotopes, lead ten, etc.

Nuclei with the same mass number A, but different atomic numbers Z are called *isobaric nuclei*. Such nuclei are, for example ${}_{18}^{40}\mathrm{Ar}$ and ${}_{20}^{40}\mathrm{Ca}$.

Nuclei with an equal number of neutrons $N = A - Z$ are called *isotone nuclei*. These include, for example, ${}_6^{13}\mathrm{C}$ and ${}_7^{14}\mathrm{N}$.

There are also radioactive nuclei with the same values of Z and A that differ in their half-life. Such nuclei are called *isomers*. For example, there are two isomers of the nucleus ${}_{35}^{80}\mathrm{Br}$. One has a half-life of $18\,min$, and the other 4.4 hours.

About 1 500 nuclei are known, which differ either in their Z or A values, or both. Approximately 1/5 of them are stable nuclei, while the rest are radioactive. Many nuclei have been artificially produced through nuclear reactions.

In nature, there are elements with atomic numbers Z from 1 to 92, except for technetium (Tc, $Z = 43$) and promethium (Pm, $Z = 61$). Plutonium (Pu, $Z = 94$) was first obtained artificially, and later it was established that it occurs in small amounts in nature. Other transuranium elements have been produced artificially using various nuclear reactions.

20.2 Nuclear Binding Energy and Mass Defect

It has been experimentally shown that the mass of the nucleus is smaller than the sum of the masses of the free nucleons that make up the nucleus. This means that the energy of the nucleus is less than the sum of the energies of the free nucleons that form it (the energy E of mass m is $E = \Delta mc^2$). Therefore, a certain amount of energy is required to separate the nucleus into nucleons, which will be converted into their mass. Similarly, when a nucleus is formed from free nucleons, the same amount of energy is released.

The energy E released during the formation of a nucleus, that is the energy that binds nucleons within the nucleus, is called *nuclear binding energy*.

If a nucleus consists of Z protons and $A - Z$ neutrons, then the binding energy of such a system is

$$E = [Zm_p + (A - Z)m_n - m_j]c^2 = \Delta mc^2, \tag{20.3}$$

where m_p, m_n, and m_j are the masses of the proton, neutron, and nucleus, respectively, and c is the light speed in a vacuum.

$$\Delta m = Zm_p + (A - Z)m_n - m_j, \tag{20.4}$$

is called the mass defect of the nucleus or the nuclear mass defect, or just mass defect.

In addition to the binding energy of the nucleus E, the average binding energy per nucleon, E_v, is also defined, which represents the ratio of the binding energy E to the mass number A:

$$E_v = \frac{E}{A}. \tag{20.5}$$

20.3 Nuclear Forces. Nuclear Models

An attractive force acts between the nucleons in the nucleus, whose intensity is significantly greater than the repulsive Coulomb forces between protons. Such forces are called *nuclear forces*.

Experiments on nucleon scattering on nuclei and nuclear reactions have shown that nuclear forces are significantly stronger in intensity than gravitational and electromagnetic forces. Nuclear forces belong to the group of strong interaction force, and act over short distances on the order of $10^{-15}\,m$. Nuclear forces do not distinguish between protons and neutrons, i.e., they act equally on protons and neutrons, showing a non-electrical nature.

Nuclear forces demonstrate the phenomenon of saturation, meaning that each nucleon in the nucleus acts only on a limited number of its closest nucleons.

Nuclear forces depend on the orientation of the spins of the nucleons that make up the nucleus (it should be noted that nucleons also have spin). For example, a proton and a neutron form deuterium only when their spins are oriented parallel to each other.

The complex nature of nuclear forces and the difficulties in precisely determining the motion of all nucleons that constitute the nucleus have so far prevented the development of a unified theory of the atomic nucleus. Therefore, in addressing this complex problem, partial consideration of the nucleus components is applied along with the introduction of certain assumptions that allow for a mathematical interpretation of existing methods. This leads to the creation of various models of nuclear structure through which their properties and phenomena are interpreted.

Some models explain the stability of the nucleus, the saturation of nuclear forces, the magnitude of spins, magnetic moments, and similar aspects, while others explain the interaction of the nucleus with other particles. In these interpretations, very often different, and even contradictory, assumptions about the structure and laws governing the nucleus are taken as a starting point. However, each of the existing models, either alone or through mutual supplementation, leads to a successful explanation of certain problems related to the structure of the nucleus.

There are several models that have mainly developed in two seemingly opposite directions. In some models it is assumed that there are strong mutual interactions among nucleons, while in others the opposite is true, i.e., nucleons are weakly bound and move almost freely within the nucleus: *liquid drop model, nuclear shell model, collective model, Fermi gas model, optical model and alpha particle model.*

20.4 Natural Radioactivity

Natural radioactivity refers to the spontaneous transformation of one atomic nucleus into another, accompanied by the emission of various forms of radioactive radiation (alpha particles and/or beta particles, and gamma rays). In addition to natural radioactivity, there is also *artificial radioactivity*, where the nucleus also spontaneously transforms into another nucleus by emitting one or two protons along with alpha, beta, and gamma radiation.

When radiation emitted by certain elements passes through a magnetic field that is perpendicular to the plane of the drawing (Fig. 20.2), one part of it will deflect (bend) from its original path in the same way that positively charged particles would in this field. Another part of the beam will deflect as negatively charged particles would, while the third part of the beam will not deflect.

If a beam of radiation from a radioactive source is passed through the electric field of a capacitor with parallel plates, it also splits into three parts, one of which deflects toward the negative electrode, another toward the positive electrode, and the third does not change its direction of motion (Fig. 20.3).

These experiments also show that the beam from a radioactive source consists of three types of radiation. Radiation that behaves in electric and magnetic fields like a beam of positively charged particles is called alpha (α) radiation. Radiation that behaves in these fields like a beam of negatively charged particles is called beta (β) radiation, while radiation that does not change its direction of motion is called gamma (γ) radiation.

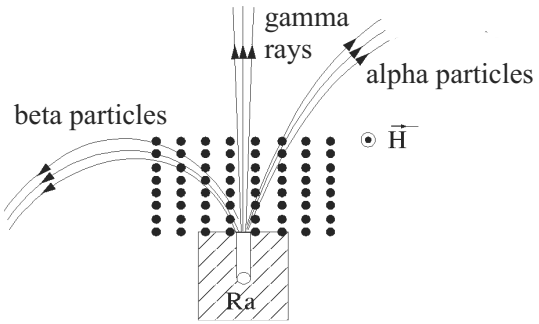

Figure 20.2: Splitting of a beam of radioactive radiation in a magnetic field perpendicular to the plane of the diagram

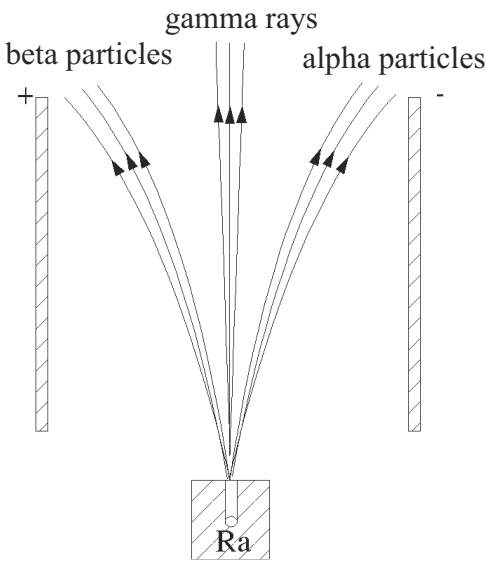

Figure 20.3: Splitting a beam of radioactive radiation in the electric field of a capacitor

20.5 Radioactive Decay Law

Radioactive nuclei decay (disintegrate) with a certain probability, independently of each another. The number of radioactive nuclei that decay dN in a very short time interval dt is proportional to this time interval dt and the number

of available radioactive nuclei (atoms) N at a moment t representing the beginning of that time interval dt ($dN \sim dtN$; Fig. 20.4).

The dN is equal to

$$dN = -\lambda N dt, \tag{20.6}$$

where λ is the radioactive decay constant and represents a characteristic value for each radioactive element. The minus

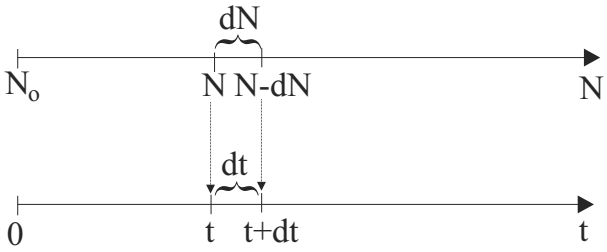

Figure 20.4: Along with the derivation of the expression for radioactive decay

sign is used because during radioactive decay the number of nuclei of a given element decreases. $(-dN/dt)$ represents the rate of decay of radioactive nuclei (radioactive decay rate).

The expression (20.6) can be written in the following form:

$$\frac{dN}{N} = -\lambda dt. \tag{20.7}$$

Integration of this expression within the given limits

$$\int_{N_o}^{N_t} \frac{dN}{N} = -\lambda \int_o^t dt \tag{20.8}$$

gives

$$N_t = N_o \exp(-\lambda t), \tag{20.9}$$

where N_o is the number of nuclei at the start of the observation ($t = 0$), and N_t is the number of undecayed nuclei after time t. Expression (20.9) represents *the radioactive decay law.* It shows that the number of undecayed nuclei decreases over time according to the exponential law.

Figure 20.5 shows the dependence of $\ln(N_t/N_o)$ on time t. The slope of the given line: $\ln(N_t/N_o) = -\lambda t$, determines the radioactive decay constant λ.

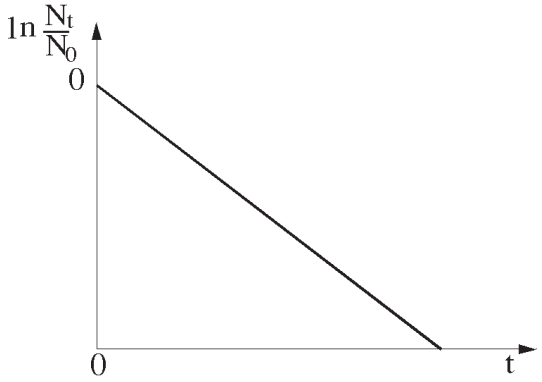

Figure 20.5: The dependence of $\ln(N_t/N_o)$ on time t

The number of nuclei that decay over time t is determined by the expression

$$\Delta N_t = N_o - N_t = N_o[1 - \exp(-\lambda t)]. \tag{20.10}$$

The time T in which half of the original number of nuclei decays is called the half-life. It is determined from the condition that for $t = T \Rightarrow N_T = N_o/2$, so equation (20.9) gives

$$\frac{N_o}{2} = N_o \exp(-\lambda T), \tag{20.11}$$

i.e.

$$T = \frac{\ln 2}{\lambda} = \frac{0,693}{\lambda}. \tag{20.12}$$

The half-life of all elements known to date ranges from $3.10^{-7}\ s$ to 5.10^{15} years.

From expressions (20.9) and (20.12), we obtain for N_t

$$N_t = N_o \exp(-\frac{\ln 2}{T} t). \tag{20.13}$$

The dependence of the number of undecayed nuclei N_t on time t is shown in Figure 20.6, and the number of undecayed nuclei of a radioactive element asymptotically approaches zero over time. Practically, it can be considered that the complete decay of a radioactive element occurs after $10\,T$, that is when only one-thousandth of the total initial number of nuclei remains undecayed: $N_t(t = 10\,T) = N_o e^{-\frac{\ln 2}{T} 10\,T} = N_o e^{-10 \ln 2} = N_o e^{\ln 2^{-10}} = N_o 2^{-10} \approx 0,001 N_o$.

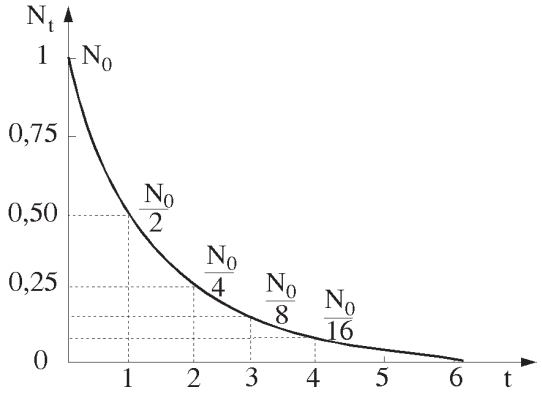

Figure 20.6: Number of undecayed nuclei N_t as a function of time t

It is known that the radioactive decay of the nucleus of a radioactive element follows a statistical law. This means that it is impossible to predict the time after which a particular nucleus will decay. Since the radioactive nuclei of the same element have different lifetimes ranging from zero to infinity, the average lifetime of a radioactive element is defined as the statistical average of the lifetimes of most nuclei.

20.6 Activity of a Radioactive Source

The activity of a radioactive source A represents the decay rate of radioactive nuclei

$$A = -\frac{dN}{dt}. \tag{20.14}$$

From the expression (20.6), it follows

$$A = -\frac{dN}{dt} = \lambda N. \tag{20.15}$$

If the initial activity of a radioactive source is A_o (activity at time $t = 0$), and the activity after a certain time t is A_t, then

$$A_t = \lambda N_t = \lambda N_o \exp(-\lambda t) = A_o \exp(-\lambda t), \tag{20.16}$$

where N_t is the number of nuclei at time t, and $A_o = \lambda N_o$ (N_o is the number of nuclei at time $t = 0$).

Expression (20.16) shows that the activity of a radioactive source A_t decreases over time according to the same (exponential) law by which the number of radioactive nuclei N_t decreases.

The unit of activity is s^{-1} and is called the Becquerel (Bq), which represents one decay per second.

20.7 Radioactive Equilibrium

As mentioned earlier, the nuclei of a radioactive element decay with the emission of alpha, beta, and gamma radiation (all or some of them can be emitted). The nuclei of the newly formed element (daughter nuclei) can be either stable or radioactive. If the daughter nuclei are radioactive and decay at the same rate as the parent element's nuclei, radioactive equilibrium occurs. According to equation (20.15), the decay rate of the parent element's nuclei, denoted by subscript 1, is equal to $\lambda_1 N_1$, where N_1 is the number of nuclei of the parent element, and λ_1 is its radioactive decay constant. The decay rate of the daughter element is $\lambda_2 N_2$, where N_2 is the number of daughter nuclei and λ_2 is its radioactive decay constant. When equilibrium is reached, the rate of formation of the daughter from its parent element will be equal to its decay rate:

$$\lambda_1 N_1 = \lambda_2 N_2. \tag{20.17}$$

Since the daughter element also decays, it represents the element of a new daughter. If radioactive equilibrium is reached again, there is a relationship completely analogous to equation (20.17):

$$\lambda_2 N_2 = \lambda_3 N_3. \tag{20.18}$$

From these considerations, it follows that there may be a series of successive decays, in which the general condition for equilibrium is given by the expression:

$$\lambda_1 N_1 = \lambda_2 N_2 = \lambda_3 N_3 = \lambda_4 N_4 = ..., \tag{20.19}$$

where N_1, N_2, N_3,... are the numbers of nuclei of the radioactive elements 1, 2, 3,... that are in radioactive equilibrium, and λ_1, λ_2, λ_3,... are their corresponding radioactive decay constants.

Accordingly, for any two members A and B of a radioactive series, regardless of whether one is a parent and the other a daughter, or several generations separate them, the equality holds

$$\lambda_A N_A = \lambda_B N_B. \tag{20.20}$$

Equation (20.20) can be written in the following form:

$$\frac{N_A}{N_B} = \frac{\lambda_B}{\lambda_A} = const. \tag{20.21}$$

Since λ_A and λ_B are constants, the ratio λ_B/λ_A is also constant, and therefore the ratio of the numbers of nuclei of any two members that decay in the process of radioactive equilibrium is constant.

20.8 Alpha and Beta Particles

As mentioned earlier, *alpha particles* are helium nuclei ${}_2^4He$, which are doubly positively charged (α^{++}). The emission of an alpha particle from a nucleus (known as alpha decay) takes place according to the following scheme:

$${}_Z^AX \rightarrow {}_{Z-2}^{A-4}Y + {}_2^4He, \tag{20.22}$$

where the letter X denotes the chemical symbol of the nucleus before alpha-particle emission (parent nucleus), and Y denotes the chemical symbol of the newly formed nucleus (daughter nucleus) after alpha-particle emission. Alpha decay is usually accompanied by the emission of a gamma quantum by the daughter nucleus. From this reaction, it can be seen that by emitting an alpha particle, the atomic number of the daughter nucleus decreases by 2, and its mass number decreases by 4.

Beta particles (beta radiation) are composed of a beam of electrons (β^- particles) or positrons (β^+ particles). These two particles (electron and positron) have the same masses and the same amount of electric charge but with opposite signs. Also, they differ in their lifetimes. While electrons can remain free indefinitely, the lifetime of a positron is on the order of a nanosecond. When a positron captures an electron, two gamma quanta with energies of $0,51\,MeV$ each are produced. This process is called *annihilation* and can be used for positron detection.

In beta decay, the kinetic energy of an electron or positron can range from zero to a certain maximum value, E_{max}. The spectrum of these particles is continuous and similar for most radioactive elements, as shown in Fig. 20.7. The curve of beta particle energy distribution for most radioactive elements has a maximum at approximately $E = E_{max}/3$.

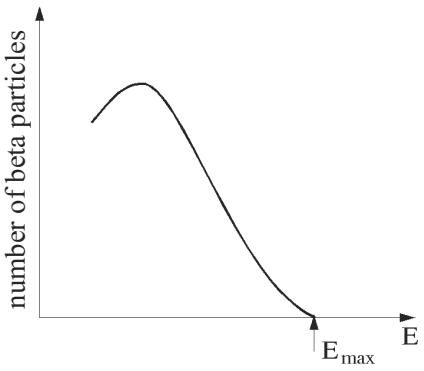

Figure 20.7: General form of the distribution of beta particles over energies (beta particle spectrum)

The continuous energy distribution of beta particles confused many scientists. Namely, by studying alpha particles and the causes of gamma radiation emission, it was concluded that discrete energy states of the nucleus exist. Since both the parent nucleus and the daughter nucleus have precisely defined internal energies, the question arises as to how beta particles can have different energy values.

Another difficulty is the question of how electrons (or positrons) can be emitted from a nucleus that consists only of protons and neutrons.

Pauli found the solution to the problem associated with the continuous spectrum. According to Pauli's hypothesis, along with the emission of a beta particle, another particle is emitted, which he called *the neutrino*. This hypothetical particle is electrically neutral and its mass is about four orders of magnitude smaller than that of the electron. During beta decay, the neutrino carries the difference between the maximum energy and the energy of the beta particle.

The distribution of energy between the beta particle and the neutrino has a statistical character. Sometimes, in a single decay event, most of the energy is carried by the beta particle, and sometimes by the neutrino. Very rarely, all the energy is carried by the beta particle. There are also cases where all or almost all the energy is carried by the neutrino, but these are also very rare. The first case corresponds to the energy maximum in Fig. 20.7, and the second to the vicinity of the origin of the same curve. It should be noted that the distribution between the electron and the neutrino is in good agreement with the continuous spectrum of beta particles.

The hypothesis of neutrino emission during beta decay not only explains the difficulty regarding the continuous distribution of beta particles, but also the difficulty concerning the conservation of spin. Namely, beta particles have

a spin of 1/2. The spontaneously emitted neutrino also has a spin of 1/2, which means the total angular momentum changes by 1 or 0. This fully satisfies the law of spin conservation, which, according to quantum theory, can only change by an integer.

After the theoretical postulates about the existence of neutrinos, an attempt was made to detect this particle. However, since the neutrino has a very small mass and is uncharged, it easily passes through matter without causing any ionization. Therefore, it is very difficult to detect.

Fermi developed a theory about two particles, the neutrino and the antineutrino, according to which these two particles relate to each other like the electron and the positron, where the antineutrino is the antiparticle of the neutrino. He also assumed that electrons and positrons are produced by the decay of protons and neutrons, thereby *resolving the dilemma of the origin of electrons and positrons in the nucleus.*

According to Fermi's theory, nuclei that contain a greater number of neutrons than necessary to be stable transform according to the following scheme:

$$ {}^{1}_{o}n \rightarrow {}^{1}_{1}p + {}^{o}_{-1}e + {}^{o}_{o}\tilde{\nu}, \tag{20.23} $$

i.e. the electron emission occurs along with an antineutrino (β^- particles are often denoted as ${}^{o}_{-1}e$, because they are negatively charged electrons). The upper number represents the number of nucleons, and the lower number represents the amount of elementary charge e. It can be seen that the law of conservation of the total number of nucleons and the law of conservation of electric charge hold.

The general formula for the process of electronic nuclear decay can be presented in the following form:

$$ {}^{A}_{Z}X \rightarrow {}^{A}_{Z+1}Y + \beta^- + {}^{o}_{o}\tilde{\nu}. \tag{20.24} $$

If a nucleus has an excess of protons, β^+ particle emission occurs (β^+ can be denoted as ${}^{o}_{+1}e$) along with a neutrino ${}^{o}_{o}\nu$ (the above-mentioned laws of conservation apply):

$$ {}^{1}_{1}p \rightarrow {}^{1}_{o}n + {}^{o}_{+1}e + {}^{o}_{o}\nu. \tag{20.25} $$

The general formula for positron decay is

$$ {}^{A}_{Z}X \rightarrow {}^{A}_{Z-1}Y + \beta^+ + {}^{o}_{o}\nu. \tag{20.26} $$

During beta decay, the daughter nucleus can be in one of the excited states. When it transitions from the excited state to the ground state, a gamma quantum is emitted. Therefore, the emission of a beta particle is accompanied by the emission of a gamma quantum, and the total energy of the nuclear transmutation is distributed among the electron (or positron), the antineutrino (neutrino), and the gamma quantum.

20.9 Gamma Radiation

Unlike alpha and beta particles, gamma radiation is a wave and belongs to the group of electromagnetic waves. Gamma radiation consists of photons (quanta of energy) that are carriers of the electromagnetic field. Photons have no rest mass, only moving mass, and their energy can be calculated as:

$$ E = h\nu = h\frac{c}{\lambda}, \tag{20.27} $$

where ν is the frequency of the wave of gamma radiation, c is its speed in vacuum, and λ is its wavelength.

It has been experimentally shown that gamma radiation does not appear as an independent form of radiation, but accompanies the emission of alpha and/or beta particles, and also occurs in nuclear reactions.

The gamma radiation spectrum is line spectrum. Such a gamma radiation spectrum represents one of the most important evidence for the existence of discrete energy states of the atomic nucleus.

It has been established that gamma radiation is emitted by the daughter nucleus, not by the parent nucleus. The daughter nucleus, at the moment of its formation after the emission of the alpha and/or beta particles from the parent nucleus, remains in an excited state for a duration of 10^{-14} to 10^{-13} s seconds. After this period, the nucleus transitions to the ground state, during which a gamma quantum is emitted.

Excited nuclei of certain radioactive elements, when returning to the ground state, can pass through a series of excited states. At each of these transitions from a higher to a lower excited state, a gamma quantum is emitted. Accordingly, the gamma radiation of a given radioactive element can contain several gamma quanta that differ from each other in energy.

When gamma rays pass through a material, they undergo absorption and scattering, which affects the reduction of their intensity:

$$ I = I_o \exp(-\mu d), \tag{20.28} $$

where I_o is the gamma ray intensity from a given source before passing through the material, I is the intensity after passing through a material of thickness d, and μ is the linear absorption coefficient.

It has been shown that the basic processes occurring during the interaction of gamma radiation with matter are *the photoelectric effect*, *the Compton effect*, and *electron-positron pair creation.*

The photoelectric effect, or photoelectric absorption of gamma radiation (gamma photons), is a process in which an atom absorbs a gamma quantum and then emits an electron. In this case, the gamma quantum transfers all its energy to the electron. Since the electron is released from one of the atom's inner energy levels, the vacant spot is filled by an electron from higher energy levels, leading to the emission of characteristic X-ray radiation. The photoelectric effect is the dominant mechanism of emission in the low-energy range of gamma quanta ($h\nu < 100\,keV$). The probability of the photoelectric effect occurring is higher in materials with larger atomic numbers Z. As Z increases, absorption rises rapidly and is proportional to Z^3.

When the energy of a gamma quantum is greater than $500\,keV$, the Compton effect occurs. In this process, a photon collides with an electron in an atom, the electron leaves the atom, and the photon changes direction and loses part of its energy. The Compton effect is evidence of the particle-wave duality of the photon, as the photon undergoes a collision with an electron, which is a characteristic of particles.

When the energy of a gamma-ray photon $h\nu$ is greater than $1,02\,MeV$, it is possible for an electron-positron pair to be created. Specifically, when such a photon encounters a heavy nucleus, it transforms into an electron-positron pair, consuming $1,02\,MeV$ of its energy. The remaining energy of the photon appears as the kinetic energy of the electron and positron. The occurrence of a scattered quantum $h\nu'$ is also possible (Fig. 20.8).

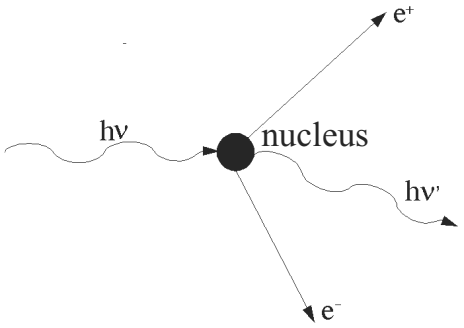

Figure 20.8: Electron-positron pair formation during photon scattering

20.10 Electron Capture

Electron capture is a phenomenon in which the nucleus absorbs an electron from the K, L or M shell. Most commonly, this absorption occurs from the K-shell, causing a proton to convert into a neutron and emitting a neutrino:

$$ {}^{1}_{1}p + e^{-} \rightarrow {}^{1}_{o}\,n + {}^{o}_{o}\,\nu. \tag{20.29} $$

The newly formed nucleus is usually in an excited state. When it transitions to the ground state, a gamma ray (gamma quantum) is emitted.

The general diagram illustrating how the electronic capture is performed can be shown as follows:

$$ {}^{A}_{Z}X + e^{-} \rightarrow {}^{A}_{Z-1}\,Y + {}^{o}_{o}\,\nu. \tag{20.30} $$

The electron levels in the atom from which the affected electron was removed are filled by an electron from higher levels, resulting in characteristic X-ray radiation.

20.11 Acceleration of Charged Particles

The interaction of high-energy particles with the nuclei of atoms of certain elements has explained many properties of nuclei and elementary particles. This also explains the production of a larger number of products in artificial transmutation (transformation) of the nucleus. Since the energy of particles emitted by natural radioactive sources does not exceed $20\,MeV$, their use is considerably limited. However, by applying modern methods of accelerating charged particles, their energy can be increased by several orders of magnitude compared to the energy of particles emitted by natural radioactive sources.

There are several devices for accelerating charged particles, most commonly called accelerators. Here, two types of these devices will be covered.

20.11.1 Linear Accelerators

High, rapidly fluctuating voltages are used to accelerate particles to high energies, and the devices that use these voltages are linear accelerators and cyclotrons. A diagram of a linear accelerator is shown in Fig. 20.9. It consists of a long tube containing multiple electrodes in the form of hollow cylinders, whose axes are aligned. The odd-numbered cylinders are connected to one pole (A), and the even-numbered ones to the other pole (B) of the high-frequency, high-voltage generator. In this way, at any given moment, the even and odd cylinders have opposite potentials. For instance, during a given half-period of the generator, all odd cylinders will have a positive potential, and the even ones

a negative potential. In the next half-period, the potentials will reverse, with the odd cylinders becoming negative and the even ones positive, and so on.

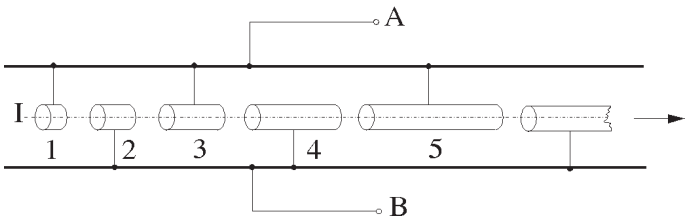

Figure 20.9: Schematic representation of a linear accelerator

Let positive ions from the source I move along the common axis of the cylinders from left to right toward the negatively charged first cylinder and enter it. While moving through the first cylinder, the speed of the ions remains unchanged. At the moment the ions leave the first cylinder, the first cylinder becomes positively charged, and the second cylinder negatively charged, and the ions are accelerated in the space between the two cylinders as they exit the first cylinder. Then the positive ions enter the negatively charged second cylinder and move through it at a constant speed higher than the initial speed. The length of the cylinder is adjusted so that the potential changes sign at the moment the ions reach the exit of the second cylinder.

At the moment when ions exit the second cylinder, the second cylinder becomes positively charged and the third cylinder becomes negatively charged, causing the positive ions to accelerate again between these cylinders. By successively increasing the length of the cylinder, due to the increase in ion velocity, it is possible to keep the ions exactly in phase with the change of potential polarity. This results in an increase in the energy of the ions each time they move between the two cylinders.

This type of linear accelerator is primarily used for accelerating heavy particles such as protons, deuterons, and alpha particles. For accelerating electrons, a linear accelerator of a different design is used than that for accelerating positive ions, because electrons, due to their small mass, at energies around $2\,MeV$ move at speeds close to that of light, and relativistic effects occur. The operating principle of that accelerator is based on the fact that electrons can be forced to travel with high-frequency radio waves and absorb energy from these waves.

20.11.2 Cyclotron

To accelerate particles to high energies using a linear accelerator, a large number of electrodes is required, which makes this device very bulky. For this reason, another type of accelerator is often used for accelerating charged particles, called a cyclotron. Its schematic diagram is shown in Fig. 20.10.

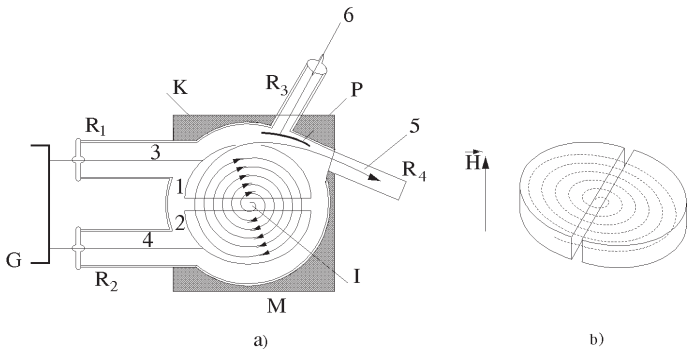

Figure 20.10: Schematic representation of a cyclotron

Between the poles of magnet M (part of one pole is shaded in Fig. 20.10a)is placed a vacuum chamber K of cylindrical shape with extensions R_1, R_2, R_3 and R_4. Inside the chamber are the duants 1 and 2, which are shown in detail in Fig. 20.10b. The duants are made by forming a hollow cylinder from a conductive material, then cutting it along the axis into two parts and separating the resulting pieces. The duants are connected to the high-frequency and high-voltage generator G via conductors 3 and 4. A particle source I is placed between the duants. The charged particles emitted from the source start moving in the electric field between the duants and deflect in the magnetic field, moving in a circular path due to the Lorentz force. The radius r of the circular path of a charged particle, at its speed v that is much less than the speed of light, can be determined from the equality of the Lorentz and centrifugal forces

$$qvB = \frac{mv^2}{r}, \tag{20.31}$$

where q is the charge of the particle, m its mass, v its velocity, and B the magnetic induction. From equation (20.31) it follows that the radius of the particle is

$$r = \frac{mv}{qB}. \tag{20.32}$$

The velocity of a particle along a circular path (peripheral or tangential velocity) v is related to the angular velocity ω and the rotation frequency ν:

$$v = \omega r = 2\pi\nu r. \tag{20.33}$$

From expressions (20.32) and (20.33) , it follows

$$\nu = \frac{qB}{2\pi m}. \tag{20.34}$$

Due to the alternating electric field, charged particles will move from one electrode (duant) to another, which results in their acceleration. Let source I emit positive ions. If, at the moment of emission of such an ion, duant 1 is at a negative potential and duant 2 at a positive potential, the ion will move toward duant 1 along a semicircular trajectory at a constant speed and, after some time, reach the gap between the duants. The oscillation frequency of the high-frequency generator is adjusted so that at the moment the ion reaches the gap, the polarity of the duants switches. This allows the charged particle to end up in duant 2 but with a higher speed than it had at the beginning of its motion. Because of that, the circular trajectory in duant 2 under the influence of the magnetic field will have a larger radius r, as follows from expression (20.32).

One of the important characteristics of a cyclotron is that the time required for a charged particle to traverse a given semicircular distance does not depend separately on the particle's speed and the radius. Therefore, an increase in distance length due to a larger radius is compensated by an increase in the particle's speed. The distance length traveled by an ion in one duant is πr, while the time T to cover this distance at speed v is

$$T = \frac{\pi r}{v}. \tag{20.35}$$

From expressions (20.32) and (20.35), it follows

$$T = \frac{\pi m}{qB}. \tag{20.36}$$

Accordingly, when the frequency of the generator is adjusted depending on the charge of the particle and the strength of the magnetic field, the motion will be in phase with the changes in electric potential on the cyclotron's duants. Each time the particle passes through the gap between duants 1 and 2, it will receive a certain amount of energy due to the change in the potential sign on them. The particle's energy continuously increases, and it describes a spiral in the plane with an ever-growing radius. After a certain time, the ion reaches the periphery of the duants. Then, using the deflecting plate P, which is negatively charged, the positively charged particle is guided out of the circular trajectory and exits through the opening R_4 along a straight-line trajectory.

If the potential difference on the duants at the moment a charged particle passes is U, and the number of passes is N, then the energy it receives is given by the expression

$$E = qNU. \tag{20.37}$$

A cyclotron is used to accelerate heavy particles such as protons, deuterons, alpha particles, or ions of heavy atoms. It is suitable for accelerating particles to energies on the order of $100\,MeV$.

20.12 Nuclear Fission

Nuclear fission is the splitting of a nucleus and it can occur spontaneously (spontaneous fission) or under the influence of a neutron. When the nucleus of heavy elements, such as uranium, plutonium, etc., is bombarded, it can split into two nuclei called fragments, and neutrons and gamma quanta are also emitted. Such nuclear reactions can be induced by bombarding heavy nuclei with neutrons, protons, alpha particles, deuterons, as well as gamma quanta.

Fission was first discovered in uranium. In the spontaneous fission of uranium ${}^{238}_{92}\mathrm{U}$, the uranium nucleus splits into two nuclei that are near the middle of the periodic table of elements. The half-life of the uranium ${}^{238}_{92}\mathrm{U}$ nucleus is 10^{16} years. Other heavy nuclei, such as plutonium, californium, fermium, and others, also undergo spontaneous fission, and their half-lives are also long.

The element uranium is found in nature as a mixture of isotopes ${}^{234}_{92}\mathrm{U}$, ${}^{235}_{92}\mathrm{U}$ and ${}^{238}_{92}\mathrm{U}$. Uranium ore mostly contains the isotope ${}^{238}_{92}\mathrm{U}$ at $99,282\%$, then ${}^{235}_{92}\mathrm{U}$ at $0,712\%$, and ${}^{234}_{92}\mathrm{U}$ at $0,006\%$. All three isotopes transform under the influence of neutrons, but in different ways. Uranium fission can be achieved with both thermal (slow) and fast neutrons. Fission with thermal neutrons is dominant in ${}^{235}_{92}\mathrm{U}$, while fast fission is dominant in ${}^{238}_{92}\mathrm{U}$. Later research showed that fission can occur when heavy nuclei are bombarded with other particles as well.

A characteristic accompanying phenomenon of fission is the emission of neutrons. Namely, during the splitting of a nucleus, among other things, neutrons are also emitted, which under certain conditions can be used to split other nuclei. This phenomenon is characteristic, for example, of isotopes uranium ${}^{235}_{92}\mathrm{U}$ and uranium ${}^{233}_{92}\mathrm{U}$, as well as plutonium ${}^{239}_{94}\mathrm{Pu}$.

The fission process occurs according to the following scheme:

$${}^{A}_{Z}X + {}^{1}_{o}n \rightarrow {}^{A_1}_{Z_1}Y_1 + {}^{A_2}_{Z_2}Y_2 + k{}^{1}_{o}n + Q, \tag{20.38}$$

where Y_1 and Y_2 are fragments of fission, $Z = Z_1 + Z_2$, $A \approx A_1 + A_2 + k + Q/c^2$, and k represents the number of neutrons released during the fission of the nucleus. During fission, in addition to neutrons, a large amount of energy Q and several gamma quanta are also released.

20.13 Nuclear Chain Reaction

Since neutrons are released during the fission of a nucleus, they can induce the fission of new nuclei, resulting in a self-sustaining fission called *nuclear chain reaction*. This type of nuclear reaction is exothermic, i.e., it is accompanied by the release of a large amount of energy. For example, the energy released during the fission of $1\,g$ of uranium is about $10^{11}\,J$ or $20\,MWh$. This energy is about 10^9 times greater than the energy released by burning $1\,g$ of coal.

A schematic representation of a chain reaction is given in Fig. 20.11. However, not all neutrons are captured by the nuclei of the fission material. Besides the fission material, there are other materials present that are unable to undergo fission upon neutron capture. Furthermore, some neutrons leave the area where the fission material is located, so they are lost for fission. In addition, not all neutrons that strike the nuclei of fission materials succeed in causing fission. The excess energy received from neutrons can be released by the nucleus, for example, through the emission of a gamma quantum.

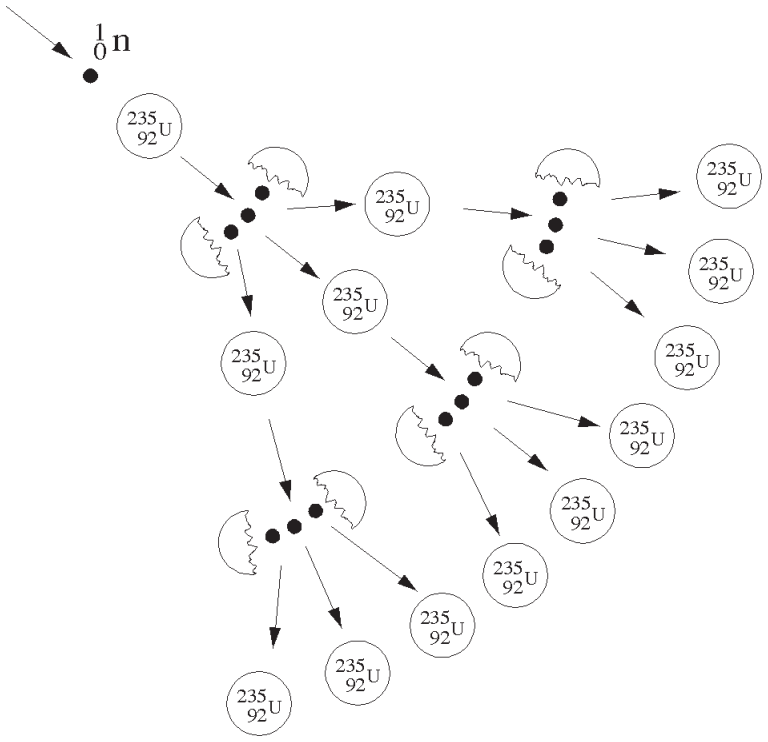

Figure 20.11: Nuclear chain reaction of uranium $^{235}_{92}$U

20.14 Nuclear Fusion

A large amount of energy can be released not only during the fission of heavy nuclei, but also during the fusion (synthesis) of light nuclei into a nucleus of greater atomic mass. This process is known as *nuclear fusion*. For fusion to occur, the repulsive electrostatic forces between the nuclei involved in the process must be overcome. This requires that the nuclei come close enough so that the attractive nuclear forces exceed the repulsive electrostatic forces. These distances are on the order of $10^{-15}\,m$. To overcome the repulsive electrostatic forces, the nuclei that are fusing must have high energies, or high velocities, to reach these distances. The fusion process can be achieved in two ways: either by accelerating the nuclei with an accelerator or by heating them to very high temperatures.

Fusion of hydrogen nuclei, as well as the nuclei of other light elements, is only possible if all electrons are removed from the atomic shell. For example, at a temperature of $2 \cdot 10^5\,K$ hydrogen atoms are completely ionized, meaning that at this temperature the gas consists of free protons and electrons. Since the average energies of protons and electrons are equal, they do not change in their mutual collisions. This state of matter is called *high-temperature plasma*. Because the number of protons is equal to the number of electrons in any part of the observed volume of such plasma, the average charge density is zero. The probability that a neutral hydrogen atom is formed during a collision between a proton and an electron in high-temperature plasma is low, and thus can be neglected.

When a proton with an energy of about $1\,MeV$ passes through a high-temperature plasma, the conditions for a fusion process known as *thermonuclear fusion* are met. Fusion can also occur at lower proton energies due to the tunneling effect.

The fusion of light nuclei into a nucleus with a higher atomic mass is accompanied by the release of a large amount of energy. The following reactions are given only as examples (they do not need to be memorized):

$$\begin{aligned} {}^{2}_{1}H + {}^{2}_{1}H &\rightarrow {}^{3}_{1}H + {}^{1}_{1}p + 4 \text{ MeV} \\ {}^{2}_{1}H + {}^{2}_{1}H &\rightarrow {}^{3}_{2}He + {}^{1}_{0}n + 3{,}3 \text{ MeV} \\ {}^{2}_{1}H + {}^{3}_{1}H &\rightarrow {}^{4}_{2}He + {}^{1}_{0}n + 17{,}6 \text{ MeV} \\ {}^{2}_{1}H + {}^{6}_{3}Li &\rightarrow {}^{4}_{2}He + {}^{4}_{2}He + 22,4 \text{ MeV} \end{aligned} \tag{20.39}$$

The energy released per nucleon during fusion is significantly larger than that released during the fission of heavy nuclei. The energy released in the fission of a uranium nucleus ${}^{235}_{92}\mathrm{U}$ is $195\,MeV$, which corresponds to approximately $0,82\,MeV$ per nucleon. In the fusion of deuterium and tritium, as seen from reaction (20.39), the energy per nucleon amounts to $17,6/5 \approx 3,5\,MeV$. It is assumed that thermonuclear reactions are one of the sources of energy for the Sun and stars.