


A.V. AGANOV, K.S. USACHEV

MEDICAL PHYSICS

Part 1

Mechanics. Molecular physics


KAZAN FEDERAL UNIVERSITY

A.V. AGANOV, K.S. USACHEV

MEDICAL PHYSICS

Part 1

Mechanics.
Molecular physics

KAZAN
2022

УДК 53:60
БКК 22.3.:5

Printed on recommendation of Educational-methodical commission
Institute of Physics Kazan (Volga Region) Federal University
Protokol № 3, November 29, 2022

Reviewers:
Professor **D.A. Tayurskii**
Professor **Yu.A. Chelyshev**

Edited by Aganov A.V.

Aganov A.V., Usachev K.S.
Medical Physics / A.V. Aganov, K.S. Usachev – Kazan.: Kazan University Press, 2022. – Part 1. Mechanics. Molecular physics – 280 p.

The textbook is a short course in general physics for biomedical specialties under the title "Medical Physics". It is based on the third edition of the textbook of the same name in Russian (author prof. A.V. Aganov) and the course of the same name, read at Kazan Federal University by prof. K.S. Usachev for English-speaking foreign students of the Institute of Fundamental Medicine and Biology of KFU. It takes a more detailed look at some of the hard-to-learn sections of the general physics course, important in terms of understanding the place of physics in medicine, especially the part that is close to everyday life. The textbook is addressed primarily to students of medical and biological specialties, but can also be recommended to students of physics (undergraduate "Physics" and "Biotechnical systems and technologies") who are focused on continuing their studies in the "Medical Physics" master's program. Most of the material can be recommended as a theoretical introduction to the laboratory work of the general (for students of medical and biological specialties) and specialized (in the master's program "Medical Physics") workshops on medical physics. The first part consists of the following sections: mechanics, oscillations and waves, the bases of the Kinetic Molecular Theory of Matter, the principles of thermodynamics, physical systems and phenomena (processes). The Introduction considers the stages of the formation of modern natural science in the context of the place of physics and physical methods of research in the development of fundamental and clinical medicine.

The material is presented in a concise and comprehensible form. The book is well illustrated. A major list of foreign literature is included, but there are no borrowings. Also removed Appendices, which go beyond the scope of the Program of the discipline. References to only English books are given. Where appropriate, terms from foreign textbooks are given in brackets.

The authors are grateful to Professor D.A. Tayurskii who took an active part in forming the concept of the course and in preparing the Russian-language edition; Dr. I.A. Latfullin for advice and constructive comments on the coverage of medical and biological issues; Associate Professors: L.F. Galiullina, A.V. Khaliullina and A.R. Yulmetov for careful review of the manuscript and helpful comments on the material of the general physics course; N.F. Galiullina (Engineer, Department of Medical Physics) for support in the preparation of the illustrations and proofreading the manuscript for publication, Associate Professor A.N. Turanov for review the English version of this book and read a Medical Physics course.

Table of contents

FOREWORD

This textbook represents the first part of a course "Medical physics" (a course in general physics for students of medical and biological faculties) which assumes that it considers the interrelationship of physics and medicine. It is based on the third revised edition of the "Medical physics" textbook (A.V. Aganov, "Medical physics. Part 1") in Russian and the course of lectures for foreign students in English. The main principles of the general physics course are presented in various textbooks published by foreign authors. There are also textbooks where the main sections of a general physics course and their applications to various medical and biologic tasks are more or less fully considered. We have not used them, but references to them are given.

The implementation of competency-based approach has changed the structure of curricula and the organization of the educational process so that a significant part of the material should be mastered by students in practice-oriented classes and self-study. It completely corresponds to the students' wishes, expressed in the results of questionnaire replies, to shift the accents towards a more practical orientation of mastering the discipline in the diadrome "to know – to be able". However, the volume of the theoretical course of the general physics for nonphysical specialties, as a rule, is reduced to such a degree (34 hours for a full course) that the integrity of perception of a material far not simple for mastering is lost. The clipo-digest method of learning disciplines with the use of Internet-resources is typical for the majority of students of any specialty – the new Internet generation. Meanwhile, physics and physical methods in different forms and degrees not only enter the practice of doctor-researcher and clinician, but become almost the main tool of medical examination (for example, complex: doppleroscopy, cardiography, ultrasound, CT, MRI, etc.) and medical and biological research. The course "Medical Physics" – which contains only 18 lectures from mechanics to nuclear physics, covers only basic sections of general physics with varying degrees of coverage of the relationship of a particular phenomenon of physics to medical applications. The part of the theoretical course is in addition mastered on practical lessons in the laboratory of the general practical work "Medical physics", where performance of laboratory works on a course of the general physics of a medical-biological orientation (120 hours a year) is provided. Naturally, the presented material is quite enough for the preparation of the student for the attestation on this course. But the main purpose of this course – to arouse the interest of future physician in physics and to realize the importance of its study for deeper mastering of the chosen profession, both by independent study of material, and during the discussion, which is simple enough to organize in the hours allocated for laboratory practice. The selected material is the most striking, from the author's point of view, examples of the relationship between physics and medicine, the most important and useful information for anyone, and those fragments of the general physics course that are traditionally poorly understood by students or are not covered in the lecture course.

We believe that the textbook will be useful both for undergraduate students of "Physics", "Biosystems and biotechnologies", oriented for continuing education in master's degree program "Medical physics", and for those, who have already master's degree program "Medical physics", as a brief theoretical introduction to carrying out laboratory works of the special practical work "Medical physics".

INTRODUCTION

Physics is the science of the most general properties and forms of motion of matter, the basis of all natural science. The influence of natural phenomena, the use of various physical influences on people (sunlight, thermal waters, etc.) for therapeutic purposes has been known since ancient times. As a field of knowledge, physics and medicine have been developing, mutually enriching each other for several centuries. But there was no clear distinction between them, and our great predecessors were multifaceted natural scientists. The division into separate fields of knowledge appeared only when the evidential basis (experimental) and mathematical apparatus were created, on the basis of which the quantitative relations – laws were determined.


M. Born
(1882–1970)

The evolution of ideas about the phenomena of nature and the methodology of their cognition is described by Max Born[*] in his book (collection of articles) "Physics in My Generation": “*The doors through which Nature imposes her presence on us are the senses. Their properties determine the extent of what is accessible to sensation or to intuitive perception. The further we go back in the history of the sciences, the more we find the natural picture of the world determined by the qualities of sense. Older physics was subdivided into mechanics, acoustics, optics and theory of heat. We see the connections with the organs of sense, the perceptions of motion, impressions of sound, light, and heat. Here the qualities of the subject are still decisive for the formation of concepts. The development of the exact sciences leads along*

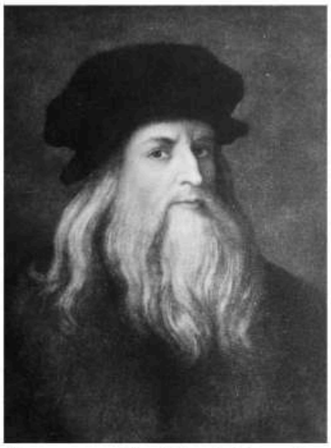
---

[*] ***Max Born,*** (11.12.1882, Breslau – 5.01.1970, Göttingen) – German and British theoretical physicist and mathematician, one of the creators of quantum mechanics. Winner of the Nobel Prize in Physics (1954). Member of several scientific societies of the world, including a foreign member of the Academy of Sciences of the USSR (1934) (Born, M. Physics in my generation [Text] / M. Born. Springer-Verlag Berlin Heidelberg, 1969. 172 p.).

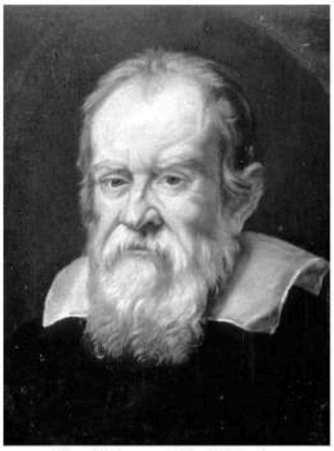
a definite path from this state to a goal which, even if far from being attained, yet lies clearly exposed before us: it is that of creating a picture of nature which, confined within no limits of possible perception or intuition, represents a pure structure of concepts, conceived for the purpose of depicting the sum of all experiences uniformly and without inconsistencies.

Nowadays mechanical force is an abstraction which has only its name in common with the subjective feeling of force. Mechanical mass is no longer an attribute of tangible bodies but is also possessed by empty spaces filled only by ether radiation. The realm of audible tones has become a small province in the world of inaudible vibrations, distinguishable physically from these solely by the accidental property of the human ear which makes it react only to a definite interval of frequency numbers. Modern optics is a special chapter out of the theory of electricity and magnetism, and it treats of the electro-magnetic vibrations of all wavelengths, passing from the shortest 7-rays of radioactive substances (having a wavelength of one hundred millionth of a millimeter) over the X-(Röntgen) rays, the ultraviolet, visible light, the infra-red, to the longest wireless (Hertzian) waves (which have a wave-length of many kilometers). In the flood of invisible light that is accessible to the mental eye of the physicist, the material eye is almost blind, so small is the interval of vibrations which it converts into sensations. The theory of heat, too, is but a special part of mechanics and electrodynamics. Its fundamental concepts of absolute temperature, of energy, and of entropy belong to the most subtle logical constructions of exact science, and, again, only their name still carries a memory of the subjective impression of heat or cold.

Inaudible tones, invisible light, imperceptible heat, these constitute the world of physics, cold and dead for him who wishes to experience living Nature, to grasp its relationships as a harmony, to marvel at her greatness in reverential awe.” (P. 1–2).

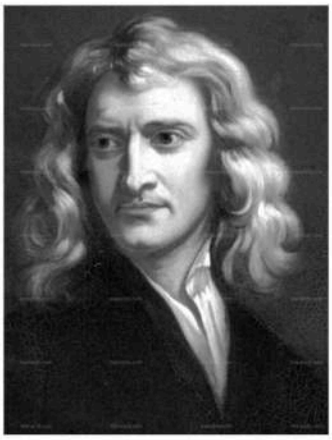


Leonardo Da Vinci
(1452–1519)



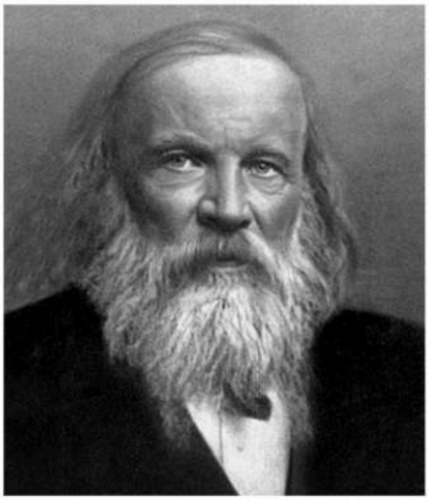
Galileo Galilei
(1564–1642)

I. Kepler
(1571–1630)



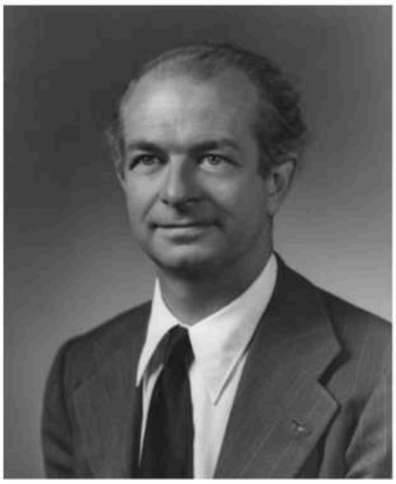
I. Newton
(1643–1727)

It is well known that scientific knowledge (modern natural science) begins with observational astronomy, the success of which is due to the appearance of the telescope, which «opened a window on the Universe». Its creation is connected with the names of I. Kepler and Galileo Galilei – the founders of experimental physics. However, it is necessary to note that the first sketches of a telescope were made by one of the greatest representatives of Renaissance Leonardo da Vinci a hundred years earlier. The first scientific description of the results of astronomical observations was the empirical laws of planetary motion, established by I. Kepler in 1609–1619. Based on these laws, I. Newton derived the law of universal gravitation. In his major work "Mathematical Principles of Natural Philosophy" (1686) he practically offered a complete mathematical model, which included the law of gravitation, laws of motion, and a system of methods for mathematical investigation (mathematical analysis). This was the beginning **of classical physics.**



D.I. Mendeleev
(1834–1907)

Chemistry. A century later A. Lavoisier introduced the concept of chemical elements (1789). At the beginning of the 19th century, atomic-molecular concepts were formed, according to which each chemical element was considered as a separate kind of atom (D. Dalton and others). The logical consequence of this was the discovery by D.I. Mendeleev[*] of the periodic dependence of the properties of chemical elements on their mass (the Periodic Table of Mendeleev). In 1912–1914 a principal amendment was made to Mendeleev's periodic law: the periodicity of change in the properties of elements is determined not by the atomic weight (mass), but by the charge of the nucleus equal to the atomic number and the number of electrons surrounding the nucleus. Their distribution determines the chemical properties of the elements. This amendment was based on G. Moseley's law (1913), according to which the frequency of the characteristic X-ray radiation (i.e., the radiation spectrum of the atom of a chemical element) is related to its atomic number.



L. Pauling
(1901–1994)

A. Leeuwenhoek
(1632–1723)

But the article[**] gives a later date of the birth of modern chemistry – 1923 – the publication of L. Pauling's article devoted to the X-ray structure

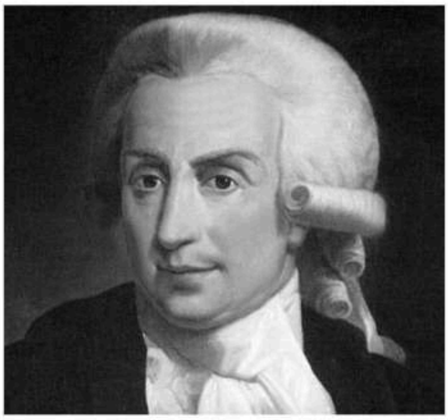
---

[*] ***D.I. Mendeleev*** was a Russian scientist and encyclopedist: chemist, physicist, metrologist, economist, technologist, geologist, meteorologist, oilman, teacher, balloonist, instrument maker. He was professor at St. Petersburg University; corresponding member of the Imperial St. Petersburg Academy of Sciences in the Physics field. Among his most famous discoveries was the periodic law of chemical elements, which is one of the fundamental universal laws, essential to the whole of natural science. He is the author of the classic book "Fundamentals of Chemistry".

[**] ***Borman,* S.** Chemistry Crystallizes Into Modern Science. The past 75 years have marked profound changes in the content, scope, and direction of the field [Text] / S. Borman, R. Dagani, R.L. Rawls, P.S. Zurer // Chemical & Engineering News. 1998. Copyright 1998 by the American Chemical Society. http://pubs.acs.org/cen/hotarticles/ cenear/980112/crystal2.html.

analysis of molybdenite crystals[*]. This coincided in time with the formation of quantum physics and quantum ideas in chemistry.

Biology and Medicine. Here, first of all, we should mention the invention of the microscope and its application to the study of living organisms. This outstanding achievement belongs to A. Leeuwenhoek[**], who combined the qualities of a mechanic and a naturalist. After mastering the craft of grinding, Leeuwenhoek became a very skilled and successful lens maker. By fitting his lenses into metal frames, he assembled a microscope (in total he created about 500 lenses and at least 25 microscopes, 9 of which survive today – they give 275x magnification) and used it to conduct the most advanced research at the time. A. Leeuwenhoek sketched the observed objects and described his observations in letters, which he sent to the Royal Society of London and some scientists for more than 50 years. In 1673 one of them was first published in the journal of the Royal Society of London *"Philosophical Transactions"*. This is most likely the beginning of evidence-based medicine. He discovered red blood cells, described bacteria (1683), yeast, protozoa, lens fibers, skin epidermis scales, sketched spermatozoa (1677), etc.[***]



L. Galvani
(1737–1798)

---

[*] ***Linus Carl Pauling***, (February 28, 1901, Portland, Oregon – August 19, 1994, Big Sur, California) – American chemist, crystallographer, winner of two Nobel Prizes in Chemistry (1954) and Peace Prize (1962) and the International Lenin Prize "For Consolidation of Peace between Nations" (1970).

[**] ***Antoni van Leeuwenhoek***, (Thonius Philips van Leeuwenhoek; 24.10.1632 – 26.08.1723, Delft, Netherlands) – naturalist, constructor of microscopes, founder of scientific microscopy. There is information that the microscope was created by H. Jansen and Z. Jansen also lived in the Netherlands, around 1590.

[***] ***R. Hooke,*** who observed living cells (cell) in his improved microscope, also made a significant contribution. The term "cell" was introduced by R. Hooke (see Micrographia, 1666).

The 18th century was marked by the creation of the doctrine of electricity and electro-physiology. One of the founders was L. Galvani*.

The first studies of electrical phenomena during muscle contraction ("animal electricity") belong to him. In 1791, his "Treatise on the Forces of Electricity in Muscular Motion" described this famous discovery. Galvani suggested that the muscle is a kind of battery of Leiden jars, continuously excited by the action of the brain, which is transmitted by nerves. Galvani's name also survives in the name of some so-called galvanic apparatus and processes. Among Galvani's followers was A. Volta. It is noteworthy that Galvani's work was preceded by the work of English anatomist J. Walsh, who proved the electrical nature of the stingray's impact, and Scottish surgeon and anatomist J. Hunter, who gave an exact description of this marine animal's electrical organ (1773). L. Galvani also discovered the emergence of a potential difference when different kinds of metal and electrolyte came into contact. At first sight it seems strange that the main law of nature – the law of conservation of energy was formulated by physician R. Mayer (19th century). Note that M. Lomonosov was the first in establishing the law of conservation of matter only (18th century) (these issues are discussed in detail in this book). But this is only at first sight: the life of primitive man depended on fire (heat – external energy) and food (internal energy delivered to the human body with food – substance). Just as it is today.

Physicians and physics contributions to the development of science are a separate topic beyond the scope of this book. Consider the major later discoveries in physics that "came from physics" and influenced the development of biology and medicine.

At the turn of the 19th and 20th centuries, two great discoveries were made that had a revolutionary impact on the development of science and humanity: the discovery by W.K. Röntgen** of a new type

---

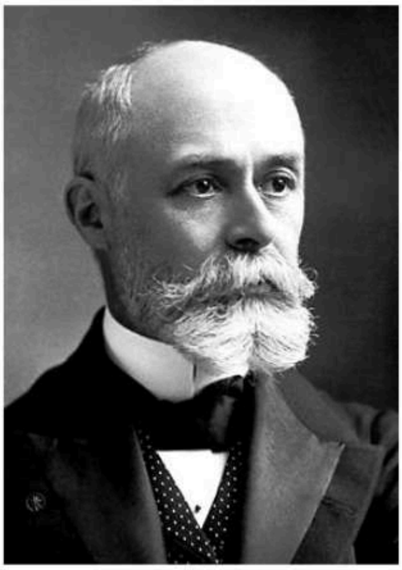
* ***Luigi Galvani***, (09.09.1737 – 4.12.1798) – Italian physician, anatomist, physiologist and physicist, one of the creators of electrophysiology and the doctrine of electricity, the father of experimental electrophysiology.

** ***Wilhelm Conrad Röntgen,*** (27.03.1845 – 10.02.1923) – German physicist, first Nobel Prize winner in the history of science (1901).

of electromagnetic radiation, which he called X-rays and the discovery by A. Becquerel* of spontaneous radioactivity.


W. Röntgen
(1845–1923)

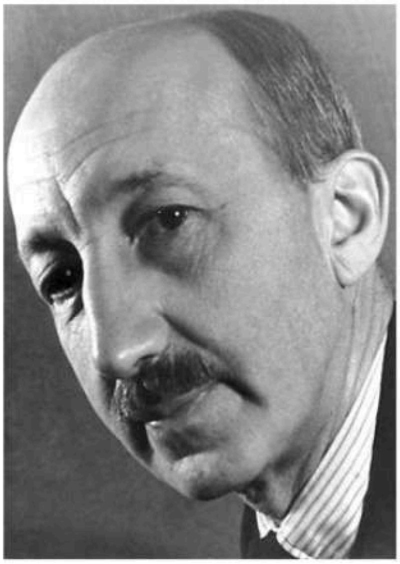
The first discovery is dated December 22, 1895, when W.K. Röntgen made a public announcement about it at the Physics Institute of Würzburg University. And on December 28, 1895 in the journal of the Würzburg Physico-Medical Society his article "On a new type of rays" was published. Already in early 1896, attempts to find an answer to the question of the nature of X-rays led A.A. Becquerel to the accidental, as sometimes written, discovery of radioactivity. And soon Pierre and Marie Curie (Skłodowska-Curie) began to investigate the spontaneous radiation of matter (nuclei).


A. Becquerel
(1852–1908)

X-rays are electromagnetic waves in the range of ~10^3 to ~10^{-2} Å (~10^{-7} to ~10^{-12} m). Due to its high energy (this scale of electromagnetic waves corresponds to photon energies in the range from ~10 eV to ~1 MeV, i.e. between the energies of ultraviolet and gamma ray quanta), X-rays are able to penetrate deep into materials that are opaque in the visible range. This property of X-rays was already evident in the first experiments, and the practical possibilities of fluoroscopy, including in medical diagnostics (known as X-ray examination, fluoroscopy, radiography), were evident immediately, but W.K. Röntgen refused to sign any commercial contract related to the use of his discovery.

---

* ***Antoine Henri Becquerel,*** (15.12.1852 – 25.08.1908) – French physicist, Nobel Prize in Physics "in recognition of the extraordinary services he has rendered by his discovery of spontaneous radioactivity" awarded jointly with Pierre and Marie Curie (1903).



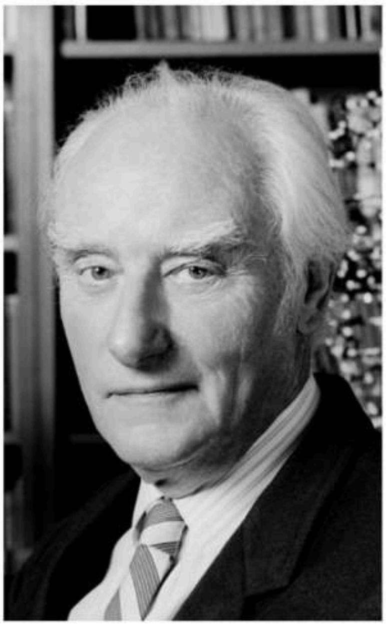
D. de Hevesy
(1885–1966)

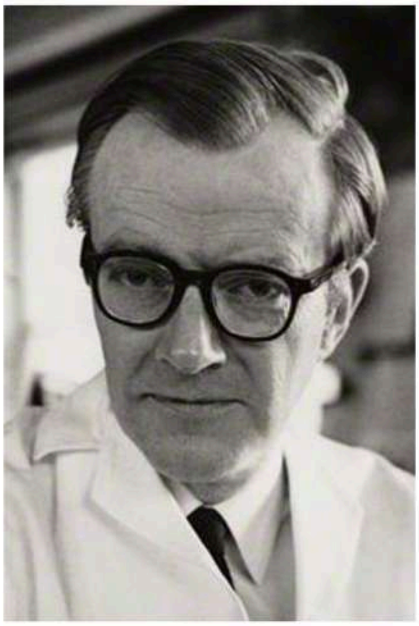
M. von Laue
(1879–1960)

He held the view that his discovery belonged to all mankind, and should not be subject to patenting, licensing and contracting. The result was undoubtedly the widespread availability of cheap X-ray sources. X-rays are ionizing. As subsequently discovered, such exposure to tissue of living organisms can cause radiation sickness, radiation burns and malignant tumors, and is mutagenic. Unfortunately, the dangerous effects of radioactive radiation on living organisms were understood many years after its discovery. Intensive studies of its properties began only in the middle of the 20th century. Practical applications of radioactive radiation formed a new field of medicine – radiation therapy – the forerunner of nuclear medicine, which today is rapidly developing all over the world. Nuclear medicine is a branch of clinical medicine, which deals with the use of radionuclide pharmaceuticals in diagnosis and treatment. Sometimes nuclear medicine also includes methods of distant radiation therapy. The founder of radioisotope diagnostics is considered to be the Hungarian D. Hevesy, who in 1913 suggested using the method of labeled atoms in biological research. Note that as early as 1901, French physicists A. Danleau and E. Bloch first used radium to treat cutaneous tuberculosis. The American inventor of the telephone, A. Bell proposed in 1903 the use of radium to treat tumors. In the USSR, its use began in 1923 to relieve joint pain. In 1929 E. Lawrence invented the cyclotron (resonant cyclic ion gas accelerator), which became the main tool for producing radionuclides. Another important direction in the use of X-rays is the creation of the method of X-ray structure analysis (XRD) based on the phenomenon of X-ray diffraction, discovered in 1912 by M. Laue[*] (the first X-ray spectra were called lauegrams). Many major discoveries have been made using the XRD method, of which about 20 have been awarded Nobel prizes. One of them has already been mentioned above. Let us note only those of them that have had a fundamentally important influence on the development of biology and medicine.

The first discovery in this series, the establishment of the three-dimensional structure of deoxyribonucleic acid (DNA – known as the double helix), the molecule containing all hereditary information, was made in 1953 by J. Watson, F. Crick and M.Wilkins[**].


J. D. Watson
(1928)


F. Crick;
(1916–2004)


M. H. F. Wilkins
(1916–2004)

The importance of the discovery for which they were awarded the Nobel Prize in Physiology and Medicine (1962) is emphasized

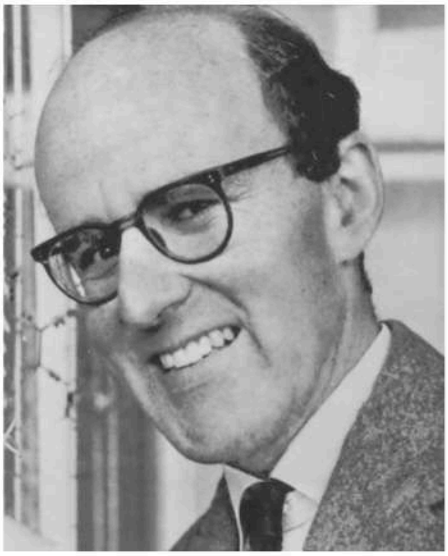
---

[*] ***Max von Laue,*** (09.10.1879, Koblenz, German Empire – 24.04.1960, West Berlin) –German physicist, Nobel Prize in Physics (1914) “for discovery of the diffraction of X-rays by crystals”.

[**] ***James Dewey Watson,*** 06.04.1928 – American biologist.
Francis Harry Compton Crick; (08.06.1916 – 28.07.2004) – British molecular biologist, biophysicist and neurobiologist.
Maurice Hugh Frederick Wilkins, (15.12.1916 – 05.10.2004) – English physicist and molecular biologist.

by the wording itself: "for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material". In 1959, three-dimensional structures of myoglobin (J. Kendrew) and hemoglobin (M. Perutz) proteins were determined*.

J. Kendrew
(1917–1997)

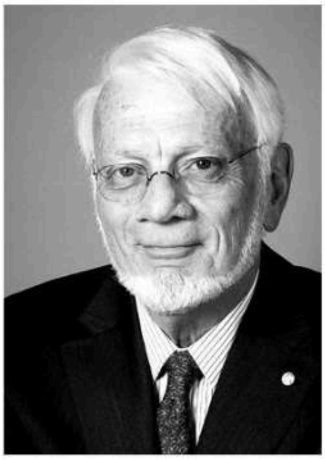


M. Perutz
(1914–2002)

As it is known, myoglobin is an oxygen-binding protein of skeletal muscles and heart muscle. Myoglobin's function is to create an oxygen reserve in muscles, which is consumed as needed to compensate for temporary oxygen shortages. Hemoglobin is a complex iron-containing protein in animals with blood circulation that can reversibly bind to oxygen, ensuring its transfer to tissues. In vertebrates it is found in red blood cells, while in most invertebrates it is dissolved in blood plasma (erythrocruorin) and may be present in other tissues. That is, myoglobin performs approximately the same functions as erythrocyte hemoglobin in the blood. These discoveries of the middle of the last century marked the beginning (the formal date is considered to be 1953) of modern genetics and molecular (structural) biology.

*** *Sir John Cowdery Kendrew*,** (24.03.1917, Oxford, England – August 23, 1997, Cambridge, England) – English biochemist, specialist in molecular biology

Max Ferdinand Perutz, (19.05.1914, Vienna – 06.02.2002, Cambridge) – English biochemist of Austrian origin, a specialist in molecular biology. Awarded the Nobel Prize in Chemistry "for studies of the structures of globular proteins" (1962).



T. Steitz
(1940-2018)

V. Ramakrishnan
(1952)

A. Yonath
(1939)

K. Wüthrich
(1938)

The second half of the 20th century was marked by two more major discoveries in molecular biology. One of them was the determination by XRD of the structure and function of the most important component of the protein-synthesizing apparatus of the human cell, the ribosome. Ribosome is a non-membrane organoid (organelle) of the living cell serving for protein biosynthesis from amino acids according to a given matrix based on genetic information provided by matrix RNA (mRNA). Ribosome is a target for more than 40% of all known antibiotics used in the clinic. Many research groups worked in this direction, but the priority of the discovery was given to T. Steitz, V. Ramakrishnan and A. Yonath.

Another outstanding achievement is the development by K. Wüthrich* (1985) of the NMR spectroscopy method for determining the three-dimensional structure of biological macromolecules in solution. (For the discovery of NMR (1945-1946), F. Bloch and E. Parcell were awarded the 1952 Nobel Prize in Physics). Note that this was preceded by the discovery of magnetic resonance as well, but in the form of electron paramagnetic

---

* ***Kurt Wüthrich***, (born in. 04.10.1938, Arberg, Switzerland) is Swiss chemist, winner of the Nobel Prize in Chemistry in 2002.

resonance (EPR), made by E.K. Zavoisky at Kazan University in 1944. Both methods have found wide application in medical and biological research.


F. Bloch
(1905–1983)


E. Parcell
(1912–1997)

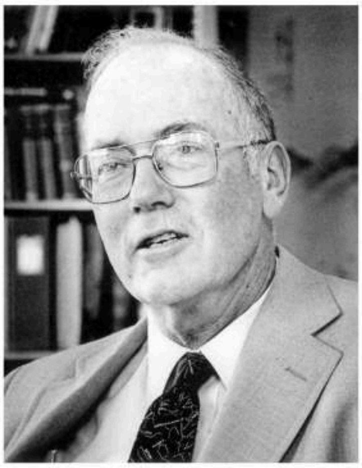

E.K. Zavoisky
(1907–1976)

It is known that in order to perform X-ray diffraction analysis it is necessary to obtain a protein in a crystalline form. But obtaining a crystalline protein is a difficult task and is not always possible. NMR spectroscopy also makes it possible to establish the spatial structure of proteins, both in crystalline form and in solution, i.e. in its most natural form in a living organism, but also makes it possible to study the dynamics of protein molecules, complexation processes, etc.

Finally, the creation of lasers and laser medicine is another breakthrough discovery. Laser, or optical quantum generator (laser, acronym from "**l**ight **a**mplification by **s**timulated **e**mission of **r**adiation") is a device that converts pump energy (light, electrical, thermal, chemical, etc.) into the energy of coherent, monochromatic, polarized and narrowband electromagnetic radiation. Since their creation, lasers have established themselves as "ready-made solutions to yet unknown problems". The first microwave generator (maser) on ammonia was created in 1954. Ch. Townes, N.G. Basov, and A.M. Prokhorov[*] were awarded the 1964 Nobel Prize in Physics.

---

[*] Winners of the 1964 Nobel Prize in Physics "for fundamental work in the field of quantum electronics, which has led to the construction of oscillators and amplifiers based on the maser-laser principle":



Ch. Townes
(1915–2015)

N.G. Basov
(1922–2001)

A.M. Prokhorov
(1916–2002)

In medicine lasers are widely used in surgery as bloodless scalpels, used in the treatment of ophthalmic diseases (cataract, retinal detachment, laser vision correction, etc.). Lasers are also widely used in cosmetology (laser hair removal, treatment of vascular and pigmented skin defects, laser peeling, removal of tattoos and pigment spots).

Almost three centuries after the creation of the optical microscope, an electron microscope was created to image objects with a maximum magnification of up to 10^6 times. In 1931, R. Rudenberg received a patent for a transmission electron microscope (other modifications later appeared), and in 1932 M. Knoll and E. Ruska[*] built the first prototype of the modern device.

The transmission electron microscope uses an electron beam instead of a light flux to form an image. The electron beam is created by means of a cathode. The resulting electron beam is accelerated usually to 80–200 keV

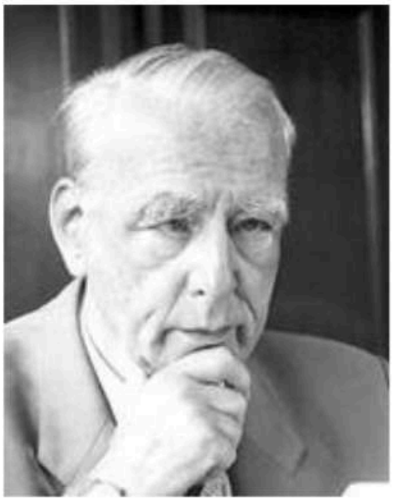
Charles Hard Townes, (28.07.1915, Greenville, South Carolina – 27.01.2015) – American physicist, member of the National Academy of Sciences (1956), foreign member of the Russian Academy of Sciences (1994).

Nikolai Gennadievich Basov, (14.12.1922, Usman', Tambov Province – 1.07.2001) – Soviet and Russian physicist, awarded the Lenin Prize (1959) and the USSR State Prize (1989), twice Hero of Socialist Labor (1969, 1982).

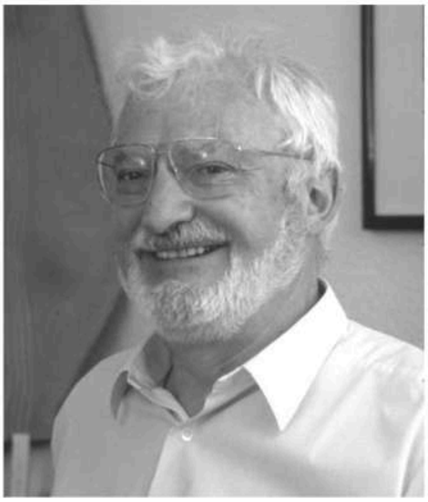
Alexander Mikhailovich Prokhorov (11.07.1916, Atherton, Queensland, Australia – 8.01.2002, Moscow) – the Soviet physicist, one of the founders of the most important direction of modern physics – quantum electronics.

[*] ***Ernst August Friedrich Ruska***, (25.12.1906, Heidelberg – 27.05.1988, West Berlin) – creator of the electron microscope, winner of the 1986 Nobel Prize in Physics.

(various voltages from 20 kV to 1 MV are used), focused by a system of magnetic lenses (sometimes electrostatic lenses), passes through the sample so that some of the electrons scatter on the sample and some do not. Thus, the electron beam passing through the sample carries information about the structure of the sample. The beam then passes through a system of magnifying lenses and forms an image on a fluorescent screen, photoplate, or CCD[*] camera. The resolution of an electron microscope is 1000-10000 times better than that of a traditional light microscope and for the best modern instruments can be less than one Angstrom ($1\text{Å} = 10^{-10}$ m = 0.1nm).


E. Ruska
(1906–1988)


G. Binnig
(1947)


H. Rohrer
(1933–2013)

The use of the transmission electron microscope for scientific research began in the late 1930s, when the first images of some viruses (tobacco mosaic virus and bacteriophages) were obtained. Currently, electron microscopy has found the most widespread application in cytology, microbiology and virology, causing the creation of new branches of science. In electron microscopy of biological objects, special methods of preparation are used. This is necessary to identify the individual components of the objects under study (cell, bacterium, virus, etc.), as well as to preserve their structure in a high vacuum under the influence of a flow

---

[*] A CCD is a specialized analog integrated circuit consisting of light-sensitive photodiodes made of silicon, using CCD technology – charge-coupled devices.

of electrons. Electron microscopy is used to study the external shape of the object, the molecular organization of its surface, and the internal structure of the object is examined using ultrathin sections.

Electron microscopy combined with biochemical, cytochemical methods of investigation, immunofluorescence, and X-ray structural analysis provide information on the composition and function of the structural elements of cells and viruses. In 1981, G. Binnig and H. Rohrer[*] created the scanning tunneling microscope (STM) in its modern form (the principles of this class of instruments were laid down earlier by other researchers), the first in the series of so-called probe microscopes. In STM, a sharp metal needle is brought to the sample at a distance of about 0.1 nm. When a small potential relative to the sample is applied to the needle, a tunneling current is induced. The value of this current depends exponentially on the sample-needle distance: typical current strength is 11000 pA (picoampere, 1 pA = 10^{-12} A) at a sample-needle distance of about 0.1 nm.

During the scanning process the needle moves along the sample surface, the tunnel current is kept stable by the feedback action, and the readings of the tracking system change depending on the surface topography. Such changes are recorded, and an elevation map is constructed based on them. Another technique involves moving the needle at a fixed height above the surface of the sample. In this case, changes in the tunnel current value are recorded and the surface topography is plotted based on this information. A significant limitation of scanning tunneling microscopy was the requirement for electrical conductivity of the samples under study. With the invention of the scanning atomic force microscope the possibilities of scanning probe microscopy became almost unlimited in terms of objects, up to investigations of the structure of proteins, nucleic acids and other biological molecules in liquid media. Probe tunneling microscopy has expanded the possibilities of nanotechnology, allowing not only visualization of individual atoms, but also their

---

[*] ***Gerd Karl Binnig***, 20.07.1947 – German physicist, and ***Heinrich Rohrer***, (06.06.1933, Buchs – May 16, 2013, Wallerau) – The Swiss physicist, along with ***E. Ruska***, were awarded the 1986 Nobel Prize in Physics.

manipulation. Other outstanding results in biology and medicine based on physics are known, as well as methods of research, analysis and diagnosis. But in the framework of the Introduction to the course it is not possible even to describe them briefly.

Let's consider only a number of physical methods of research, which revolutionized medical diagnosis – turned medical diagnosis from art to science and today form the basis of modern clinical diagnostics. These are methods of imaging human internal organs. Currently there are many complementary methods of imaging (introscopy) in medicine. Medical introscopy solves two, to a certain extent independent, tasks. The first one is imaging of internal organs (determining their shape, size, location, etc.). The second one is determination of physical, physical-chemical and other properties of biological tissues and character of their functioning. Advantages of each of the methods are shown when solving the second task, in particular, when analyzing specifics of interaction of radiation used for imaging with various biological tissues in an organism. All types of imaging are based on the physics of interaction between radiation and matter. It is necessary for radiation to penetrate into the body and be partially scattered by it. That is, the body must be semi-transparent to radiation. This condition becomes evident when we consider two limiting cases. The first case is that the body is completely opaque to longwave optical radiation, which, therefore, cannot be used to obtain information about internal structures. The second is that the neutrino flux, for which the body is completely transparent, can hardly be used to obtain images.

The simplest type of imaging is obtaining a flat image of a volumetric body, when the object is located between a stationary emitter, such as an X-ray tube and a stationary screen-receiver, such as a photographic plate. In this case, the elementary surface contains averaged information on the interaction of radiation with the substance, obtained by passing through the entire thickness of the object. This method of imaging has been successfully used in medicine to this day since the creation of the above-mentioned method of fluoroscopy, as well as in ultrasonic defectoscopy and diagnostics, etc. The next level is tomography itself – a method of studying the internal structure of various objects, which consists in a layer-by-layer image of the object when it is irradiated (tomos – from Greek – layer, cross-section). There are: X-ray, radiation, optical, magnetic resonance, ultrasound tomography, etc. We shall cite only some of them. The technique for imaging individual layers is varied. These issues are discussed in special literature. The most perfect images are obtained by computer (computational) tomography. And here it all started with X-rays. The first clinical tests of computed X-ray tomography made by G. Hounsfield and A. Cormack* are dated 1972–1973.

G. Hounsfield
(1919–2004)

A. Cormack
(1924–1998)

Computer tomography (CT) is based on the fact that when an object with a complex internal structure is exposed to X-rays, the information about its structure can be reconstructed by calculating the spatial distribution of the intensity of radiation that has passed through the object. Thus, the differential distribution of parameters inside the body is obtained as a result of solving the inverse problem of multibeam irradiation of the sample, which gives the initial integral characteristics. The creation and development of positron emission tomography (PET) (another name

---

* ***Godfrey Newbold Hounsfield***, (28.08.1919, Nottinghamshire, UK – 12.08.2004, UK) – British electrical engineer and ***Allan McLeod Cormack***, (23.02.1924, Johannesburg, South Africa – 07.05.1998, Massachusetts, USA) – South African and American physicists received the 1979 Nobel Prize in Physiology and Medicine for "Development of Computed Tomography".

for it is two-photon emission tomography) is related to advances in nuclear physics. This is a radionuclide tomographic method of examining the internal organs of humans or animals. The method is based on recording gamma ray pairs arising from the annihilation of positrons with electrons. Positrons are produced by positron beta decay of a radionuclide that is part of a radiopharmaceutical that is injected into the body before the study. When positrons annihilate with electrons in body tissues, two gamma quanta are almost always produced. Most positrons penetrating tissues lose energy very quickly and annihilate with the medium's electrons while already at rest. Therefore, annihilation gamma-quanta have zero total momentum. That is, they are emitted in mutually opposite directions, but with the same energy, 511 keV. Thus, if two gamma ray detectors connected by a coincidence scheme simultaneously absorb a gamma ray with an energy of 511 keV, we should expect that the annihilation point is located on the line connecting these two detectors, on the so called "off-click" line. Using a large set of detectors arranged around the object under study (or by moving a pair of detectors around the object), it is possible to construct many such lines in space. They will all pass through the points where the annihilation took place (i.e. through the points where the decayed radionuclide nucleus is located – to the very short range length of positrons in the tissue). This allows a three-dimensional reconstruction of the radionuclide distribution in the scanned object.

Magnetic resonance imaging (MRI)* – a method of obtaining tomographic medical images to study internal organs and tissues using the NMR phenomenon, is widely used not only in diagnostics, but also in biomedical research, including the study of brain structure and function. There are several varieties of MRI: micro- and minitomography (imaging of very small and small animals) and human tomography.

---

* "For the Invention of Magnetic Resonance Imaging" ***Peter Mansfield*,** (09.10.1933, London, UK – 08.02.2017, Nottingham, UK) – British physicist, (Nottingham) and ***Paul Christian Lauterbur***, (06.05.1929, Sydney, Ohio – 03.27.2017, Urbana, IL) – American chemist, were awarded the 2003 Nobel Prize in Medicine.

P. Mansfield
(1933–2017)

P. Lauterbur
(1929–2007)

E. Schrödinger
(1887–1961)

In the late 1950s, D.E. Kuhl, L. Chapman, and R. Edwards developed the principle of emission tomography. Later their work led to the design of several tomographic instruments as well. In 1975, tomographic methods were further developed by M. Ter-Pogosian together with J. Eugene-Robinson, C. Sharp Cooke.

Ultrasound imaging (tomography) techniques, which were developed at the end of the last century, were perhaps the most widely used in medical diagnostics. The first publications of numerous authors, including clinical applications, date back to 1980-1982. The use of ultrasound in medical diagnosis and therapy will be discussed in detail in this course. Naturally, the cited material reflects only one, although extremely important, aspect of the Life Sciences. This was pointed out half a century ago by M. Born in one of his articles in the collection cited above: “We are convinced that the physical and chemical aspect is by no means sufficient to depict the facts of life, let alone the facts of thinking”.

***

The major achievements of physics that have influenced the development of biology and medicine, given in the Introduction, are essentially means to answer the main question, which was formulated in the middle of the last century by E. Schrödinger[*]: "What is life from the point of view of physics".

In popular form, the question is: "How can physics and chemistry describe the processes occurring in living organisms?" Subsequently, this book had a great influence on the development of biophysics and molecular biology. The understanding of the fact that physics has a special role in the development of medicine, both fundamental and clinical, led to the creation of the discipline "Medical Physics" and the allocation of its proper place in both physical and biomedical education.

This course consists of two parts:

Part 1. "Mechanics. Molecular Physics".

Part 2. "Electricity and Magnetism. Optics. Fundamentals of Atomic and Nuclear Physics".

Part 1 is presented in this edition.

---

[*] ***Schrödinger E.*** What is Life? The Physical Aspect of the Living Cell.1945. URL: http://scisne.net/a- 868. Erwin Schrödinger is one of the creators of wave quantum mechanics, winner of the Nobel Prize in Physics.

MECHANICS

Mechanics is a branch of physics, which studies the motion of bodies relative to each other or parts of the same body relative to each other (deformation), – the basis of an independent scientific discipline and the field of knowledge "Biomechanics".

Biomechanics is a branch of natural sciences that studies the mechanical properties of living tissues, individual organs and systems or the body as a whole and the mechanical phenomena occurring in them using models and methods of mechanics. Biomechanical studies cover different levels of organization of living matter: biological macromolecules, cells, tissues, organs, organ systems, as well as whole organisms and their communities. Most often the object of study of this science is the movement of animals and humans, as well as mechanical phenomena in tissues, organs, and systems. Mechanical motion refers to the movement of the entire biosystem, as well as the movement of individual parts of the system relative to each other – the deformation of the system. All deformations in biological systems are associated with biological processes that play a decisive role in the movements of animals and humans. These include muscle contraction, the deformation of tendons, bones, ligaments, fascia, and joint movements. A separate branch of biomechanics is biomechanics of the respiratory apparatus, its elastic and inelastic resistance, kinematics (that is, the geometric characteristics of movement) and dynamics of respiratory movements, as well as other aspects of the respiratory apparatus in general and its parts (chest, lungs); biomechanics of circulation studies the elastic properties of vessels and heart, the hydraulic resistance of blood flow to vessels, the spread of elastic vibrations in the vascular wall, blood movement, heart work, etc. Human biomechanics is a complex science. It includes a wide variety of knowledge from other sciences, such as: mechanics and mathematics, functional anatomy and physiology, age-related anatomy and physiology, pedagogy, theory of physical education. Sufficiently complete general information can be found in [1].

In accordance with this definition, the sections 1-3 of general physics presented in this textbook, together with medical and biological applications, can be considered as a brief Introduction to the course "Biomechanics".

SECTION 1. MECHANICS OF SOLIDS

Mechanics, to a certain extent conventionally, is divided into kinematics, dynamics, and statics.

Kinematics is a description of motion out of connection with the causes of this motion.

Dynamics is the establishment of the relationship between the impact on the body and the result of this impact (change in the state of motion).

Statics is a special case of dynamics – the determination of the conditions of equilibrium of bodies.

When describing physical processes, physical models (approximations) are widely used. Let us introduce two of them, which are necessary for the description of mechanical motion: "material point" and "absolutely rigid body" (rigid object). A material point (particle) is the simplest physical model in mechanics, an abstract body of infinitely small size. Practically, a material point is a body possessing mass, the size and shape of which can be neglected when describing motion.

An absolutely solid body is a set of material points (particles), the distances between which are preserved in the course of any movements performed by the body. That is, it is assumed that an absolutely rigid body (rigid object) does not change its shape when it is affected (it does not deform).

1.1. Kinematics

1.1.1. Kinematics of a material point (particle)

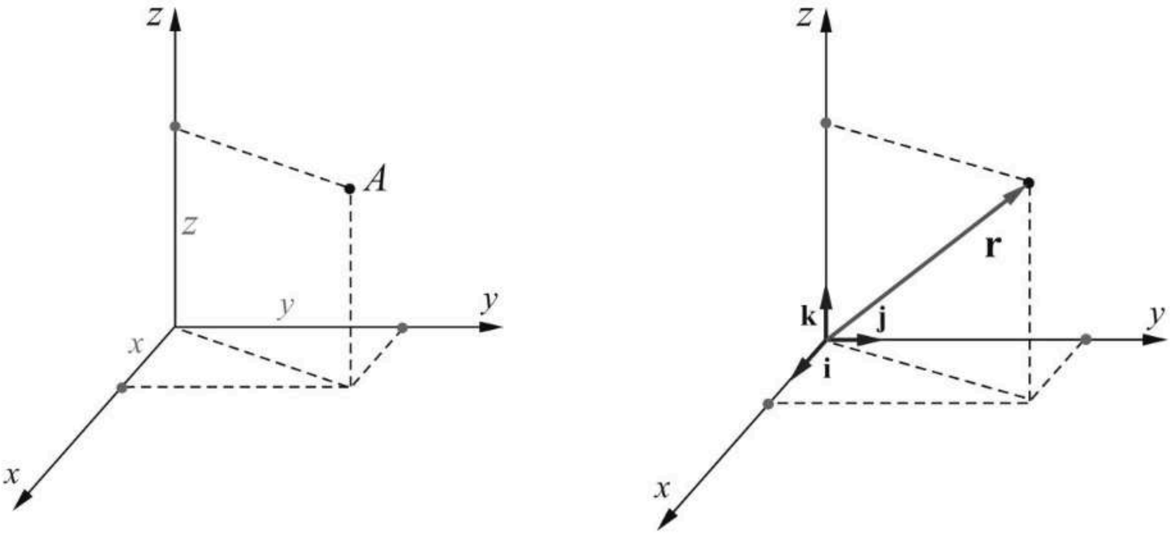
To describe the motion it is necessary:

- enter a reference point and "bind" the coordinate system to it;
- determine the way to measure the time.

How do we define the position of a material point (particle) in space? In the most common Cartesian coordinate system, this can be done either by the x, y, z coordinates or by the radius-vector **r** *(Fig. 1.1a, b), defined as*

$$\mathbf{r} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z,$$

where **i***,* **j***,* **k** *– unit vectors (orts) directed along the coordinate axes (other designations are also used:* $\mathbf{e}_x$*,* $\mathbf{e}_y$*,* $\mathbf{e}_z$*).*



a

b

Fig. 1.1. Determining the position of a point in the Cartesian coordinates: a) – by means of x, y, z – coordinates; b) – by means of radius-vector **r**

Let us introduce basic kinematic concepts.

A trajectory is a line defining the movement of a material point (particle) in space. In the general case, the equation of trajectory has the following form:

$$f(x, y, z) = 0.$$

A distinction is made between straight-line and curvilinear trajectories (in a particular case, circular motion), etc.

The equation of motion is an equation (function) that determines the position of a material point (particle) in space at any point in time:

$$f(x, y, z, t) = 0, \text{ or } x = x(t),\ y = y(t),\ z = z(t) \text{ (parametric form),}$$
$$\text{or } \mathbf{r}(t) = \mathbf{i}x(t) + \mathbf{j}y(t) + \mathbf{k}z(t).$$

Velocity. To describe the displacement let us use the time dependence of the vector $\mathbf{r}(t)$ (Fig. 1.2a). Thus, the displacement over the time interval Δt, defined as $\Delta\mathbf{r} = \mathbf{r_2} - \mathbf{r_1}$, ($|\Delta\mathbf{r}| = \Delta\mathrm{r}$), (Fig. 1.2b), will be written in the form:

$$\Delta\mathbf{r}(t) = \mathbf{i}\Delta x(t) + \mathbf{j}\Delta y(t) + \mathbf{k}\Delta z(t).$$

The average velocity of movement is determined as follows:

$$\frac{\Delta\mathbf{r}}{\Delta t} = \langle\boldsymbol{\upsilon}\rangle = \boldsymbol{\upsilon_{av}}. \tag{1.1}$$

Note that the modulus (numerical value) of the increment of the radius-vector is not equal to the increment of the length of the radius-vector, i.e.:

$$|\Delta\mathbf{r}| \neq \Delta|\mathbf{r}|.$$

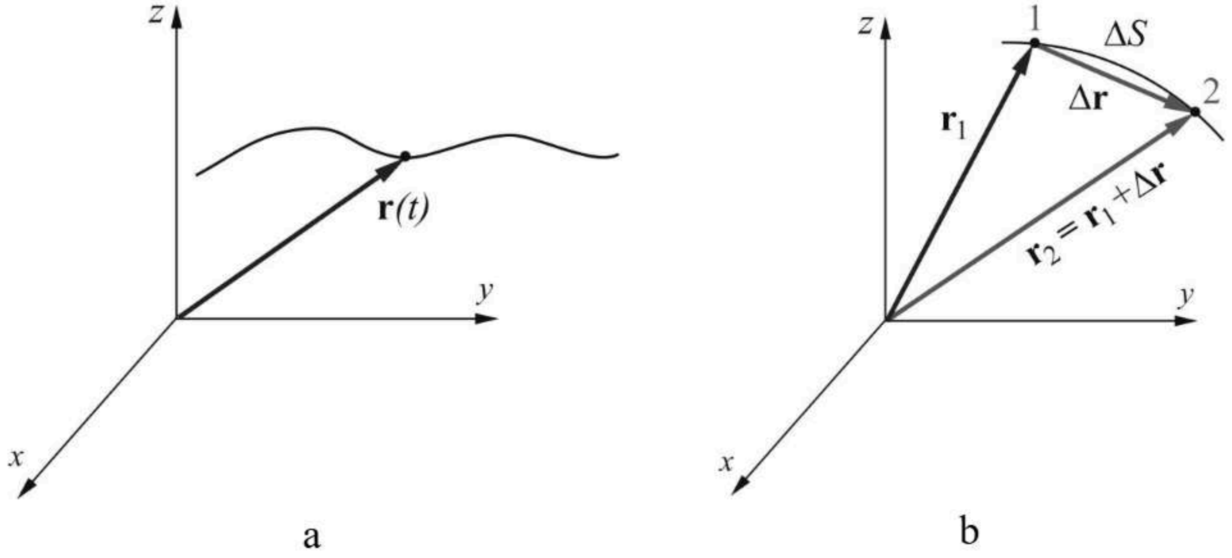


Fig. 1.2. Graphical representation of motion of a material point (particle) along a trajectory by means of a radius-vector $\mathbf{r}(t)$:
a) – position of the material point (particle) on the trajectory at an arbitrary moment of time; b) – $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$ – coordinates of the material point (particle) at time moments t_1 and $t_2 = t_1 + \Delta t$, respectively;
$\Delta\mathbf{r}$ – increment of radius-vector $\mathbf{r}$ and path ΔS at time Δt

The velocity at a given moment of time (instantaneous), i.e. when the time interval Δt becomes the infinitesimal value dt, is determined by the expression:

$$\boldsymbol{\upsilon} = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{r}}{\Delta t} = \frac{d\mathbf{r}}{dt}. \tag{1.1a}$$

As seen in Fig. 1.3, at $\Delta t \to dt$, $\alpha \to 0$, $\Delta\mathbf{r} \to d\mathbf{r}$, the velocity vector (equation 1.1a) coincides with the tangent to the trajectory at this point (at this time). Since $|\Delta\mathbf{r}| \cong \Delta S$, the numerical value of speed is:

$$\upsilon = \lim_{\Delta t \to 0} \frac{\Delta S}{\Delta t} = \frac{dS}{dt}. \tag{1.1b}$$

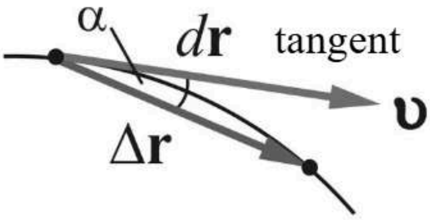


Fig. 1.3. Geometric representation of the instantaneous velocity $\boldsymbol{\upsilon}$

Since $\Delta\mathbf{r}(t) = \mathbf{i}\Delta x(t) + \mathbf{j}\Delta y(t) + \mathbf{k}\Delta z(t)$, the velocity components will be written as:

$$\lim_{\Delta t\to 0}\frac{\Delta x}{\Delta t} = \frac{dx}{dt} = \upsilon_x, \quad \lim_{\Delta t\to 0}\frac{\Delta y}{\Delta t} = \frac{dy}{dt} = \upsilon_y, \quad \lim_{\Delta t\to 0}\frac{\Delta z}{\Delta t} = \frac{dz}{dt} = \upsilon_z,$$

and the velocity vector can be represented as follows:

$$\boldsymbol{\upsilon} = \mathbf{i}\upsilon_x + \mathbf{j}\upsilon_y + \mathbf{k}\upsilon_z. \tag{1.1c}$$

(The first time derivative is sometimes denoted as follows: $\dot{x} = \frac{dx}{dt}$, $\dot{y} = \frac{dy}{dt}$, $\dot{z} = \frac{dz}{dt}$, $\dot{f} = \frac{df}{dt}$ etc.).

Acceleration. If the velocity of the trajectory from point 1 to point 2 (Fig. 1.4a) changes by $\Delta\boldsymbol{\upsilon} = \boldsymbol{\upsilon}_2 - \boldsymbol{\upsilon}_1$ during the time Δt, then the average value of its change, i.e. the average value of acceleration is

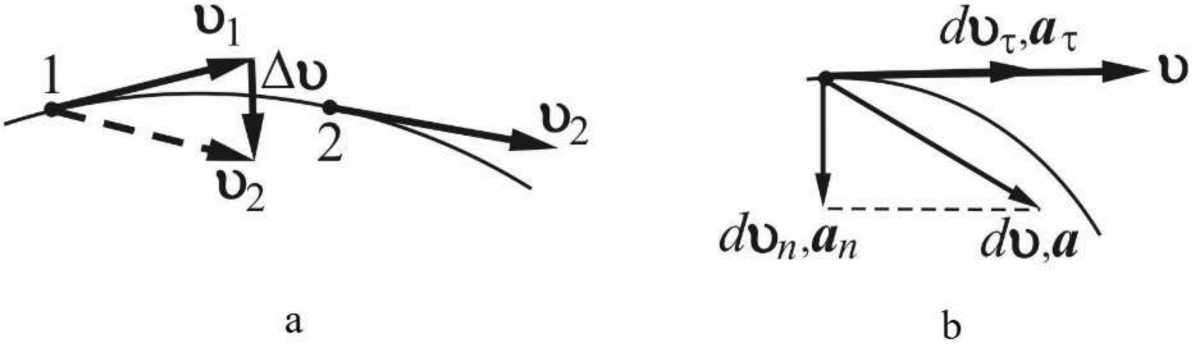
$$\frac{\Delta\boldsymbol{\upsilon}}{\Delta t} = \boldsymbol{a}_{\boldsymbol{av}}. \tag{1.2a}$$

In the limit with infinitesimal interval Δt we obtain expression (1.2b) for the instantaneous value of acceleration:

$$\lim_{\Delta t\to 0}\frac{\Delta\boldsymbol{\upsilon}}{\Delta t} = \frac{d\boldsymbol{\upsilon}}{dt} = \boldsymbol{a} = \frac{d^2\mathbf{r}}{dt^2}, \tag{1.2b}$$

i.e.,
$$\boldsymbol{a} = \mathbf{i}\frac{d^2x}{dt^2} + \mathbf{j}\frac{d^2y}{dt^2} + \mathbf{k}\frac{d^2z}{dt^2}. \qquad (1.2c)$$

(The second time derivatives are denoted in the same way: $\ddot{x}$, $\ddot{y}$, $\ddot{z}$, $\ddot{f}$ etc.).



a

b

Fig. 1.4. a) – Change in velocity $\boldsymbol{\upsilon}$; b) – corresponding instantaneous value of acceleration $\boldsymbol{a}$

Tangential and normal accelerations. The velocity can change both in magnitude and direction (when moving along a curvilinear trajectory), while $\Delta\boldsymbol{\upsilon}$ and $d\boldsymbol{\upsilon}$ are not always $\parallel \boldsymbol{\upsilon}$. In the general case

$$d\boldsymbol{\upsilon} = d\boldsymbol{\upsilon}_\tau + d\boldsymbol{\upsilon}_n, \qquad (1.2d)$$

where $d\boldsymbol{\upsilon}_\tau$ – is the tangential (i.e., directed along the tangent to the trajectory at a given point) increment of velocity, characterizing the change of velocity in magnitude, and $d\boldsymbol{\upsilon}_n$ – is the normal component of velocity change, (directed perpendicularly "$\perp$" to the tangent). Then for the instantaneous value of acceleration we can write:

$$\boldsymbol{a} = \frac{d\upsilon}{dt} = \frac{d\upsilon_\tau}{dt} + \frac{d\upsilon_n}{dt} = \boldsymbol{a}_\tau + \boldsymbol{a}_n, \qquad (1.3)$$

where $\boldsymbol{a}_\tau$ – is the tangential acceleration, $\boldsymbol{a}_n$ – is the normal acceleration (Fig. 1.4b).

Path. If there is an equation of the trajectory $f(x, y, z,) = 0$, then the traveled path – the length of the trajectory segment between points 1 and 2 (Fig. 1.5), is determined by the formulas of analytical geometry.

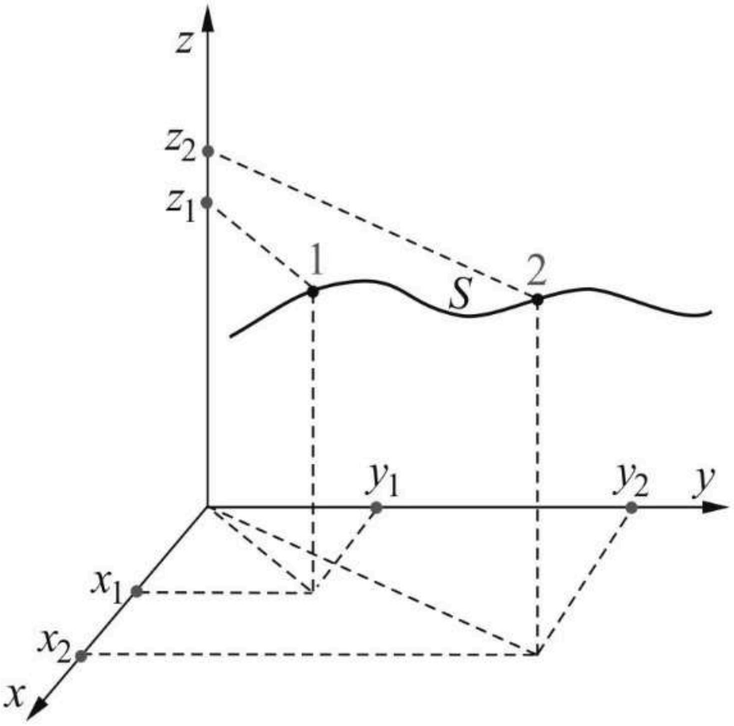


Fig. 1.5. Graphical representation of the traveled path from point 1 to point 2 in Cartesian coordinate system

If there is an equation of motion, for example, in the form of the time dependence of velocity $\upsilon(t)$ (Fig. 1.6), then the distance traveled can be determined as follows. On the curve we mark a short time interval Δt_i, within which the velocity is assumed to be equal to υ_i. The elementary path is equal to $\Delta S_i = \upsilon_i \Delta t_i$ (in Fig. 1.6 it is shown as a shaded rectangle). Traversed path in time interval from t_1 to t_2 is determined by the sum of ΔS_i, which in the limit for infinitesimal intervals Δt_i is an integral of function $\upsilon(t)$ in time interval from t_1 to t_2:

$$S = \sum \Delta S_i \Rightarrow \sum_{\Delta t_i \to 0} \upsilon_i \Delta t_i \Rightarrow S = \int_{t_1}^{t_2} \upsilon(t) dt\,. \qquad (1.4)$$

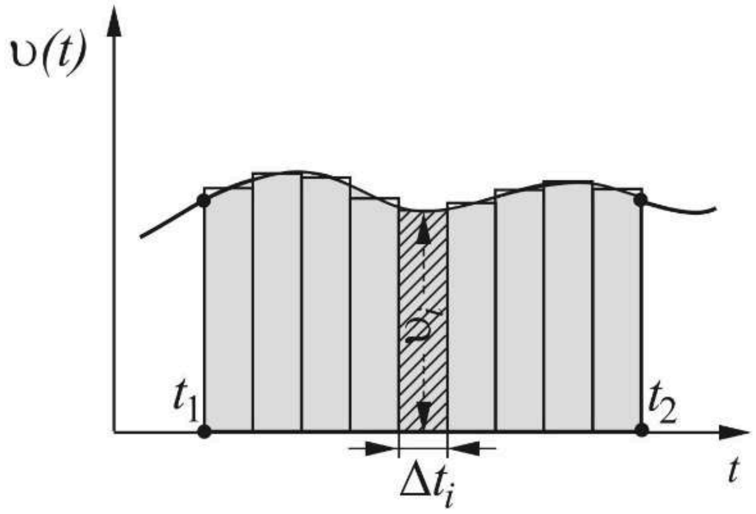


Fig. 1.6. Time dependence of the velocity of movement $\upsilon(t)$.
The traveled path is represented by a shaded figure

In translational uniformly accelerated motion, the velocity at an arbitrary moment of time is $\upsilon_0 \pm at$. Here the sign "+" corresponds to an increase in velocity (acceleration), the sign "–" to a decrease in velocity (deceleration), where υ_0 – is the initial velocity. The path S traveled during time t is defined as

$$S = \int_0^t (\upsilon_0 + at)dt = \upsilon_0 t \pm \frac{at^2}{2}\,. \qquad (1.4a)$$

Uniform movement on a circle. A particular but important case is circular motion at a constant velocity.

In this case only the direction of velocity changes. If $\Delta\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ during time Δt along a circular trajectory from point 1 (marked A) to point 2 (B) the velocity will change by value $\Delta\boldsymbol{\upsilon} = \boldsymbol{\upsilon}_2 - \boldsymbol{\upsilon}_1$ (Fig. 1.7). Let's make a parallel transfer of the velocity vector $\boldsymbol{\upsilon}_2$ from point *B* to point *A*. As can be seen from the figure, the triangles *AOB* and *CAD* are similar (Δ *AOB*~Δ *CAD*): they are two isosceles triangles with mutually perpendicular sides.

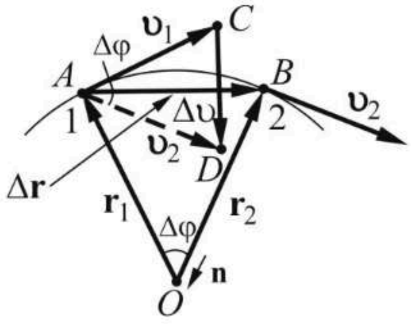


Fig. 1.7. Geometric representation of the movement of a point on a circle with a constant in magnitude velocity

And since $|\mathbf{r}_1| = |\mathbf{r}_2| = R$, $\quad |\boldsymbol{\upsilon}_1| = |\boldsymbol{\upsilon}_2| = \upsilon$, then:

$$\frac{\Delta\upsilon}{\upsilon} = \frac{\Delta r}{R} \ (\approx \Delta\varphi).$$

Then $\quad \Delta\upsilon = \upsilon\frac{\Delta r}{R}$, or $\frac{\Delta\upsilon}{\Delta t} = \frac{\upsilon}{R}\frac{\Delta r}{\Delta t}$ and

$$\lim_{\Delta t\to 0}\frac{\Delta\upsilon}{\Delta t} = \frac{\upsilon}{R}\lim_{\Delta t\to 0}\frac{\Delta r}{\Delta t} = \frac{\upsilon}{R}\frac{dr}{dt} = \frac{\upsilon^2}{R} = a_n \Rightarrow \boldsymbol{a}_n = a_n\boldsymbol{n}\,. \tag{1.5}$$

Here a_n is the numerical value of the centripetal acceleration, $\boldsymbol{n}$ – is the unit vector directed to the center of rotation along the radius of rotation, and $a_n\boldsymbol{n}$ is the centripetal acceleration vector. When moving along an arbitrary curvilinear trajectory (Fig. 1.8), the instantaneous value of centripetal acceleration is determined by formula 1.5, in which R is the radius of curvature (the radius inscribed in the trajectory of a circle at a given point of the trajectory, i.e. at a given time).

Fig. 1.8. Radius of curvature at arbitrary points of the curvilinear trajectory. The dotted line indicates an inscribed circle

1.1.2. Kinematics of a solid body (rigid object)

If a line arbitrarily chosen in the body maintains its orientation in space when a solid body moves in space or if any points in the body describe the same (similar) trajectories, then such motion of the body can be regarded as translational motion of a material point (particle) (Fig. 1.9). This condition is not satisfied if the body rotates around an internal axis (i.e., an axis which passes through the body). According to the fundamental principle of mechanics, if a body makes a complex motion and its components do not depend on each other, the resulting displacement will be determined by the vector sum of each of them. Therefore, it is sufficient to consider the kinematic characteristics of rotational motion as a component of this complex motion.

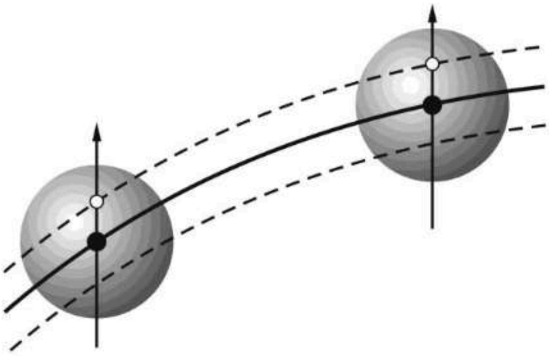


Fig. 1.9. The illustration of translational motion of a body (rigid object) along a curvilinear trajectory

In the simplest case, when a symmetric body, for example, of spherical shape, rotates about an axis passing through the center of symmetry O (Fig. 1.10), trajectories of any point of the body are concentric circles, whose centers O, O_1, O_2, etc., lie on the axis of rotation. In this case, the radius of rotation R_i at the same time interval Δt is rotated by the same angle $\Delta\varphi_1 = \Delta\varphi_2 \ldots = \Delta\varphi_i = \Delta\varphi$. This value serves as a measure of rotation of the body. The direction of the rotation angle vector $\Delta\boldsymbol{\varphi}$ is determined by the rule of the right-hand rule (screw) – it coincides with the direction of movement of the screw head along the rotation axis during clockwise circular motion of the point on its handle.

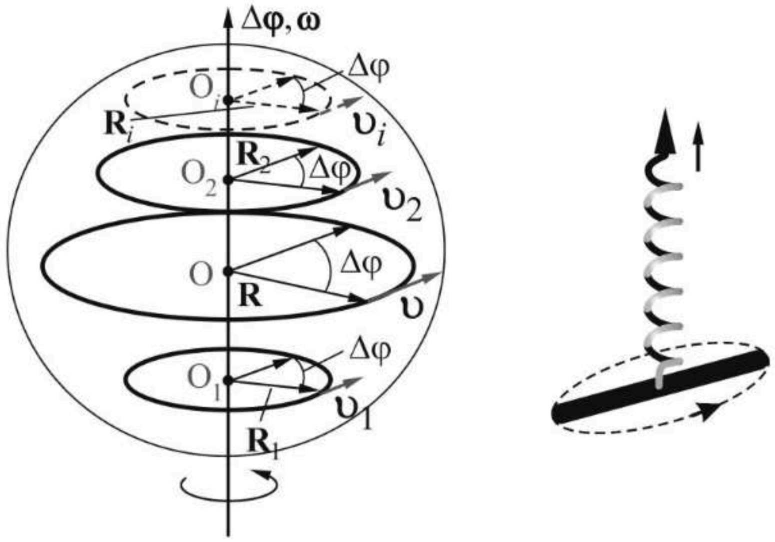


Fig. 1.10. Rotation of the body (rigid object) relative to a stationary axis

The average $\boldsymbol{\omega}_{\mathrm{av}}$ and instantaneous angular velocity of rotation $\boldsymbol{\omega}$ are defined by relations (1.6) and (1.6a):

$$\boldsymbol{\omega}_{\mathrm{av}} = \frac{\Delta\boldsymbol{\varphi}}{\Delta t}\ , \tag{1.6}$$

$$\boldsymbol{\omega} = \lim_{\Delta t \to 0} \frac{\Delta\boldsymbol{\varphi}}{\Delta t} = \frac{d\boldsymbol{\varphi}}{dt}. \tag{1.6a}$$

The dependence of the linear velocity $\boldsymbol{\upsilon}$ on the circular trajectory on the angular velocity $\boldsymbol{\omega}$ and radius of rotation $\mathbf{R}$ (Fig. 1.11) is determined by the vector multiplication rule:

$$\boldsymbol{\upsilon} = [\boldsymbol{\omega}\mathbf{R}]. \tag{1.7}$$

The numerical value of the velocity is: $\upsilon = \omega R \mathrm{Sin}\widehat{\boldsymbol{\omega}\mathbf{R}}$ (in this case $\upsilon = \omega R$, because the angle between the vectors $\boldsymbol{\omega}$ and $\mathbf{R}$, denoted as $\widehat{\boldsymbol{\omega}\mathbf{R}}$, is 90°), and the direction of velocity is determined by the right-hand rule: the vector $\boldsymbol{\upsilon}$ is directed perpendicularly from the plane in which the vectors

$\boldsymbol{\omega}$ and **R**, are, when turned in the direction from vector $\boldsymbol{\omega}$ to vector **R** (clockwise), that is, toward the smaller angle (Figure 1.11).

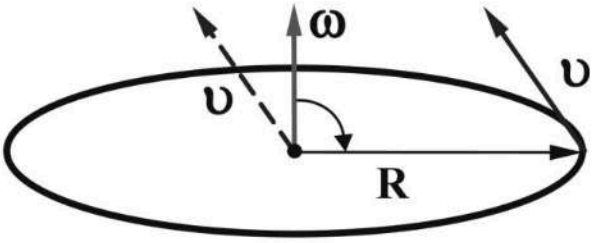


Figure 1.11. The illustration of vector multiplication of vectors – (right-hand rule) vectors **υ** and **R** lie in the plane ⊥ to the vector $\boldsymbol{\omega}$

In the simplest case of rotation with variable velocity at a constant orientation of the axis of rotation, the mean and instantaneous values of the angular acceleration $\boldsymbol{\beta}_{av}$ and $\boldsymbol{\beta}$ are equal, respectively:

$$\boldsymbol{\beta}_{\mathrm{av}} = \frac{\Delta\boldsymbol{\omega}}{\Delta t}, \tag{1.8}$$

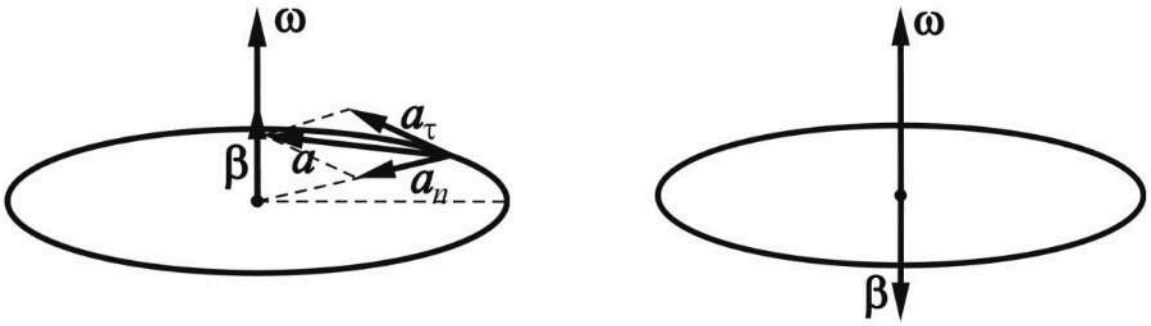
$$\boldsymbol{\beta} = \lim_{\Delta t \to 0} \frac{\Delta\boldsymbol{\omega}}{\Delta t} = \frac{d\boldsymbol{\omega}}{dt}. \tag{1.8a}$$

The directions of vector $\boldsymbol{\beta}$ with increasing and decreasing speed of rotation are shown in Fig. 1.12. The numerical values of the tangential a_τ, normal a_n and total a accelerations of the material point are equal, respectively:

$$a_\tau = \beta R, \tag{1.9}$$

$$\mathrm{a}_n = \frac{\upsilon^2}{\mathrm{R}} = \omega^2 R, \tag{1.9a}$$

$$a = \sqrt{\beta^2 R^2 + (\omega^2 R)^2}. \tag{1.9b}$$



a b

Fig. 1.12. The illustrations: a) – accelerated rotation; b) – decelerated rotation of a body (rigid object)

1.2. Dynamics of a solid body (rigid object)

Dynamics is a branch of mechanics in which the relationship between an action on a body and the result of that action is established. As noted in the introduction, statics is a particular case of dynamics; it is the definition of the equilibrium condition of bodies. But in fact, statics is a separate section of mechanics, because its key provisions underlie the calculations of mechanical structures and constructions, etc. This separation is justified because, after some relatively short time, mechanical (deformation) processes in assembled structures (constructions, etc.) seem to stop and do not manifest themselves externally (but static stresses remain). When applied to living systems, such a division is incorrect. When lifting a weight a human makes a mechanical movement, all parts of the body are involved. As a result, the person has fixed the load in some static position. The mechanical processes have stopped. But the human body continues to work, biochemical and physical processes proceed intensively in it, energy is spent. It is well known and used in sports medicine: muscle mass and therefore strength can be increased both by dynamic and static loads (isotonic method of training).

1.2.1. Dynamics of material point (particle)

Classical mechanics is based on three laws formulated by I. Newton. The 1st law states: "Each body is in a state of rest or uniform straight-line motion until the action of other bodies causes it to change its state. (In an inertial frame of reference, an object either remains at rest or continues to move at a constant velocity, unless acted upon by a force).

"Rest" and "uniform straight-line motion" are essentially the same state! It is the natural state of the body, since nothing needs to be done to make it so.

The property of a body to retain a state of rest or velocity in the absence of influence from any other body is called inertia. Reference frames where these properties are preserved are called inertial reference frames. The assertion of the existence of such reference frames is the essence of the first law of mechanics, often referred to as Newton's law of inertia (Newton-Galileo).

Mass m is used as a measure of inertia, which is also a measure of the amount of substance contained in a certain volume. As a measure of impact, a force **F** is used – a vector quantity characterized by its numerical value (vector modulus), direction and point of application. Graphically, the force is represented by an arrow originating from the point of its application.

According to Newton's 2nd law, "The acceleration gained by a body (particle) is directly proportional to the acting force and inversely proportional to its mass:

$$\boldsymbol{a} = {}^{\mathbf{F}}\!/_{m} \qquad \text{or} \qquad (\boldsymbol{a}m = \mathbf{F}). \qquad (1.10)$$

According to the fundamental principle of mechanics already mentioned above, the principle of independence of action of forces (motions), if several forces act simultaneously on a material point, then each force gives it acceleration as if there were no other forces (equivalent statement:

"if a body participates in a complex motion, then the resultant motion will be determined by the vector sum of motions"). In other words, the displacement will occur under the action of the resultant force (equidomotive force) $\mathbf{F} = \mathbf{F}_1 + \mathbf{F}_2$, determined by the vector sum (according to the parallelogram rule, Fig. 1.13).

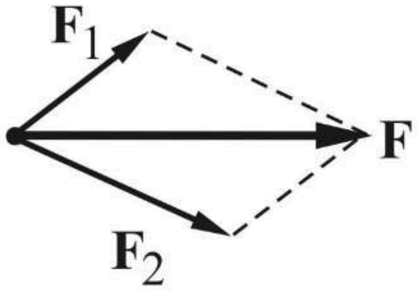


Fig. 1.13. The illustration of the resultant force vector

After some simple manipulations, equation 1.10 is transformed into equation 1.11

$$m\boldsymbol{a} = m\frac{d\upsilon}{dt} = \frac{d\mathbf{p}}{dt} = \mathbf{F}, \qquad d\mathbf{p} = \mathbf{F}dt. \qquad (1.11)$$

Here $\mathbf{p} = m\upsilon$ – momentum of a particle, $\mathbf{F}dt$ – force pulse. This is another statement of Newton's 2nd law, known as the "law of conservation of momentum in an isolated system": in the absence of external influence ($\mathbf{F} = 0$) the momentum of the system does not change ($d\mathbf{p} = 0$). This is one of the three basic laws of conservation (integrals of motion) in mechanics.

Newton's 3rd law is formulated as: "The force of the action is equal to the force of the counteraction" (Fig. 1.14).

$$\mathbf{F}_{1,2} = -\mathbf{F}_{2,1} \qquad (1.12)$$

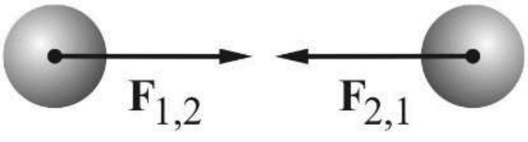


Fig. 1.14. The illustration of Newton's 3rd law

It is important to note that these forces are applied to different bodies (material points).

1.2.2. Dynamics of translational motion of a solid body (rigid object)

At first, consider the translational motion of a set of material points whose relative position does not change (Fig. 1.15).

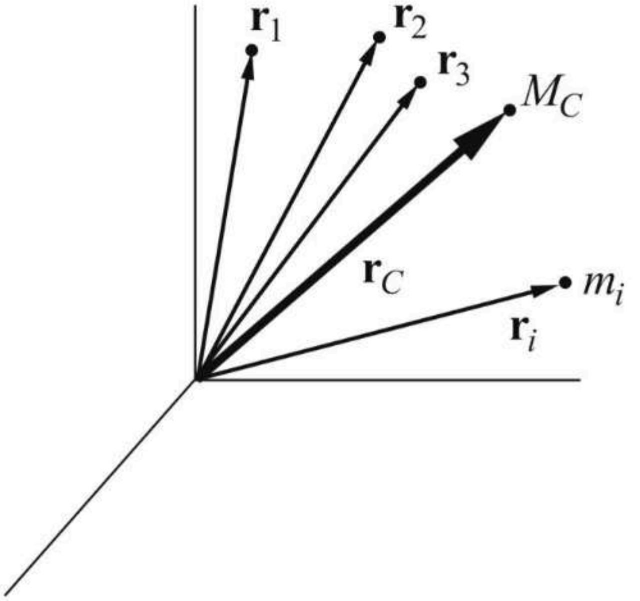


Fig. 1.15. The illustration of the motion of material point (particle) system, $\mathbf{r}_C$ is the center of inertia (masses) coordinate, M_C is the total mass of material points (particle)

The momentum of a material point (particle) m_i, moving with velocity $\boldsymbol{\upsilon}_i$, is given as $\mathbf{p}_i = m_i\boldsymbol{\upsilon}_i$. The total momentum of this set of material points (particles) is

$$\mathbf{p} = \sum_{i=\mathbf{1}}^{N} \mathbf{p}_i \,,$$

i.e. is defined as the vector sum of impulses of *N* material points (particles) (additively); hereafter, the summation is performed from 1 to N.

The coordinate of the center of inertia (masses) $\mathbf{r}_C$ is defined as follows:

$$\mathbf{r}_C = \frac{m_1\mathbf{r}_1+\cdots+m_i\mathbf{r}_i}{m_1+m_2+\cdots+m_i} = \frac{\sum m_i\mathbf{r}_i}{\sum m_i=M_C} . \qquad (1.13)$$

By differentiating expression 1.13 in time, we obtain the velocity of displacement of the center of mass (inertia):

$$\boldsymbol{\upsilon}_C = \frac{d\mathbf{r}_C}{dt} = \frac{\sum m_i{}^{d\mathbf{r}_i}/_{dt}=\sum m_i\upsilon_i}{M_C} = \frac{1}{M_C}\sum \mathbf{p}_i = \frac{1}{M_C}\mathbf{p}_C. \qquad (1.14)$$

In other words, the motion of this set of *N* particles can be considered as the motion of a particle with mass M_C, located at the center of inertia (masses) $\mathbf{r}_C$, with velocity $\boldsymbol{\upsilon}_C$, i.e. having momentum $\mathbf{p}_C$:

$$\mathbf{p}_C = M_C\boldsymbol{\upsilon}_C. \qquad (1.14a)$$

By analogy we obtain Newton's 2nd law (equation of motion) for translational motion of a solid body (rigid object) (Fig. 1.16).

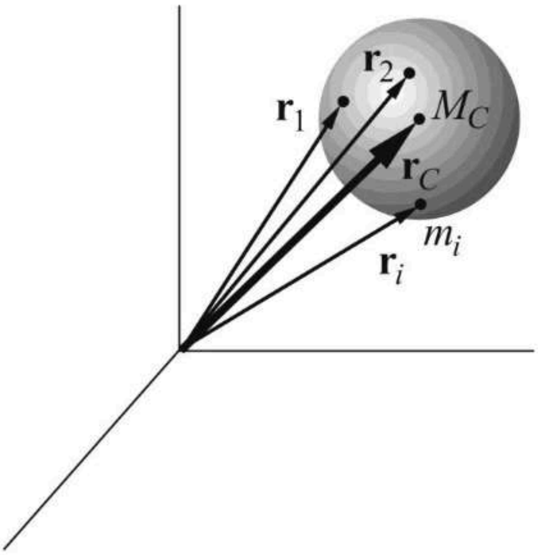


Fig. 1.16. The illustration of translational motion of a rigid object as a set of an infinite number of material points (particles)

Let's write the equation of motion for an elementary mass m_i under the influence of internal ($\mathbf{F}_{ij}^{int}$) and external ($\mathbf{F}_i^{ext}$) forces:

m_i in response to internal ($\mathbf{F}_{ij}^{int}$) and external ($\mathbf{F}_i^{ext}$) forces:

$$m_i \boldsymbol{a}_i = \mathbf{F}_{ij}^{int} + \mathbf{F}_i^{ext}. \qquad (1.15)$$

Here $\boldsymbol{a}_i$ – is the acceleration of elementary mass m_i (index j refers to the impact of elementary mass m_j on elementary mass m_i).

Summing up the left and right parts by all indices *i* and *j*, we obtain

$$\sum m_i \boldsymbol{a}_i = \sum \mathbf{F}_{ij}^{int} + \sum \mathbf{F}_i^{ext}. \qquad (1.15a)$$

The sum of all internal forces according to Newton's 3rd law is zero ($\sum \mathbf{F}_{ij}^{int} = 0$). Thus, the motion of this infinite set of material points, i.e. a solid body, is determined only by the action of external forces:

$$\sum m_i \boldsymbol{a}_i = \sum \mathbf{F}_i^{ext}. \qquad (1.15b)$$

Let's use the definition of the coordinate of the center of mass of the system of material points $\mathbf{r}_C = \sum m_i \mathbf{r}_i / M_C$ (1.13), where m_i refers to the i-mass element of the body, the sum of which is equal to the mass of the body M_C. Since the acceleration of the element of mass m_i is equal to $\boldsymbol{a}_i = \frac{d^2\mathbf{r}_i}{dt^2}$, the acceleration $\boldsymbol{a}_C$ of the center of mass of the body can be written in the form

$$\boldsymbol{a}_C = \frac{d^2\mathbf{r}_C}{dt^2} = \frac{\sum m_i \frac{d^2\mathbf{r}_i}{dt^2}}{M_C} = \frac{\sum m_i \boldsymbol{a}_i}{M_C}.$$

Then considering equation (1.15b), we obtain:

$$\boldsymbol{a}_C M_C = \sum \mathbf{F}_i^{ext} = \mathbf{F}, \tag{1.16}$$

i.e. the center of inertia (masses, gravity) moves as a material point (particle) with a total mass $M = \sum m_i$ which would move under the action of all the applied forces (resultant force **F**).

1.2.3. Inertial and non-inertial reference frames

It follows from Newton's 1st law that the equation of motion of a material point depends on the properties of the reference frame. It is satisfied in the so-called inertial reference frames. It is easy to show that there are an infinite number of such systems. To do this, consider the simplest case: the system k' moves along the coordinate x with a constant velocity υ_{0x} with respect to the inertial reference system k (Fig. 1.17). In classical physics, it is assumed that at speeds much smaller than the speed of light, time in both coordinate systems flows equally: $t = t'$.

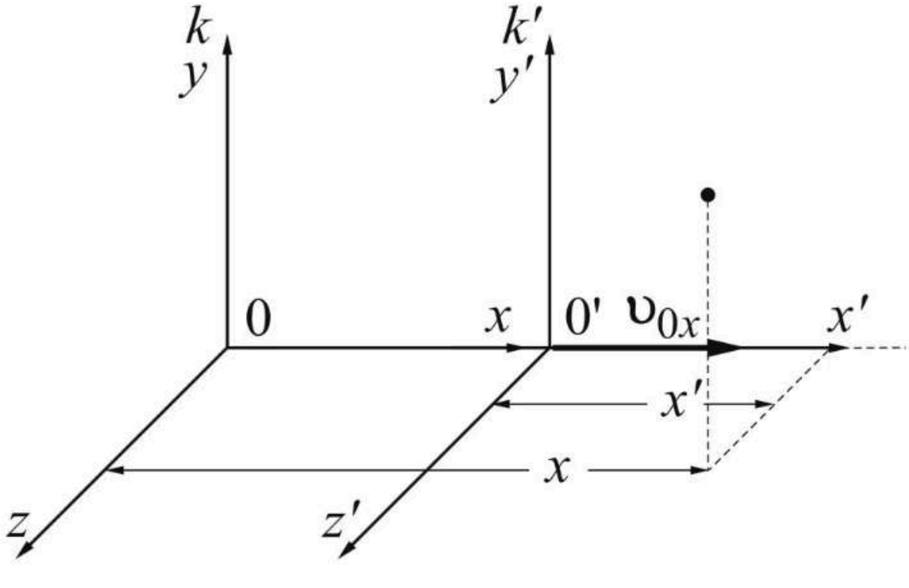


Fig. 1.17. The illustration of the construction of inertial reference frames

The relationship between the coordinates of a material point in these coordinate frames is established by a set of four equations

$$x = x' + \upsilon_{0x} t, \;\; y = y', \quad z = z', \quad t = t', \qquad (1.17)$$

called Galilean transformations. By differentiating the coordinate functions with respect to time, we obtain the relations for the velocity components:

$$\frac{dx}{dt} = \frac{dx'}{dt} + \upsilon_{0x}\,, \qquad \frac{dy}{dt} = \frac{dy'}{dt}\,, \qquad \frac{dz}{dt} = \frac{dz'}{dt}\,;$$

$$\upsilon_x = \upsilon_x{}' + \upsilon_{0x}\,, \qquad \upsilon_y = \upsilon_y{}'\,, \qquad \upsilon_z = \upsilon_z{}'\,. \;(1.18a)$$

Similar relations can be written when moving only along *y* or *z*:

$$\upsilon_y = \upsilon_y{}' + \upsilon_{0y}\,, \qquad \upsilon_x = \upsilon_x{}'\,, \qquad \upsilon_z = \upsilon_z{}', \qquad (1.18b)$$

$$\upsilon_z = \upsilon_z{}' + \upsilon_{0z}\,, \qquad \upsilon_x = \upsilon_x{}'\,, \qquad \upsilon_y = \upsilon_y{}'\,. \qquad (1.18c)$$

Then the scalar relations (1.18a, 1.18b, 1.18c) for the velocity components can be represented in vector form (graphically as in Fig. 1.18):

$$\boldsymbol{\upsilon} = \boldsymbol{\upsilon}' + \boldsymbol{\upsilon}_0. \qquad (1.19)$$

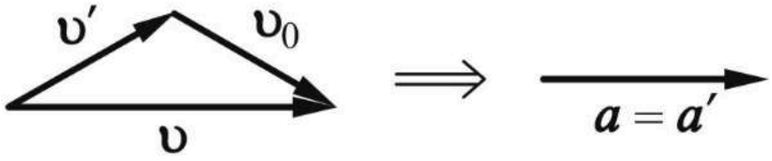


Fig. 1.18. The representation of velocity vectors (left) of and acceleration (right) a material point (particle) in inertial reference frames

It is obvious that relation (1.19) is satisfied for any choice of mutual directions of the coordinate axes, while, for example, relation (1.18a) is satisfied only for orientation of the coordinate axis as shown in Fig. 1.17. Differentiating with respect to time leads to the relation

$$\frac{d\upsilon}{dt} = \frac{d\upsilon'}{dt}, \qquad \boldsymbol{a} = \boldsymbol{a}'. \qquad (1.19a)$$

Multiplying both parts of the equality by the mass of the material point m, we obtain

$$\boldsymbol{a}m = \boldsymbol{a}'m, \qquad \mathbf{F} = \mathbf{F}'. \qquad (1.19b)$$

Thus, we conclude that the laws of dynamics are invariant with respect to Galileo's transformations (i.e. they do not change when passing from one inertial reference frame to another). From a mechanical point of view, all inertial frames are equivalent. This manifests itself in the fact that no mechanical experiments within an inertial frame can determine whether a body is moving uniformly straight or is at rest. This statement is called Galileo's principle of relativity.

Now consider two of the simplest examples of the motion of a material point in non-inertial coordinate frames.

1. The coordinate frame k' moves progressively relative to the inertial frame k along the x axis with acceleration a_{0x} (Fig. 1.19).

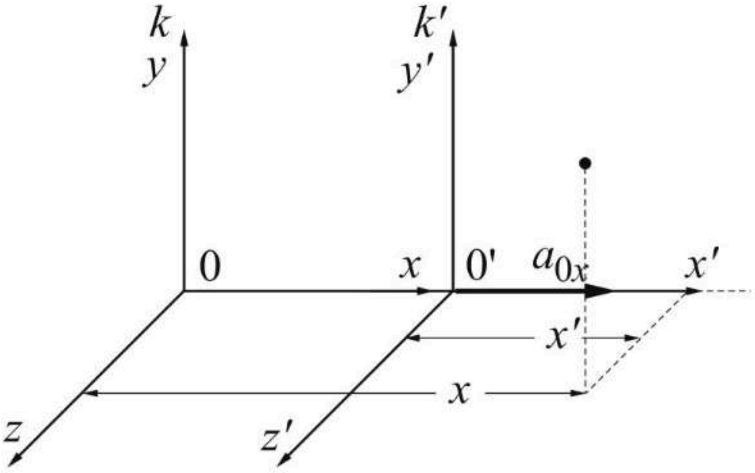


Fig. 1.19. The illustration of constructing a non-inertial coordinate frame

The relationship between the coordinates of a material point in an inertial system and the reference frame moving with acceleration a_{0x} is established by relation (1.20) by analogy with equations (1.17):

$$x = x' + \frac{a_{0x}t^2}{\mathbf{2}}, \qquad y = y', \quad z = z'. \tag{1.20}$$

Then the relation between the components of acceleration of a material point in the inertial and non-inertial a_x, a_y, a_z and non-inertial reference frame $a_{x'}$, $a_{y'}$, $a_{z'}$ is defined as follows (1.21):

$$\frac{d^2x}{dt^2} = \frac{d^2x'}{dt^2} + a_{0x}, \;\; a_x = a_{x'} + a_{0x}, a_y = a_{y'}, \; a_z = a_{z'}, \tag{1.21}$$

i.e. the z- and y-components of the acceleration remain unchanged.

Similar relations can be written down when the reference k' moves along the y- or z-axis.

In general, the acceleration $\boldsymbol{a}$ of a material point in an inertial coordinate system is related to the acceleration $\boldsymbol{a}'$ in a coordinate system whose axes are oriented arbitrarily relative to the coordinate axes in system k and moves with acceleration $\boldsymbol{a}_0$, by equation (1.21a):

$$\boldsymbol{a} = \boldsymbol{a}' + \boldsymbol{a}_0. \tag{1.21a}$$

Multiplying the left and right parts of equation (1.21a) by m, we obtain an expression for Newton's 2nd law in the form

$$\boldsymbol{a}m = \mathbf{F} = \boldsymbol{a}'m + \boldsymbol{a}_0 m, \tag{1.22}$$

from which follows the equation of dynamics of a material point in a non-inertial reference frame k':

$$\boldsymbol{a}'m = \mathbf{F}' = \mathbf{F} + \mathbf{F}_{in}(= -\boldsymbol{a}_0 m). \qquad (1.23)$$

Thus, when moving to a noninertial reference frame, a fictitious force, the so-called force of inertia, whose value is determined by the relative acceleration $\boldsymbol{a}_0$: $\mathbf{F}_{in} = -\boldsymbol{a}_0 m$ must be added to the equation of motion of the material point.

Fig. 1.20 shows the manifestation of the force of inertia in a noninertial frame of reference.

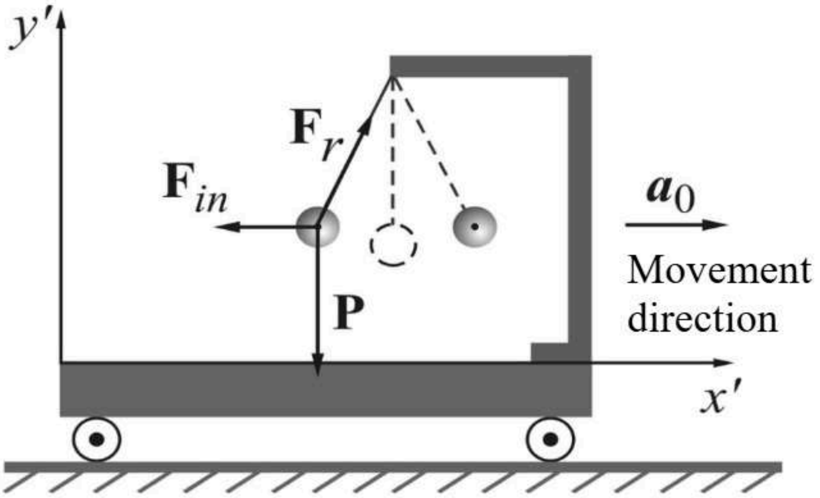


Fig. 1.20. The illustration of the inertial force during the accelerated straight-line motion of a cart on a horizontal surface (**P** – gravity; $\mathbf{F}_r$ – thread reaction; $\boldsymbol{a}_0$ – acceleration; $\mathbf{F}_{in}$ – force of inertia). During sharp braking (acceleration equals $-\boldsymbol{a}_0$) the body (object) deflects to the right

In a uniform straight-line motion of a cart on a horizontal surface, the thread on which a load of mass *m* is hanging is in a vertical position. Under the conditions of this experiment, any coordinate system fixed on the surface could be considered inertial. Now let us bring the cart in motion along the same direction with acceleration $\boldsymbol{a}_0$. The load will deflect to the left, as shown in Fig. 1.20, or to the right (when breaking sharply). Relative to the reference frame associated with the cart, the load rests in an inclined position, although the sum of forces $\mathbf{F}_r + \mathbf{P}$ is not equal to zero (here $\mathbf{F}_r$ is the reaction force of the thread acting from the side of the support – the plumb bob). To explain why the load is at rest in an accelerated moving cart, we need to introduce the force of inertia $\mathbf{F}_{in} = -\boldsymbol{a}_0 m$, i.e. the deflection of the load will correspond to the condition:

$$\boldsymbol{a}_0 m = \mathbf{F}_r + \mathbf{P}. \tag{1.24}$$

This situation is familiar to anyone who has been in a moving vehicle when it suddenly stops or when rapidly gaining speed.

2. The coordinate frame (e.g., associated with a horizontally placed disk) rotates with a constant angular velocity ω, relative to the vertical axis, as shown in Fig. 1.21.

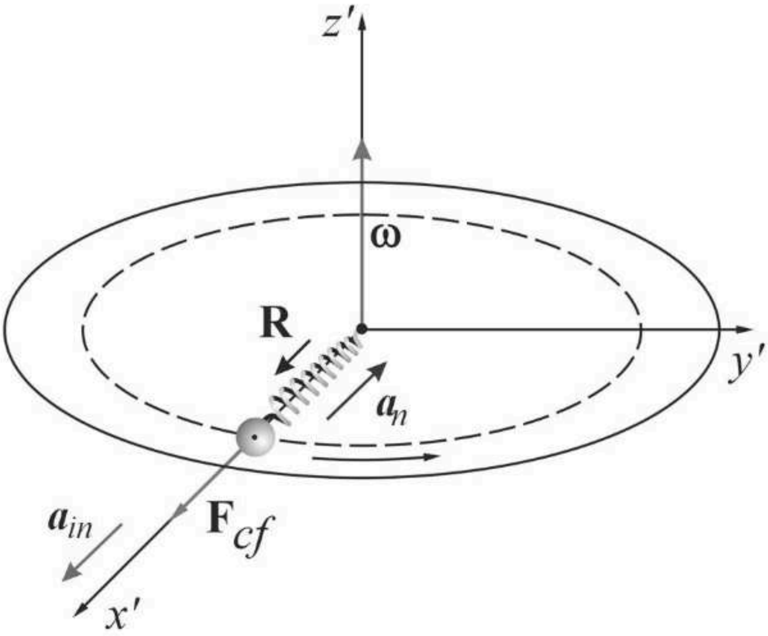


Fig. 1.21. Centrifugal forces of inertia $\mathbf{F}_{cf}$ in a rotating frame of reference

A body (rigid object) (e.g. a ball) of mass m is fixed by a spring in the center of a rotating frame of reference and can slide without friction along a spoke passing through the center of the ball. The vector **R** is counted from the center of rotation. For an external observer everything is easy to explain: the body rotates on a circular orbit with centripetal acceleration

$$\boldsymbol{a}_n = -\omega^2\mathbf{R}. \tag{1.25}$$

It is held on this trajectory by the elastic force of the stretched spring $\mathbf{F}_{\boldsymbol{str}}$

$$\mathbf{F}_{\boldsymbol{str}} = -\omega^2\mathbf{R}m. \tag{1.26}$$

For an observer located in the frame k', the stretched state of the spring can be easily explained if he physically feels the rotation of the disk (for example, being in the chair of a rotating carousel). Such a position of the body (ball) is described by the introduction of the centrifugal force of inertia:

$$\mathbf{F}_{in} = \mathbf{F}_{cf} = m\omega^2\mathbf{R}. \tag{1.27}$$

This force must always be considered in a rotating frame of reference, whether the body is moving relative to it or not. A simple experiment is a good illustration. A plumb line is placed on the rotating disk at different distances from the center of rotation **R** (Fig. 1.22).

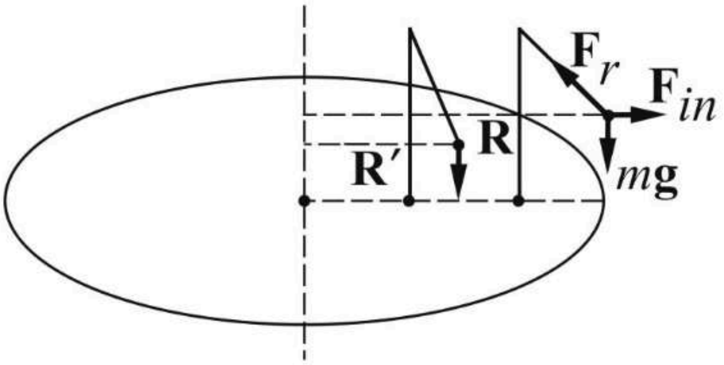


Fig. 1.22. Centrifugal forces of inertia acting on the plumb line on a rotating disk. The farther the plumb bob is located, the greater the centrifugal force of inertia

The inclined position of the plumb bob in a non-inertial system corresponds to the equation $\mathbf{F}_r + m\mathbf{g} + \mathbf{F}_{in} = 0$, the angle of inclination is determined by the value $\mathbf{F}_{in}$, depending on the mass of the plumb bob, angular velocity of the disk rotation and distance **R**.

When a body moves in a rotating coordinate frame, the so-called Coriolis force of inertia arises. On the scale of the Earth, its effect on the motion of bodies is quite significant. It is manifested, for example, in the erosion of the right bank of rivers in the northern hemisphere or the left bank of rivers in the southern hemisphere, displacement of air masses, tides, etc. But it has no effect on a person moving at the usual speed of 5–6 km per hour. When you are in a fast-moving vehicle, it is not possible to distinguish the effect of Coriolis forces from other forces. For this reason, we will omit consideration of this type of forces of inertia.

The entire variety of physical phenomena is described within the framework of the four types of fundamental interactions: gravitational, electromagnetic (electric and magnetic interactions are special cases of electromagnetic interaction), strong and weak. When describing mechanical phenomena, three types of forces are used: the forces of classical gravitational interaction, frictional forces, and elastic forces.

1.2.4. Gravitational interaction. Gravity and weight

Gravitational interaction of two material points m_1 and m_2, located at a distance r from each other, is described by the so-called law of universal gravitation [1, 2] (Fig. 1.23a), which in scalar form is represented as

$$F = \gamma \frac{m_1 m_2}{r^2}, \tag{1.28}$$

where γ is the gravitational constant ($\gamma = 6{,}67 \cdot 10^{-11}$ m^3/kg$\cdot$s^2).

For a body of mass m located at height h above the Earth's surface (Fig. 1.23 *b*), this interaction is written in the form

$$F = \gamma m \frac{M}{(R+h)^2}, \tag{1.29}$$

where M and m are the masses of the Earth and the body, respectively, R is the radius of the globe (approximately equal to 6400 km).

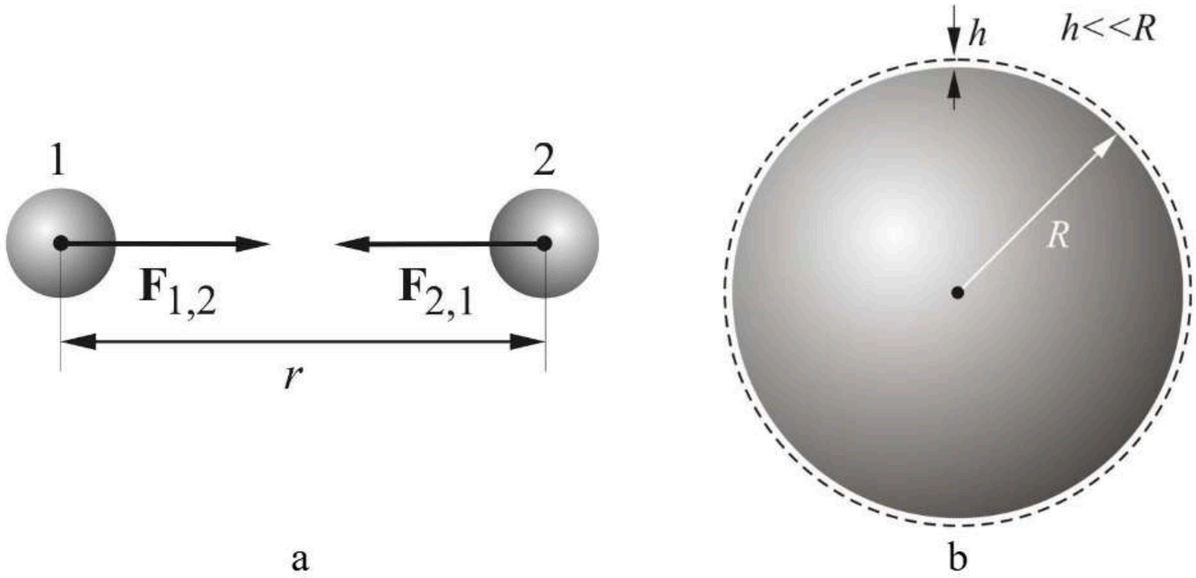


Fig. 1.23. a) – Forces of gravitational interaction of two material points; b) – the illustration of gravitational interaction of the Earth and a body located at height h above the surface of the Earth

Let us assume that the habitat of the animal world extends up to 10 km (the altitude ceiling belongs to geese: during the flight they rise to heights of about 9.5 km). It is not difficult to calculate that the differences in the Earth's gravitation near the surface and at a height of $h = 10$ km will be small (about 3%). According to Newton's second law $\boldsymbol{a}m = \boldsymbol{F}$. Since the force of the Earth's gravity is practically constant, the acceleration, under the action of which the bodies fall to the ground, can be assumed constant. This acceleration is called the acceleration of free fall and is denoted by **g**[*], and the force of gravitation-

---

[*] The free-fall acceleration g at the Earth's surface varies from 9.780 m/s^2 at the equator to 9.832 m/s^2 at the poles. This difference is due to the presence of centrifugal forces of inertia due to the Earth's daily rotation. The standard value adopted in the construction of systems of units is g = 9.80665 m/s^2. It was determined as an "average" of the values at various points on the Earth's surface. It is approximately equal to the acceleration of gravity at latitude 45.5° at sea level. The altitude difference is as follows: at sea level g is 9.80066 m/s^2, and at an altitude of 10 km it is 9.7759 m/s^2. In approximate calculations, the value of the free fall acceleration g is taken to be 9.81, 9.8 or 10 m/s^2.

al interaction is called the gravitational force and is denoted by $\mathbf{F}_g$. Now the equation of motion of the falling body is written in the form:

$$m\mathbf{g} = \mathbf{F}_g. \tag{1.30}$$

Body weight **P** is the force with which a body (e.g., the ball in Fig. 1.24) acts on a horizontal support (highlighted in gray) or a suspension.

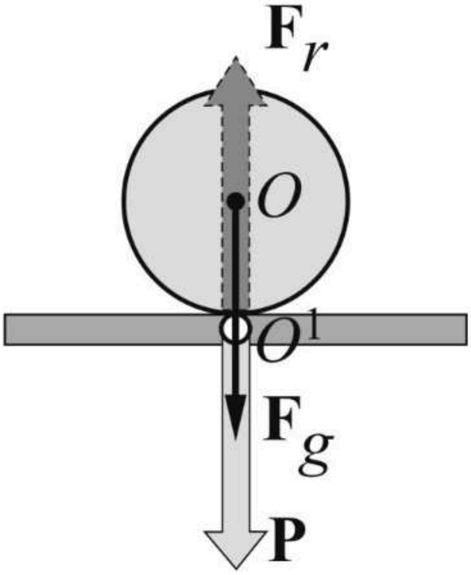


Fig. 1.24. Directions and points of application of forces in the body-support system located in the field of Earth's gravity

The support reaction force $\mathbf{F}_r$, (indicated by the dark gray arrow bordered with a dotted line and pointing upward), i.e., the force with which the support acts on the body, is equal in module, but opposite in direction to the vector of body weight, according to Newton's third law:

$$\mathbf{P} = -\mathbf{F}_r. \tag{1.31}$$

If the body is at rest on a horizontal surface, i.e. only gravity $\mathbf{F}_g$, (is indicated by an arrow applied to the center of mass of the body – point O) and the support reaction force $\mathbf{F}_\mathrm{r}$ act on it, then from the equality

$$\mathbf{F}_g = -\mathbf{F}_r \qquad (1.32)$$

it follows that

$$\mathbf{F}_g = \mathbf{P}. \qquad (1.33)$$

Thus, the weight **P** of the body is equal to the force of gravity $\mathbf{F}_g$, but these forces are applied to different bodies. The vector denoting the weight of the body is applied to the fulcrum[*] (O^1), and the vector of gravity, as already noted, is applied to the center of mass of the body (point O) lying on the fulcrum. The elastic reaction force $\mathbf{F}_r$ has no fixed point of application to the body.

According to Newton's second law (***am*** **= F)** the vertical motion of bodies is described by the equation

$$\boldsymbol{a}m = \mathbf{F}_g + \mathbf{F}_r, \qquad (1.34)$$

where m is the mass of the body, ***a*** is the acceleration with which the support is moved. From equations (1.31) and (1.34) for the weight **P** we obtain:

$$\mathbf{P} = \mathbf{F}_g - m\boldsymbol{a} = m\mathbf{g} - m\boldsymbol{a} = m(\mathbf{g} - \boldsymbol{a}). \qquad (1.35)$$

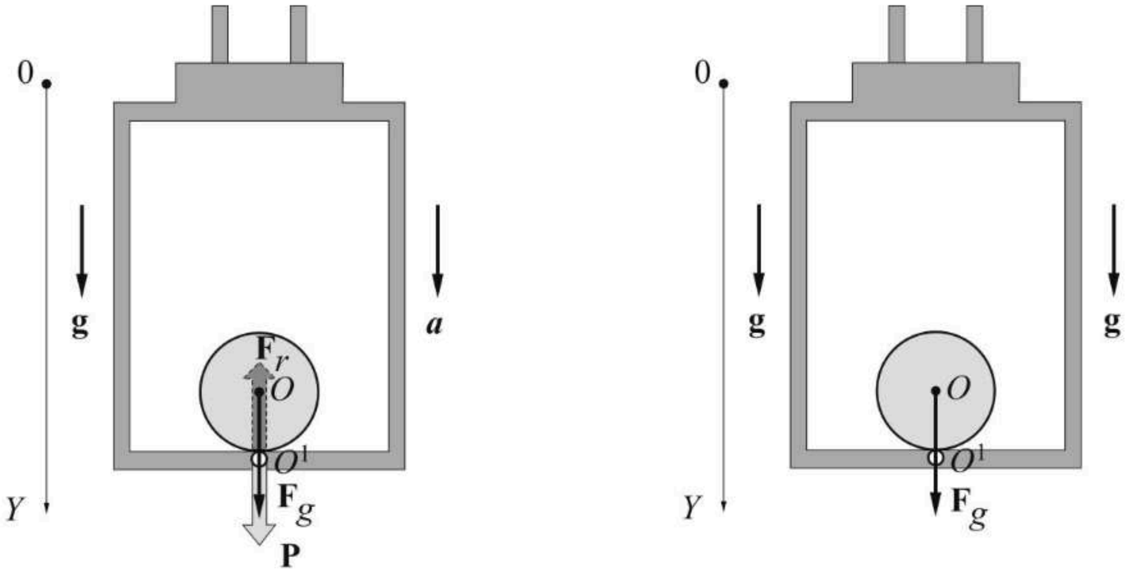
1.2.5. Weightlessness and overload

Let's consider the case of accelerated motion of the elevator (Fig. 1.25a), when the acceleration ***a*** is directed vertically downwards. This physical situation occurs, for example, if the rope holding the elevator has broken, but the braking device has worked, preventing the elevator from falling in the shaft with the acceleration of a free fall. The downward force of gravity is partially compensated by the friction force of the braking device. Obviously, in this case~~,~~ the acceleration with

---

[*] The fulcrum is a model concept. In reality, it is a contact surface.

which the elevator-body system will fall will be determined by the difference between gravity and the frictional force created by the braking device.



a b

Fig. 1.25. Downward motion of the elevator: a) – with acceleration ***a***, b) – with acceleration ***g*** (state of weightlessness of the body and elevator)

Thus, when the elevator moves with acceleration, the weight of the body in the elevator will be different from the weight measured in the elevator stationary relative to the Earth. Indeed, if the coordinate axis $0Y$ is directed vertically downward, the vectors ***P***, ***g*** and ***a*** are parallel to the axis $0Y$, and their projections are positive. Then equation 1.35 will take the form:

$$\mathbf{P}_y = m(\mathbf{g}_y - \boldsymbol{a}_y) \qquad (1.36)$$

or in scalar form

$$P = m(g - a). \qquad (1.37)$$

It follows from the equation (1.37) that the weight of a body whose direction of acceleration coincides with the direction of free fall acceleration is less than the weight of a resting body. This is the case shown in Figure 1.25a, where a body is placed inside the elevator. The same expression clearly shows that if the body and the elevator move down with the acceleration equal to the acceleration of a free fall, then there is no force of elastic reaction of the support (elevator) on the body (Fig. 1.25b). Such a situation can occur when the rope on which the elevator is suspended breaks and the braking device does not work. So, if the body together with the support falls freely, then $\boldsymbol{a} = \boldsymbol{g}$, and from formula (1.37) it follows that $\mathbf{P} = 0$. Disappearance of weight when the support moves with the acceleration of free fall under the action of gravity $\mathbf{F_g}$ is called weightlessness.

Outside the Earth's atmosphere, when the jet engines are turned off (i.e. when the thrust force of the engines is zero), only the gravitational interaction force acts on the spacecraft. Under the action of this force, the spacecraft and all the bodies in it move with the same acceleration, so the phenomenon of weightlessness is observed on the spacecraft.

When the body and the support move with acceleration $\boldsymbol{a}$ vertically upwards (Fig. 1.26), the weight of the body is greater than the force of gravity acting on it. This situation occurs when the elevator moves vertically upwards for a short time after switching on, while the winding speed of the cable holding the elevator increases in time, e.g. according to the linear law $(\upsilon = \upsilon_0 + at)$. After reaching a certain speed, the elevator moves uniformly until it begins to decelerate before the next stop. In this situation, the force leading to the accelerated upward motion of the elevator is transmitted through the rope to the elevator. The result is an increase in the resultant force of the elastic reaction of the support by the amount of this force. Both of these forces are directed vertically upward, and the increase in the resultant force leads to an increase in body weight (in high-speed elevators it is clearly felt as a bending of the legs at the beginning of the rise).

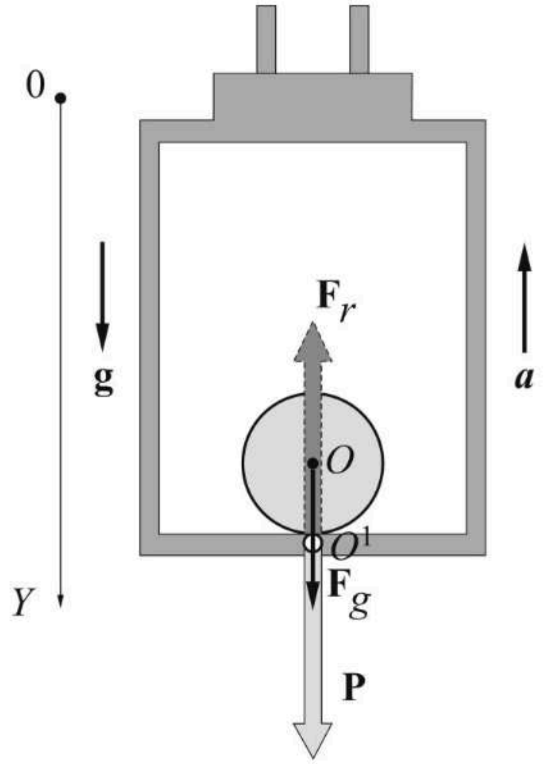


Fig. 1.26. Upward motion of the elevator with acceleration ***a***

So in this case the projections $\mathbf{P}_y$ and $\mathbf{g}_y$ are positive, and the projection $\boldsymbol{a_y}$ is negative (for the chosen direction of the $0Y$ axis, see figure 1.26). Therefore, for the weight modulus we obtain the expression

$$P = m(g + a). \tag{1.38}$$

The increase in body weight caused by the accelerated motion of the support or suspension is called an *overload.*

Units of measurement.

Let's introduce the units of measurement of physical quantities. Basic units are a set of independent units of measurement of physical quantities. Derived units are the units of measurement of physical quantities that are linked together by definite laws (formulas) using basic units.

As a rule, two systems of units are used: SI – International System of Units and CGS (centimeter-gram-second). But, following established traditions, off-system units are also used.

Basic Units:

	SI	**CGS**
length, l, L	m	cm
mass, m, M	kg	g
time, t, T	s	s

Derivative units:

	SI	**CGS**
speed υ	m/s	cm/s
acceleration a	m/s^2	cm/s^2
force f, F	N (Newton) kg·m/s^2; kg·m·s^{-2}	dyn g·cm/s^2; g·cm·s^{-2}
	1N = 10^5 dyn	

In engineering the metric system (MKS) of units is used, in which the basic units are: the unit of length – meter (m), the unit of force – kilogram-force (kgf), the unit of time – second (s). Kilogram force is defined as the force which impacts acceleration to a mass of 1 kg equal to the acceleration of free fall g. From this definition, it follows that 1 kgf = 9.81 N. It is this unit of force that is used in everyday practice, for example, to determine weight. The so-called rules of dimensionality [1] are used to form derived units, some of which will be given, as appropriate, in the following sections of the course.

Obviously, a change in the basic units entails a change in the derivative units. To construct them, we use the dimension formula, a relation that defines the relationship between the basic units $A_1 A_2$ … and the derivative A:

$$[A] = [A_1]^{\alpha} [A_2]^{\beta} [A_3]^{\gamma} \ldots [A_i]^{\tau}.$$

Here A is an alphabetic notation of a unit (symbols $\alpha, \beta, \ldots$ mean any number, including fractional, $2^{\frac{1}{2}} = \sqrt{2}$), $[A]$ is its dimensionality. In mechanics, the unit of time t corresponds to $[t] = T$, length l to $[l] = L$, mass m to $[m] = M$.

For example, for force:

$F = ma \Rightarrow [F] = [m][a] = [m][\upsilon][t]^{\mathbf{-1}} = [m][l][t]^{\mathbf{-1}}[t]^{\mathbf{-1}} = [m][l][t]^{\mathbf{-2}} = MLT^{-2} \Rightarrow$ 1 kg·m·s^{-2} (1N) or 1 g·cm·s^{-2} = 1 dyn.

The seemingly cumbersome dimension formula avoids errors in calculations, formula derivations, etc. If all calculations are done correctly, the magnitudes in the left and right parts of the resulting ratio should coincide.

1.2.6. Physiological aspects of the effects of weightlessness and overloading on humans

The proportions and sizes of the human body, the strength of muscles and the strength of bones are adapted to living in the gravitational field of the Earth. Therefore, if a person finds himself or herself in a situation where the force of gravity significantly exceeds the force of gravity or, conversely, is in a state of weightlessness or close to it, he or she has difficulty in performing the most normal movements.

1.2.6.1. *Weightlessness*

Human work in conditions of weightlessness, i.e. when the body movement is performed under the influence of a single force – the force of gravitational interaction, requires special skills, and a long stay in a state of weightlessness has a negative effect on the physical condition of humans and animals. The main factors include:

- disturbance of cerebral circulation, blood rush to the head;
- decrease of the volume of blood circulating in the cardiovascular system;
- trophy of certain muscle groups responsible for the counteraction of gravity;
- reduction of calcium in bone tissue, etc.*

---

* However, it should be noted that the effect of weightlessness on the physical condition of humans and animals with the allocation of specific factors is very relative. It is

To work under conditions of weightlessness and reduced gravity (for example, gravity on the Moon is 6 times less) it is necessary to understand the essence of these phenomena and, of course, to be able to move properly. Knowledge about human movement activity under these conditions is accumulated in the course of special medical and biological experiments, which widely use biomechanical methods.

Similar experiments have shown, for example, that at reduced gravity the rate and energy expenditures of human locomotor (*loko moveo* (Latin) – to budge) movements are reduced; locomotion and human state are characterized by increased amplitude of bending large joints; the way of movement changes – not by steps but by jumps. A brief state of weightlessness in terrestrial conditions can be created in an airplane moving along a special trajectory. This circumstance is used in the training of cosmonauts. In addition, to simulate reduced gravity special simulators have been developed. Methods of biomechanics are also used to develop means to facilitate human movements in unusual conditions.

When performing standard exercises or actions a person in the gravitational field develops certain movement stereotypes that ensure unconscious attainment of the required result. At the same time the person necessarily interacts with the support, to which gravity presses it. The usual interaction with the support disappears in non-consciousness. This is why the standard performance of exercises or actions leads to significant side effects. For example, a person who throws an object begins to move in the opposite direction according to the law of conservation of momentum.

Before the throw of the object of mass m_1 by the person (mass m_2) the momentum of the system was 0 and their momentum were also 0 (the person and the object are at rest, i.e. the total momentum $\mathbf{p} = 0$). It should remain unchanged even after throwing the object with velocity $\boldsymbol{\upsilon}_1$.

---

composed of stress factors and the influence of overloads during the flight before transition to weightlessness and the influence of weightlessness itself. These questions belong to the section of space medicine "Gravitational Pathophysiology".

Consequently, the person will start moving with some velocity υ_2 in the opposite direction, according to the condition $\mathbf{p} = m_1\upsilon_1 + m_2\upsilon_2 = 0$.

So if you jump into the water from the bow or stern of the boat, the boat will move in the opposite direction[*].

As was shown above, natural on Earth physical loads on the human body in space are absent. Therefore, during prolonged stay in weightlessness osteodystrophy (bone tissue dystrophy) occurs. It is caused by disruption of tissue metabolism, mainly phosphorus-calcium metabolism. It is characterized by rearrangement of the bone structure with replacement of bone elements by fibrous tissues. As a result, the resistance (resilience) of the human bone-supporting apparatus to the action of shock loads decreases. The main reason of changes in biomechanical properties of the bone tissue is first of all exudative inflammation of the epidermis in which the serous liquid is accumulated in the intercellular space and as a result the intercellular lacunas become larger which leads to decrease of its mineral density or saturation. The mineral density decreases the tensile strength and the modulus of elasticity. Negative calcium balance and decreased mineral dens1ity of bone tissue of some skeletal elements are mainly manifested in weightlessness conditions. However, the mineral density varies differently depending on the height of the skeleton. Starting from the level of lumbar vertebrae and below, the mineral density of bone tissue decreases. The recovery time of lumbar vertebral mineral density after the flight can be 2-3 times longer than the duration of the flight. This fact makes it possible to plan a post-flight rehabilitation regime for cosmonauts. It was found that conditions of weightlessness in terms of mineralization could be simulated. It turns out that calcium losses in spaceflight conditions correspond to those observed during prolonged bed rest. The fact is that in the horizontal position the load on the spine is minimal, which negatively affects all functions of the body, a manifestation

---

[*] This is only a qualitative comparison, because in this case the system is not closed – there are frictional forces. Jumping off the side of the boat is not recommended: the boat may tip over or scoop up water with the side.

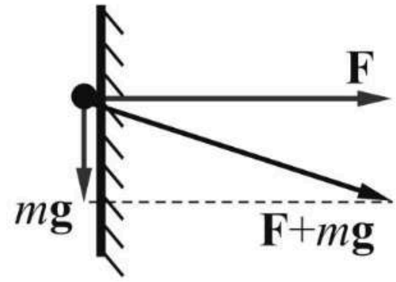
of which is the loss of calcium. *This allows us to consider bed rest as an adequate model of weightlessness in relation to the bone system.*

Adverse effects of real and simulated bedded weightlessness on bone mechanical characteristics were confirmed by experiments with rats on biosatellites and experiments with bone tissue biopsy specimens[*] taken from volunteers after prolonged hypokinesia (restricted motion).

As a means of preventing bone atrophy, artificial loading can be used to provide a level of tension in the skeleton that corresponds to terrestrial gravitational stresses. Or, for example, sufficiently long, on the order of one hour of quiet standing (with bed rest the rest of the time) prevents a negative calcium balance. Of course, regular exercise contributes to this.

Prolonged stay in weightlessness leads to under-loading of muscles and musculoskeletal system. For this reason astronauts must perform special physical exercises, wear special suits that make it difficult to move, etc. However, as experience shows, all this is not enough. A radical solution to the problem can only be achieved by creating artificial gravity.

1.2.6.2. Overloading



For example, astronauts experience the effect of overload both during takeoff of a space rocket (if the spacecraft starts with an acceleration of 4 *g*, the weight of the body in the spacecraft equals 5 mg, i.e. the astronaut feels a fivefold overload) and during braking of the spacecraft during re-entry into the dense layers of the atmosphere. Braking occurs both as a result of the action of braking engines and frictional forces arising from the motion of the spacecraft with high negative acceleration (deceleration) in the atmosphere. The same

---

[*] Biopsy is a method of research that involves taking cells or tissues from the body *in vitro* for diagnostic or research purposes.

situation takes place, for example, during acceleration or braking of a car. Let's write down Newton's second law in differential form to estimate how the value of short-term overload depends on the time of braking (acceleration): $\mathbf{F} = m\frac{\mathrm{d}\upsilon}{\mathrm{d}t}$. A resting or uniformly straight horizontally moving body has weight $m\mathbf{g}$. Then the value of the horizontal component of overload in relative values is $\frac{F}{mg} = \frac{m}{mg}\frac{\mathrm{d}\upsilon}{\mathrm{d}t} = \frac{1}{g}\frac{\mathrm{d}\upsilon}{\mathrm{d}t}$. (In general, the direction and magnitude of the overload are determined by the direction and modulus of the total vector $\mathbf{F} + m\mathbf{g}$; the vertical component due to the body's weight $m\mathbf{g}$ causes the body to move along the surface of the obstacle).

Thus, the value of short-term overload is the ratio of acceleration arising during braking to the acceleration of free-fall, and it is more significant the shorter the time of braking (acceleration). But it is measured in units of **g** (the value of free-fall acceleration **g** in the field of terrestrial gravity is taken as a norm: n-fold overload means that overload equals $n\mathbf{g}$). With sharp, instantaneous braking, it is a shock with a devastating effect even at a low speed of motion. *Remember the cases in life: often a glass jar, even with a minor impact on a concrete floor, breaks, which rarely happens when in contact with the floor, covered with a relatively soft material, such as linoleum. In the first case, the interaction time between the can and the hard floor (cushioning time) is very short and the overload is transformed into an impact.* Seat belts and airbags, shock absorbing bumpers are designed to increase the interaction time of the driver's body and the obstacle in the way of his movement. (See articles on the internet using the keyword "crash test"). (Not to be confused with the term "crash syndrome" or "prolonged pressure syndrome" – this condition occurs when a person is crushed by a heavy object for an extended period of time, for example in rubble. Note that the limit of overload endurance (the limit of acceptable tolerance) for a person is 50-60 g. The endurance limit for an unprotected body is the kinetic energy (see Section 1.4.1.2) absorbed by the body moving directly in the vehicle, corresponding to a sudden stop at a speed of approximately 15 km/h. At 50 km/h the energy exceeds the permissible value by about 10 times. Consequently, the task is to reduce the acceleration of the human body before the collision with the obstacle and after it has occurred, due to continuous deformations of the front part of the car body, deformation of airbags[*] etc.

In the absence of active safety equipment already at a speed of 50 km/h passengers sitting on different seats in the car would experience overloads of 90 g to 200 g. If safety equipment is used overloads would be in the range of 5-10 g.

Overloads are experienced by the pilot taking the airplane out of a dive (Fig. 1.27).

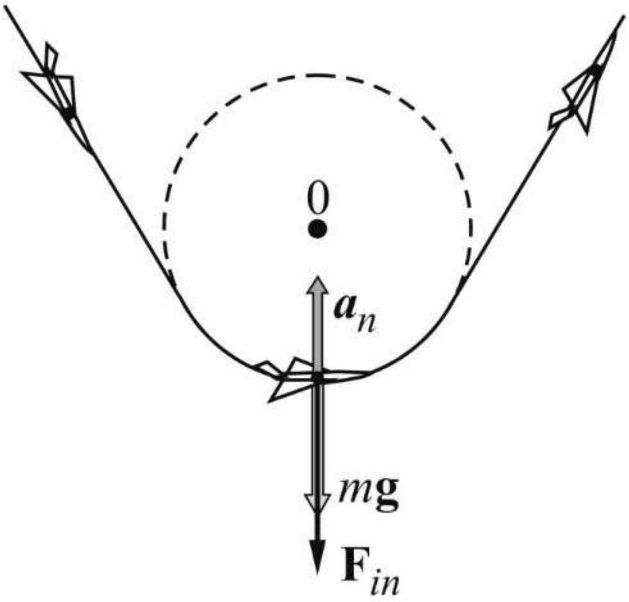


Fig. 1.27. Overloads occurring during aircraft withdrawal from a dive

The instantaneous numerical value of the centripetal acceleration when a material point moves along a curvilinear trajectory is equal to (see equation 1.5)

---

[*] In automobile racing accidents, data acquisition sensors often record overloads of the order of 80g. A record overload was recorded during the Silverstone race in Great Britain (1977) – about 180g. The sensors recorded 6 stops of the pilot's heart. As a result of the accident he suffered 29 fractures and three dislocations.

$$a_n = \frac{\upsilon^2}{R}, \tag{1.39}$$

where R is the radius of curvature and υ is the linear velocity, respectively, at a given point in time (here at the lower point of the trajectory). The airplane can be taken as a rotating non-inertial coordinate system, whose instantaneous value of centripetal acceleration is equal to the above value, and its direction is shown in Fig. 1.27. The pilot's body rests relative to the airplane (seat), so it presses on the seat with force **F**, equal to the sum of gravity force $m\mathbf{g}$ and centrifugal force of inertia $\mathbf{F}_{in}$ (absolute values – modules of force vectors are given):

$$P = mg + \frac{m\upsilon^2}{R}. \tag{1.40}$$

According to Newton's 3rd law, the same effect the seat has on the pilot.

Special training is necessary to prepare a person for work in conditions of considerable overload. For this purpose, a centrifuge is used, which is a cabin rotating in a horizontal plane on a long rod (Fig. 1.28). (Another application of the centrifuge is the separation of heterogeneous systems (components of different masses, such as blood components) in a field of centrifugal forces – the so-called centrifugal field).

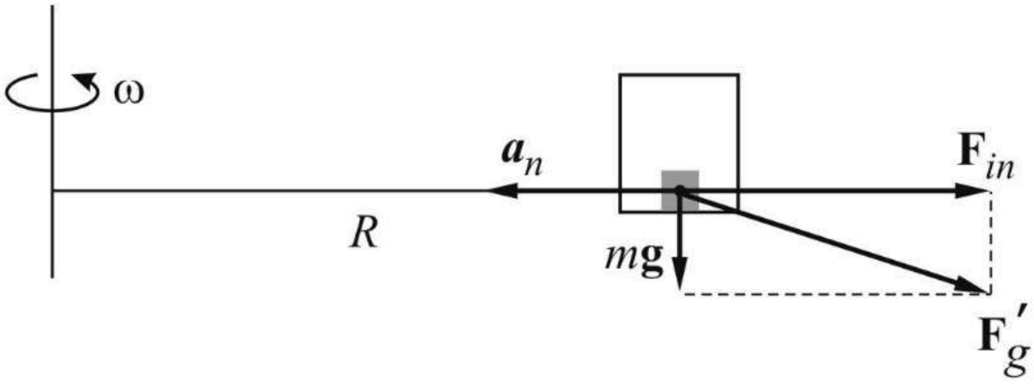


Fig.1.28. The principle of creating overloads on a centrifuge

Denote by R the total distance (length of the centrifuge rod + distance of the body's center of gravity to the edge of the booth). If the booth (hence the body) rotates with angular velocity ω, then the modulus of centripetal acceleration a_n is

$$a_n = \omega^2 R \tag{1.41}$$

and the *centrifugal force of inertia* $\mathbf{F}_{in}$ acts on the body inside the cabin, which is a non-inertial co-ordinate system (see Section 1.2.3):

$$\mathbf{F}_{in} = -m\boldsymbol{a}_n\,, \quad \mathbf{F}_{in} = m\omega^2 R. \tag{1.42}$$

The force $\mathbf{F}_\mathbf{g}'$, acting on the body is determined by the vector sum of the force of inertia $\mathbf{F}_{in}$ and the force of gravity near the Earth's surface $m\mathbf{g}$:

$$\mathbf{F}_g' = \mathbf{F}_{in} + m\mathbf{g}. \tag{1.43}$$

The modulus F_g' is equal to

$$F_g' = \sqrt{F_{in}^2 + (mg)^2} = \sqrt{(mg)^2 + (m\omega^2 R)^2} = m\sqrt{g^2 + \omega^4 R^2}. \tag{1.44}$$

The magnitude of the overload is determined by the ratio:

$$\frac{F_g'}{mg} = \sqrt{1 + \frac{\omega^4 R^2}{g^2}}. \tag{1.45}$$

Thus, by changing the angular velocity of rotation, it is possible to set the necessary overload. Table 1.1 shows the values of overloads that occur under some conditions, and Table 1.2 shows the values of short-term overloads that a person can bear.

To be able to bear significant overloads, special devices are used: ejection and shock-absorbing chairs, harness systems, protective helmets, etc.

Table 1.1

Values of overloads (in *g* units) occurring under different conditions

Overload conditions	Overload values
Overloading a still standing person	1
Passenger during airplane takeoff	Up to 1,5
Skydiver during parachute opening at a drop speed of 30 m/s	1,8
Skydiver during parachute opening at 40 m/s	3,3
Skydiver during parachute opening at fall speed of 50 meters per second	5,2
Pilot during parachute ejection from the aircraft	Up to 16
Overloads during the descent of Vostok spaceship	Up to 8–10
Overloads during descent of Soyuz spaceship	Up to 3–4

Table 1.2

Short-term overload (in *g* units), relatively painlessly tolerated by a trained person

Direction of local gravity	Magnitude overloads
in the back-thorax and chest-back direction	Up to 30
in the head-foot direction	Up to 20
in the head-to-toe direction	Up to 8

The magnitudes of overloads may vary within tolerance limits, but in all cases they should not interfere with the blood supply to the brain. Numerous studies have shown that acceleration in the head-foot direction causes blood outflow from the head and leads to noticeable disturbances in brain activity. Acceleration in the "chest-back" direction is much easier to endure a person, the blood supply to the brain, if disturbed, then to a much smaller extent.

Overloading impairs coordination of voluntary movements. Thus limits of infringement depends on the condition of the vestibular apparatus of the person and training (general and in the conditions of overloads).

1.3. The deformations and elastic forces. Mechanical properties of biological tissues

1.3.1. The deformations and elastic forces

When mechanical action is applied, bodies are deformed, i.e., they change their shape and size. There are several types of deformation: stress (compression), shear, torsion and bending. The basic relations describing deformations can be illustrated by the example of tensile deformation (Fig. 1.29).

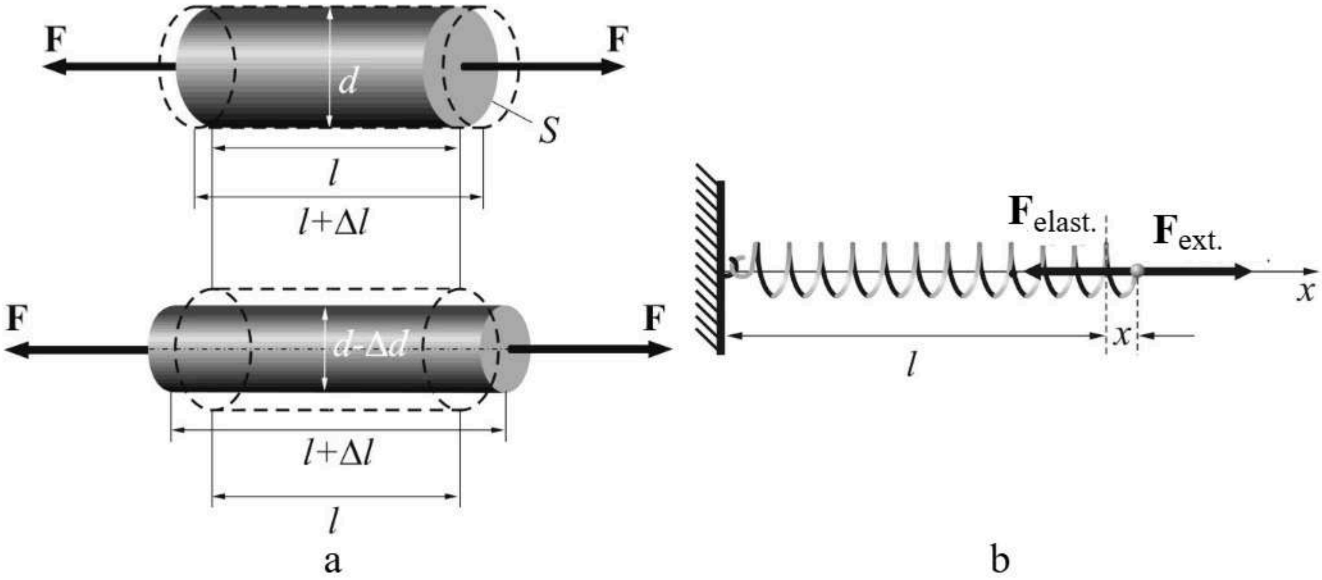


Fig. 1.29. Schematic representation of tensile deformation of a body: a) top – hypothetical deformation without changing the cross-section, at the bottom – with the change of cross-section; b) – deformation of tensile strain of an elastic spring

According to Hooke's law[*], the elongation of a rod Δl is directly proportional to the force of external action. Hooke's law is usually written in the form: $F_{\text{ext.}} = k\Delta l$, or:

$$F_{\text{ext.}} = kx, \tag{1.46}$$

where x $(= \Delta l)$ is the elongation along the conditional coordinate x, k is the stiffness (elasticity) factor, Fig. 1.29b.

Under the external action, obviously, the so-called elastic force $\mathbf{F}_{\text{elastic}}$ appears in the body, numerically equal to the force of external action $\mathbf{F}_{\text{ext.}}$, but directed in the opposite direction, i.e. $F_{\text{elastic}} = -kx$ in accordance with Fig. 1.29b.

The force acting on a unit of the rod (body) cross section σ, called *tension* (sometimes *surface force*, or *surface load*), is taken as a measure of the impact

$$\sigma = \frac{F}{S}. \tag{1.47}$$

Relative elongation $\varepsilon = \Delta l/l$ is used as a measure of strain. Here l is the initial length of the rod, Δl is the elongation. Taking into account the introduced notations, equation (1.46) is transformed into the equation:

$$\sigma = E\frac{\Delta l}{l} = E\varepsilon, \tag{1.48}$$

where E is tensile (compression) elasticity modulus – Young's modulus. The value of E depends on material properties. The relative tension (compression) of a rod is accompanied, as shown in Fig. 1.29a (bottom), by its transverse contraction $d - \Delta d$ (expansion $d + \Delta d$), the relative measure

---

[*] ***Robert Hooke***, (18 (28) July 1635 – 3 March 1703) – English naturalist and inventor. Member of the Royal Society of London (1663). Hooke is called one of the fathers of experimental physics, in many areas of science he owned some of the first fundamental works and many discoveries.

of which is the value *Δd/d*. These two strain characteristics are related by a relation called Poisson's ratio $\mu^* = \left|\frac{\Delta d}{d}\right| / \left|\frac{\Delta l}{l}\right|$.

Figure 1.30 shows a typical tensile strain curve of a solid, describing the dependence of stress σ on relative elongation ε. It highlights characteristic areas: the ОП area is the area of proportionality, where elongation increases in proportion to the stress and further up to the point «У» somewhat faster. The point «У» is called the elastic limit: with further increase in stress, the first signs of residual deformation, i.e. the deformation remaining in the body after the stress is removed, appear. This is where Hooke's law is fulfilled. The УТ region is the region of plastic deformation. In the horizontal section of TT′, the strain increases without increasing the stress (yield region) and then the point of rupture (fracture) «P» – the ultimate strength.

1.3.2. Mechanical properties of biological tissues

Mechanical properties of living tissues and organs determine their reactions to mechanical impacts. Living tissues, like any real body, deform under mechanical action.

As applied to the tissues of living organs, only the areas of elastic and residual deformation are considered.

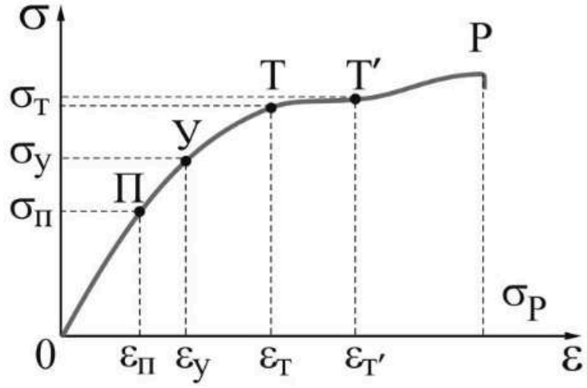


Fig. 1.30. Typical dependence of the stress σ on the relative elongation ε for a solid body

---

[*] It is equal to the ratio of modules (numerical values) of relative transverse contraction (compression) to elongation (shortening)

The curve describing the dependence of the stress σ on the relative elongation after the stress is removed beyond elasticity does not coincide with the curve in the OУ section, because, as noted above, the body shape does not return to its initial state. This phenomenon is called a hysteresis, and the graph itself is called a hysteresis loop. The area inside the hysteresis loop is proportional to the energy ΔU irreversibly lost during the process of stretching and subsequent return to the initial state. It is dissipated in the form of heat. The hysteresis is determined by residual internal stresses. Fig. 1.31 shows such dependence for isolated (i.e. obtained in laboratory conditions) atelectasis of human lungs. (Atelectasis is a pathological condition in which lung tissue loses its airiness and shrinks, reducing (sometimes significantly) its respiratory surface).

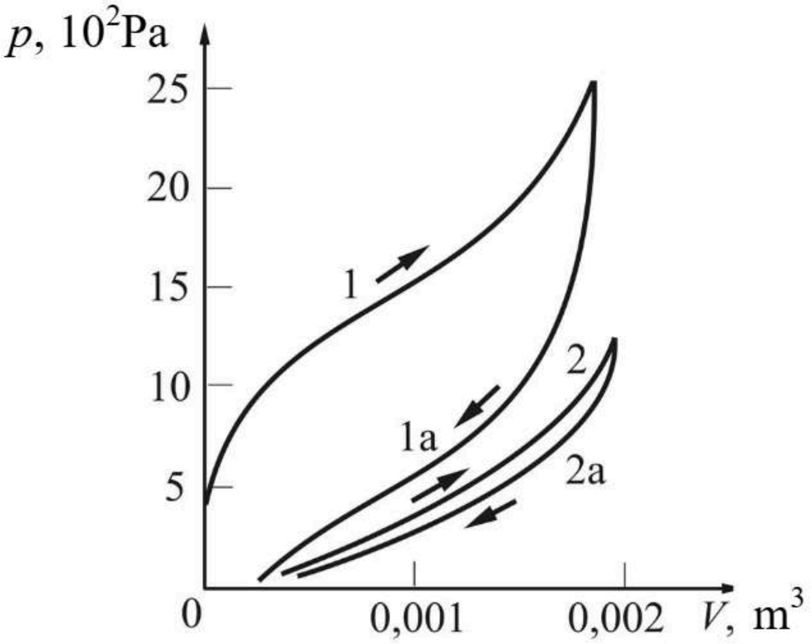


Fig. 1.31. The diagrams of isolated atelectic human lungs. 1, 1a – stretching and their subsequent compression (respectively) in air medium; 2, 2a – stretching and their subsequent compression (respectively) in a physiological solution; V – lung volume, p – pressure

Hooke's law for body tissues works only for short-term relatively small deformations, even for bones. The short-term application of force reveals mainly the elastic properties of living tissue, because the reactions of its viscous components do not have time to develop. Smooth muscles and other components of hollow organ walls (gallbladder, urinary bladder, stomach, intestines, venous and lymphatic vessels, etc.) are usually subjected to gradually increasing tensile forces for a long time. Only at the very beginning of the force they tense, showing their weak elastic properties, but then their tension gradually weakens (Fig. 1.32a). This mechanical property of living tissues is called plasticity. A quantitative measure of plasticity is the ratio of relative elongation to tension ε/σ. In the circulatory system it is inherent in the veins. Due to plasticity, the venous vessels hold more than two-thirds of all human blood. The inverse ratio is a measure of the elasticity of the material. Arteries (Fig. 1.32b), for example, are characterized by high elasticity and can withstand high loads during cardiac contractions (see section 2 for details).

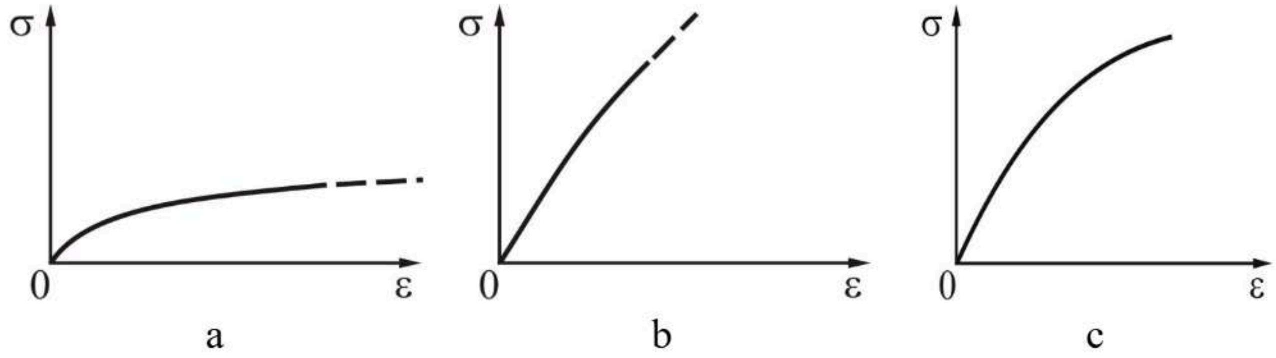


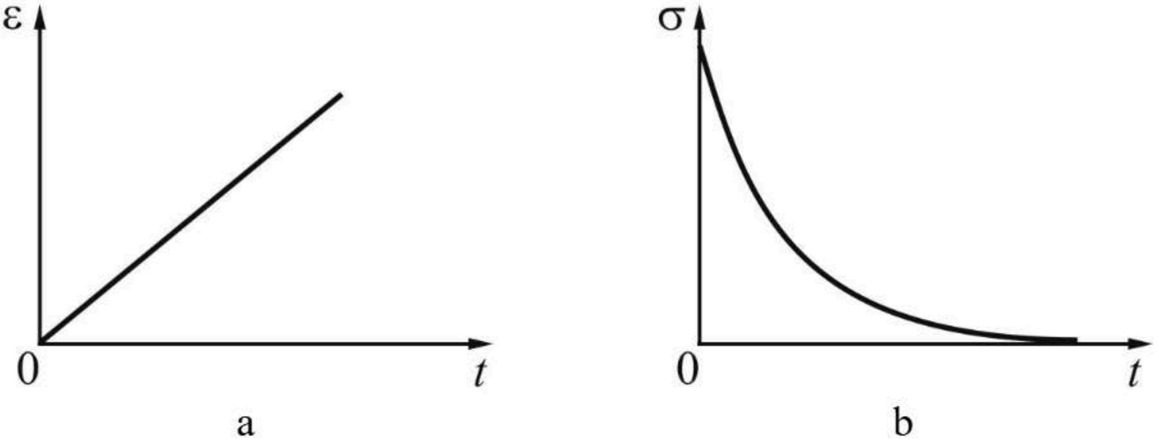
Fig. 1.32. Characteristic view of stress σ vs. relative tension ε: a) – plastic tissue, b) – elastic tissue, c) – bone tissue

As a technical material, biological tissue is a composite material; it is formed by a volumetric combination of chemically heterogeneous components. The mechanical properties of biological tissue differ from the mechanical properties of each individual component.

Bone tissue. Bone are the main material of the musculoskeletal system. About 2/3 of the mass of the compact bone tissue (0.5 volume) is an inorganic material, the mineral substance hydroxyapatite in its crystalline form. The rest of the bone consists of organic material, mainly collagen (high molecular weight compound, fibrous protein with high elasticity). Hydroxyapatite crystals are located between the collagen fibers (fibrils). The density of bone tissue is about 2400 kg/m^3. Its mechanical properties depend on many factors, including age, individual growth conditions and, of course, the site of the body. The composite structure of bone gives it the required mechanical properties: hardness, elasticity and strength. The dependence of stress σ on relative tension ε for bone tissue has the characteristic form shown in Fig. 1.32c, i.e. it is similar to the similar dependence for a solid body; Hooke's law is fulfilled at small deformations. The Young's modulus is about 10 GPa, the ultimate strength 100 MPa. These data can be compared with data for glass-reinforced capron (E = 8 GPa, tensile strength 150 MPa).

Skin. Skin consists of 60% of practically incompressible water, then collagen fibers, elastin (a fibrous protein just like collagen) and the main tissue, the matrix. Collagen makes up about 75% of dry mass, and elastin about 4% (Young's modulus 10÷100 Pa, tensile strength 100 Pa). Elastin (0.5 MPa and 5 MPa, respectively) stretches very strongly (up to 200÷300%), like rubber. Collagen can stretch up to 10%, which corresponds to the elastic properties of capron fiber.

Muscles. Muscles contain connective tissue consisting of collagen and elastin fibers. Therefore, mechanical properties of muscles are similar to those of polymers. As noted above, smooth muscles can be stretched considerably without much tension (they are plastic). Fig. 1.33 shows the dependences of relative strain (a) and stress (b) on the duration of exposure.



a

b

Fig. 1.33. Characteristic view of the dependences: a) – relative strain ε; b) – stress σ on the duration of exposure for smooth muscles

Blood vessel wall tissue (vascular tissue). Mechanical properties of blood vessels are mainly determined by properties of collagen, elastin and smooth muscle fibers. The content of these components of vascular tissue changes along the course of the blood system: the ratio of elastin to aortic collagen in the common carotid artery is 2:1, and in the femoral artery 1:2. The proportion of smooth muscle fibers increases with distance from the heart; in arterioles they are already the main component of vascular tissue.

In a detailed study of the mechanical properties of vascular tissue, it is distinguished which way the sample is cut out of the vessel (along or across the vessel). It is possible, however, to consider vessel deformation as a whole, as a result of liquid (blood) pressure action from inside on an elastic cylinder (fig. 1.34).

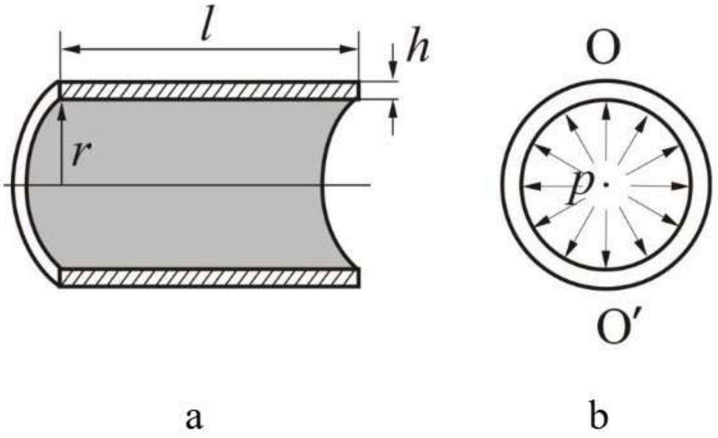


Fig. 1.34. The model of a blood vessel; the explanations in the text

Consider a cylindrical part of a blood vessel of length l, thickness h and radius of the inner part r. Sections along (shaded) and across the vessel axis (highlighted in gray) of the cylinder are shown in Fig. 1.34. Two halves of the cylindrical vessel interact with each other along the sections of the cylinder walls (shaded areas in Fig. 1.34a). (The methodology of the following calculation is based on the use of the model of surface tension in liquids [1]). The total area of this "interaction cross section" is $2hl$.

If there is a mechanical stress σ in the vessel wall, the elastic interaction force of the two vessel halves (see definition of σ, formula 1.46) is equal:

$$F = \sigma \cdot 2hl. \qquad (1.49)$$

This force is balanced by the forces of pressure on the cylinder from the inside (they are shown by arrows in Fig. 1.34b), which stretch them along the walls of the vessel, which leads to the aforementioned elastic forces – the elastic interaction). The forces are directed at different angles to the horizontal plane (perpendicular to the plane of the figure). To find their equilibrium, the horizontal projections must be summed up. However, it is easier to find the equal-action forces if we multiply the pressure by the projection of the semi-cylinder's area on the vertical plane HO'. This projection is $2rl$. Then the expression for the force through pressure looks like this:

$$F = p \cdot 2rl, \qquad (1.50)$$

where p is the pressure in the vessel. By equating expressions 1.49 and 1.50 we obtain:

$$p \cdot 2rl = \sigma \cdot 2hl. \qquad (1.51)$$

It follows from so-called Lame's equation

$$\sigma = \frac{p \cdot r}{h}. \qquad (1.52)$$

Let's assume that when stretching the vessel, the volume of its wall does not change (the wall area increases and the thickness decreases), i.e. the cross-sectional area of the vessel wall does not change. Then

$$2\pi rh = \text{const}, \text{ or } \quad rh = b = \text{const}.$$

Thus, formula (1.51) is transformed to

$$\sigma = \frac{p \cdot r}{h} = \frac{p \cdot r^2}{rh} = \frac{p \cdot r^2}{b}. \tag{1.53}$$

From the expression obtained, an important conclusion follows that in capillaries (i.e., at r→0) the mechanical stress tends to a minimum value (σ→0).

1.4. Frictional forces

When touching bodies or parts of the same body move relative to each other, frictional forces arise. In the first case they are called external, in the second case – internal. This is the movement of liquids and gases, the movement of a solid body in a liquid or gas, because the layers of liquid or gas immediately adjacent to the body move with the velocity of the body, and the motion of the body is affected by friction between adjacent layers. If there is no interlayer between the bodies in contact, e.g. in the form of lubricant, then the friction is called dry friction. If there is such a layer, then the friction between the body and the medium, as well as the friction between the layers of the medium itself is called viscous friction. A distinction is made between dry kinetic friction and dry rolling friction.

Kinetic friction forces are directed tangentially to the rubbing surfaces and counteract their relative motion. They act on both bodies as shown in Fig. 1.35 ($\mathbf{F}_{fr} = -\mathbf{F}'_{fr}$), which shows a body sliding on the horizontal surface of a fixed body 2 under the action of force **F**.

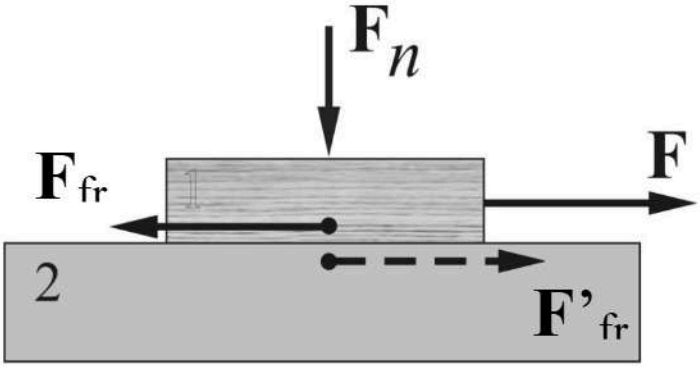
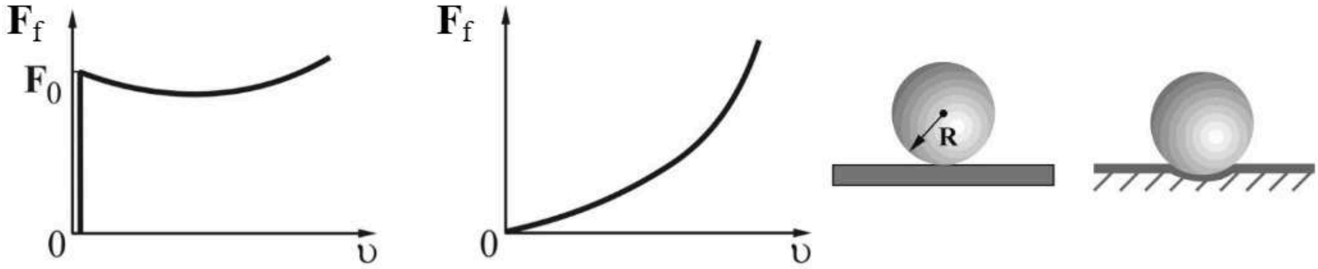


Fig. 1.35. To the illustration of the occurrence of frictional forces when a body slides on the surface of another

Body 1 is pressed against a body 2 with some force $\mathbf{F}_n$ (for example, the force of gravity), directed along the normal (perpendicular) to the surface of the contact. This force is called the normal pressure force, which determines the friction force. But frictional force arises not only in motion, but also when we try to move the body from its place. It increases from the moment the force is applied, as shown in Fig. 1.36a, to a certain value $\mathbf{F}_0$, after which the body begins to move, i.e., in the range of values $0 - F_0$, the body is at rest. This means that the force **F** needed to move the body is balanced by some force equal in magnitude to it, but oppositely directed. This is the so-called rest friction force. Obviously, the kinetic friction force $\mathbf{F}_{\text{fr}}$ is equal to the maximum value of the rest friction force $\mathbf{F}_0$. The kinetic friction force does not remain constant as the velocity of motion increases: at first, it decreases slightly and then begins to increase as shown in Fig. 1.36a (but with smooth surfaces, the friction force is virtually independent of the velocity of motion of the body).



a b c

Fig. 1.36. a) – The dependence of dry friction force on velocity of movement; b) – dependence of viscous friction on velocity of body movement in liquid flow; c) – the illustration of rolling friction force appearance: the deeper the body is pressed into underlying surface, the greater friction force (explanations in the text)

The magnitude of the friction force is determined by the ratio (1.54)

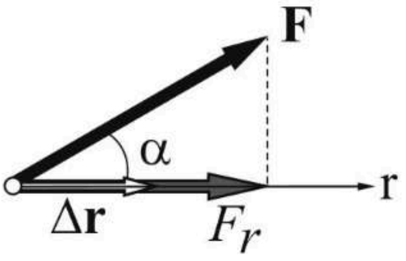
$$F_{fr} = kF_n, \tag{1.54}$$

where k is the coefficient of kinetic friction (at rest) – a value that does not depend on the area of the contacting surfaces. It depends on the physical characteristics of the rubbing surfaces and may depend on the velocity of movement. Body 1 will move with acceleration $\boldsymbol{a}$ under the action of the resultant force: $m\boldsymbol{a} = \mathbf{F} + \mathbf{F}_{fr}$ (i.e., the magnitude of the acceleration is determined by the relation $a = (F - kF_n)/m$)). The dependence of viscous friction on the velocity of the body in the medium is shown in Fig. 1.36b. At low velocities, the friction force is proportional to the first degree of velocity, and at high velocities, to the square of the velocity of body movement. There is no quiescent friction in viscous friction, i.e., body movement begins at the minimum force. This issue will be discussed in more details in the next section. Consideration of rolling friction forces is beyond the scope of this course. We will only note the general patterns. The friction force is smaller the greater the radius of the rolling body, the higher the hardness of the same-type materials in the pair "rolling body and underlying surface". For example, in the pair hardened steel – hardened steel, the coefficient of friction is about 0.01 mm (unlike the dimensionless value of the coefficient of kinetic friction, the rolling friction coefficient is of length unit dimension). In rubber-concrete pairs it is very high (from 15 to 35 mm) and depends on elastic properties of rubber. To a certain extent, such a pair is simulated by an air-inflated wheel moving on a hard surface. Naturally, here the elastic properties of the wheel are determined to a greater extent by the elasticity of the pumped air. The rolling friction force is always lower when the hardness of the underlying surface is higher than the hardness of the rolling body.

The frictional force plays a dual role. It is sought to be reduced in various mechanisms in order to reduce the cost of overcoming it. On the other hand, without it, it is impossible to transmit forces in mechanical units (e.g., friction belt-pin (wheel) type transmissions, which replace gears). Perhaps the most obvious examples are given by our everyday life in winter: ice and its consequences for pedestrians and vehicles. In medical practice, various types of friction forces in combination with elastic forces are used primarily in physical therapy procedures (massage, rubbing the skin, rubbing various ointments into the skin for therapeutic or cosmetic purposes, etc.), are considered in the design of various means of transportation, etc.

1.5 Work and energy. Dynamics of rotational motion

1.5.1. Work and Energy. The law of the conservation of energy



Work is a measure of the transfer of motion or energy from one body to another. Energy is a quantitative measure of motion in all forms of that motion.

The work ΔA to move a body under the action of a constant force $\mathbf{F}$ is defined by the relation

$$\Delta A = (\mathbf{F}\Delta\mathbf{r}) = F\Delta r\cos\alpha = F_r\Delta r\,, \qquad (1.55a)$$

where α is the angle between the vector directions of the acting force $\mathbf{F}$ and the displacement $\Delta\mathbf{r}$. The displacement along an arbitrary trajectory l (Fig. 1.37) can be represented in the form of small (elementary) displacements $\Delta\mathbf{r}$, which are performed under the action of a constant force $\mathbf{F}_i$, where the elementary work is written as

$$\Delta A_i = (\mathbf{F}_i\Delta\mathbf{r}_i) = F_i\Delta r_i\cos\alpha_i = F_{\Delta r_i}\Delta r_i. \qquad (1.55b)$$

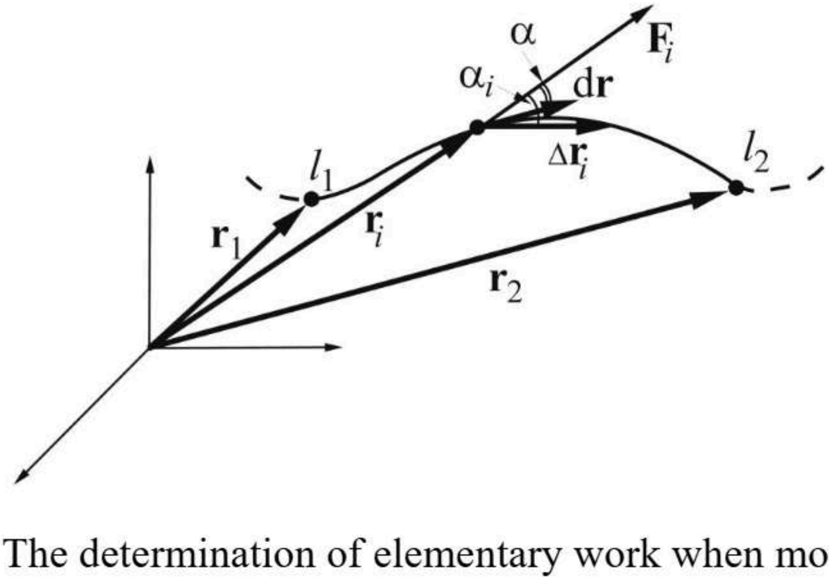


Fig. 1.37. The determination of elementary work when moving along an arbitrary trajectory l

The work on the segment 1-2 is approximately equal to:

$$A_{1,2} = \sum \Delta A_{\boldsymbol{i}} = \sum \mathbf{F}_i \Delta \mathbf{r}_i = \sum F_{\Delta r_i} \cdot \Delta r \cos\alpha_i . \qquad (1.55c)$$

With infinitesimal displacement along the trajectory $\Delta r_i \to dr$, $\cos\alpha_i \to \cos\alpha$, $F_{\Delta r_i} \to F_i$. Consequently, infinitesimal work can be written as $\delta A = \mathbf{F}d\mathbf{r} = F_r dr = F_l dl$. Here F_l is the projection of the force on a tangent to the trajectory at an arbitrary point r_i, dl is the infinitesimal displacement along a curvilinear trajectory (equivalent to ds, see Section 1.1.1). Then expression (1.55c) is transformed into expression (1.56), which determines the work of displacement on a segment 1-2 of an arbitrary trajectory l:

$$A_{1,2} = \int_1^2 \mathbf{F} d\mathbf{r} = \int_1^2 F_r dr = \int_1^2 F_l dl . \qquad (1.56)$$

The concept of *power* is introduced to determine the rate (speed) at which work is produced (done). The average power is defined as the work done in the time interval Δt

$$\langle N \rangle = \Delta A / \Delta t \qquad (1.57a)$$

and the instantaneous power, i.e. the power at a given moment in time, is determined by the ratio

$$N = \lim_{\Delta t \to 1} \frac{\Delta A}{\Delta t} = \frac{\delta A}{dt} \quad \text{or}$$

$$N = \mathbf{F}\frac{d\mathbf{r}}{dt} = \mathbf{F} \cdot \boldsymbol{\upsilon} = F\frac{dr}{dt}\cos\alpha = F\upsilon\, cos\alpha, \qquad (1.57b)$$

where υ is the velocity at which the work δA is done in an infinitesimal time interval dt (i.e., the velocity of the body being moved).

1.5.1.1. Potential energy

Consider the work of gravity (work in the field of gravity). The space where gravity acts is called the gravity field (gravitational field).

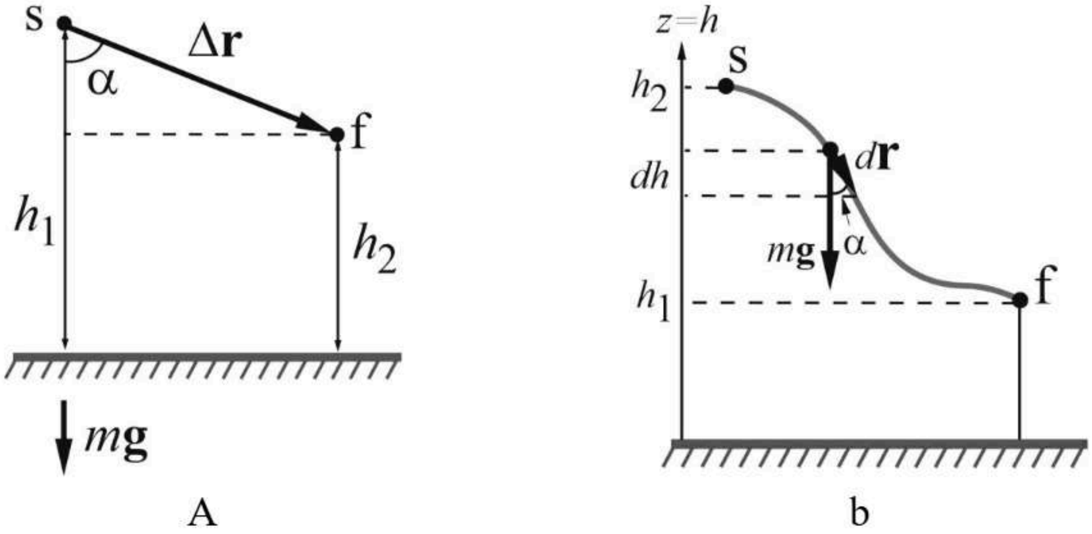


Fig. 1.38. To the illustration of the work in the field of gravity at displacement along straight (a) and curvilinear trajectories (b)

The work A_{12} to move a body of mass *m* at a distance **Δr** under the action of gravity along a straight-line trajectory is equal to

$$A_{s-f} = mg\Delta r\cos\alpha = -mg(h_1 - h_2) = mg(h_2 - h_1). \quad (1.58)$$

To determine the curvilinear trajectory work, we use equation (1.56), replacing the generalized force **F** by the force of gravity $m\mathbf{g}$:

$$A_{s-f} = \int_{s}^{f} m\mathbf{g}d\mathbf{r} = m\int_{s}^{f} g\cos\alpha\, dr = -mg\int_{s}^{f} dh = -mg(h_2 - h_1). \quad (1.58a)$$

Thus, work in the field of gravity depends only on the initial and final position of the body in this field. Such fields are called potential and forces are called conservative.

Let us rewrite the expression (1.58a) as

$$\Delta A_{s-f} = E_{P_f} - E_{P_s}, \text{ where } E_{P_s} = mgh_1, \qquad E_{P_f} = mgh_2. \qquad (1.59)$$

The value $E_P = mgh$ is called potential energy. Thus, the potential energy of a body is the ability of a body to perform work due to the body being in a field of conservative forces, a potential field. Work is done by the loss of potential energy.

The same conservative forces are elastic forces. Consider a spring made of elastic material as a model of an elastic body (Fig. 1.39).

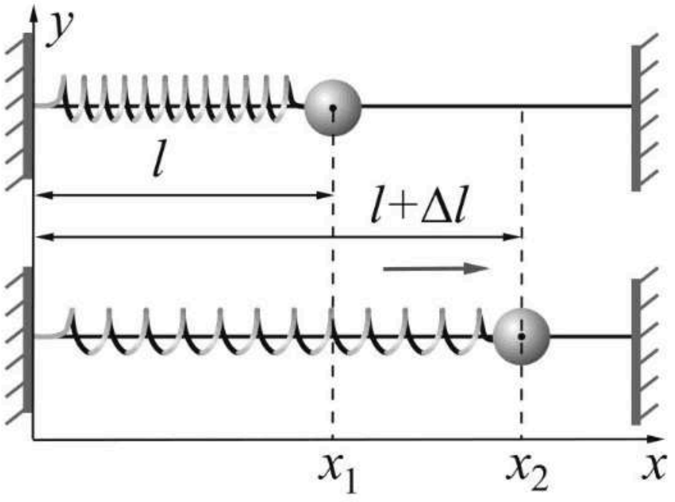


Fig. 1.39. To the illustration of the elastic forces work

The work *dA* of stretching (compression) a spring by an infinitesimal amount *dx* is equal to *dA* = *Fdx* = *kxdx* and, when stretched within elasticity limits, is determined by

$$A_{12} = \int_{s}^{f} dA = \int_{s}^{f} -kxdx = -\frac{k}{2}(x_2^2 - x_1^2) = \frac{k}{2}(x_1^2 - x_2^2). \qquad (1.60a)$$

Here *x* is the current value of the coordinate fixing the elongation of the spring *dx*. When the spring returns to its initial state, the elastic force *F* = – *kx* performs work

$$A_{21} = \int_{f}^{s} -kxdx = -\frac{k}{2}(x_1^2 - x_2^2) = E_{P_2} - E_{P_1}. \tag{1.60b}$$

Let's take the value $x_1 = 0$ as the origin of the coordinate, which corresponds to the position of the end of an undeformed spring. Then $E_P = \frac{kx^2}{2}$ – is the potential energy of elastic deformation (i.e., the energy stored in a certain volume of the deformed spring-body). The density of energy of deformation, i.e. the energy stored in a unit volume of an elastic deformed body, is defined by the relation

$$u = \frac{\varepsilon^2 E}{2}, \tag{1.61}$$

where ε^2 is the relative elongation of the spring, E – is the Young's modulus of compression (tension).

1.5.1.2. Kinetic energy

Consider again the simplest case (Fig. 1.40).

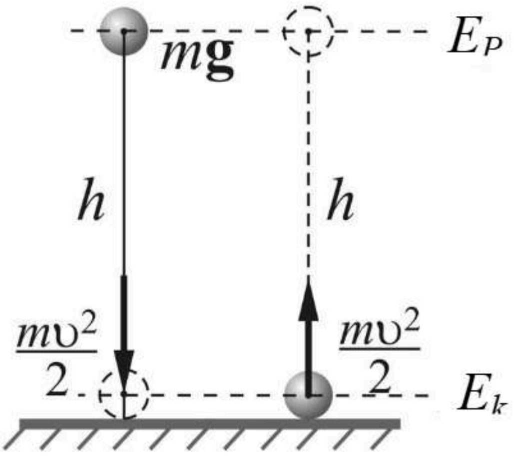


Figure 1.40. To the illustration of the potential energy transformation into kinetic energy and vice versa

A body of mass m falls freely under the action of gravity from height h, where its potential energy was equal to $E_P = mgh$, with acceleration of free fall g. Having flown distance $h = \frac{gt^2}{2}$, by the moment of contact with a horizontal surface the body has gained velocity $\upsilon = gt$ (t is time of falling; initial velocity υ_0 is assumed to be 0), see formula (1.4a). It follows that

$$mgh = m\frac{g^2t^2}{2} = \frac{m\upsilon^2}{2} = E_k. \qquad (1.62)$$

The value $\frac{m\upsilon^2}{2} = E_k$ is called kinetic energy. That is, kinetic energy is the energy a body possesses due to its motion. Thus, all of the potential energy has changed to kinetic energy. This kinetic energy will either do destructive work when it hits an obstacle, or it will do positive work to move the other body it hits (this is how piles are driven into the ground). If the impact with the surface occurs without any loss of energy (i.e., absolutely elastic), the kinetic energy of the falling body is completely transformed into the potential energy of the elastic deformation. Further, due to the potential energy of the elastic deformation, the body will gain initial velocity (kinetic energy at the moment of rebound), sufficient for it to rise to the same height h from which it fell. This is the model case, in which the interconversion of potential energy into kinetic energy and vice versa takes place. Naturally, in an arbitrary phase of body motion, the total energy $E_k + E_P$ remains unchanged (we assume that there is no energy loss, i.e. this interconversion of energy occurs in a closed system).

Now let us consider the general case. Let us use Newton's second law of motion for a material point

$$m\boldsymbol{a} = \mathbf{F} \text{ or in the form } m\frac{d\upsilon}{dt} = \mathbf{F}. \qquad (1.63)$$

Multiplying the left and right parts by the infinitesimal displacement *dr*, we obtain:

$$m\frac{d\upsilon}{dt}\cdot d\mathbf{r} = \mathbf{F}\cdot d\mathbf{r}. \tag{1.63a}$$

But $\frac{d\mathbf{r}}{dt} = \boldsymbol{\upsilon}$, then

$$m\boldsymbol{\upsilon}d\boldsymbol{\upsilon} = \frac{md(\upsilon\upsilon)}{2} = \frac{md\upsilon^2}{2} = d\left(\frac{m\upsilon^2}{2}\right) = \mathbf{F}d\mathbf{r}. \tag{1.63b}$$

It follows from this expression that if the force $\mathbf{F} = 0$, i.e. the system is closed, then the kinetic energy is conserved (kinetic energy is the integral of motion in an isolated system!). In the presence of force $\mathbf{F}$ kinetic energy is expended to do work $dA = \mathbf{F}d\mathbf{r}$. Let's integrate equation 1.63b and obtain

$$\int_1^2 \frac{dm\upsilon^2}{2} = \frac{m\upsilon_2^2}{2} - \frac{m\upsilon_1^2}{2} = E_{k_2} - E_{k_1} = \int_{\mathbf{1}}^{\mathbf{2}} \mathbf{F}d\mathbf{r} = A_{12}. \tag{1.64}$$

That is, the increase in kinetic energy is due to the work of external forces. On the other hand, the work in the potential field, as already noted, is due to the loss of potential energy:

$$A_{12} = E_{k_2} - E_{k_1}, \qquad\qquad A_{21} = E_{P_2} - E_{P_1}.$$

Since $A_{12} = -A_{21}$, we can write:

$$E_{P_1} - E_{P_2} = E_{k_2} - E_{k_1} \quad \text{or} \quad E_{P_1} + E_{k_1} = E_{P_2} + E_{k_2} = E, \tag{1.65}$$

i.e. the total energy is conserved. This statement is the law of conservation of mechanical energy: in a closed (isolated) system, total mechanical energy is conserved (i.e., in a given system, total energy is the integral of motion).

Obviously, if the system is unclosed, its total energy changes by an amount equal to the work of external forces:

$$\Delta E = A_{ext.forces} \cdot \tag{1.65a}$$

1.5.2. Dynamics of rotational motion

A real body moves as a material point (particle) only if the point of application of force O' and the center of inertia (gravity, masses) are on the same line, coinciding with the direction of force ***F*** (Fig. 1.41 a).

Fig. 1.41. Dependence of body motion on direction and point of force application. a) – central strike – translational motion; b) – component $\mathbf{F_r}$ of force **F** determines rotational motion, while component $\mathbf{F_t}$ determines translational motion. As a result, the body will move along a complex trajectory

In all other cases, the force action results in a rotation around some instantaneous axis of rotation passing through the point O (the center of inertia of the body) moving in space (Fig. 1.41b), and the dynamics of motion are defined by more complex equations.

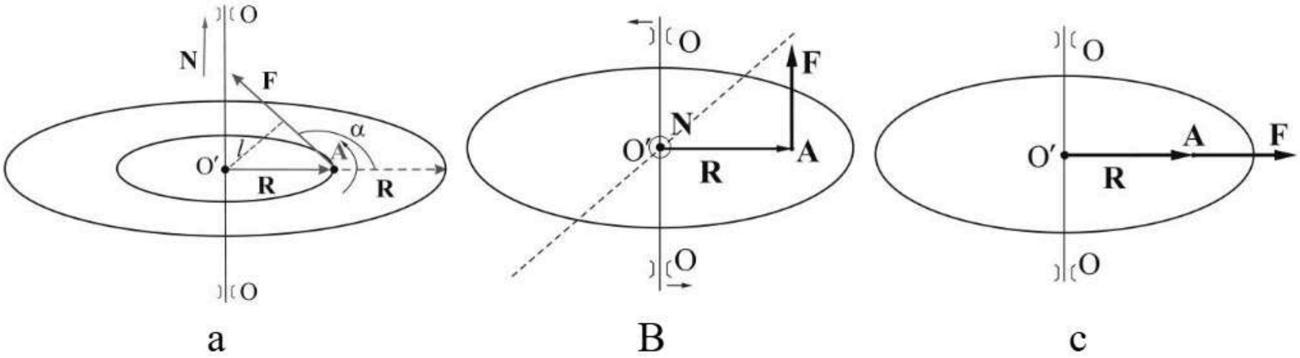


Figure 1.42. To the illustration of the effect of a multi-directional force on a body with a fixed axis of rotation. Schematically shown is a disk attached to the axis of OO. The point of force **F** is located at a distance **R** from the rotational center O'. Three special cases are presented: a) – the force **F** is located in the plane of the disk and leads to the rotation of the disk relative to the axis OO; b) – the force **F** acts ⊥ in the plane of rotation; c) – the force F is directed along R (explanation in the text)

1.5.2.1. Momentum of force (torque)

When describing rotational motion, to characterize the action of force **F**, a quantity called the momentum of force **N** is introduced as

$$\mathbf{N} = [\mathbf{RF}], \tag{1.66a}$$

whose numerical value (modulus) is equal to

$$N = F \cdot l, \tag{1.66b}$$

where l is the arm of the acting force **F** equal to the shortest distance from the center of rotation O to the direction of force action, i.e. $l = R\text{Sin}\alpha$ (Fig. 1.42a). The direction of the torque vector N is determined by the right-hand rule.

In Fig. 1.42a, the vector of the acting force is located in a plane perpendicular to the axis of rotation. If the force is directed at an angle to this plane, only that component of the force **F**, that lies in the plane ⊥ to the axis of rotation causes rotation. If the force **F** is ⊥ to the plane of rotation, it creates a torque **N**, that tends to rotate the axis of rotation, as shown in Fig. 1.42b (the icon ⊙ means that the vector N is directed toward the observer). If the force **F** lies in the plane of rotation, i.e., it is ⊥ to the axis of rotation, but it is directed along the vector **R** (Fig. 1.42c), no rotational moment arises, and the applied force tends to displace the axis of rotation along its direction of action, i.e., along **R**. In all cases, it is assumed that the axis of rotation is fixed (not free, the places of fixation are designated as «⟆ ⟅»).

1.5.2.2. The equation of dynamics of rotational motion of a body

Now let's get the equation of motion in the case of rotation of the body (rigid object) relative to a stationary axis (Fig. 1.43).

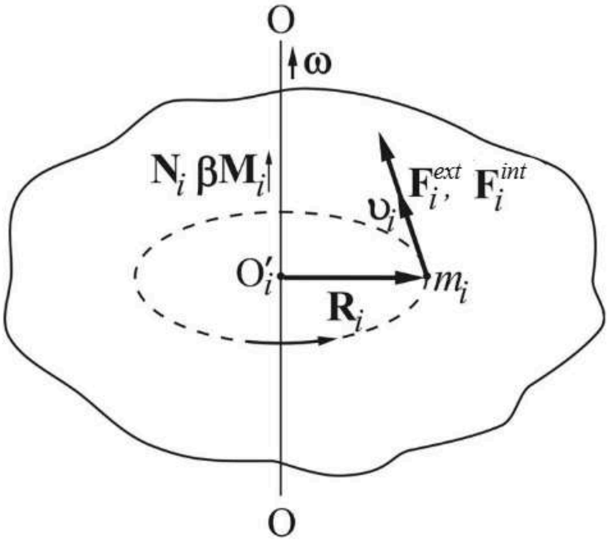


Fig. 1.43 Kinematic and dynamic quantities, characterizing the rotation of a solid body (rigid object) relative to the stationary axis of rotation O. The coordinate system is located at an arbitrary point of rotation O_i' of the element of mass m_i. The vectors of acting forces, radius of rotation $\mathbf{R}_i$ and linear velocity $\boldsymbol{\upsilon}_i$ lie in the rotation plane. The vectors of the momentum of forces **N**, angular momentum **M** are directed along the axis of rotation, as well as the vector of angular velocity $\boldsymbol{\omega}$ and angular acceleration $\boldsymbol{\beta}$ (in the case of increasing angular velocity)

Let us simplify the task as much as possible and consider only the components of forces directed along the tangent to the circle along which the elementary mass m_i (material point) moves with linear velocity $\boldsymbol{\upsilon}_i$, the vector of which also points along this tangent (the coordinate system is located at the rotation center of the elementary mass m_i). Let us write the equation of its motion in the form

$$\frac{dm_i\upsilon_i}{dt} = \mathbf{F}_i^{ext} + \mathbf{F}_{ij}^{int}, \tag{1.67}$$

where $\mathbf{F}_{ij}^{ext}$ is the internal forces acting on the material point (forces of interaction between the elements of mass m_i and m_j), $\mathbf{F}_i^{ext}2$ is the external force. Since $\boldsymbol{\omega}_i = \boldsymbol{\omega}$ (all mass elements rotate with the same angular velocity $\boldsymbol{\omega}$), and $\boldsymbol{\upsilon}_i = [\boldsymbol{\omega}\mathbf{R}_i]$ (see Equation 1.7), Equation 1.67 can be rewritten as

$$\frac{dm_i\upsilon_i}{dt} = m_i\left[\frac{d\boldsymbol{\omega}}{dt}, \mathbf{R}_i\right] = m_i[\boldsymbol{\beta}\mathbf{R}_i] = \mathbf{F}_i^{ext} + \mathbf{F}_{ij}^{int}, \tag{1.67a}$$

where $\boldsymbol{\beta}$ is the angular acceleration $\frac{d\boldsymbol{\omega}}{dt}$.

Let us multiply both parts of equation (1.67a) by $\mathbf{R}_i$:

$$m_i\big[\mathbf{R}_i[\boldsymbol{\beta}\mathbf{R}_i]\big] = [\mathbf{R}_i\mathbf{F}_i^{ext}] + \big[\mathbf{R}_i\mathbf{F}_{ij}^{int}\big]. \tag{1.67b}$$

According to the rule of triple vector, the product of vectors

$$\big[\mathbf{R}_i[\boldsymbol{\beta}\mathbf{R}_i]\big] = \boldsymbol{\beta}(\mathbf{R}_i\mathbf{R}_i) + \mathbf{R}_i(\mathbf{R}_i\boldsymbol{\beta}) = \boldsymbol{\beta}{R_i}^2,$$

since the scalar product of vectors $(\mathbf{R}_i\mathbf{R}_i) = R_i^2$, and the scalar product of vectors $(\mathbf{R}_i\boldsymbol{\beta}) = 0$. Thus, equation (1.67b) is transformed to

$$\boldsymbol{\beta}m_iR_i^2 = [\mathbf{R}_i\mathbf{F}_i^{ext}] + \big[\mathbf{R}_i\mathbf{F}_{ij}^{int}\big]. \tag{1.67c}$$

Summing over all indices *i* and *j*, we obtain

$$\boldsymbol{\beta}\sum m_i R_i^2 = [\mathbf{R}_i\mathbf{F}_i^{ext}]. \tag{1.67d}$$

Since the sum of internal force moments is equal to zero (according to Newton's third law, the internal forces are equal in magnitude, but oppositely directed, hence force moments created by them are also equal in magnitude and act in opposite directions), the equation (1.67g) turns into the equation (1.68)

$$\mathrm{I}\boldsymbol{\beta} = \mathbf{N}. \tag{1.68}$$

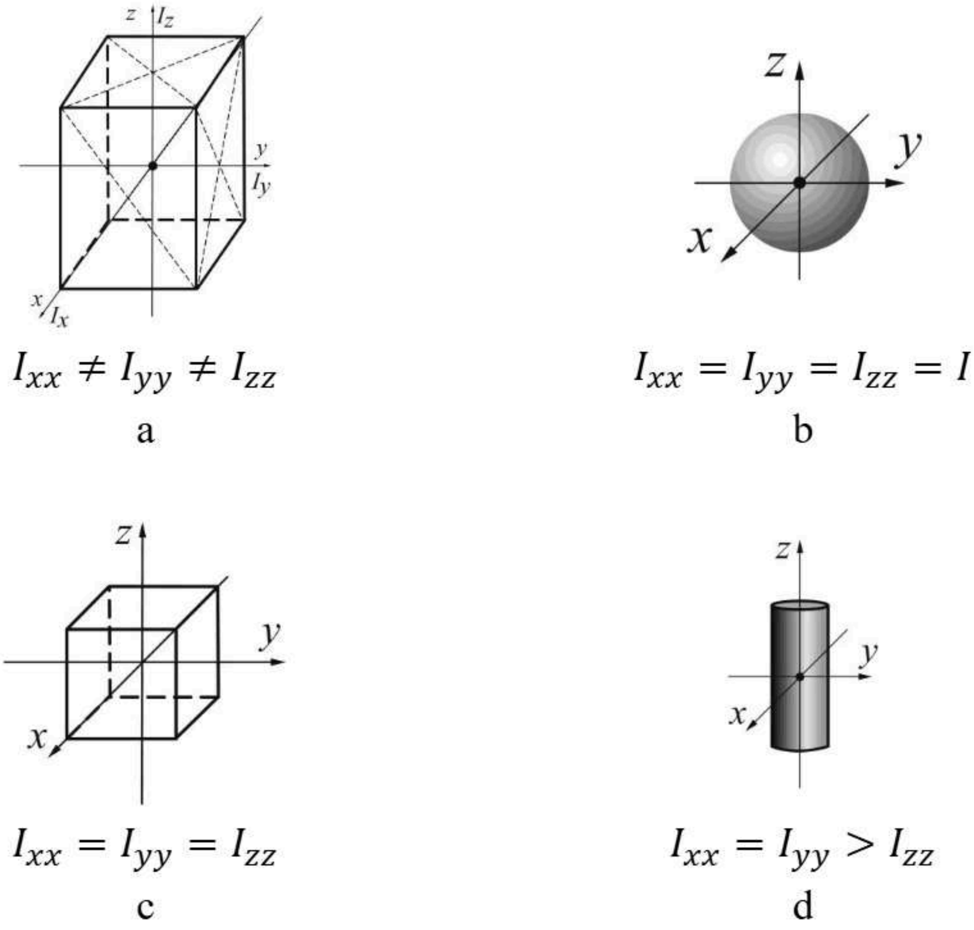
Here $I = \sum m_i R_i^2$ is the moment of inertia of the body relative to the chosen axis of rotation (analog to the mass of the body in translational motion of the body), $\sum[\mathbf{R}_i\mathbf{F}_i^{ext}] = \mathbf{N}$ is the total moment of external forces. It is obvious that angular velocity $\boldsymbol{\omega}$, angular acceleration $\boldsymbol{\beta}$ (with increasing angular velocity) and momentum of forces **N** have the same directions, as presented in Fig. 1.43.

The equation (1.68) is the equation of motion for a rotating solid body (i.e. analog of Newton's 2nd law for translational motion)[*].

Obviously, the values of moments of inertia depend on the shape of the body (symmetry) and the choice of the axis of rotation. For any body there are three mutually perpendicular axes passing through the center of inertia (mass) of the body, which can serve as free axes of inertia, i.e., such axes of rotation that retain their position in space in the absence of external influence on them. Such axes of inertia are called principal axes of inertia (shown in Fig. 1.44). The moments of inertia relative to the principal axes of rotation are called principal moments of inertia.

---

[*] In the rigorous derivation of equation (1.68), it is assumed that the position of an element of mass is determined by the radius-vector $\mathbf{r}_i$, measured from the general center of coordinates located on the axis of rotation. Subsequent mathematical operations on vectors lead to the same results.



$I_{xx} \neq I_{yy} \neq I_{zz}$

a

$I_{xx} = I_{yy} = I_{zz} = I$

b

$I_{xx} = I_{yy} = I_{zz}$

c

$I_{xx} = I_{yy} > I_{zz}$

d

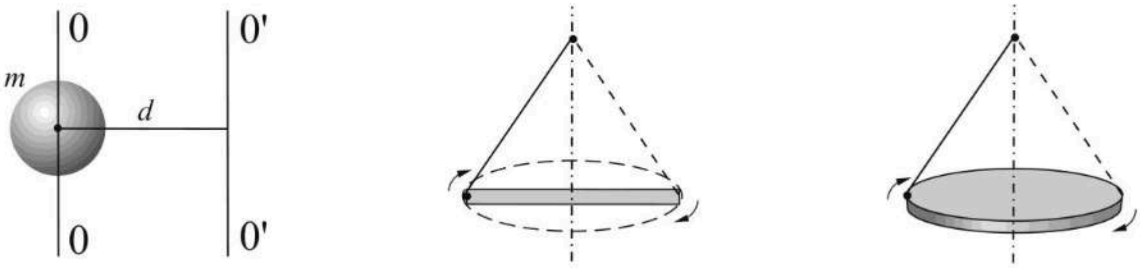
Fig. 1.44. The moment of inertia of homogeneous bodies of different symmetries: a) – parallelepiped; b) – sphere (infinite number of rotation axes with equal values of the moment of inertia); c) – cube (cubic symmetry); d) – cylinder (cylindrical symmetry): main value of moment of inertia relative to longitudinal rotation axis is less than equal main values of moment of inertia relative to infinite number of main rotation axes perpendicular to longitudinal rotation axes

The moment of inertia of a body with respect to an arbitrary axis O′O′ is equal to the sum of the moments of inertia of this body with respect to the axis OO passing through the center of mass of this body and the product of the body mass *m* by the square of the distance *d* between the axes O'O' and OO (see Fig. 1.45a), i.e.

$$I_{O'O'} = I_{OO} + md^2. \qquad (1.69)$$

This relation is known as Huygens-Steiner's theorem (Theorem of Parallel Axes).

The nature of rotation with respect to free axes of rotation (i.e., preserving the directions of the axis of rotation in space in the absence of support) and non-free axes of rotation (for example, one end of the axis of rotation is fixed) differs significantly. In the first case, rotations around rotational axes, relative to which the main values of the moment of inertia have the maximum or minimum values, will be stable. In the second case, the rotations will be stable around the major axis, which corresponds to the maximum value of the moment of force (for example, a rod or disk suspended on a thread will rotate as shown in Fig. 1.45b, c).



a B c

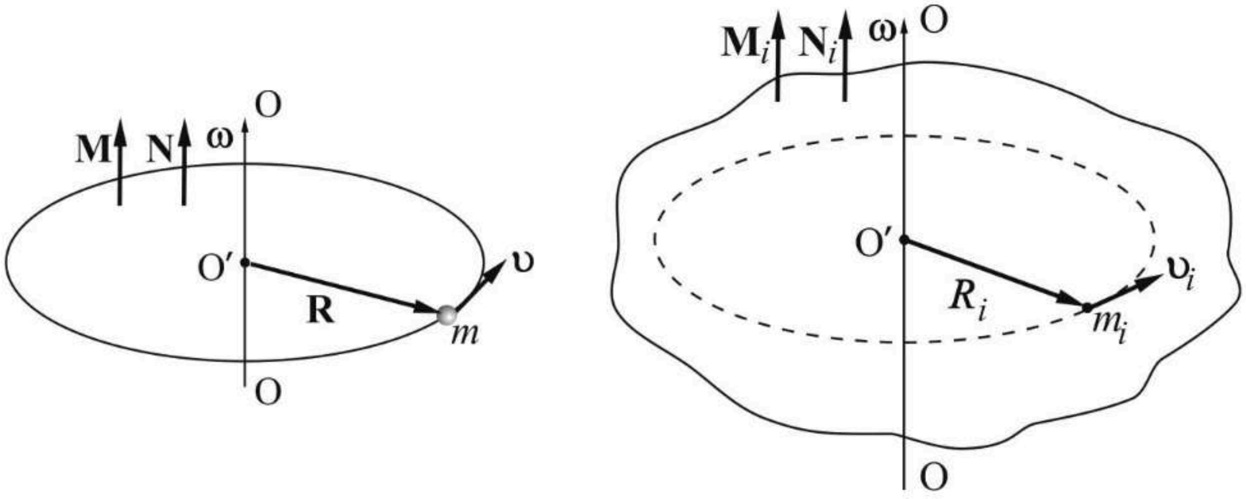
Fig. 1.45. a) – The explanation of Huygens-Steiner's theorem, b) – and c) – the illustration of a steady rotation with a fixed axis of rotation

1.5.2.3. Law of angular momentum conservation. The equation of moments

The momentum of a particle can be considered as a dynamic characteristic of a material point (particle). Its analogue in the rotation of a material point (particle) is the momentum vector **M** (angular momentum)

$$\mathbf{M} = m[\mathbf{R}\upsilon]. \qquad (1.70)$$

Its modulus is equal to $m\upsilon R$, and its direction is also determined by the Right-hand rule (see Fig. 1.46a).



a b

Fig. 1.46. The illustration of rotation with respect to the fixed axis of rotation of OO: a) – particle, b) – body (rigid object)

For an elementary mass m_i the numerical value of the angular momentum is $M_i = m_i R_i \upsilon_i$. Since $\upsilon_i = \omega R_i$, then $M_i = \omega m_i R_i^2$. Summing over all absolute values of the angular momentum M_i gives the numerical (absolute) value of the angular momentum of the rotating body M, expressed through the moment of inertia I and angular velocity ω:

$$M = \sum M_i = \sum \omega m_i R_i^2 = \omega \sum m_i R_i^2 = I\omega \quad (1.71)$$

or in vector form:

$$\mathbf{M} = I\boldsymbol{\omega}. \quad (1.71a)$$

The change of angular velocity of rotation, and correspondingly angular momentum, is possible only under external influence, i.e. under the influence of external force momentum **N** (*torque*), and this relationship is established by the relation (with a stationary axis of rotation)

$$\frac{d\mathbf{M}}{dt} = I\frac{d\boldsymbol{\omega}}{dt} = I\boldsymbol{\beta} = \mathbf{N}. \quad (1.72)$$

This is the equation of moments, or the law of angular momentum conservation: the angular momentum of the body $\mathbf{M}$ in an isolated system, i.e., at $\mathbf{N} = 0$, does not change.

The kinetic energy of a rotating body is defined as follows:

$$E_\text{k} = \sum \frac{m_i \upsilon_i^2}{2} = \sum \frac{m_i \omega^2 R_i^2}{2} = \frac{I\omega^2}{2} \tag{1.73}$$

and the work dA, which must be expended to rotate the body by an infinitesimal angle, is determined by the relation

$$dA = N d\varphi. \tag{1.74}$$

Here, N is the numerical value of the momentum of external forces relative to the axis of rotation.

Thus, three basic laws of conservation (integrals of motion) determine the motion of a body in an isolated system: the laws of conservation of momentum, angular momentum and total mechanical energy.

In conclusion, here is a table of kinematic and dynamic quantities and equations characterizing translational and rotational motion of a body (Table 1.1).

Table 1.1

Comparison between the formulas of the rotational motion mechanics (relative to an arbitrary axis of rotation passing through the center of inertia (masses) of the body) with analogous formulas of translational mechanics (mechanics of a material point)

Translational motion	**Rotational motion**
$\boldsymbol{\upsilon}$ – linear velocity	$\boldsymbol{\omega}$ – angular velocity
$\boldsymbol{a} = \frac{d\upsilon}{dt} = \dot{\boldsymbol{\upsilon}}$ – linear acceleration	$\boldsymbol{\beta} = \frac{d\omega}{dt}$ – angular acceleration
m – mass	I – moment of inertia
$\mathbf{P} = m\boldsymbol{\upsilon}$ – momentum	$\mathbf{M} = I\,\boldsymbol{\omega}$ – angular momentum
$\mathbf{F}$ – force	$\mathbf{N}$ [or N] – momentum of force (torque)

Continuation Table 1.1

Translational motion	Rotational motion
$\mathbf{F} = \frac{d\mathbf{p}}{dt}$	$\mathbf{N} = \frac{d\mathbf{M}}{dt}$
$\boldsymbol{a}m = \mathbf{F}$	$\boldsymbol{\beta} I = \mathbf{N}$
$E_{\boldsymbol{K}} = \frac{1}{2} m\upsilon^2$	$E_{\boldsymbol{K}} = \frac{1}{2} I\omega^2$ (for the stationary axis of rotation)
$dA = F_s ds (\equiv F_l dl)$	$dA = Nd\varphi$

1.5.3. Muscle biomechanics

Human reflex movements follow the laws of dynamics even in sleep. This is particularly well illustrated by rising out of bed or out of a chair from a "sitting" position. The biomechanics of this process are very complicated, so we will limit ourselves to its schematic representation (Fig. 1.47) as an illustration of the role of moments of gravity arising when getting up from a chair. Let's consider 4 arbitrary axes of rotation, respectively located: in the ankle joint – O^1 (it is motionless in this approximation); in the knee joint – O^2, hip joint – O^3 and cervical joint – O^4. Let us also assume that the body moves only in the plane of the figure. Let us distinguish four basic elements of mass with conditional centers of inertia (masses): tibia (O_t), femur (O_f), torso (O_t, we assume that arms are immobile relative to torso) and head (O_h.). The ideal position of the body, as a set of mechanical parts, is on a chair with rectangular geometry (Fig. 1.47a) and standing position (Fig. 1.47c). In this case, the centers of masses are on the same vertical line passing through the fulcrum O^1. In both positions, the moments of gravity of the body fragments are balanced.

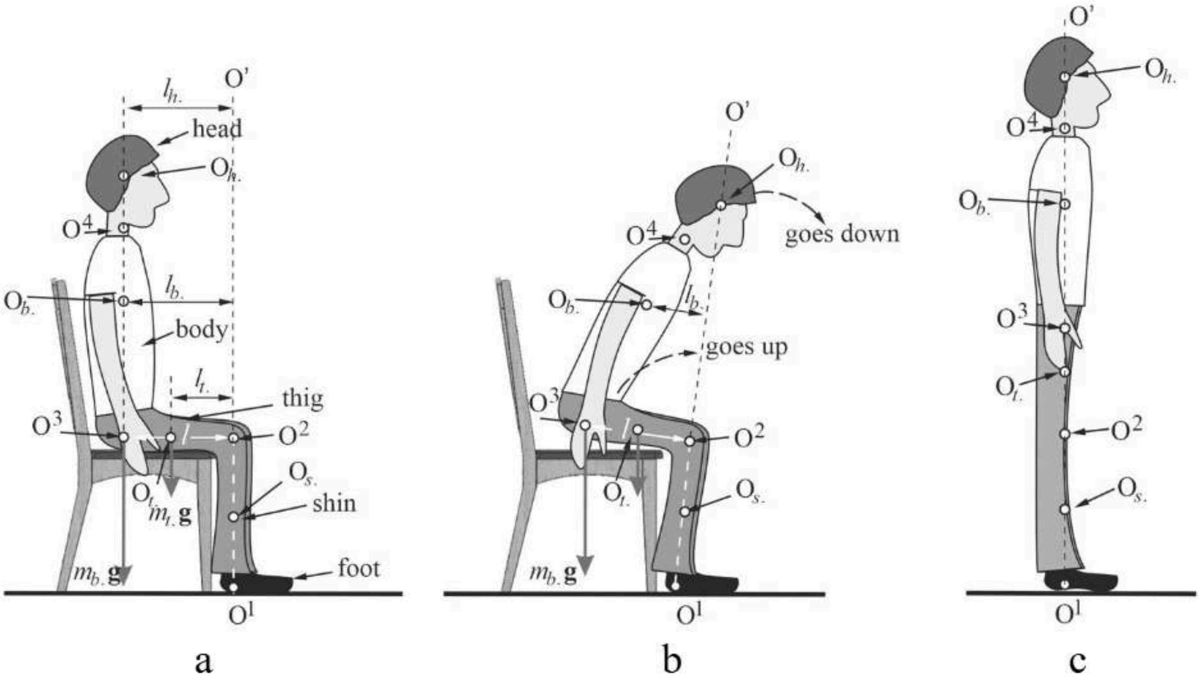


Fig. 1.47. Schematic representation of the transition from "sitting" to "standing" position: a) – mechanical model of stable equilibrium in sitting position; b) and c) – body positions in intermediate and final states

But the human is not a mechanical device, and both positions are not physiological: to hold the body in these positions requires the efforts of the entire human muscular corset. This is a completely different topic, only partially touched upon in this textbook. Let us formulate the question in this form: how to get from a "sitting" position to a "standing" position. There are many instructions and recommendations on the Internet (unfortunately, without graphics), but they all boil down to the following. A person moves slightly forward on the chair, then bends forward the trunk and head, reducing the arms of gravity forces acting on the body parts: head – l_h, trunk – l_t, hips – l_h (the tibia is not shown because its contribution is the lowest). Then the muscles pull the thighs away from the chair, and gradually the trunk is lifted and straightened. Muscular efforts to overcome the forces of gravity of body fragments are minimized along the way.

Skeletal muscles perform mechanical work, with the source of energy being biochemical processes in the body. The work of any muscle in the human musculoskeletal apparatus is based on the muscle's ability to contract. At the moment of muscle contraction, the muscle shortens, and the two points of attachment to the bones become closer to each other. In this case the moving point (*insertion* – the place of muscle attachment) begins to approach the initial stationary point of the muscle attachment, which is accompanied by the relative movement of body parts. Thus, by applying different degrees of muscle effort a certain mechanical work is performed (lifting a weight, moving a limb of the body with a load, etc.). The size of a muscle contraction is determined by the length of a muscle fiber. Joints of bones and interaction with muscle groups are arranged in the form of a mechanical lever, which allows performing the simplest work of lifting and moving objects.

We know from everyday practice that the farther from the fulcrum (instant axis of rotation) the force is applied, i.e., the larger the lever, the less effort can be used to perform the work. So it is in biomechanics – the further from the fulcrum a muscle is attached, the more effectively its force will be used. P.F. Lesgaft[*] classified the muscles into strong muscles, which have an attachment point farther away from the fulcrum, and fast or agile muscles, which have an attachment point near the fulcrum.

The movement around one fulcrum requires two muscles on opposite sides of the fulcrum. Movement directions in biomechanics have their own definitions: flexion and extension, adduction and abduction, horizontal adduction and horizontal abduction, medial rotation and lateral rotation.

The muscle which during contraction generates a moment of force and takes the brunt of the load is called an *agonist*. Each contraction

---

[*] ***P.F. Lesgaft*** (September 20, 1837 – November 22, 1909) was a biologist, anatomist, anthropologist, physician and teacher. He is best known as the creator of theoretical functional anatomy and a scientific system of physical education. He was a progressive public figure. In 1869-1871 he worked as an Extraordinary Professor of Kazan Imperial University.

of an agonist muscle causes a complete relaxation of the opposing *antagonist* muscle. If we are doing an elbow bend the agonist will be biceps-elbow flexor muscle (biceps brachii muscle), and the antagonist at this moment will be triceps-elbow extensor muscle (triceps brachii muscle). After the end of the movement, the momentum of forces created by these muscles will be balanced and the muscles themselves will be in a slightly stretched state. This phenomenon is called muscle tone. Muscles that help the agonist muscle perform the movement and act in the same direction as the agonist muscle, but have a lesser load and lesser degree of contraction are called *synergists*[*]. The muscles that provide stability and balance to a particular joint during movement are called *fixators*. In addition to the fixators, a significant role in the training process is played by the *stabilizer* muscles, which work as elements of body balance when the center of gravity shifts and the overall force load increases. In addition, the stabilizer muscles are involved in everyday human life in ensuring that the body parts are in a balanced position relative to each other regardless of the force load.

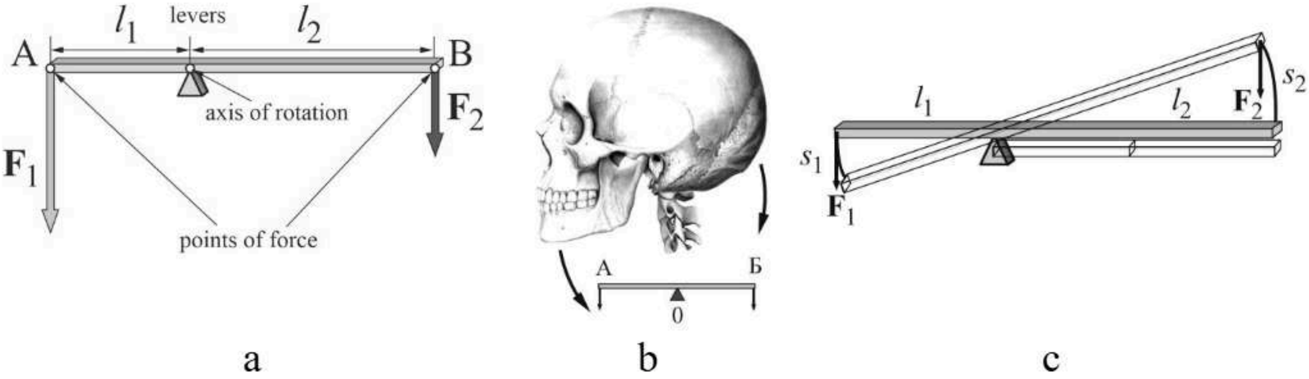
At any given moment of movement, the bones form mechanical levers, following the muscular forces. Biomechanics distinguishes 2 types of biomechanical levers – a lever of type 1, where the force application points are located on opposite sides of the axis, and a lever of type 2, where the force application points are located on the same side of the axis, but at different distances from it.

In biomechanics, a lever of the first kind is called an "equilibrium lever" (Fig. 1.48a). Since the fulcrum is located between the two points of force application, the lever is also called "double-armed" (the condition of equilibrium is equality in magnitudes of momentums of forces acting

---

[*] Synergists (from Greek συνεργοσ (synergos) – acting together) in anatomy and physiology are muscles (or groups of muscles) acting jointly and functionally homogeneously; they take part in the same movement. For example, in mammalian animals and humans, such synergists as the external intercostal and intercartilaginous muscles, as well as the diaphragm muscles, contracting simultaneously, provide the act of breathing. The phenomenon of synergy is provided by the regulating activity of the central nervous system.

in opposite directions, $F_1l_1 = F_2l_2$). Such a lever takes place, for example, in the connection of the spine and cranium (Fig. 1.48b). If the torque of the force acting on the occipital part of the skull is equal to the torque of gravity acting on the rear part, the differently directed action of the force moments is not felt and the muscles are not strained. This circumstance is used in everyday life – for example, in the East, girls carry water jugs on their heads. The so-called "golden rule of mechanics" (Fig. 1.48c) can be demonstrated by the example of the first kind of lever. A weightless board is placed on a support: different in magnitude forces F_1 and F_2 act on the differently distant ends of the board (l_1 and l_2).



a b c

Fig. 1.48. a) –, b) – Levers of the first kind; c) – the illustration of the "golden rule of mechanics

In order to lift the load located at a distance l_2 (let's denote the weight of the load as counteracting force $\mathbf{F}_2$), it is necessary to apply force $\mathbf{F_1}$ at the opposite end of the lever at a point located at a distance l_1 from the support, which is determined from the condition $F_1l_1 > F_2l_2$. After completing the ascent to a certain height, points 1 and 2 will move by distances s_1 and s_2 , respectively. Obviously, the work (see section 1.5) of lifting the load (overcoming resistance of force $\mathbf{F}_2$) is equal to the work done by force $\mathbf{F_1}$. It follows that $F_2s_2 = F_1s_1$. That is, if we win in strength, we lose in distance, and vice versa, if we lose in strength, we win in distance. This is the Golden Rule of Mechanics. The time of motion of both ends of the board is the same, hence the weight at the end of the long arm travels at a greater speed than the point of force at the end of the shorter lever. Therefore, the short lever is called the "speed lever," the long lever – "force lever". This rule is widely and very effectively used in a variety of situations. *The famous Archimedes quote: "Give me a fulcrum and I will move the earth," gave rise to the term "Archimedean lever," used to mean: propulsive force in general.*

A type 2 lever is shown in Fig. 1.49. In biomechanics it is subdivided into two kinds: "force lever" and "speed lever".

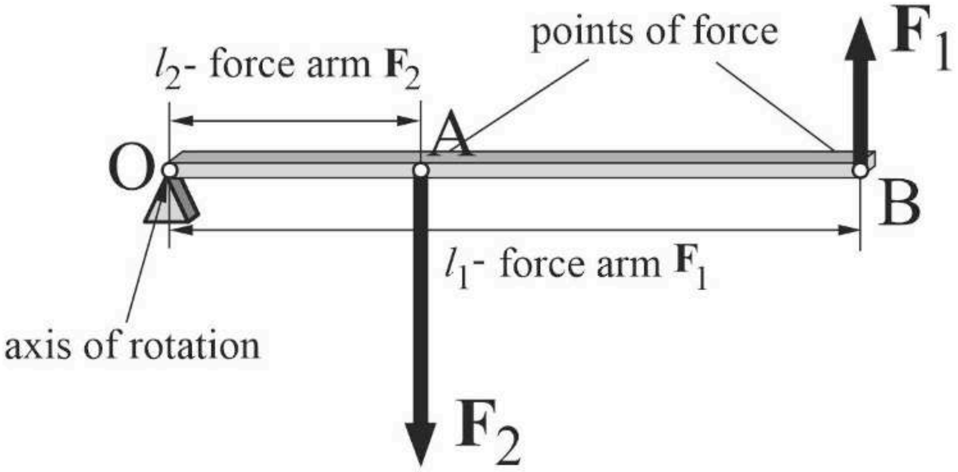


Fig. 1.49. The lever of type 2

The name and action of these levers depend on where the load is applied, but for both types of levers the point of force application and the point of resistance application are on the same side of the axis of rotation, so both levers are "single-arm". A *lever of force* is formed if the muscle force application arm is longer than the gravity (resistance) force application arm. As an example, we can demonstrate the human foot (this type of leverage is shown schematically in Fig. 1.50a).

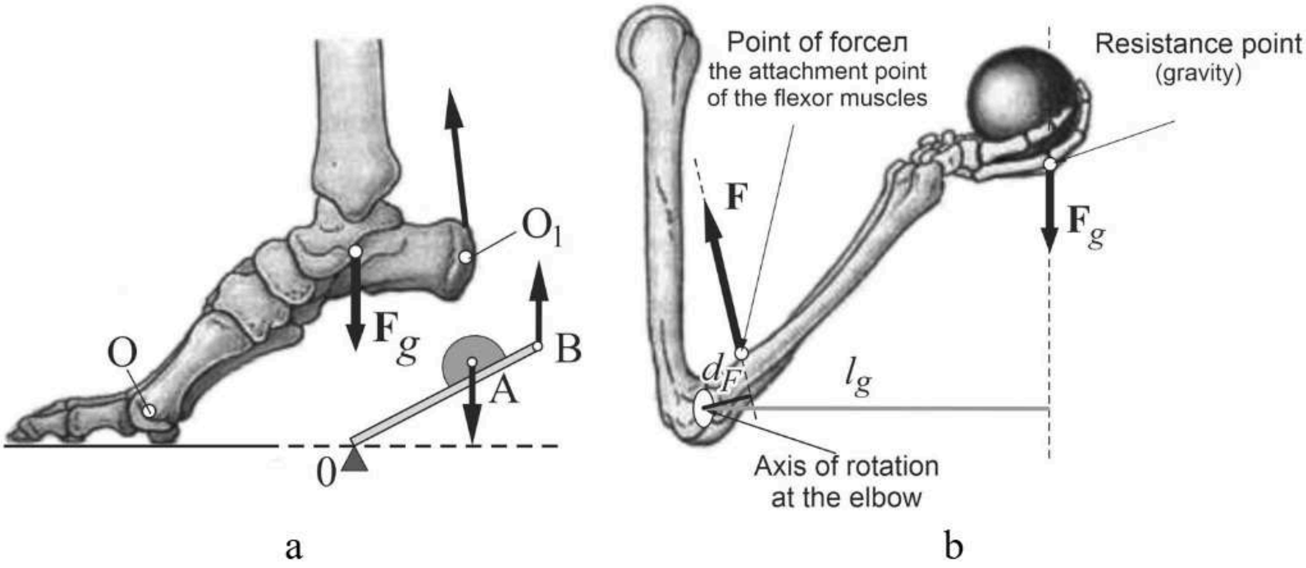


Fig. 1.50. a) – Force lever, b) – speed lever

The "axes of rotation" here are the heads of the metatarsal bones (conditional point O), the heel bone (point O1) serves as the point of application of the ankle joint muscle force, and the body gravity force $\mathbf{F}_g$ forms the resistance force in the ankle joint. There is a gain in force due to the longer arm of muscular force and a loss in speed. As you can see from Fig. 1.50a, the shoulder of the counterforce (gravity) is shorter than the shoulder of the muscle force. An example of a speed lever is the work of the flexor muscles at the elbow joint (shown schematically in Fig. 1.50b). The biceps are attached near the axis of "rotation" (elbow joint), and with such a short arm you need extra force from the flexor muscle. Here there is a gain in speed and stroke, but a loss in strength. We can conclude that the closer the muscle is attached to the place of support, the shorter the arm of the lever will be and the more significant will be the loss in strength (short-armed, long-armed are common concepts). When two bone pairs are connected, a biokinetic pair is formed, the nature of movement in which is determined by the structure of the bone articulation (joint) and the work of muscles, tendons, and ligaments. Mobility in a joint can depend on numerous factors. First of all, it is anatomical – the shape of the bones largely determines the direction and scope of the bone in the joint. Next, the central-nervous regulation of muscle tone, as well as the tension of the antagonist muscles. Namely, from the ability to relax the muscles being stretched and to tense the muscles being contracted. In other words, the degree of intermuscular coordination. Naturally, joint mobility depends on gender and age.

In order to optimally and correctly take the initial position for performing physical actions, it is necessary to be guided by the properties of the first and second type levers. If we change the position of the limb or torso, the length of the arm of the limb or torso will change in a certain way. Obviously, in any case, the starting position is always chosen in such a way that the initial period of training is accompanied by less stressful positions of the limbs and trunk. Later on, depending on the physical condition and form of the trainee (as well as on the type of physical activity, such as a sport), you can gradually increase the length of the lever arm, to increase the impact on a particular muscle group. Increasing the force of the counter force at the same time as lengthening the lever arm, in turn, puts even more emphasis on strengthening the strength of a particular muscle group or single muscle.

In order to perform a technically correct movement during physical activity, it is necessary to know how the joint that connects the active muscle group works.

The human musculoskeletal system consists of various bones connected to each other through joints, the structure of which determines the ability to perform various movements. Types and names of joints can be found in the relevant anatomy literature, and types and descriptions of rotational axes and anatomical planes that determine the ability to perform a particular movement can be found in kinesiology textbooks [3].

2.1. Pressure and compressibility

Let's introduce basic definitions necessary to describe the mechanics of liquid and gas. Fig. 2.1 shows a scheme of a simple experiment: a cylinder with gas (liquid) under the piston.

Under the influence of the piston, an elastic force arises that is numerically equal to the external force. The liquid or gas particles act on each other with a force equal to the compression force. This action is called pressure. Obviously, the pressure is uniformly distributed in all directions, which can be easily verified by observing liquid jets flowing under pressure through holes of the same diameter in the cylinder walls.

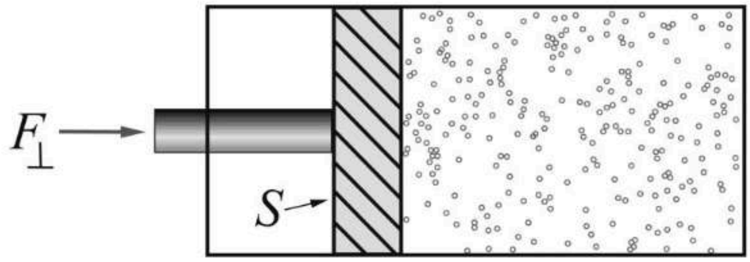


Fig. 2.1. Schematic diagram of an elementary apparatus for determining of gas (liquid) characteristics

A quantitative measure of pressure is the ratio of the numerical value of the force acting perpendicularly to the area of the selected volume to the size of that area

$$p = \frac{F_\perp}{S}. \qquad (2.1)$$

Pressure Units:
SI: 1 n/m^2 = 1 Pa (pascal).
CGS: 1 dyn/cm^2 = 1 bar.

1 Pa = 10^{-5} bar.
Off-system units: 1 bar = 10^5 Pa.
Technical atmosphere: 1 at = 1 kgf/cm^2 = 0.98 bar.
Physical atmosphere: 1 atm = 1.01 bar.
1 mm Hg = 1 tor; 760 mm Hg = 1 atm;
1 mm Hg = 13.6 mm a.c.; 1 mm a.c. = 9.8 Pa.
To determine the elastic characteristics, the notion of compressibility is introduced

$$\chi = -\frac{\Delta V}{V\Delta p}. \qquad (2.2)$$

The symbol "-" (minus sign) means that the volume decreases with increasing pressure, which corresponds to positive values for the compressibility coefficient $\chi > 0$. Let's estimate the value of pressure Δp, required for relative compression by 1% (i.e. $\frac{\Delta V}{V} = 10^{-2}$), for example for water, for which $\chi = 5\cdot10^{-10}$ m^2/N:

$$\Delta p = \frac{\Delta V}{V}\Big/\chi = \frac{10^{-2}}{5\cdot10^{-10}} = 0{,}2\cdot10^{8}\ \text{Pa} = 200\ \text{atm}.$$

For air, the compressibility under normal conditions is higher by an order of magnitude.

The pressure in a resting liquid (hydrostatic pressure), as we know from a secondary school physics course, depends on the height of the liquid (gas) column (Fig. 2.2)

$$\Delta p = \rho g h_1 - \rho g h_2 \quad , \qquad (2.3)$$

or

$$p_2 - p_1 = \rho g(h_2 - h_1), \quad p = p(h). \qquad (2.4)$$

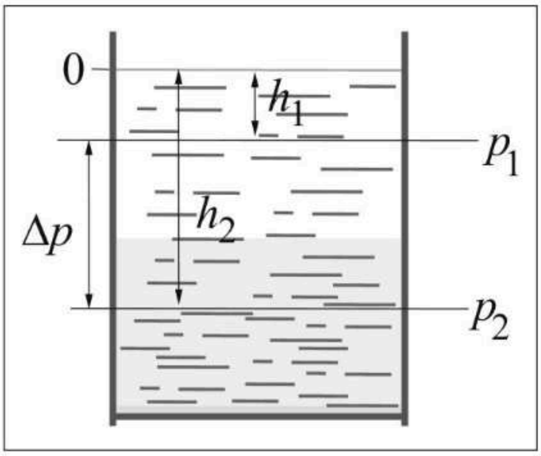


Fig. 2.2. Pressure in a resting liquid

2.2. Elements of hydro(aero)dynamics

Hydrodynamics is a branch of continuum mechanics, which studies the motion of incompressible fluids (gas) and their interaction with solid bodies. To a certain extent, the main provisions of fluid dynamics are applicable to the description of gas flow. In the context of this textbook, we will describe the motion of liquids. If necessary, differences in liquid and gas flows will be noted. The basic provisions of fluid dynamics are based on the jet continuity theorem, Bernoulli's equation and similarity criterion (Reynolds). In the theoretical model, a moving incompressible fluid is characterized by a field of velocity vectors of fluid particles. Flow motion is depicted by so-called current lines, tangents to which at each point coincide with directions of velocity vectors (Fig. 2.3).

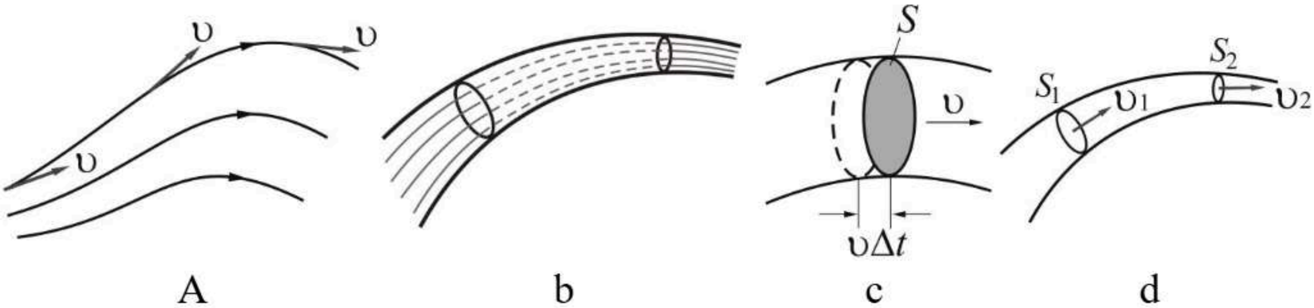


Fig. 2.3. a) – Current flow lines, b) – current flow tube; c) and d) – illustrations to the Bernoulli’s theorem

At steady (stationary) flow of liquid the velocities of particles at any point in the flow (current tube cross-section) do not change with time. Naturally, in this case current lines coincide with trajectories of liquid particles

Any part of the flow confined by current lines can be represented by a current tube, the cross-section S of which is so small that the flow velocity across the cross-section is the same (Fig. 2.3b). At layered (laminar) motion of liquid, the current lines do not break, do not cross and do not go beyond the current tube. Under such conditions, the same amount of liquid flows through any cross-section of the current tube (perpendicular to the current line) for the same time Δt (see Fig. 2.3c, d)

$$\Delta Q_1 = S_1 \upsilon_1 \Delta t = \mathit{\Delta} Q_2 = S_2 \upsilon_2 \Delta t, \qquad (2.5)$$

where $\mathit{\Delta} Q_1$ and $\mathit{\Delta} Q_2$ are the volumes of liquid flowing through sections S_1 and S_2 with velocities υ_1 and υ_2, respectively, equal throughout the section. It follows

$$S_1 \upsilon_1 = S_2 \upsilon_2. \qquad (2.5a)$$

This is the so-called Bernoulli’s theorem[*]. (The value $S\upsilon$ is sometimes called the volume velocity (volume flow rate) of the moving liquid).

Using this equality, we can obtain a complete dynamic flow characteristic in the form of Bernoulli's equation valid for any current tube (i.e., any cross-section, Fig. 2.4) for an ideal fluid (i.e., a fluid with no internal friction):

---

[*] In the general case, the continuity equation of the jet is written in the form: $\rho_1 S_1 \upsilon_1 = \rho_2 S_2 \upsilon_2$, where ρ_1 and ρ_2 are liquid or gas densities. But condition (2.5a) is fulfilled in a wide range of flow characteristics.

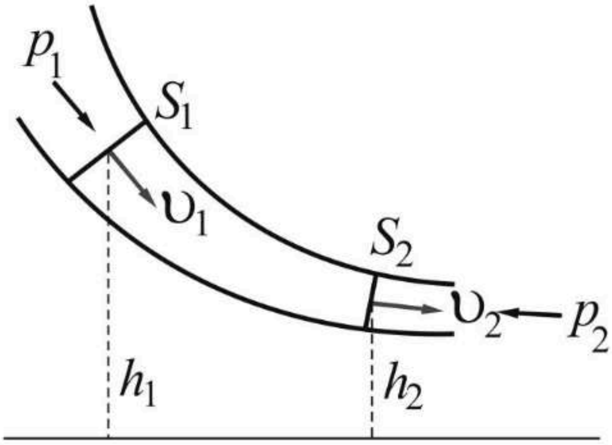


Figure 2.4. The illustration of the derivation of Bernoulli's equation

$$p_1 + \rho\upsilon_1^2/2 + \rho g h_1 = p_2 + \rho\upsilon_2^2/2 + \rho g h_2, \qquad (2.6)$$

where ρ – density of liquid, p_1 and p_2 – static pressure of liquid in two sections of the current tube; υ_1 and υ_2 velocity of liquid in these sections; $\rho\upsilon_1^2/2$ and $\rho\upsilon_2^2/2$ dynamic pressure of liquid in the same sections, respectively; h_1 and h_2 height of selected section above some level, taken as zero level; $\rho g h_1$ and $\rho g h_2$ values of hydrostatic (hydraulic, weight) pressure, created by weight of vertical liquid column in height h. This equation is valid for any cross-section of the current tube, so in any cross-section of the current tube the condition is fulfilled:

$$p + \frac{\rho\upsilon^2}{2} + \rho g h = \text{const.} \qquad (2.6a)$$

In the case when both sections are at the same height ($h_1 = h_2$),

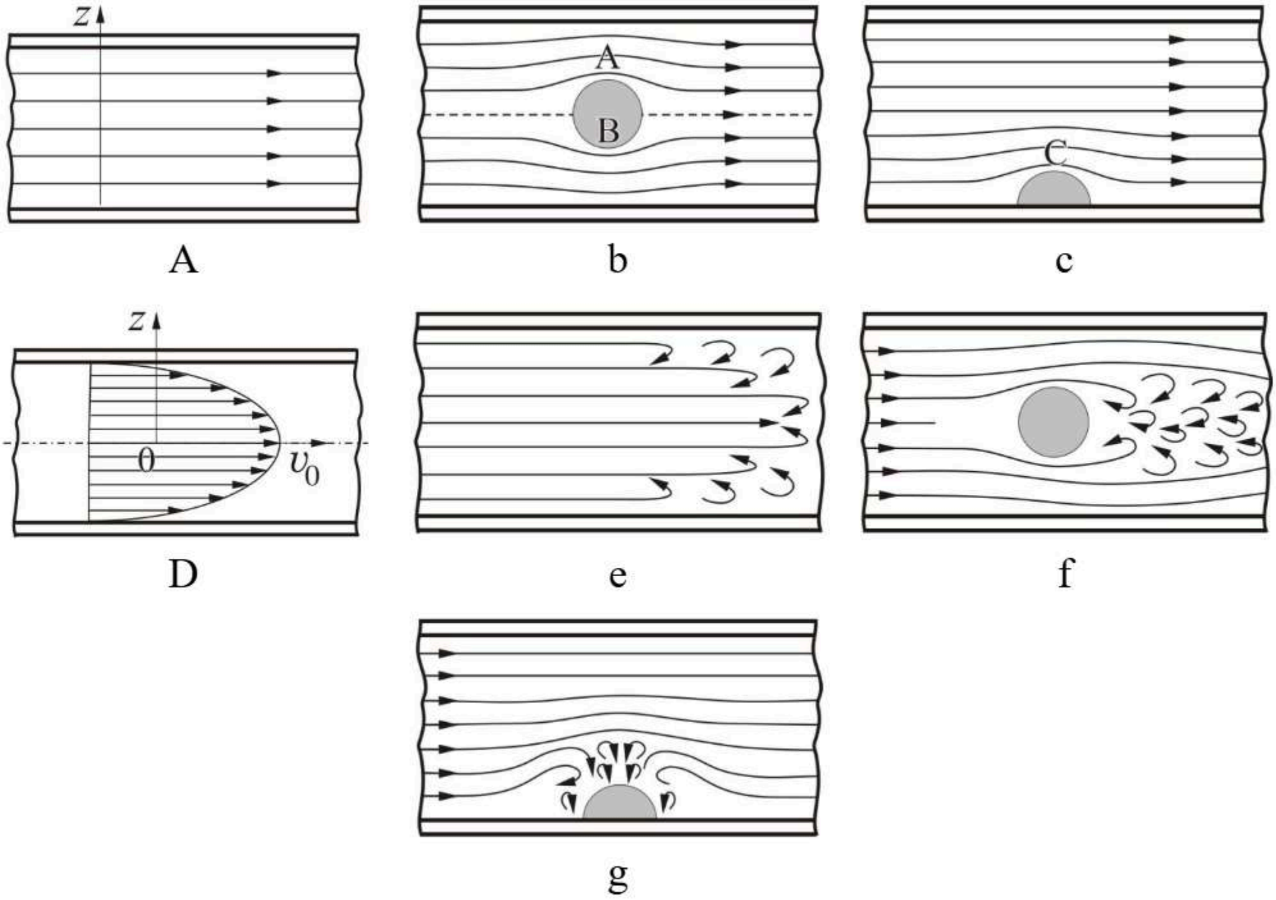
$$p + \frac{\rho\upsilon^2}{2} = \text{const.} \qquad (2.6b)$$

Static pressure p is the pressure that adjacent layers of liquid exert on each other. It can be measured with a manometer that moves with the liquid.

The dynamic pressure $\rho\upsilon^2/2$ is caused by the motion of the liquid. The hydrostatic pressure ρgh, as noted above, is created by the weight of a vertical column of liquid of height h. So, Bernoulli's equation is formulated as follows:

In the stationary flow of an ideal fluid, the total pressure equal to the sum of static, dynamic and hydrostatic pressures is the same in all cross-sections of the current tube. Bernoulli's equation is also applicable to the motion of a real fluid in a limited geometry, for example, in a tube under certain conditions.

Let's imagine a smooth tube in which the flow of a non-viscous fluid is moving. The current lines will be equidistant parallel lines, as shown in Fig. 2.5a, repeating the profile of the tube. The velocities of the flow particles are the same at any cross-section of the flow. If there is, for example, a ball in the middle of the flow or a streamlined bulge on the pipe in the path of the non-viscous fluid, the current lines will change as shown in Fig. 2.5b, c: the flow areas A, B and C will have higher velocities than the other areas. In real liquid due to forces of intermolecular interaction "liquid – tube surface" a boundary layer of liquid is formed (thin layers of liquid "stick" to the tube walls), which practically does not move inside the tube. But since there is also an intermolecular interaction between the liquid layers, there is a flow velocity gradient $\left(\frac{d\upsilon}{dz}\right)$ in the flow: the velocity in the middle of the flow is maximum, and near the pipe surface it is almost zero (Fig. 2.5d).



A b c

D e f

g

Fig. 2.5. Schematic representation of fluid flow: a) – non-viscous fluid; b) –, c) – non-viscous fluid in the pipe when flowing around obstacle, d) – Newtonian fluid in laminar flow, e) – with turbulence in pipe without obstacle; f) –, g) – with flow around obstacle

Thus, the viscous fluid flow moves, overcoming resistance – the force of viscous (internal friction), which is defined by Newton's formula (2.7)

$$F_{\text{fr}} = \eta^{*} \left|\frac{d\upsilon}{dz}\right| \text{S}. \qquad (2.7)$$

(hence the name – Newtonian fluid). Similarly, a boundary layer is formed on the surface of the ball (protrusion), creating a velocity gradient

---

[*] η – coefficient of internal friction (coefficient of viscosity, dynamic viscosity). The value η/ρ, where ρ is the density of the fluid, is called kinematic viscosity.

in the liquid flow, causing the so-called drag to the flow in the presence of an obstacle (or drag to a body moving in the flow). Under limited geometry, fluid motion is determined by Reynolds number.

The Reynolds number for a fluid flow characterized by dynamic viscosity η and density ρ, flowing at an average velocity $\langle \upsilon \rangle$ in long tubes with a diameter d, is defined by the relation

$$\mathrm{Re} = \frac{\rho \langle \upsilon \rangle d}{\eta}. \tag{2.8}$$

Critical Reynolds number for fluid flow in long tubes $Re_{\mathrm{cr}} = 2000 - 2300$. At lower values of Reynolds number fluid motion is laminar. At values $Re > Re_{\mathrm{cr}}$ turbulence may occur, and at $Re \gg Re_{\mathrm{cr}}$, fluid motion turns into turbulent (vortex) motion. Note that the jet continuity theorem and Bernoulli's equation are sufficiently well satisfied for real fluids, the viscosity of which is not very high. In this case, the correspondence of the model to the real physical system is established on the basis of the Reynolds principle of similarity: in any "pipe-liquid" systems, the flows will be the same if the Reynolds numbers are the same. Naturally, as long as laminar flow is preserved.

When a certain value of flow velocity defined by Reynolds criterion (2.7) is exceeded, the liquid particles detach from the liquid particles immediately adjacent to the surface of the obstacle. Laminar motion turns into turbulent motion (Fig. 2.5e, f, g). According to Bernoulli's equation (2.6b), increase in dynamic pressure and decrease in static pressure corresponds to increase in velocity of particles in the flow [1, 4]. If the fluid flow contains suspended particles, they will be precipitated on surfaces of bodies, contacting with the flow in turbulence zone (see Fig. 2.5e, g.) A typical example is car rear windows covered with dust when driving on a country road.

2.3. Hemodynamic issues

Hemodynamics is a branch of biomechanics, which studies blood flow in the vascular system. The first fundamental studies in this area are associated with Poiseuille[*].

2.3.1. Blood as an object of study of fluid movement

Before discussing issues related to blood movement, we should briefly describe its main characteristics as an object of physical and physical-chemical research. Blood is a mobile connective tissue of liquid medium of the body consisting of two main components: plasma and form elements suspended in it (cells leukocytes, postcellular structures – erythrocytes and thrombocytes (platelets))[**].

The sedimented blood consists of three layers: the upper layer is formed by yellowish plasma, the middle, relatively thin gray layer consists of white blood cells, and the lower red layer is formed by red blood cells. In a healthy adult, the volume of plasma reaches 50-60% of the volume of whole blood, and the volume of blood-form elements is about 40-50%. The ratio of blood cells (sometimes only red blood cells) to the total blood volume is called the hematocrit number Ht (from the Greek αιμα – blood, κριτός – index) or hematocrit.

Blood plasma (from Greek πλάσμα – something formed) is the liquid part of blood, which contains water (about 90%) and substances suspended in it – proteins (mainly albumin, globulin and fibrinogen). Inorganic substances are about 2 to 3% (cations (Na^+, K^+, Mg^{2+}, Ca^{2+}) and anions (HCO_3^-, Cl^-, PO_4^{3-}, SO_4^{2-})). Organic substances (about 9%) in blood are divided in-

---

[*] ***J. L. Poiseuille*** – French physician and physicist, member of the French Academy of Medicine (1842), studied blood circulation and respiration of animals and humans, in 1828 he first used a mercury manometer to measure arterial pressure of animals, experimentally established the law of fluid flow (laminar flow) through a thin cylindrical tube (Poiseuille law and flow).

[**] On average, the normal volume of blood in men is 5.2 liters, in women – 3.9 liters, while in newborns its amount is 200÷350 ml. Mass fraction of blood in the total human body mass for an adult is 6-8%.

to nitrogen-containing (proteins, amino acids, urea...) and nitrogen-free (glucose, phospholipids, cholesterol...). Blood plasma also contains gases (oxygen, carbon dioxide) and biologically active substances (hormones, vitamins, enzymes, mediators). Histologically, plasma is the intercellular substance of liquid connective tissue (blood).

As an object of physical and chemical research (physical chemistry), blood is a polydisperse system – a suspension[*] of red blood cells in plasma: red blood cells are in suspension, proteins form a colloidal solution, urea, glucose and other organic substances and salts form a homogeneous water solution.

Since blood is a suspension (heterophase, heterogeneous system) and, as noted above, stratifies during sedimentation by gravity, blood components are separated by centrifugation, for example (Section 1.2.6.2). This circumstance significantly distinguishes blood flow from the flow of a homogeneous viscous fluid, as the distribution of blood particles over the cross section of the vessel will depend on the mass of these particles. *Platelets, for example, move in close proximity to the walls of blood vessels. Their main function is to repair blood vessels if they are damaged. They move freely with the blood flow and only stay in places where the blood flow goes into a calmer state (spleen, liver, subcutaneous tissue).*

Blood is not a Newtonian fluid, but plasma can be considered a Newtonian fluid. It should be considered that, as will be shown in the following sections, blood moves in a tube (not to be confused with "current tube", a concept used in describing fluid flow) with elastic or plastic walls, not rigid ones. Nevertheless, the general principles of hydrodynamics are quite

---

[*] Suspension (from Latin suspensio, hanging) is a mixture of substances, where the solid substance (or insoluble component) is distributed in the form of tiny particles in the liquid substance in suspension. A suspension is a coarse-dispersed system with a solid dispersed phase and a liquid dispersion medium. Usually the particles of the dispersed phase are so large (more than 10 microns) that they settle by gravity (sediment). Suspensions, in which sedimentation is very slow due to the small density difference between the dispersed phase and the dispersion medium, are sometimes called suspensions. In concentrated suspensions dispersed structures easily arise and separate.

applicable to the description of blood flow in various vessels of the blood system. It is generally accepted that it corresponds to laminar flow conditions (at least in human blood vessels without pathologies at rest). According to different estimates Reynolds number for blood varies in the range $Re_{cr} = 500 \div 1600$. Occurrence of blood flow turbulence in certain areas of the human circulatory system is considered as one of the factors of blood clots (blood clots consisting of fibrins) formation. *Other factors include: – slowing or cessation of blood flow in the capillaries (stasis); – damage to the walls of blood vessels and inflammatory processes in them; – changes in blood composition and properties (primarily, coagulation).* Turbulence causes the appearance of noise, which allows medical diagnostics by sonic methods (see section 3).

2.3.2. The equation of the continuity of blood flow in human vessels

Blood flow depends both on the physical properties of blood and on the properties and condition of the blood vessels. The human circulatory system[*] is a complex closed system of tubes of various diameters. It includes: aorta, arteries, arterioles, capillaries, venules, veins. Blood from the heart enters the aorta, from which it is distributed through the arteries, then through the arterioles and then disperses through millions of small capillaries. Through the veins, blood returns to the heart. It should be noted that the basis of the circulatory theory was laid by C. Galen, who first suggested the existence of two circuits of blood circulation[**].

---

[*] The main function of the circulatory system is the transport of oxygen and excretion of metabolic products (metabolism) and, as a consequence: maintaining a constant body temperature and constant body composition (homeostasis), immunity.

[**] ***Claudius Galen*** (Greek: Κλαύδιος Γαληνός), 129 or 131 – ca. 200 or 217) was a Roman (Greek) physician, surgeon, and philosopher. He created the first theory of circulation, which existed until 1628, when William Garvey published his work "Anatomical Study of the Movement of the Heart and Blood in Animals". Galen contributed significantly to the understanding of anatomy, physiology, pathology, pharmacology, neurology, philosophy and logic.

Fig. 2.6 shows the general scheme of the human circulatory system, and Fig. 2.7 shows the circulatory system with indication of the functional purpose of its parts [***].

The duration of the circulatory cycle (large and small circuits) in humans is approximately 20 seconds. Small circle of circulation: right atrium – right ventricle — lungs (here blood is saturated with oxygen) – left atrium – left ventricle. Blood ejection from the left ventricle (stroke volume) is the beginning of the great circle of circulation: left ventricle – aorta – arteries – capillaries of organs of the whole body – capillary-venous anastomoses. Here blood enters the venous system and further through the vena cava (upper and lower) enters the right atrium. It should be noted that in the small circle, venous blood flows through the arteries, and arterial blood flows through the veins.

During a day, the flow of blood makes about 4,200 cycles, i.e. during a day the heart pumps up to 10,000 liters of blood through all the vessels! The speed of blood flow varies in different vessels. The main parameters of blood flow in different human blood vessels are given in Table 2.1.

Table 2.1.

Dynamic characteristics of blood flow in human vessels

Vessels	Diameter, mm	Velocity, 10^{-2} m/s	Pressure, mm Hg.
Aorta	20	30 – 50	50 – 150
Arteries	20 – 5	20 – 50	80 – 20
Arterioles	0,1 – 0,5	1 – 20	50 – 20
Capillaries	0,5 – 0,01	0,01 – 0,05	20 – 10
Venules	0,1 – 0,2	0,1 – 1,0	10 – 5
Veins	10 – 30	10 – 20	(-5) – (+5)

[***] A variety of circulatory system diagrams with descriptions are available on all Internet search platforms.

The parameters in Table 2.1 give only a general idea of the hydrodynamic characteristics of human vessels. The scatter of values given in different sources is quite large. It is quite understandable, as even for one and the same person they depend on diurnal phases of vital activity. Some parameters are obtained from experimental data, others are based on theoretical calculations. The discussion of these special issues is far beyond the scope of this course. The materials are quite fully presented in Internet search engines (key words "Characteristics of blood movement through vessels", "Hydrodynamic characteristics of the vascular channel", etc.). Let us note only the fact that these data are not always comparable, as the highlighted elements of the vascular bed do not always coincide and they are not always labeled in the same way. As for the data given in Table 2.1, it gives, in a sense, collective characteristics for a group of vascular bed fragments performing the same functions. The given values of blood particle velocity and pressure are the cross-sectional averages in the given type of vascular bed.

At first sight, it seems that the given values contradict the equation of flow continuity (2.5) – blood flow rate in thin capillaries is about 1000 times less than in arteries. However, this discrepancy is apparent. The fact is that Table 2.1 shows the diameter of a *single* vessel. This value really decreases with branching. But the total area of branching increases. So the total cross-sectional area of all capillaries (about 2000 cm^2) is hundreds of times greater than the cross-sectional area of the aorta – this explains such a low blood velocity in capillaries. Low blood flow rate in the capillaries ensures effective exchange between blood and tissues.

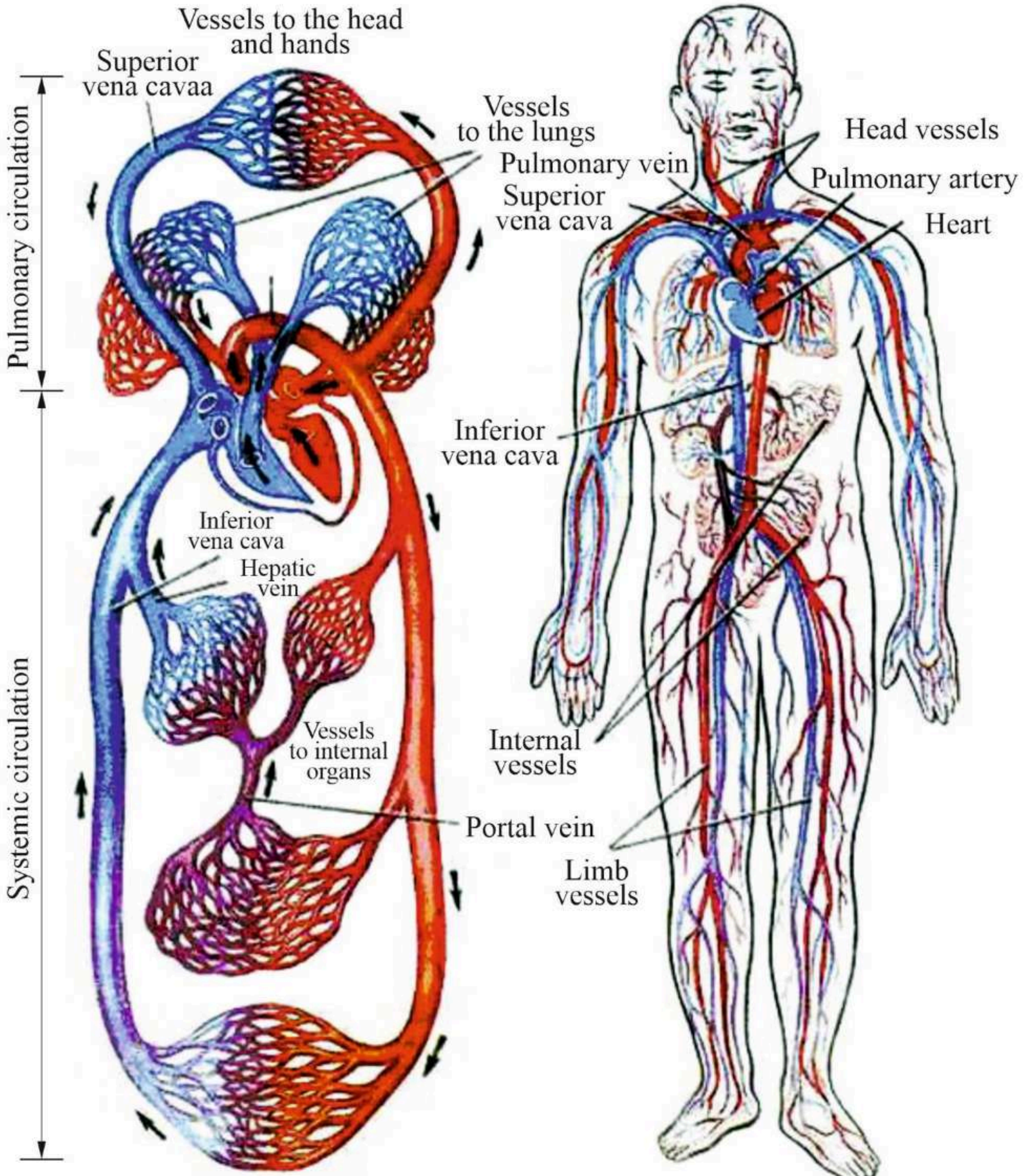
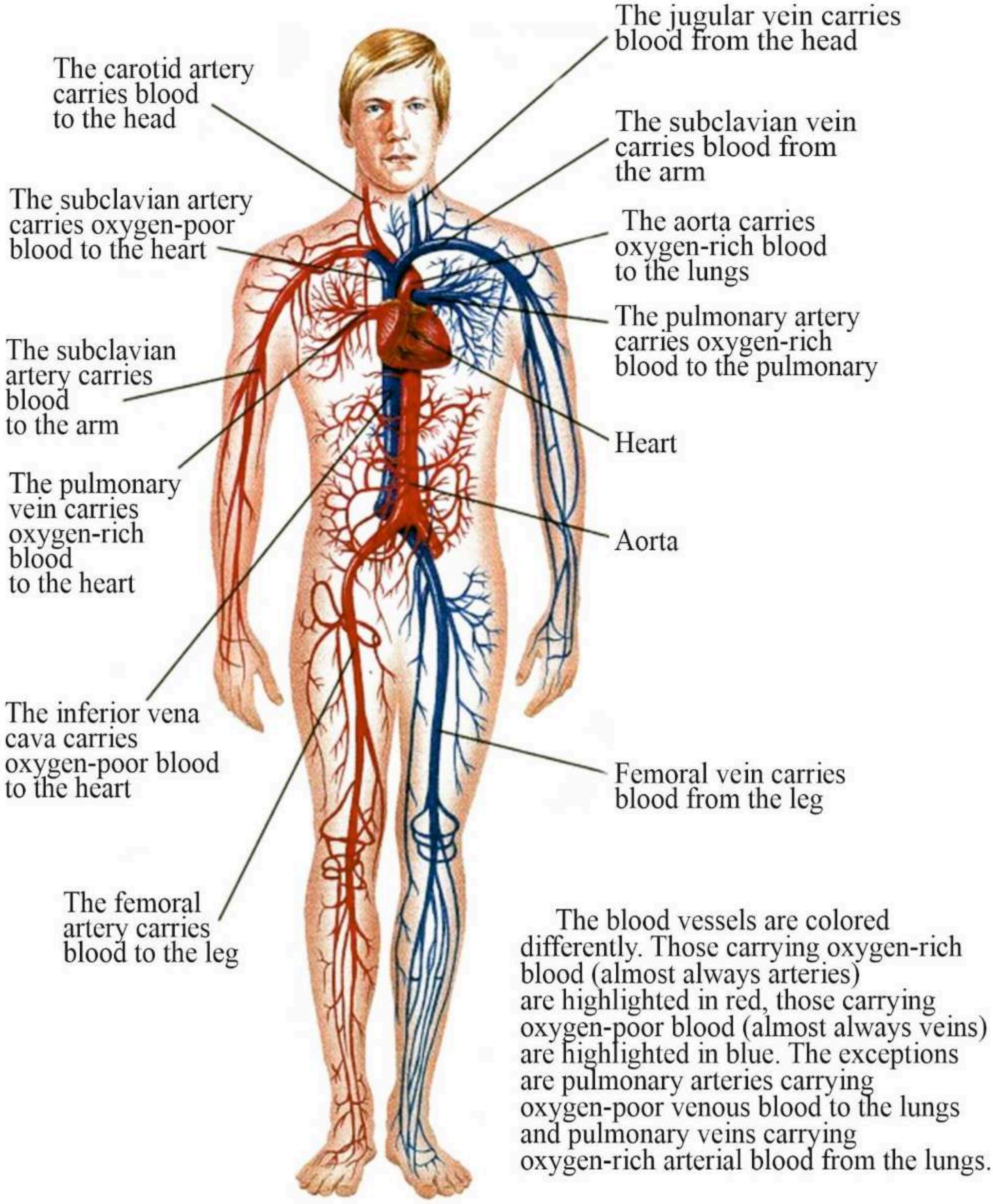


Fig. 2.6. General scheme of the human circulatory system



The blood vessels are colored differently. Those carrying oxygen-rich blood (almost always arteries) are highlighted in red, those carrying oxygen-poor blood (almost always veins) are highlighted in blue. The exceptions are pulmonary arteries carrying oxygen-poor venous blood to the lungs and pulmonary veins carrying oxygen-rich arterial blood from the lungs.

Fig. 2.7. Human circulatory system (see the caption to Fig. 2.6). The functional purpose of the circulatory system links is given

2.3.3. Heart functioning. Pulse wave

To maintain fluid motion in a closed hydrodynamic system, a "pump" is required, which creates a pressure difference to ensure overcoming hydraulic resistance. In the circulatory system, the heart plays the role of such a "pump". Its external view is shown in Fig. 2.8; Fig. 2.9 shows the cardiac conduction system and Fig. 2.10 – sectional anatomy.

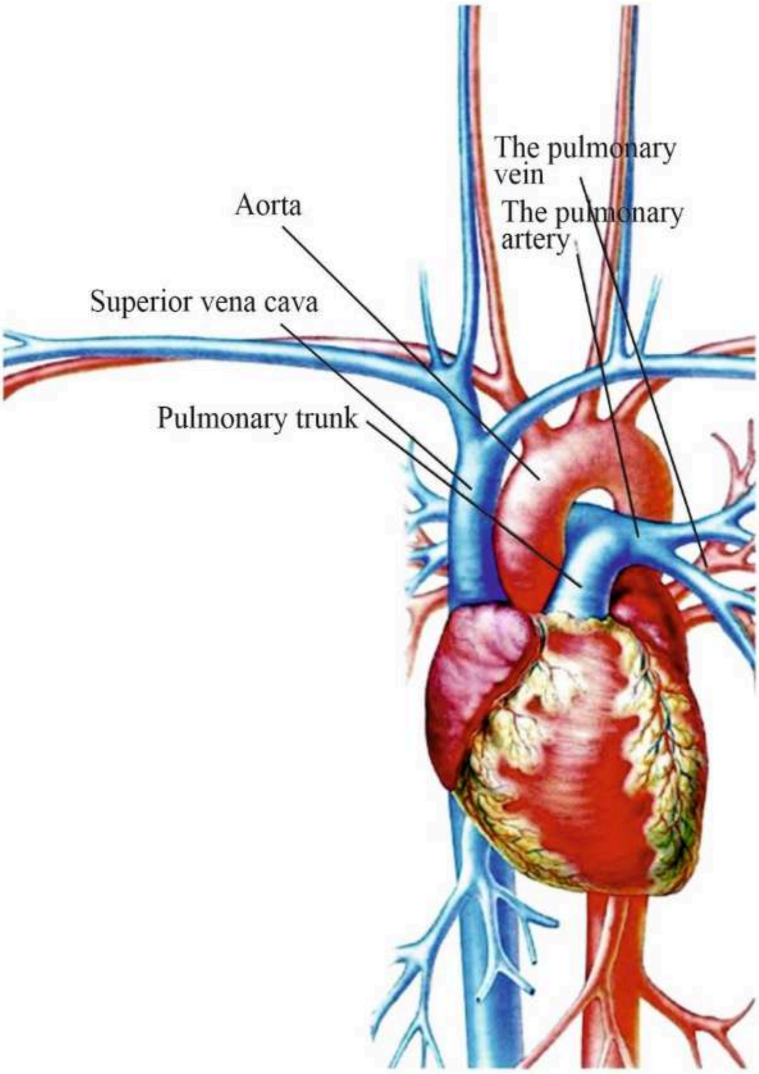
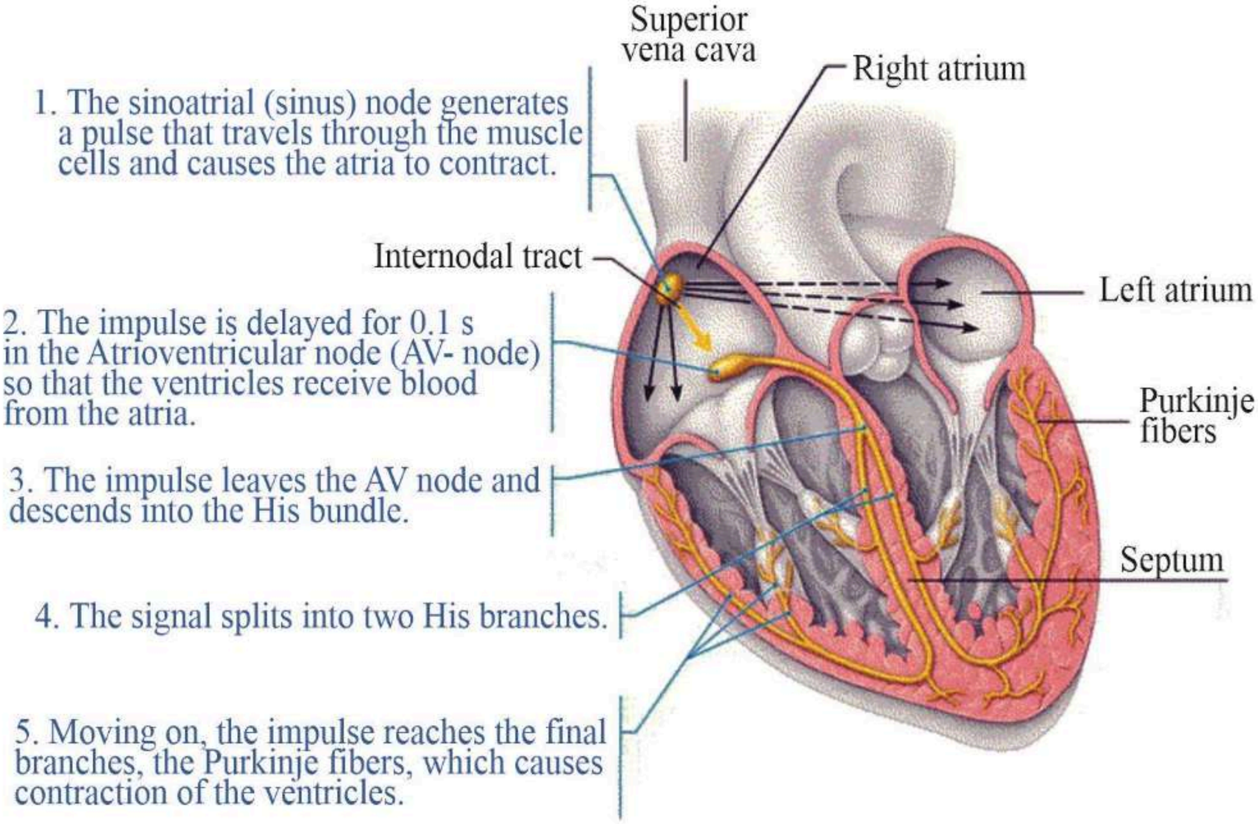


Fig. 2.8. Exterior view of the heart



1. The sinoatrial (sinus) node generates a pulse that travels through the muscle cells and causes the atria to contract.

2. The impulse is delayed for 0.1 s in the Atrioventricular node (AV- node) so that the ventricles receive blood from the atria.

3. The impulse leaves the AV node and descends into the His bundle.

4. The signal splits into two His branches.

5. Moving on, the impulse reaches the final branches, the Purkinje fibers, which causes contraction of the ventricles.

Fig. 2.9. Cardiac conductive system (frontal section) of a human

A detailed review of the heart is beyond the scope of this textbook. Let us note only the basic functions of the heart [5]:

- automatism, the ability of specialized myocardial cells[*] to produce electrical impulses that cause myocardial excitation (normally, the cells of the sinus node[**], located in the right atrium have the greatest automatism);
- conduction, the ability to conduct electrical impulses from the autonomic cells to the contractile myocardium;
- excitability, the ability of the heart to be excited by electrical impulses (conduction system cells and myocardial contractile cells have this function); during excitation, an electrical current is generated;
-contraction, the ability of the heart to contract under the influence of these electrical impulses.

---

[*] Myocardium (from the Greek μῦς - muscle and καρδία - heart) is the name of the muscular middle layer of the heart, which constitutes the main part of its mass.

[**] SN is a term formed from two concepts: sinus, "sinus," a cavity, and atrium (Latin atria).

The function of the heart, contractility, is actually the subject of study in the section "Mechanics". The "pump" that makes the heart muscles contract is "turned on" by electrical impulses.

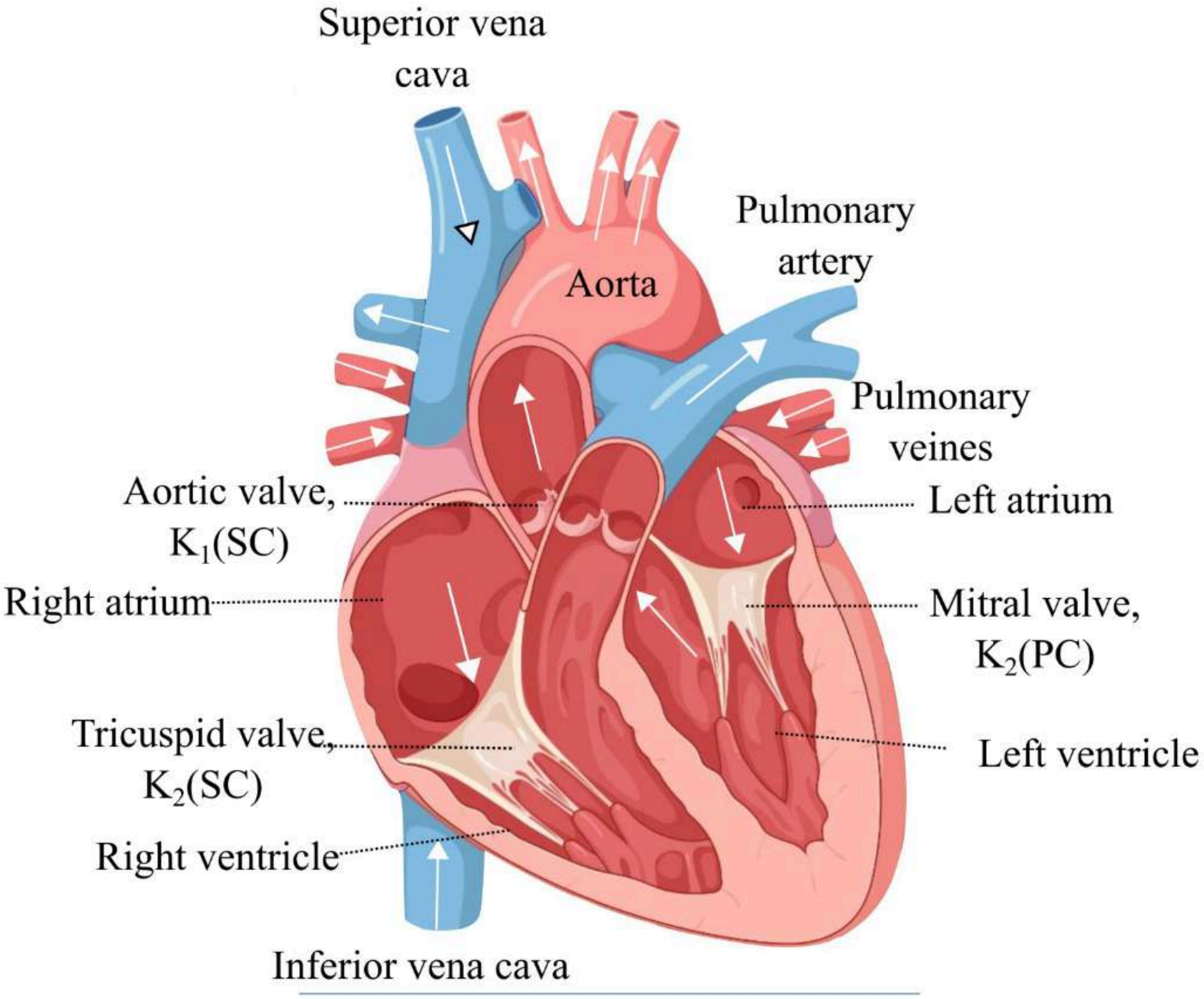


Fig. 2.10. Anatomical structure of the heart (frontal section): K_1(SC), K_2(SC), K_1(PC), K_2(PC) – valve designations as in the mechanical heart model in Fig. 2.12 a, b. (they are absent in the original). PC – pulmonary circulation, SC – systemic circulation

As seen in Fig. 2.9, the conduction system through the atrioventricular node covers both atria, the ventricular muscles, controlling their tension and relaxation through the electrical impulses generated in the sinoatrial node. Thus, the sinoatrial node acts as a generator of electrical impulses

(60-80 impulses per minute). Therefore, different phases of work are relatively clearly recorded by measuring the difference in the electric field potentials occurring on the human skin. This discovery belongs to W. Einthoven[*] (Netherlands) and is the basis of a widely used heart diagnostic technique – ECG (electrocardiography). In Russia, the first electrocardiogram of a patient was recorded by A.F. Samoilov, professor of Kazan Imperial University, in 1908.

Let us consider schematically mechanical phenomena occurring in the large circle of circulation.

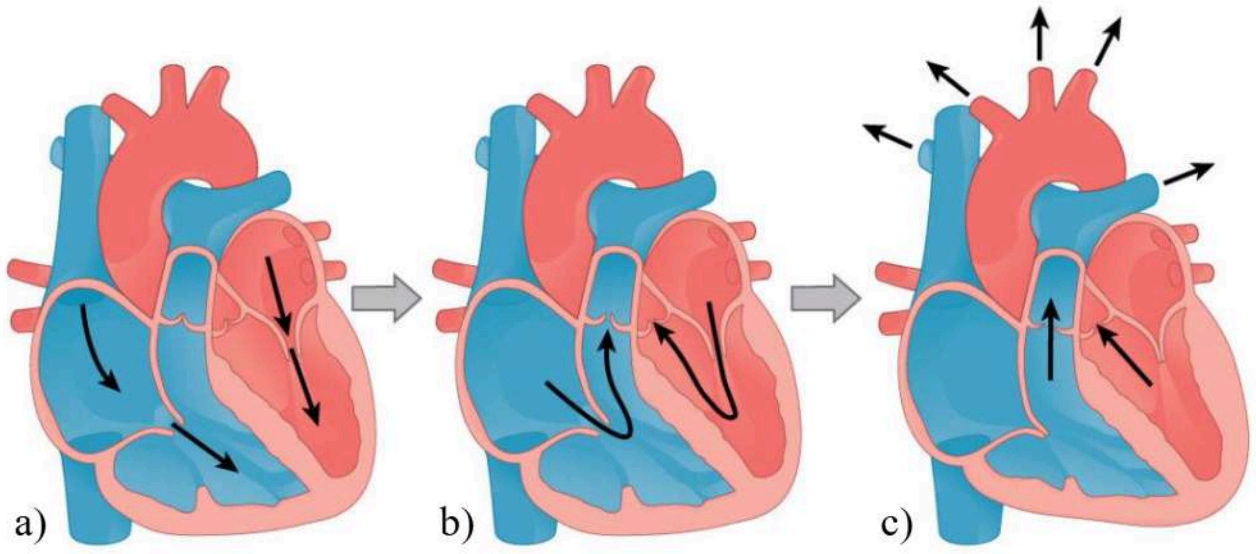


Fig. 2.11. The phases of the cardiac cycle: a) – cardiac diastole, b) – atrial systole, ventricular diastole, c) – atrial diastole, ventricular systole[**]

The cardiac cycle is shown schematically in Fig. 2.11 and consists of successive contraction and relaxation of the heart muscle with short intervals for rest. Three main phases can be distinguished in one cardiac

---

[*] ***Willem Einthoven***, (May 21, 1860, Semarang – September 29, 1927, Leiden) – Dutch physiologist, founder of electrocardiography, designed a device in 1903 to record the electrical activity of the heart, first used electrocardiography for diagnostic purposes in 1906, received the Nobel Prize in Physiology or Medicine in 1924.

[**] Picture borrowed from the Internet [https://courses.lumenlearning.com/wm-biology2/chapter/the-cardiac-cycles].

cycle. During the first phase (Fig. 2.11a), which lasts about 0.1 s in an adult, both atria contract synchronously (blood is pumped to the ventricles), and the ventricles are in a relaxed state (diastole) by this time. The atrium starts working after a short time (0.05÷0.1 s) after the excitation of the sinus node (the beginning of the electrical impulse). Atrial contraction begins from the place where veins enter it, which ensures primary compression of their mouth and impossibility of blood to flow back into the veins. The second phase follows (Fig. 2.11b) – *systole (from Latin sustolḗ – contraction), in this case the ventricles contract* (it is also about 0.1 s). The ventricles contract synchronously (working as a pressure pump) and push blood into the aorta through the left ventricle and into the pulmonary artery through the right ventricle; the atria are relaxed at this time and the valves separating the atria from the ventricles (tricuspid and mitral) close. Then comes the third and final phase (Fig. 2.11c) – *diastole (from Latin diastole – relaxation),* a pause during which there is a general relaxation of the heart (no blood in the ventricles). Its duration is about 0.4 s. This time includes the duration of depolarization (repolarization in terms of electrocardiography) of the ventricular myocardium. During this time the atria, partially also the ventricles, through the open valves, fill up with blood. Thus, the whole cardiac cycle – the duration of mechanical heartbeats (or from pulse to pulse) lasts up to 1 s. The inverse of this value is called the heart rate (HR). During one cardiac cycle the atria work approximately 10÷12.5 % of the total time of cardiac cycle, and the ventricles 30÷37.5 %. The time of atrial contraction with atrioventricular conduction and ventricular contraction takes about 0.3 seconds. The heart, as a mechanical pump pushing blood into the circulatory system, works for about one third of the total cycle time. The rest of the time, the "pump" shuts down.

Blood flow from the atria is regulated by a system of valves (Fig. 2.10), preventing blood backflow (*regurgitation*) from the ventricles and ensuring synchronous flow in the small and large circuits for simultaneous filling of both ventricles (compliance with the condition of "continuity of flow").

It should be noted that during measured heart rate, the relaxation time (diastole) is enough to fill the ventricles with blood. However, when the work of the heart becomes more frequent, for example, at high loads, the atrial systole becomes crucial, because without it the ventricles simply would not have had time to fill with blood. Blood is released into the aorta and the arteries only during contraction of the left ventricle, and this contraction is called the pulse. As you know, human pulse values are very individual and depend on a number of factors in the same person. Therefore, the above time intervals, which characterize different phases of the heart, are quite conditional. In the above scheme, the duration of the cardiac cycle is taken as 1 s. This corresponds to a heart rate of 60 beats per minute. This value is used for further evaluation of the heart as a mechanical pump, the active phase of which takes 3 seconds in total. (The normal value is the heart cycle time of 0.85 s, which corresponds to the average heart rate of 70÷72 beats per minute)[*].

Fig. 2.12a shows a simplified diagram of the large and small circuits of the blood circulation. The mechanical model of the cardiovascular system (CVS) can be represented as a closed, fluid-filled system of many branched tubes with elastic (and plastic in the vein system) walls (Fig. 2.12b). Fluid movement is under the action of a rhythmically working pump in the form of a pear with two valves.

---

[*] The total duration of the cardiac cycle can be significantly longer and its increase is associated with an increase in the duration of the third phase – diastole. A pulse rate greater than 80 beats per minute is considered abnormal (tachycardia), leading to a shorter life span, and below 40 is also abnormal (brachycardia), requiring appropriate correction, such as the installation of a pacemaker. Heart rate in animals varies greatly from species to species: elephants, for example, have about 30 beats per minute (live 60÷70 years), while rats have 300 beats per minute (according to different sources live 2÷3 years) and shrews have 300÷600 (do not live even 2 years).

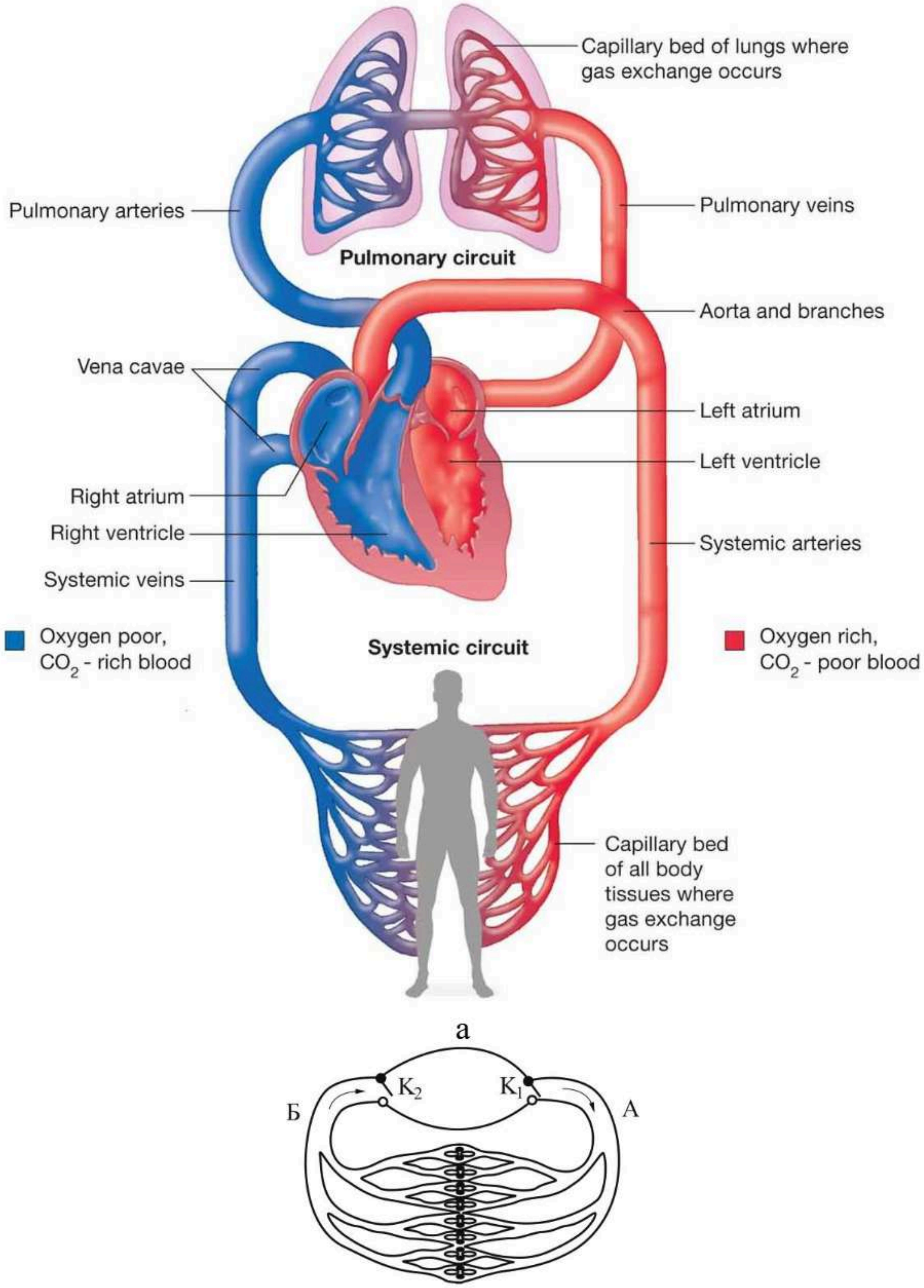


Fig. 2.12. a) – Simplified scheme of the large and small circuits circulatory system*; b) – mechanical model of blood circulation (comments in the text)

---

* Picture from https://www.pinterest.ru/pin/799951952542429574/.

When the pear is compressed (in the scheme of the systemic circulation this is a contraction of the left ventricle) the outlet valve K_1(SC) (mimicking the aortic valve) opens and the fluid (blood) contained in it is pushed out into tube A (aorta). Then the valve K_1(SC) closes. Due to the stretching of the walls, the volume of the tube increases, and it accommodates excess fluid. The walls of the aorta begin to gradually contract, pushing the excess fluid into the next link in the system (arteries). Their walls also first stretch, accepting the excess fluid, and then contract, pushing the fluid into the next part of the system. At the final stage of the circulatory cycle fluid (blood) is collected in tube B (upper and lower vena cava) and through inlet valve K_2(SC) (*tricuspid valve*) (see Fig. 2.10) from the right atrium blood flows into the right ventricle.

As a mechanical system, the heart is a rhythmically working pump, in which the working phases, systole, alternate with the idle phases, diastole. During systole, blood contained in the left ventricle is expelled into the aorta, after which the aortic valve closes. The volume of blood pushed into the aorta during one heartbeat is called *stroke* volume (60÷70 ml). Blood entering the aorta stretches its walls and the pressure in the aorta rises. This pressure is called *systolic* pressure (Systolic Blood Pressure – SBP, p_s). Increased pressure spreads along the arterial part of the co-vascular system. This spreading is caused by the elasticity of the arterial walls and is called the pulse wave.

A pulse wave is a wave of elevated (above atmospheric) pressure spreading along the aorta and arteries. It is caused by the release of blood from the left ventricle during systole. Pulse waves can be mathematically represented as a set of harmonic waves. However, with a high enough degree of approximation it is possible to take into account the dominant first harmonic. And then pulse wave can be described by harmonic damped running longitudinal wave equation (see section 3, equation 3.21, fig. 3.11). Here the circular frequency corresponds to the heart muscle contraction rate (HR): 60÷80 per min. (1÷1,3 Hz or 6,28÷8,16 $\text{rad}\cdot\text{s}^{-1}$, $\omega = 2\pi\nu$). Meeting on its way the resistance of the medium (in this case hydrodynamic), any wave attenuates if the perturbation of the medium stops. But the heart works continuously. Therefore, as blood flow moves from the center to the periphery, pulsation amplitude only decreases. Pulse wave propagates with velocity υ_p, whose value is in the range of 5÷10 m/s. The magnitude of velocity in large vessels depends on their size and the mechanical properties of the tissue walls

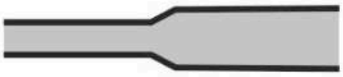
$$\upsilon_p = \sqrt{Eh/\rho d}, \tag{2.9}$$

where E is modulus of elasticity, h is thickness of vessel wall, d is diameter of vessel, ρ is density of vessel wall substance. The profile of the artery in different wave phases is shown schematically in Fig. 2.13.

Before the pulse wave passes

The pulse wave begins to pass through the artery

Pulse wave in the artery



Decrease of pulse pressure blood is pushed by the contraction of the walls

The pulse wave is finished

Fig. 2.13. Profile of an artery during pulse wave passage

After the pulse wave has passed, the pressure in the corresponding artery drops to a value called diastolic pressure (Diastolic Blood Pressure – DBP, p_d). Thus, the change in pressure in the great vessels is pulsatile in nature. This is clearly evident in the carotid artery and in the periphery, such as the radial artery. Figure 2.14 shows schematically two cycles of blood pressure changes in the brachial artery.

The pulse wave corresponds to the pulsing speed of blood flow. In large arteries the velocity is 0.3÷0.5 m/s. However, as the vascular system branches off, the vessels become thinner and their hydraulic resistance increases in proportion to R^4. This leads to a decrease in the amplitude of pressure fluctuations. In the arterioles and beyond, pressure oscillations are practically absent. As the branching process goes on, not only the amplitude of pressure fluctuations decreases, but also its average value. The nature of pressure distribution in different parts of the vascular system has the form shown in Fig. 2.15. Here we can see the excess of pressure over the atmospheric pressure.

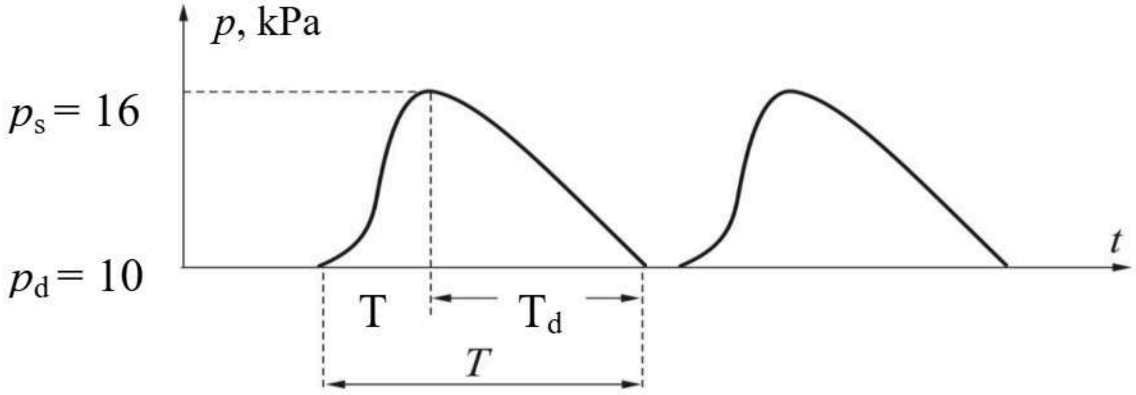


Fig. 2.14. Schematic representation of arterial pressure changes in brachial artery: T – duration of cardiac cycle; $T_s \approx 0,3T$ – duration of systole; $T_d \approx 0,7T$ – duration of diastole; p_s – maximum value of SBP; p_d – minimum value of DBP. See comments on the text

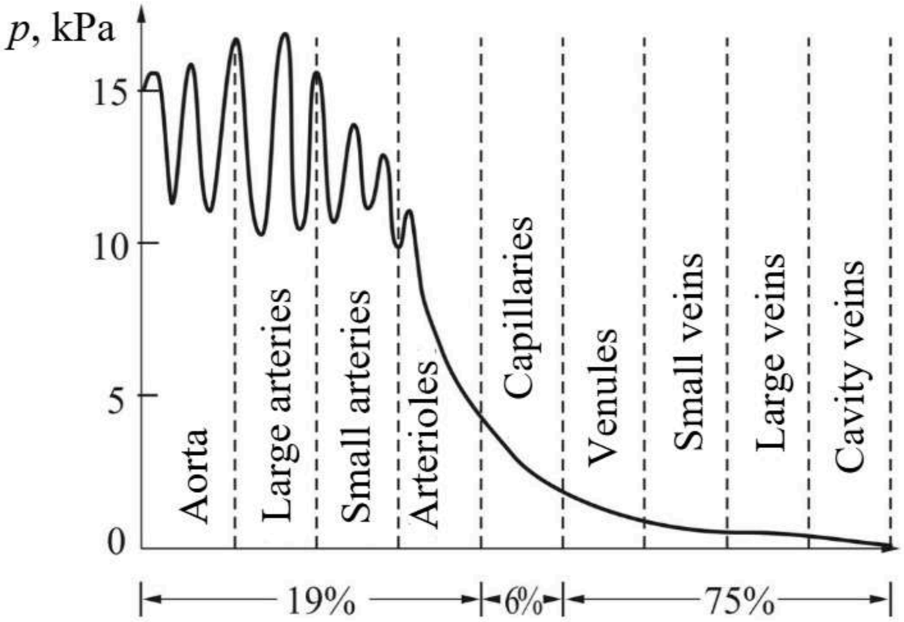


Fig. 2.15. Pressure distribution in different parts of human vascular system. The abscissa axis shows the relative proportions of the total blood volume in a given area

The cross-sections of the blood vessels of the circulatory system experience periodic changes during the day. This is due to the fact that the length of vessels is very large (100000 km) and 7÷8 liters of blood contained in blood vessels are obviously not enough for their maximum filling. Therefore, most intensively are supplied those organs, which at this moment work with maximum load. The cross-section of other vessels at this moment decreases. For example, after a meal the digestive organs function most energetically, a considerable part of blood is directed to them; for normal work of the brain it is not enough and a person feels drowsy.

A few words about the energetic characteristics of the heart. During one systole the left ventricle ejects into the aorta a shock volume of blood (60÷70 ml). Ventricular volume decreases by the same amount: $\Delta V \approx 65 \cdot 10^{-6}$ m^3. The work done by the heart muscle per one contraction can be estimated by the formula: $\Delta A = p_{av}\Delta V_{st}$, where ΔV_{st} is the average value of the stroke volume of blood, and p_{av} is the average pressure that is created inside the ventricle. It is slightly higher than the systolic pressure in the artery: $p_{av} \approx 17$ kPa. From this we get an estimate for the work of the heart muscle per contraction: $\Delta A \approx 17 \cdot 10^3 \times 65 \cdot 10^{-6} = 1,1$ J. Power developed by the heart muscle during systole $N_s = \Delta A / T_s$, where $T_s \approx 0,3$ s – systole duration. From this we get: $N_s = 1,1/0,3 = 3,7$ W. On average for a healthy person in a calm state time of one cycle of cardiac activity $T \approx 0,85$ s. Average power for the whole cycle is $N_{av} = 1,1/0,85 = 1,3$ W. In hypertension, blood pressure increases and, accordingly, the work done by the heart increases. *During life, the heart makes enough work to lift a loaded train to the highest peak in the chain of Alpine mountains – Mont Blanc (height 4810 m, Italy – France).*

Let us note that cardiac activity is found in different physical phenomena, which determine the ways of its registration and diagnostics:

- Electrical – *ECG, ventriculocardiography.*
- Acoustic – *auscultation, phonocardiography.*
- Mechanical:
    - Apex beat – *palpation, apexcardiography*;
    - Pulse wave – *palpation, sphygmography, phlebography;*

- dynamic effects – change in the center of gravity of the thorax in the cardiac cycle – *dynamocardiography;*
- ballistic effects – concussion of the body at the moment of blood ejection from the heart – *ballistocardiography*;
- changes in size, position, and shape – ultrasound, X-ray *cystography*.

Their detailed consideration is beyond the scope of this textbook. They are partially presented in the following chapters.

2.3.4. Physical basis of the clinical method of measuring blood pressure

In medicine auscultative (*ausculcultatio* – lat. – listening, (see section 3.2.4) method of arterial pressure (BP) measurement is widely used (N.S. Korotkov, 1905, Russia). A common pressure measuring device (sphygmomanometer or tonometer) consists of an occlusion (*occlusion* – lat. – locking, hiding) pneumocuff, an air inflation bulb with an adjustable valve to bleed air and a device that measures the pressure in the cuff. Mercury gauges, aneroid stick gauges, or electronic gauges are used as such a device. Auscultation is performed with a stethoscope or a membrane phonendoscope (see Section 3.2.4). The sensing head is placed at the lower edge of the cuff over the projection of the brachial artery without significant pressure on the skin.

The principle of the device is as follows. A cuff is placed around the arm between the shoulder and the elbow joint (about 2 cm higher), and air is pumped into the cuff. As soon as the pressure in the cuff exceeds the blood pressure in the brachial artery, the blood flow is stopped. When the air is deflated from the cuff the blood flow resumes first partially and then completely. The pressure at which restoration of blood flow begins and ends is recorded using a manometer. The SBP is determined by decompression of the cuff at the time of the first tones, and the DBP by the time of their disappearance. The auscultatory technique is recognized by WHO as the standard for noninvasive BP determination. (*Invasion*, one of the meanings of intervention, in this case in the body) despite slightly lower values for SBP and higher values for DBP compared to those obtained by invasive measurement (by inserting a hydrophone probe into the patient's artery to control the pressure level). The cuff is always placed in the brachial artery (Fig. 2.16).

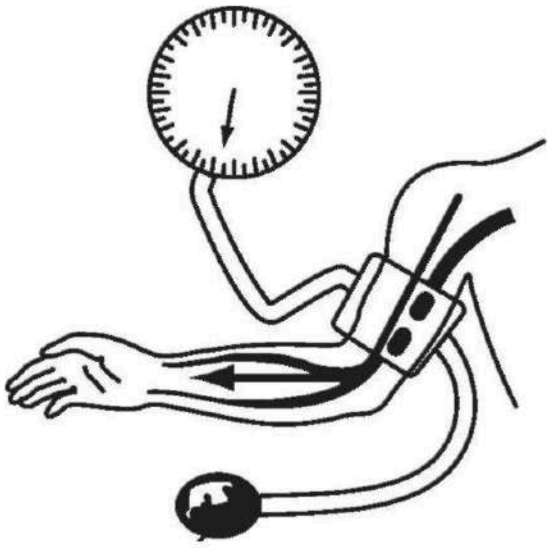


Fig. 2.16. Cuff application scheme: the cuff is always placed at the level of the heart: the patient is sitting or lying down, but sometimes it is necessary to measure the pressure when standing

This choice of cuff location allows standardization of results because the brachial artery in the lowered arm is at the level of the heart and the measured pressure coincides with the blood pressure in the part of the aorta closest to the heart. If the pressure was measured, for example, at the ankle, the measurement result would depend on the person's height and position (upright, horizontal, sitting).

Let's consider the physical basics of this method. Scheme of the processes that are consistently manifested during pressure measurements is presented in Fig. 2.17.

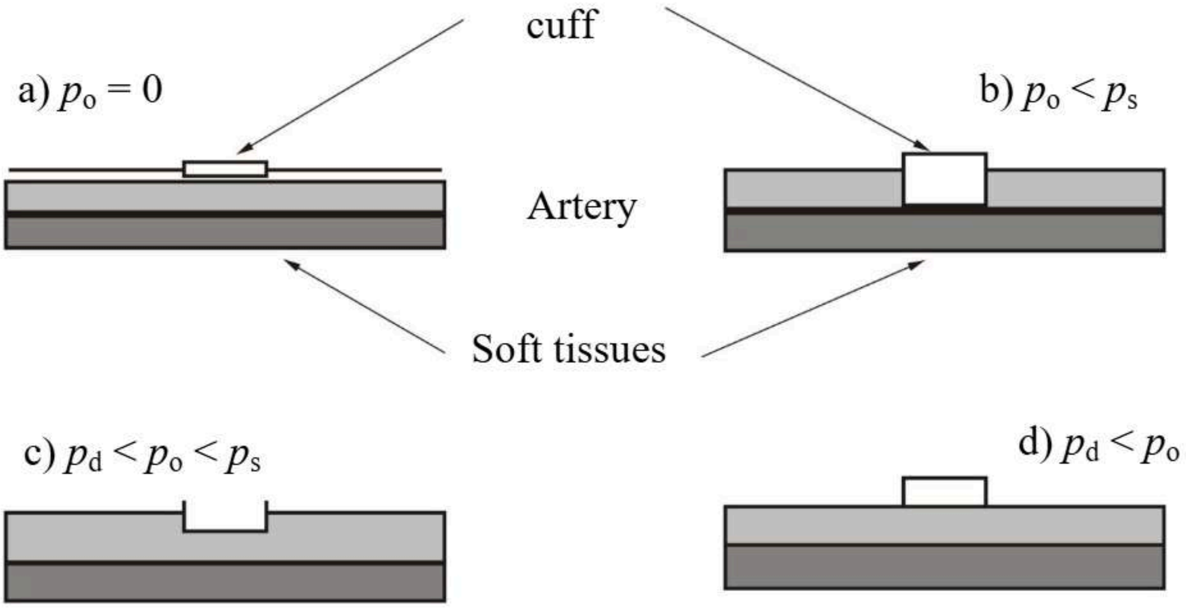


Fig. 2.17. The scheme of blood pressure measurement

1. At first, there is no overpressure p_o in the cuff and the blood flow is not interrupted.

2. As air is pumped into the cuff, it squeezes the brachial artery; when the pressure in the cuff exceeds the systolic pressure (P_s), blood flow stops.

3. Slowly deflating (venting) the air, reduce the pressure in the cuff. Once the cuff pressure is slightly less than the systolic pressure ($p_d < p_o < p_s$), blood will begin to push through the constricted artery. A flow is created in the artery, accompanied by noises that are audible through the phonendoscope. These noises are caused by the vibration of the arterial walls directly behind the cuff due to the tremors of the blood rushing through the compressed portion of the vessel. At the moment the noise appears, the systolic pressure ("upper pressure") is recorded on the manometer. In terms of time, the noise as a result of blood flow transition from laminar to turbulent appears after squeezing, i.e. it is underestimated by definition.

4. When the pressure in the cuff becomes less than the diastolic pressure p_d, the cuff stops compressing the artery. Blood flow switches from turbulence mode to laminar flow mode and noises associated with vortices

stop. At the moment of noise reduction the manometer registers diastolic pressure ("lower pressure"), but noise disappearance precedes lower pressure registration, i.e. it shows overestimation. So, the described method tends to underestimate the "upper pressure" and overestimate the "lower pressure", Figure 2.18.

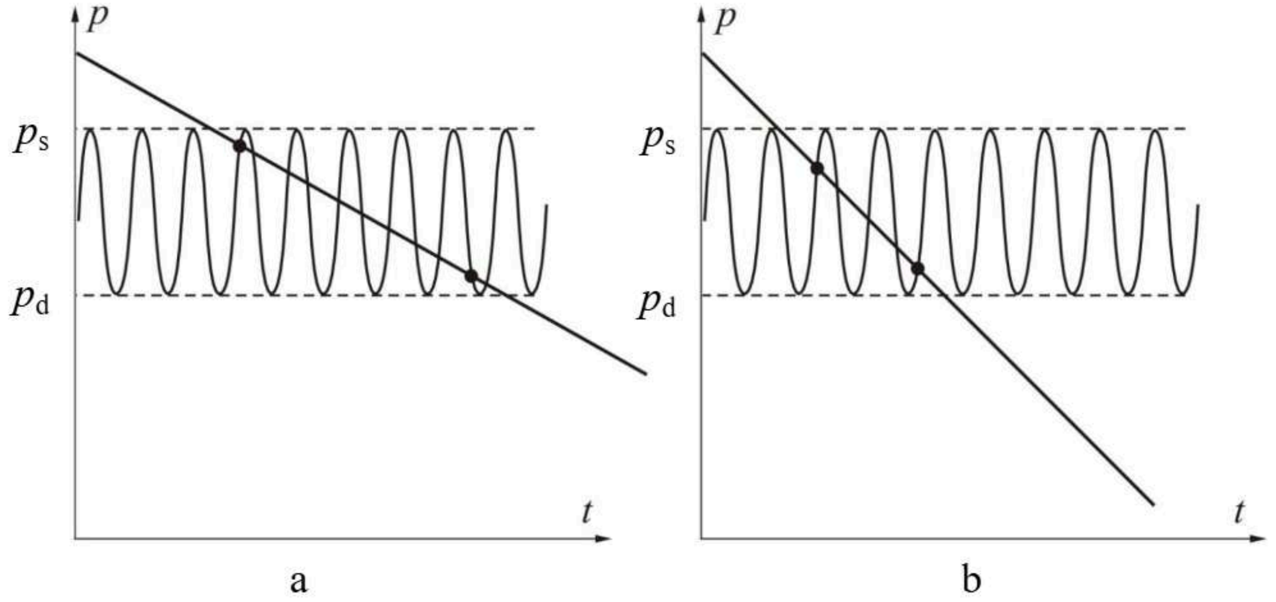


Fig. 2.18. Errors occurring when measuring Blood pressure: a) – with slow bleeding of air from the cuff, b) – with rapid bleeding of air from the cuff (explanations in the text)

The straight line shows the change in pressure in the cuff. "Sinusoid" shows the nature of the change in blood pressure. The "upper" and "lower" pressures recorded from the manometer readings are marked with bold dots. It can be seen that the faster the pressure drop in the cuff, the greater the error (mathematically speaking, the more "sinusoids" in the measurement interval, the greater the measurement accuracy). Therefore it is recommended that the pressure in the cuff be slowly released during the measurement, at intervals of 2 mm Hg.

Different types of devices are used to measure BP: manual, semi-automatic or automatic.

2.3.5. Blood pressure and vascular elasticity

Measuring blood pressure is mandatory for any medical examination. Normal blood pressure varies during each cardiac cycle from 16 kPa (120 mm Hg) to 10 kPa (80 mm Hg). These numbers indicate how much the blood pressure in an artery is higher than the atmospheric pressure. The value of arterial pressure is mainly determined by the following factors: – the force with which blood is pushed into the vascular bed during systole; – the elasticity of the vascular walls; – the resistance of the vascular bed, which depends on the size of the vascular lumen and blood viscosity.

Blood pressure level changes with muscle strain, emotional stress, changes in functional state or exposure to a number of physical factors (high-intensity sound, ultrasound, infrasound, vibration, overload, etc.). Elevated blood pressure can be true – in hypertension, or symptomatic – in other diseases. The value of blood pressure also changes in diseases associated with impaired neural regulation of the vascular lumen. In some cases there is an increase in pressure up to 220/120 mm Hg (hypertension), in other cases the pressure is lower than normal up to 90/60 mm Hg (hypotension). High blood pressure increases the strain on blood vessel walls, which can cause ruptures and internal hemorrhages. A decreased value causes problems with the blood supply to the brain. This is due to the decrease in hydrostatic pressure as height increases (see Section 2, Formula 2.3). In a person in an upright position, this decrease reaches 5 kPa (38 mm Hg). An excessive decrease in external pressure poses a serious danger to the body (see Section 4.3). With age, due to the development of sclerotic changes, vessels lose elasticity, become stiff, which significantly affects the activity of the cardiovascular system. Pulse fluctuations of the blood flow increase, the efficiency of the heart decreases and the load on it increases. In addition, the presence of sclerotic changes in the vessels significantly changes the rate of blood flow and, consequently, tissue nutrition. This follows from Poiseuille formula, according to which the volume of fluid (gas) flowing per unit time t through a long tube (volume velocity) equals

$$Q = \pi r^4 \Delta p / 8 l \eta, \qquad (2.10)$$

where r – radius of the tube; l – its length; Δp – pressure difference at the ends of the tube; η – dynamic viscosity (coefficient of internal friction) of the liquid. Increase of viscosity leads to decrease of flow velocity, i.e. volume of flowing liquid.

At the same time, various regulatory actions have significantly changed. Thus, instead of dilation, vessels are often permanently narrowed, which leads to decreased blood supply (blood flow volume is proportional to the fourth degree of the radius of the vessel lumen). This leads to deterioration of functions of the brain, heart and other organs. In the area of sclerotic changes there is not only narrowing of the vessel lumen but also occurrence of blood flow turbulence, which can be accompanied by changes in rheological blood properties with subsequent health complications.

Fig. 2.19 shows some types of possible disturbances of normal blood flow in human blood vessels. An example of a possible disturbance of the vascular wall in the bifurcation area is shown in Fig. 2.20.

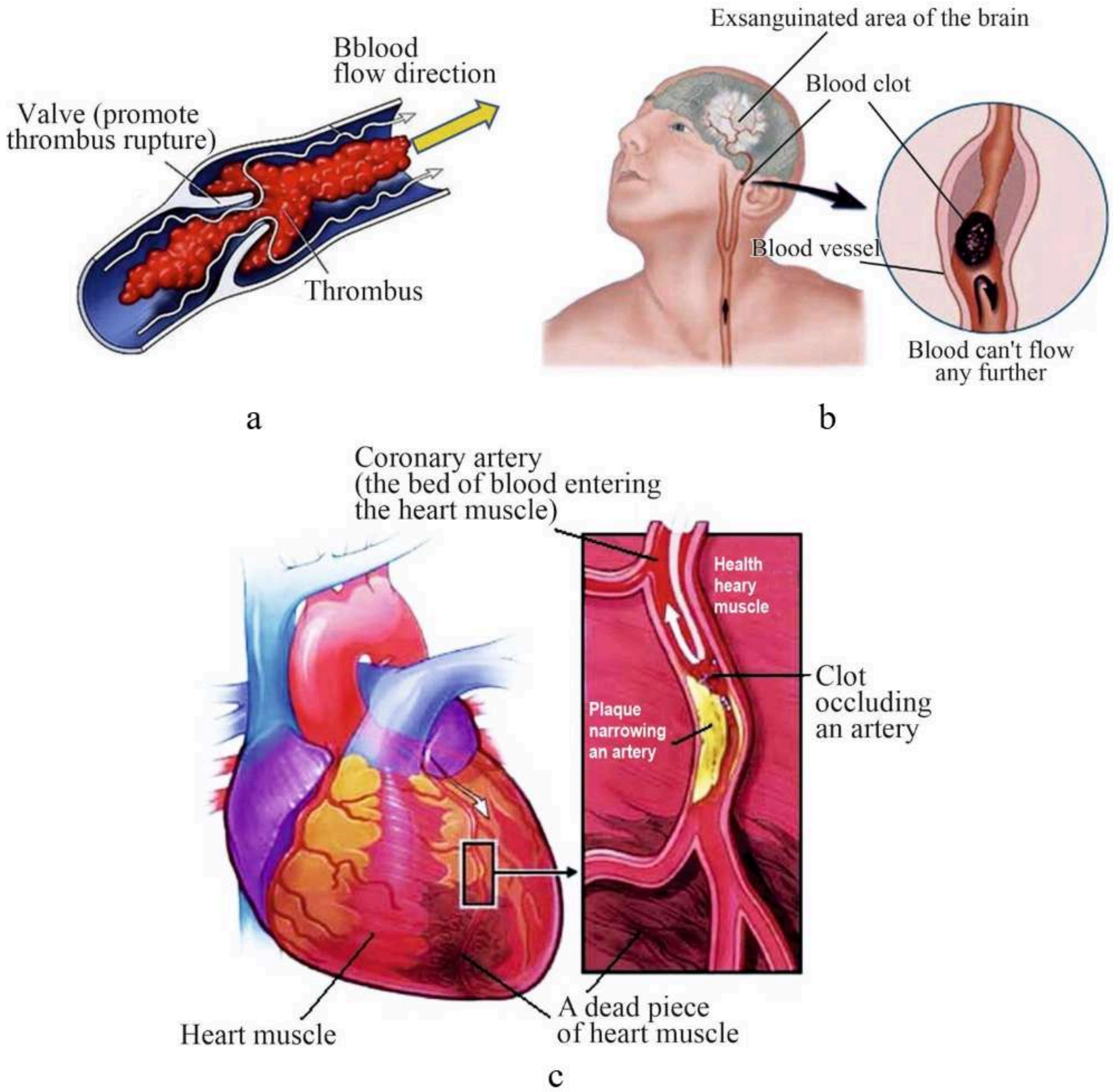


Fig. 2.19. The examples of possible consequences of changes in blood flow in human blood vessels (illustrations taken from various Internet sources)

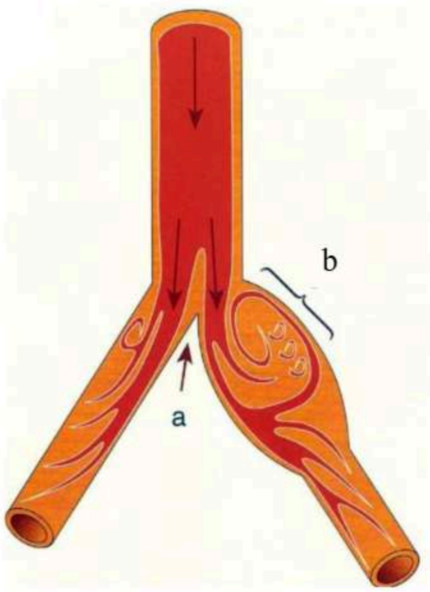


Fig. 2.20. Intravascular blood pressure distribution at the arterial bifurcation site, where blood flow is divided into two channels: in the zone (a) the vascular wall experiences the highest pressure and maximum shear stress is created; near the opposite vessel wall (b) flow whirling is created, blood flow speed decreases, shear stress minimum

SECTION 3. OSCILLATIONS AND WAVES

Sound is a physical phenomenon that is the propagation of mechanical oscillations of particles in a solid, liquid or gaseous medium in the form of elastic waves. In a narrow sense, sound refers to oscillations in the context of how they are perceived by the senses of animals and humans. Sound is characterized by amplitude and frequency (frequency spectrum). A common person is able to hear sound oscillations in the frequency range from 16÷20 Hz to 15÷20 kHz. Sound below the range of human hearing is called *infrasound*; above: up to 1 GHz – *ultrasound*, from 1 GHz – *hypersonic*. Usually for sound generation oscillating bodies of different nature are used, causing periodic changes in the pressure of elastic medium, such as air. Examples of such vibrations are the vibrations of the loudspeaker diffuser, tuning fork, vocal cords[**], elements of most musical instruments. Exceptions are wind instruments in which the sound is generated by air flow interacting with inhomogeneities in the instrument. Before turning to sound waves, let us consider the main characteristics of oscillatory processes (oscillations) on the example of harmonic oscillations.

[**] The vocal folds are folds of the laryngeal mucosa projecting into its cavity, containing the vocal cords and the vocal muscles. The spectrum of sound waves they produce is individual and can serve as an identification feature of a person, just like the papillary lines on the fingers of the hand used in fingerprinting.

3.1. Harmonic oscillations

The oscillation or oscillatory processes are motions of a system characterized by the repetitiveness (periodicity) of its passing the equilibrium position. In mechanics, the simplest oscillating system is a pendulum. One of such pendulums (one of the model systems) – a spring pendulum – is shown in Fig. 3.1. A weight of mass m can slide without friction on a rod passing symmetrically through a weight connected to a spring, the mass of which can be neglected. Being removed from the equilibrium position (point "O" on the x-axis), the weight will periodically pass through this equilibrium position.

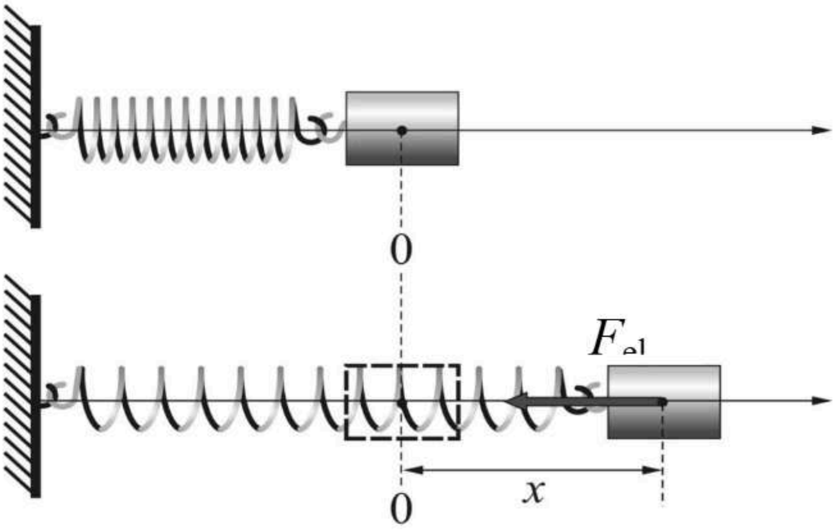


Fig. 3.1. Schematic representation of a spring pendulum

According to Hooke's law (Section 1.3.1) elastic forces F_{el}, arise in the spring, which are proportional to the displacement x: $F_{el.} = -kx$. The equation of motion of the body (weight) is written in the form of equation (3.1) or (3.1a):

$$m\ddot{x} = -kx \tag{3.1}$$

or

$$\ddot{x} + \omega_0^2 x = 0, \tag{3.1a}$$

where $\omega_0^2 = {}^{k}/_{m}$, $\ddot{x} = d^2x/dt^2 = a$.

A partial solution of equation (3.1a) is a harmonic function:

$$x = A\cos(\omega_0 t + \varphi_0) \qquad (3.2)$$

or

$$x = A\,sin\left(\omega_0 t + \varphi_0 + \frac{\pi}{2}\right), \qquad (3.2a)$$

where x is the displacement of the point from its equilibrium position, t is time, A is amplitude, ω_0 is the natural frequency, φ_0 is the initial phase of oscillations, $(\omega_0 t + \varphi_0)$ is the phase of oscillations at time t. The natural circular frequency of oscillations is related to the frequency ν_0 and the oscillation period T by the relations $\omega_0 = 2\pi\nu_0$ or $\omega_0 = \frac{2\pi}{T}$. Fig. 3.2a shows a graph of harmonic function (eq. 3.2) with an arbitrary initial phase φ_0. The value of the initial phase φ_0 will also change as the start of time changes. Let's take it as «0», and then the graph of the sine function will shift along the time axis to the reference point (Fig. 3.2b) by the value t_1.

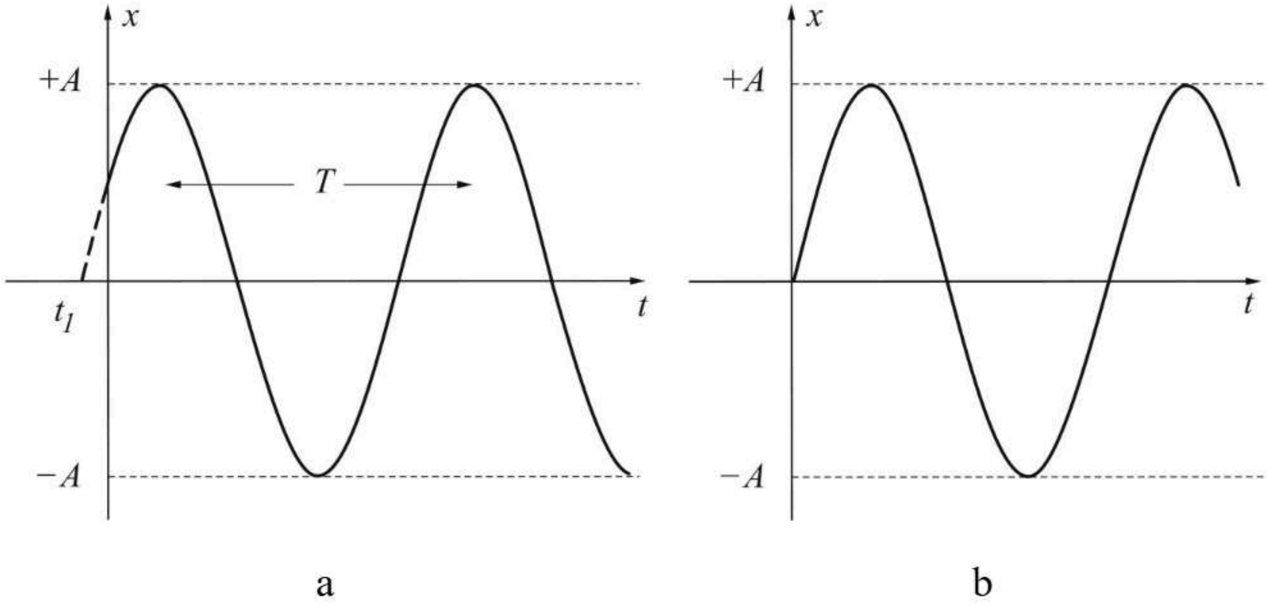
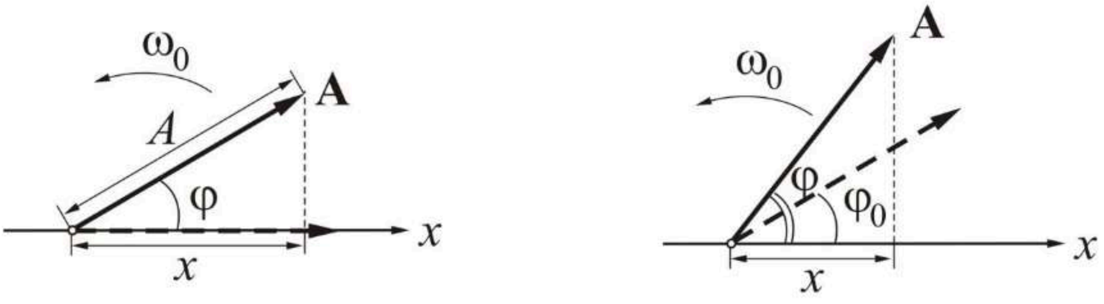


Fig. 3.2. The graph of the harmonic function (eq.3.2):
a) – with an arbitrary initial phase, b) – with zero initial phase

Harmonic oscillation can be represented as a rotation of some vector **A** with angular velocity ω_0 (Fig. 3.3).



$\varphi_0 = 0 \;\; x = A\cos(\omega_0 t)$

a

$\varphi = \omega_0 t + \varphi_0$

$x = A\cos(\omega_0 t + \varphi_0)$

b

Fig. 3.3. Vector model of harmonic oscillations: a) – oscillations with the initial phase equal to 0; b) – with the initial phase φ_0

Full rotation of the vector **A** is equal to 2π and corresponds to one period of the oscillations T. Its projection on the X-axis is determined by equation (3.2). If the rotation starts from a horizontal position (the projection of vector **A** is equal to its modulus A), it is assumed that the initial phase is 0, and if from a position inclined toward the X axis, the value of the inclination angle φ_0 is taken as the initial phase of oscillations. This angle corresponds to the time denoted in Fig. 3.2a as t_1 ($t_1 < T$). In rotational coordinates the initial phase is defined by the value $2\pi \cdot (t_1/T)$, ($2\pi = 360^{\circ}$). In any other position, it is the phase of oscillation at a given time. Thus, the phase is an angular measure of the deviation of a particle from its equilibrium position.

3.1.1. Damped oscillations

In the presence of frictional (resistance) forces, the differential equation of damped harmonic oscillations is obtained by adding the term $r\dot{x}$ to the equation of free harmonic oscillations (3.3)

$$m\ddot{x} = -kx - r\dot{x} \tag{3.3}$$

or

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = 0, \tag{3.3a}$$

here r is the drag coefficient; β is the damping coefficient $[\beta = \mathrm{r}/(2\mathrm{m})]$, ω_0 is the natural circular (angular, cyclic) frequency ($\omega_0 = \sqrt{k/m}$). When writing equations (3.3) and (3.3a), it is assumed that at low velocities of body (particle) motion, the drag force is proportional to the first degree of velocity – Stokes' law. Obviously, this assumption is true for small deviations from the equilibrium position, at which the solution of equation for undamped oscillations (3.2) is obtained. The solution of equation (3.3) is a harmonic function

$$x = A(t)\cos(\omega t + \varphi), \tag{3.4}$$

where the oscillation amplitude $A(t)$ decreases exponentially (A_0 is the amplitude at time $t = 0$) according to the relation

$$A(t) = A_0 e^{-\beta t}, \tag{3.5}$$

and the cyclic frequency ω and phase φ depend on the attenuation coefficient β

$$\omega = \sqrt{\omega_0^2 - \beta^2}. \tag{3.6}$$

The graph of the damped harmonic oscillations is shown in Fig. 3.4.

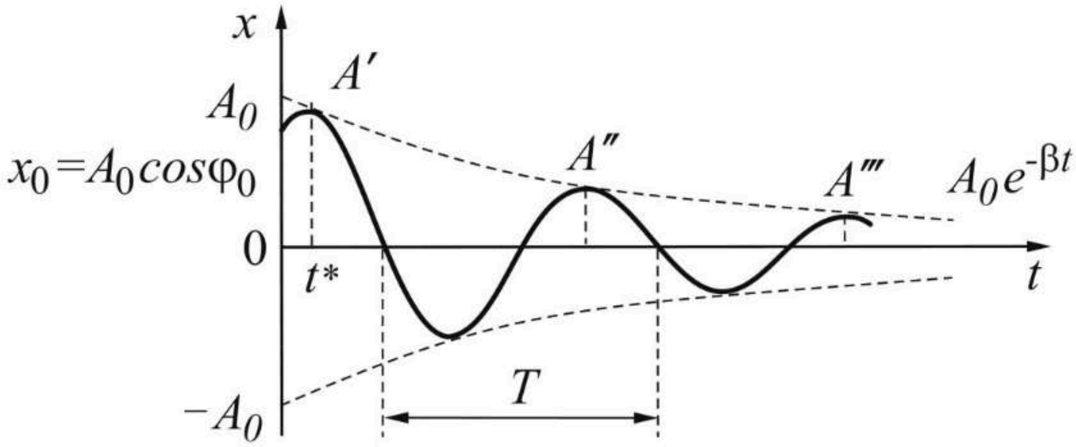


Fig. 3.4. Schematic representation of damped harmonic oscillations

To quantify the degree of attenuation, a value called the logarithmic decrement of oscillation attenuation is used:

$$\theta = \ln\frac{A(t)}{A(t+T)} = \beta T, \tag{3.7}$$

where $A(t)$ and $A(t+T)$ are the amplitudes of two successive oscillations, one period apart in time.

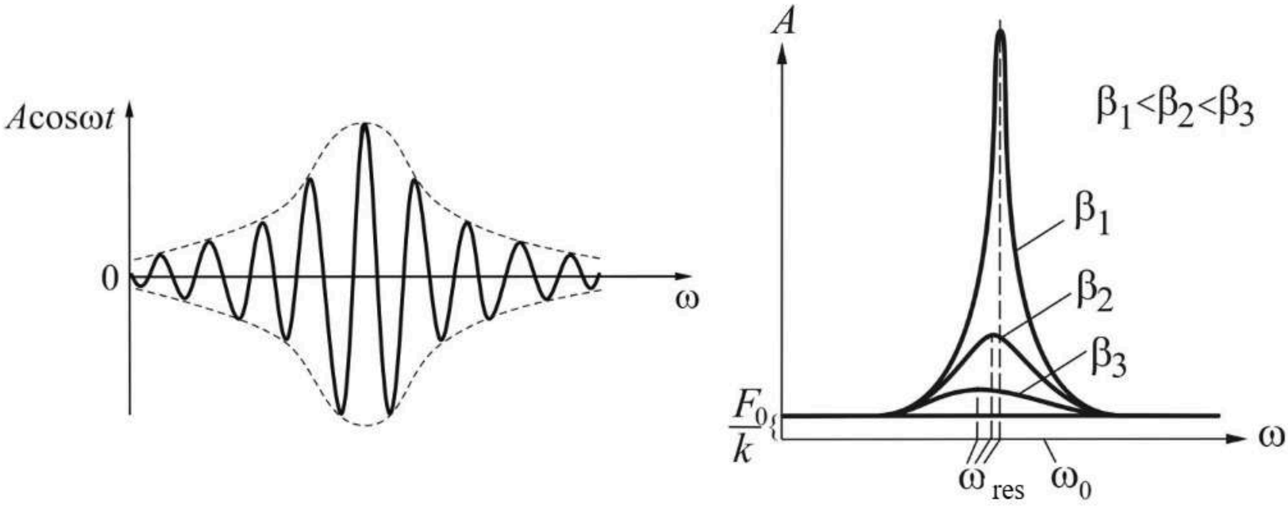
3.1.2. Forced oscillations

When a periodically varying force $F_0\cos\omega t$ is applied to the system, the equation of motion is written in the following form:

$$m\ddot{x} = -kx - r\dot{x} + F_0\cos\omega t \quad \text{or} \quad \ddot{x} + 2\beta\dot{x} + \omega_0^2 x = f_0\cos\omega t, \tag{3.8}$$

where $f_0 = {}^{F_0}/_{m}$.

Fig. 3.5a presents forced oscillations (solution of equation 3.8) in graphical form: these are also harmonic oscillations (harmonic function), but with frequency ω, the amplitude of which (the envelope of this function $A\cos\omega t$ is shown as a dotted line) depends on the frequency of natural oscillations ω_0 and the friction (resistance) forces, characterized by parameter β.



a

b

Fig. 3.5. a) – Graphic representation of forced oscillations when passing the condition of mechanical resonance; b) – the dependence of amplitude and frequency of mechanical resonance on the resistance of the medium

At a frequency ω_{res} close to the frequency of natural oscillations, namely, under the condition that

$$\omega_{\text{res}} = \sqrt{\omega_0^2 - 2\beta^2} \tag{3.9}$$

there is a sharp increase in the amplitude of oscillations. As can be seen in Fig. 3.5b, as the resistance of the medium increases, the resonant frequency decreases, the amplitude of oscillations also decreases:

$$A_{\text{res}} = \frac{F_0/m}{2\beta\sqrt{\omega_0^2-\beta^2}}. \tag{3.9a}$$

A number of phenomena associated with the superposition of oscillations will be considered when describing issues related to wave processes.

3.2. Sound waves (sound)

3.2.1. Flat longitudinal harmonic wave

The oscillations of medium particles are connected with disturbance of the equilibrium state of the system and are expressed in deviation of its characteristics from equilibrium values with subsequent return to the initial value. For acoustic oscillations such a characteristic is the pressure at a point in the medium, and its deviation from equilibrium value is the *sound pressure.*

If we make a sharp displacement of particles of elastic medium in some area of the medium, the density in this area will increase. Compaction will lead to an increase in intermolecular repulsive forces and, consequently, of pressure. As was shown in Section 2.1, due to elastic bonds between particles, neighboring particles are displaced, etc. Thus, the region of increased pressure is as though moving in the elastic medium. A region of increased pressure (compression of matter) is followed by a region of decreased pressure (rarefaction of matter), and thus a series of alternating regions of compression and rarefaction are formed, propagating in the medium as a wave. Each particle of the elastic medium in this case will perform oscillatory movements.

Since in liquid and gaseous media only compression (rarefaction) deformations take place, sound (acoustic) waves are longitudinal waves, in which the direction of oscillation of medium particles coincides with the direction of wave movement. In solids, in addition to longitudinal deformations there are elastic shear deformations, resulting from shear stress, which cause excitation of transverse (shear) waves, in which case the particles oscillate perpendicularly to the direction of wave propagation. The speed of propagation of longitudinal waves is much higher than the speed of propagation of transverse waves.

Consider the simplest type of the wave process: a longitudinal plane harmonic wave in which the particles of the medium oscillate according to the harmonic law: $\cos(\omega_0 t)$ or $\sin(\omega_0 t)$ (see Equations 3.2 and 3.2a, the initial phase is taken as 0).

Let us imagine that in an arbitrary point of space the local perturbation of the medium led to the oscillation of particles in this area, which is easily described by a harmonic function, and it propagates along the x-axis in isotropic medium without loss of oscillation energy (model case). This excitation will be transmitted from one point of the medium to another indefinitely. The deflection of oscillating points of the medium should be described by the function $\xi(x, t)$, which depends both on time and on the coordinate of the region of the medium x, to which the disturbance reached (to preserve uniformity in the description of wave processes, we use this notation for the direction of wave propagation along the x-axis). To describe wave propagation in an arbitrary direction, a radius-vector **r** is often used. Graphically, a plane longitudinal harmonic wave is shown in Fig. 3.6a. The distance between the two closest points oscillating in the same phase is called the wavelength λ. Since the excitation from one point of the medium to another is transmitted with a certain speed during the time of the complete oscillation of the particle T, we obtain an important relation between these three characteristics of the wave: $\lambda = \upsilon T$. The wave equation can be obtained as follows.

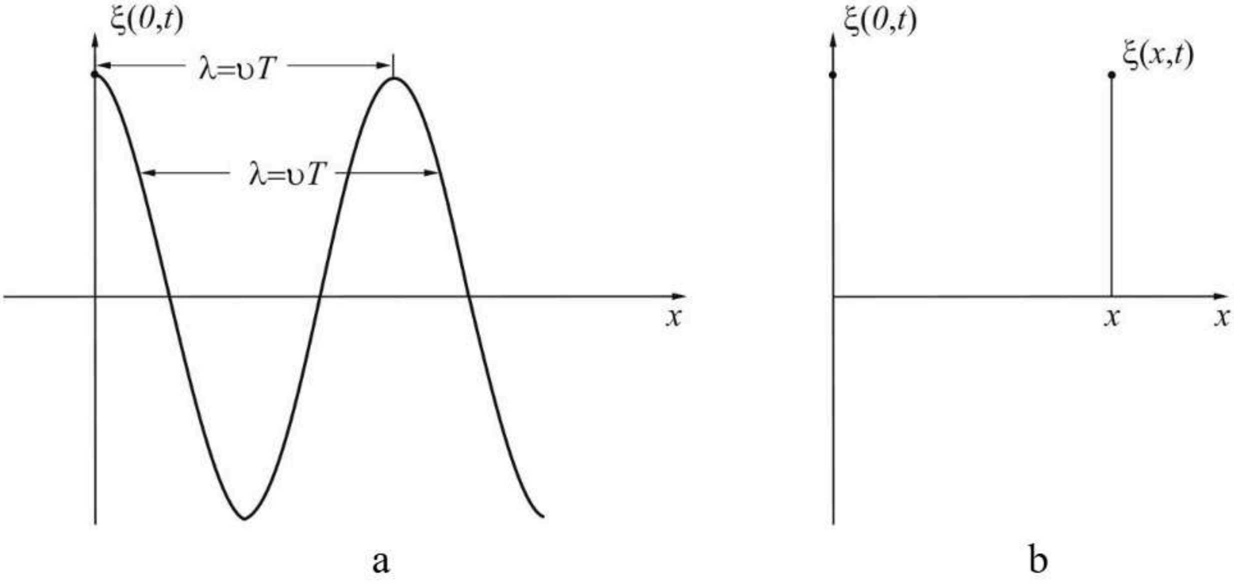


Figure 3.6. Graphical view of a cosine longitudinal plane wave

Let an oscillation described by the function $\xi(0,t) = A\cos(\omega_0 t)$ be excited at the origin (Fig. 3.6a). Propagating (without loss of energy) with velocity υ, this excitation will reach the region of space with coordinate x in time $t_1 = \frac{x}{\upsilon}$. Then the oscillation of particles of the medium at this point will be described by the equation:

$$\xi(x,t) = A\cos[\omega_0(t - t_1)] = A\cos\left[\omega_0\left(t - \frac{x}{\upsilon}\right)\right], \qquad (3.10)$$

which can be written in this form:

$$\xi(x,t) = A\cos(\omega_0 t - kx), \qquad (3.10a)$$

where k is the wave number ($k = 2\pi/\lambda$).

For a wave propagating in the opposite direction, the wave equation can be represented as:

$$\xi(x,t) = A\cos(\omega_0 t + kx)\,. \qquad (3.10b)$$

Let us fix the set of geometrical places of oscillations of particles of the medium with the same phase (they are called wave surfaces) by means of the relation $(\omega_0 t - kx) = \text{const}$. Differentiating over time leads to the equation

$$(\omega_0 - k\,dx/dt) = 0, \qquad (3.11)$$

from which it follows that

$$dx/dt = \frac{2\pi/\mathrm{T}}{2\pi/\lambda} = \upsilon. \qquad (3.12)$$

Thus, υ is the velocity of movement of wave surfaces (phases) and is therefore called phase velocity.

Figures 3.7a and 3.7b show longitudinal plane and spherical waves.

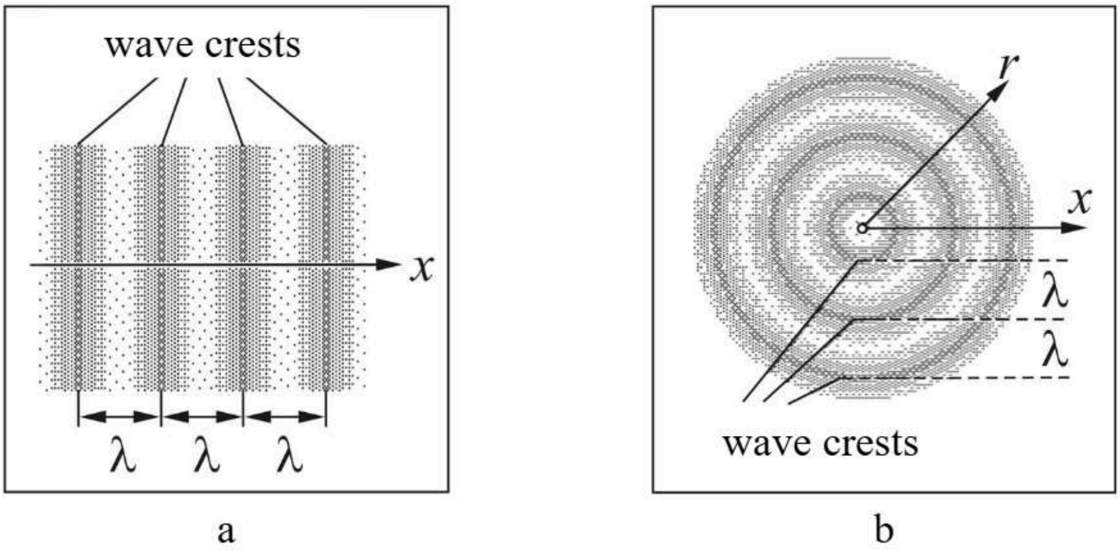


Fig.3.7. a) – Plane longitudinal wave, b) – spherical longitudinal wave; x, r – directions of wave propagation

3.2.2. Physical parameters of sound

The sound velocity is the speed of propagation of sound waves in a medium. As usual, in gases the speed of sound is less than in liquids, and in liquids the speed of sound is less than in solids, which is mainly due to the reduced compressibility of substances in these phase states. On average, the speed of sound in air is 340÷344 m/s. Sound velocity of longitudinal waves in elastic medium is calculated by formula

$$\upsilon = \sqrt{E/\rho}. \qquad (3.13)$$

Here E is the volume Young's modulus of compression (volume modulus of elasticity); ρ is the density of the medium (liquid or gas). *This formula also applies to the propagation of longitudinal waves in a thin metal rod. The speed of propagation of transverse waves in a solid is determined* *by the same formula, but instead of the Young's modulus (volume compression) E we must substitute the shear modulus G).*

In the ideal gas model $E = \gamma p$, where γ is the adiabatic index ($\gamma = \frac{c_p}{c_v}$ is the ratio of specific heat capacities of gas at constant values of pressure and volume). Then for the velocity of wave propagation we obtain the expression

$$\upsilon = \sqrt{\gamma p/\rho}. \tag{3.14}$$

Oscillational (acoustic) velocity υ_o is the speed at which the particles (infinitesimal parts of the medium), oscillating near the equilibrium position during the passage of the sound wave, move in relation to the medium as a whole. It is defined by the relation $\upsilon_o = A\omega$, where A is the amplitude of oscillations of the particles of the medium (the amplitude of the vibrational speed), through which the harmonic sound wave passes, ω is the cyclic frequency. For a sinusoidal wave, the oscillational velocity is equal to the modulus of velocity of oscillating particles of the medium at the moment of passing the equilibrium position – crossing the x - axis (the velocity is defined as the first derivative of the harmonic function 3.2 or 3.2a). The oscillational velocity should be distinguished both from the velocity of the medium itself and from the speed of propagation of a sound wave or the velocity of sound υ. Thus, $\upsilon_o < \upsilon$ in any media (gases, liquids, solids) and at any currently attainable sound intensities.

Acoustic impedance is the value that characterizes the elastic properties of the medium and is determined by the ratio of the amplitude of the sound pressure in the medium to the amplitude of the vibrational velocity of its particles. The greater the acoustic impedance, the higher the degree of compression and rarefaction of the medium at a given amplitude of oscillation of the medium's particles. Numerically specific acoustic resistance of the medium Z is equal to

$$Z = \rho\upsilon. \tag{3.15}$$

The units of acoustic resistivity: in the SI system – Pa·s/m, and in the CGS system – din·s/cm³ (1 Pa·s/m = 10^{-1} din·s/cm³).

The apparent acoustic impedance of a medium is often expressed in g/s·cm², with 1 g/s·cm² = 1 din s/cm³.

Sound or acoustic pressure is the difference between the instantaneous pressure at a given point in the medium in the presence of sound vibrations and the static pressure at the same point in the absence of them. In other words, sound pressure is a variable pressure in the medium due to acoustic oscillations. The maximum value of the variable acoustic pressure (pressure amplitude) p can be calculated through the amplitude A of particle oscillations (amplitude of a oscillational velocity):

$$p = p_0 = 2\pi\nu\rho\upsilon A, \tag{3.16}$$

where ν is the frequency of oscillations. At a distance of half the wavelength (λ/2) the sound pressure value changes its sign: positive becomes negative. The pressure difference at the two points with its maximum and minimum values (i.e. separated from each other by λ/2 along the wave propagation direction) is $2p$. The pressure units are given in Section 2.1.

Sound intensity is the time-averaged energy carried by a sound wave through a unit area perpendicular to the direction of wave propagation per unit time. It is defined by the relation

$$I = \frac{pA}{Z} = \frac{p^2}{2\rho\upsilon} = A^2\rho\upsilon(2\pi\nu)^2. \tag{3.16a}$$

It is measured in units of W/m^2 in the SI system, erg/s in the CGS system, or measured on a decibel scale (see below).

If traveling sound waves collide with an obstacle, it is not only undergoing alternating pressure, but also constant pressure. The areas of densification and rarefaction of the medium arising as the sound waves pass through create additional pressure changes in the medium relative to the ambient external pressure. Such additional external pressure is called radiation pressure. It causes that when sound waves pass through the boundary of liquid with air, fountains of liquid are formed and individual droplets are detached from the surface. This mechanism has found application in the preparation of drugs in the form of aerosols. Radiation pressure is often used to measure the power of ultrasonic vibrations in special devices – ultrasonic balances.

Sound loudness is a physiological characteristic – the subjective perception of sound intensity (the absolute value of the auditory sense). Loudness is primarily a function of sound pressure (3.16) or sound intensity (3.16a). Sound volume is also affected by its spectral composition (timbre), the duration of exposure to sound vibrations, the individual sensitivity of the human ear, and other factors. Among the audible sounds should be highlighted phonetic, speech sounds and phonemes[*] (that make up spoken speech), and musical sounds (which consists of music). Musical sounds contain not one, but several tones, and sometimes also noise components in a wide range of frequencies. The perception of sound depends naturally on the human organ of hearing, one of the five human senses – the ability to perceive the sounds of the world around us. Through hearing, a person distinguishes between sounds, recognizes their nature, locates the source of sound, and learns to speak.

The hearing organ consists of the outer, middle, and inner ear (Fig. 3.8).

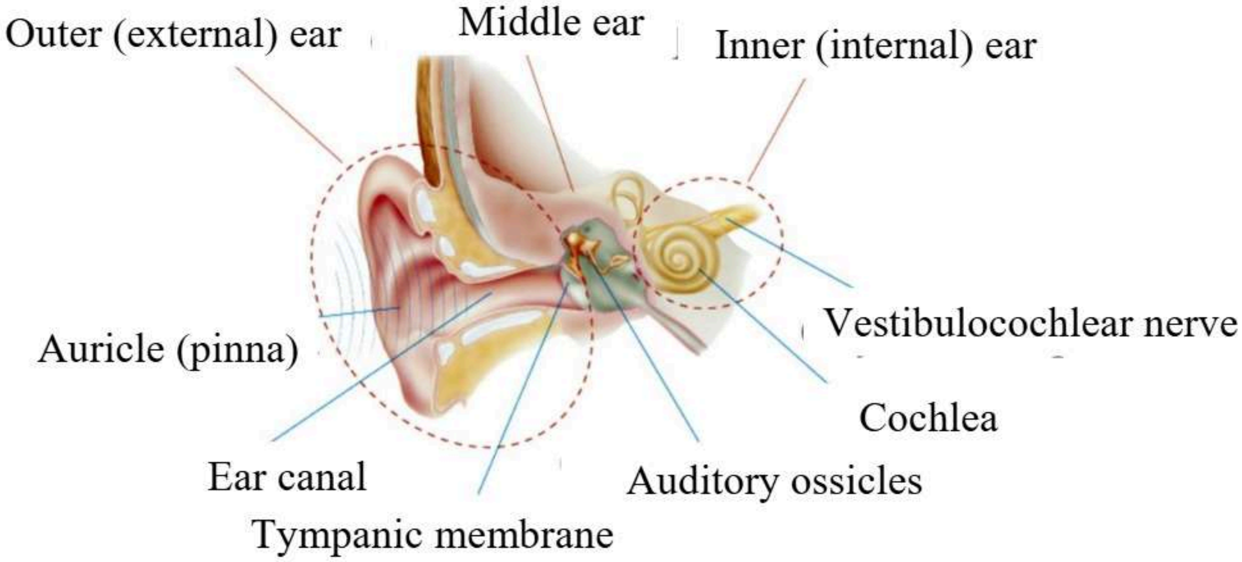


Fig. 3.8. Schematic representation of the hearing organ [5]

---

[*] Phonemes (from Greek φώνημα – sound) are the minimal meaningful units of a language.

How the healthy ear works:

- Sound signals enter the ear canal and cause the eardrum to vibrate,
- The vibrations of the eardrum move the three auditory ossicles (the other elements are omitted),
- Through the bones the vibrations are transferred to the fluid in the inner ear (cochlea),
- The movement of the fluid in the cochlea stimulates tiny hair cells,
- The hair cells convert mechanical vibrations into electrical signals,
- Electrical signals are passed through the spiral ganglion cells and the auditory nerve to the cerebral cortex.

The brain perceives and analyzes the received signals as sound. Human perception of sound is measured in decibels.

Decibel is a dimensionless unit used to measure the ratio of physical quantities: "energy" (sound intensity, power, energy, power flux density, etc.) or "power" (sound pressure, current, voltage, etc.). A logarithmic scale of attenuation and gain levels is used. The value expressed in decibels is numerically equal to the decimal logarithm of the dimensionless ratio of a physical quantity to the same physical quantity taken as a reference, multiplied by ten. If you use the ratio of "energy" quantities, in this case, sound intensity levels (sound power levels) I, then:

$$L_{dB} = 10\lg(I/I_0), \tag{3.17}$$

where I_0 is the value corresponding to the zero intensity level ($I_0 = 1\ \mathrm{pW/m^2}$).

If "power" quantities are used, in this case the sound pressure p, the ratio is used:

$$A_{dB} = 20\lg\frac{p}{p_0}, \tag{3.17a}$$

because $I \sim p^2$. Here p is the measured sound pressure value, p_0 is the basic sound pressure value ($2 \cdot 10^{-5}$ Pa at a frequency of 1 kHz is taken as the basis).

In other words, a decibel L_{dB}, A_{dB} is not an absolute value like a watt or a volt, but a relative value like a ratio or a percentage, designed to measure the ratio of two other quantities. A logarithmic scale is applied to the resulting ratio. The decibel is not an official SI unit, although the General Conference of Weights and Measures decided to allow its use without restriction together with SI units, and the International Bureau of Weights and Measures recommended to include it to the SI system.

The hearing threshold is the minimum sound pressure at which a sound of a given frequency is perceived by the human ear. The hearing threshold is expressed in decibels. The hearing threshold of a particular person depends on individual characteristics, age, physiological condition.

The threshold of pain is the amount of sound pressure at which pain occurs in the auditory organ (which is associated in particular with reaching the limit of extensibility of the eardrum). The sensation of pain determines the limit of the dynamic range of human hearing, which on average is 140 dB for a tone signal and 120 dB for noise with a continuous spectrum. Exceeding this threshold results in acoustic trauma.

An objective characteristic of the hearing organs is an *audiogram* – the dependence of sound perception on the frequency or volume level (sound pressure). Fig. 3.9 shows one of such audiograms as an example.

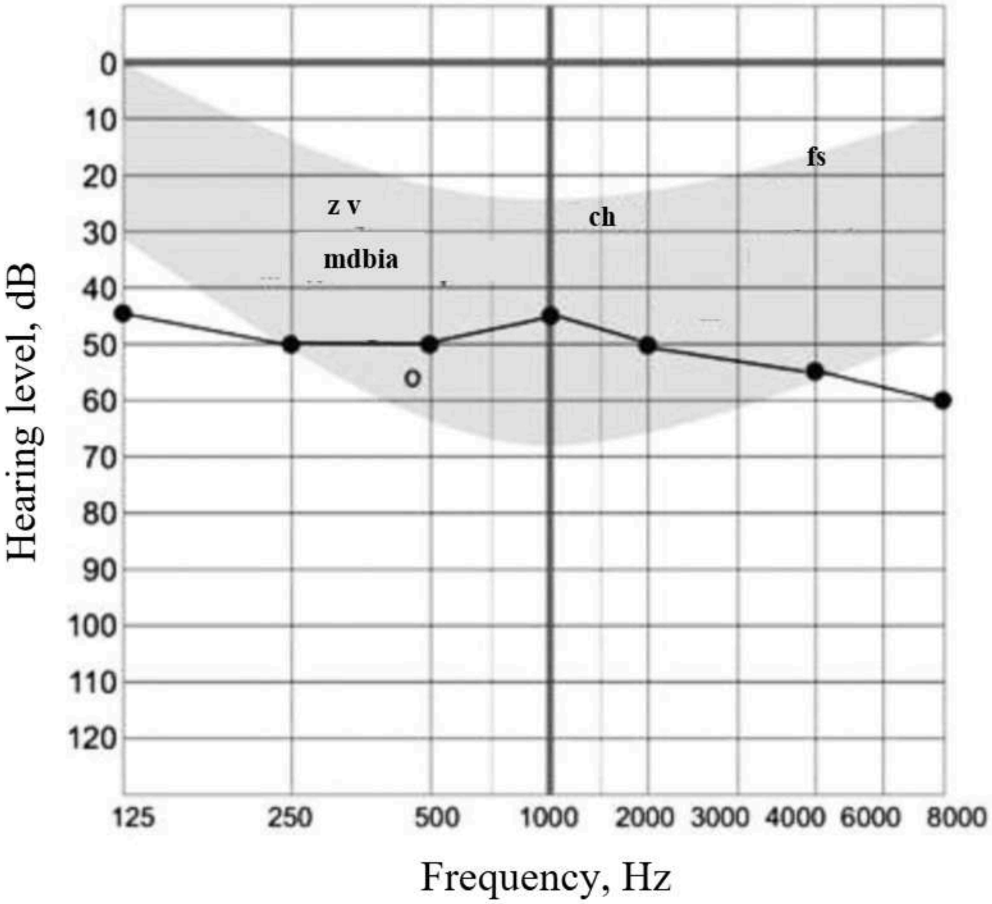


Fig. 3.9. The example of a tonal audiogram of the human hearing organ. It shows the dependence of sound perception (threshold of hearing on the frequency at the level of noise, highlighted in gray). That is, the perception of tonal noise (tonal sound), characterized by a single frequency or narrowband audibly distinguishable against the background of the general noise. The solid line is the perception of a fixed-frequency sound

3.2.3. Sound propagation

Diffraction (enveloping of obstacles by waves) takes place when the length of the sound wave is comparable (or larger) with the size of the obstacle in its path. If the obstacle is large compared to the length of the acoustic wave, there is no diffraction.

Interference is observed in the superposition of so-called coherent waves (of the same frequency and a constant phase difference), provided that the particles of the medium oscillate along one direction. It is manifested in the fact that at certain points of the medium there is an increase or decrease of oscillations, i.e. there is a redistribution of energy of the resulting oscillations of the particles of the medium. (If the frequencies do not coincide, there is a uniform distribution of the energy density of oscillations). In this case, the oscillation of points of the medium depends on the ratio of the phases of the oscillations of these waves at a given point. If the sound waves reach a certain point in the medium in the same phase (in-phase), the shifts of the particles have the same signs and the interference in such conditions leads to an increase in the amplitude of oscillations. If waves arrive to the point of medium in antiphase the displacement of particles will be differently directed that leads to decrease of vibration amplitude. These issues are described in detail in various sections of the general physics course.

When ultrasound passes through the boundary of two media in the general case there is reflection, refraction and transformation of the wave (longitudinal wave is transformed into transverse wave). *Wave transformation occurs only when transverse waves can propagate in these media (e.g., in metals). But this does not occur in the case of normal incidence to the boundary. Naturally, wave transformation does not occur when there is reflection at the boundary of two media in which transverse waves do not propagate (gas, liquid, tissue).*

Sound reflection (as applied to medicine). The phenomenon of reflection is the basis of ultrasound diagnosis and such acoustic diagnostic methods as auscultation and percussion. Reflection occurs in the border areas of skin and fat, fat and muscle, muscle and bone. If ultrasound on its way encounters organs whose dimensions are larger than the wavelength, there is refraction and reflection of ultrasound. Inhomogeneities in body tissues do not cause significant deviations: compared with the length of the ultrasonic wave (about 2 mm) probable inhomogeneities – about 0.2 mm (on any of the linear dimensions) are too small to change the direction of the wave front (the waves go around them, diffract). Moreover, they are negligibly small at longer waves in the sound range[*]. Reflectance, or reflection coefficient

$$R = \frac{I_{reflect.}}{I_{\text{falling}}} = \left(\frac{Z_1 - Z_2}{Z_1 + Z_2}\right)^2 \qquad (3.17b)$$

is determined by the differences in the values of acoustic resistance of the contacting tissues Z_1 and Z_2. Obviously, R varies from 1 to 0 (this formula is valid for the normal incidence of the sound wave on the interface). Thus, the greater the difference in the values of the acoustic impedance of the media, the less energy passes through the boundary separating the media. Air has a low density (hence low acoustic impedance). Therefore, at the boundary of media material – air there is almost complete reflection of ultrasound. (For example, on the border of steel – air reflection coefficient is equal to 1, (100%), ultrasound is reflected from the air layer, Z_2). *Formally, if we swap Z_1 and Z_2 in formula 3.17b, we get the same value of R. However, this approach is not correct, since the reflection from a less dense medium and the reflection from a denser medium occurs with different changes in the phase of oscillations, which is considered when deriving formula 3.17b).* This feature of ultrasound reflection from media with low density is the basis of the method of defectoscopy of materials. Note that the order of magnitude of the acoustic impedance of the media correlates as 1 (gases):10^3 (liquids, biotissues):10^6 (metals). Values of acoustic resistance of biotissues are close (differ only by 10-20%). However, this difference is quite enough to register the reflected ultrasonic waves. This is the basis of using ultrasound in medicine. It should be noted that ultrasonic study of tissues uses special water-based gels whose acoustic

---

[*] Sound wave length range: in air – from 1.7 m to 0.017 cm, in water – from 7.25 m to 0.073 cm. In the interaction of sound waves with inhomogeneities of any size, including biological tissues, there is also sound scattering. But to a greater extent it is observed when the size of inhomogeneity is small compared to the wavelength. For example, scattering on red blood cells (~ 7 microns). It should be noted that wave scattering on inhomogeneities is a general phenomenon characteristic of waves of any nature.

resistance differs slightly from the acoustic resistance of tissues in order to remove air from the gap between the transmitter body and the tissue. Ultrasonic methods of determining the location and size of stones in human tissues and organs are based on the reflection of ultrasound from denser media.

Ultrasound refraction. Acoustic resistance of human soft tissues is not much different from the acoustic resistance of water contained in them. Therefore, the reflection coefficient of ultrasound is small and most of it goes to the second medium (refracted if the angle of incidence deviates from the normal). This circumstance is taken into account in the development of various ultrasonic techniques. Consideration of these issues is beyond the scope of this textbook.

Standing sound waves. If there are tissues with different specific acoustic impedances on the path of propagation of ultrasonic waves, reflection from the boundary interface occurs to a greater or lesser degree. The superposition of incident and reflected waves can result in standing waves if there is no change in the frequency of the reflected wave and the distance from the transmitter surface to the reflecting surface is a multiple of half the wavelength. The standing wave equation (simply obtained by adding harmonic functions (3.10) and (3.10a)) is written in the form:

$$\xi(x,t) = 2A\cos kx \cdot \cos \omega t. \qquad (3.18)$$

That is, in the resulting wave the particles oscillate about the equilibrium position with an amplitude whose value depends on the coordinate x $(2A\cos kx)$. Beams – areas of the medium with the greatest deviation of its particles from the equilibrium position, and knots (no oscillations of the particles) are formed: the wave is localized in the region of space from the emitter to the reflector of the radiation. Hence the name "standing wave" – the excitation does not move beyond the area between the emitter and the wave reflector. The distance between the beams (standing wave length) is equal to half the running wave length. The positions of the co-oscillating points at different moments of time separated by half-period T/2 are shown in Fig. 3.10. Fig. 3.11 shows the direction of oscillation of standing wave particles.

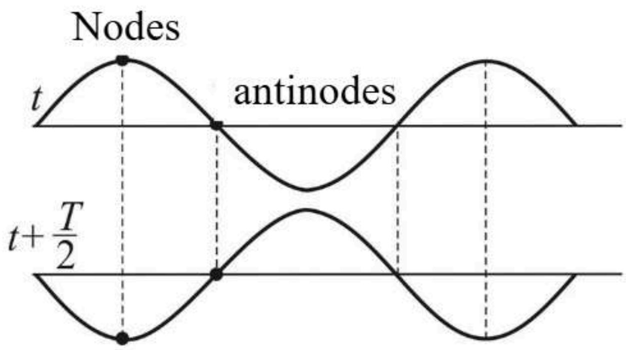


Fig. 3.10. The positions of the oscillating points of the standing wave: top – when deflections reach the maximum absolute value, bottom – after time T/2

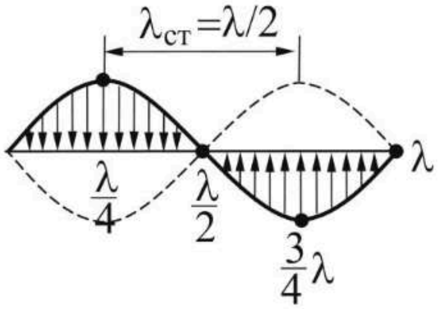


Fig. 3.11. The direction of particles oscillations in standing wave

Beats. The oscillations of particles of a medium in some area when superposing, for example, traveling and reflected waves with slightly different frequencies ($\Delta\omega \ll \omega$) can be considered as the addition of oscillations with close frequencies and written in the following form (to simplify calculations, we take the initial phases of oscillations to be zero):

$$x = A\cos(\omega t)\,, \tag{3.18a}$$

$$x = A\cos\{(\omega + \Delta\omega)t\}. \tag{3.18b}$$

When they are added together, the so-called beats of oscillations are observed. The equation of the resulting oscillations is written in the form:

$$x = 2A\cos\left(\frac{\Delta\omega}{2}t\right)\cos\left\{\left(\omega + \frac{\Delta\omega}{2}\right)t\right\} \qquad (3.19)$$

or, because the frequencies do not differ much:

$$x = \left\{2A\cos\left(\frac{\Delta\omega}{2}t\right)\right\}\cos(\omega t). \qquad (3.20)$$

Thus, the resulting oscillation is a harmonic oscillation at the frequency ω, but with an amplitude (highlighted by curly brackets) that also changes according to the harmonic law with a frequency $\frac{\Delta\omega}{2}$ (periods of $T^* = \frac{4\pi}{\Delta\omega}$), see upper Fig. 3.12. But the amplitude cannot be negative, so we take its modulus $\left|2A\cos\left(\frac{\Delta\omega}{2}t\right)\right|$, see bottom Fig. 3.12. It is perceived as an oscillation with an amplitude pulsating with the frequency Δω or with a period of beats $T_b = \frac{2\pi}{\Delta\omega} = \frac{T^*}{2}$.

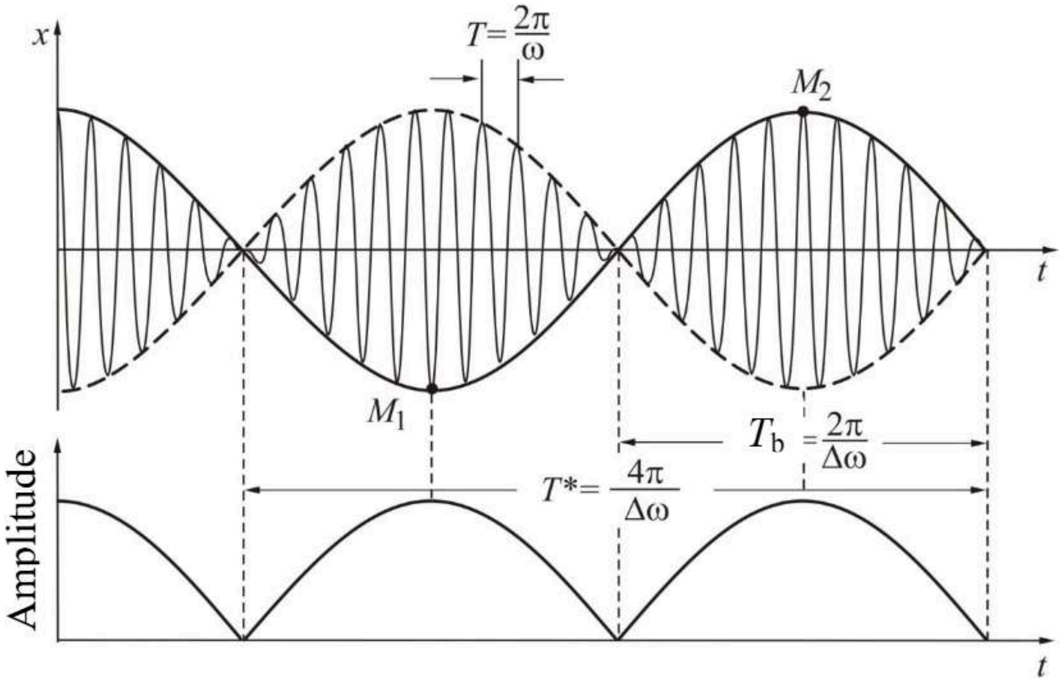


Fig. 3.12. Graphical representation of the oscillation beat

In the context of this book, this phenomenon is used in medical applications (Doppler effect, see below).

Problems related to the addition of oscillations are considered in detail in the course of general physics in the section "Oscillations and waves" (see, for example, [1, 4]).

Next, let us consider issues related to the attenuation of the wave process that occurs in a real medium.

Wave attenuation. A decrease in the intensity of a sound wave (for a harmonic wave, a decrease in amplitude) as it propagates is due to a number of reasons. The main ones are: wave divergence, scattering and absorption of sound. Attenuation due to wave divergence is due to the fact that at large distances r from the source, the flow of emitted sound energy as it propagates is distributed over an increasing area of the wave surface and, accordingly, the flow of energy per unit surface decreases. In a spherical wave, the intensity I decreases with distance in proportion to r^{-2} ($I = N/(4\pi r^2)$), where N is the radiation power of the point source (for a cylindrical case – $\sim r^{-1}$, when propagating a longitudinal harmonic plane wave, this factor can be neglected). Scattering of sound on obstacles in the medium, on its inhomogeneities, the size of which is small or comparable to the wavelength, leads to a decrease in energy flow in the original direction of sound propagation. Scattering in gases occurs on liquid droplets (fog) or particles of solids (aerosols), in liquids – on air bubbles, in solids – on various foreign inclusions or individual crystallites in polycrystals. Scattering on irregularities and inhomogeneous areas of the medium's boundary leads to a decrease in the intensity of sound reflection and its passage through the boundary, i.e. leads to a change in the reflection coefficient and wave propagation. Sound absorption – irreversible transition of sound energy into other types of energy (mainly into heat) is caused by various mechanisms. In the tissues of living organisms, viscosity and thermal conductivity of the medium play a major role. Absorption also depends on the frequency of sound vibrations – it is proportional to the square of the frequency. A detailed discussion of this issue is beyond the scope of this book.

Let us note only that at sound attenuation due to scattering and absorption, the amplitude of a plane harmonic wave decreases according to the law e^{-dr} (r is the direction of wave propagation), and the intensity of the sound wave (proportional to the square of the amplitude) decreases with distance according to the law e^{-2dr}, where d is the effective attenuation coefficient. I.e., the decrease of intensity at wave divergence occurs according to the power law, as noted at the beginning of the subsection, and is caused by scattering and absorption – according to the exponential law.

Attenuation coefficient is expressed in units of m^{-1} (cm^{-1}) or in logarithmic units Np/m, dB/m. Np is also a dimensionless logarithmic unit to measure the ratio of the two quantities. 1 Np is approximately equal to 8.6 dB, the international designation is Np.

Acoustic Doppler effect. Doppleroscopy. The acoustic Doppler effect is a change in the frequency of the perceived sound (receiver) as the sound source and receiver move in relation to each other. Probably anyone is familiar with manifestation of this effect in everyday life: change of siren sound of the object in the beginning approaching to the observer and then moving away from him (sound of siren of diesel locomotive, steam train at crossings, fire truck, ambulance, police, and just engine sound of fast moving car). Doppler effect is known for a long time, the physics of the phenomenon and its applications in detail elaborated in the special literature. General information is contained in [1]. For educational purposes, we recommend studying the material in [1-5].

There are a total of 4 variants of the Doppler effect:

1. Source of sound (transmitter) – a generator of sound waves approaching a stationary observer. Hereinafter ν is the frequency of oscillations, υ is the speed of sound propagation, ν' is the perceived frequency, $\upsilon_{trans.}$, $\upsilon_{rec.}$ are the speeds of the source and the receiver of oscillations, respectively:

$$\nu' = \frac{1}{1-\frac{\upsilon_{trans.}}{\upsilon}} \cdot \nu. \tag{3.21a}$$

2. The source of the sound (transmitter) is moving away from the resting observer:

$$\nu' = \frac{1}{1+\frac{\upsilon_{trans.}}{\upsilon}} \cdot \nu. \tag{3.21b}$$

3. The observer (receiver) approaches the resting source of sound:

$$\nu' = \left(1 + \frac{\upsilon_{rec.}}{\upsilon}\right) \cdot \nu. \tag{3.21c}$$

4. The observer (receiver) moves away from the resting source of sound:

$$\nu' = \left(1 - \frac{\upsilon_{rec.}}{\upsilon}\right) \cdot \nu \quad . \tag{3.21d}$$

Let's imagine a blood vessel, on some part of which we need to determine the speed of blood movement $\upsilon_{b.}$ (Fig. 3.13).

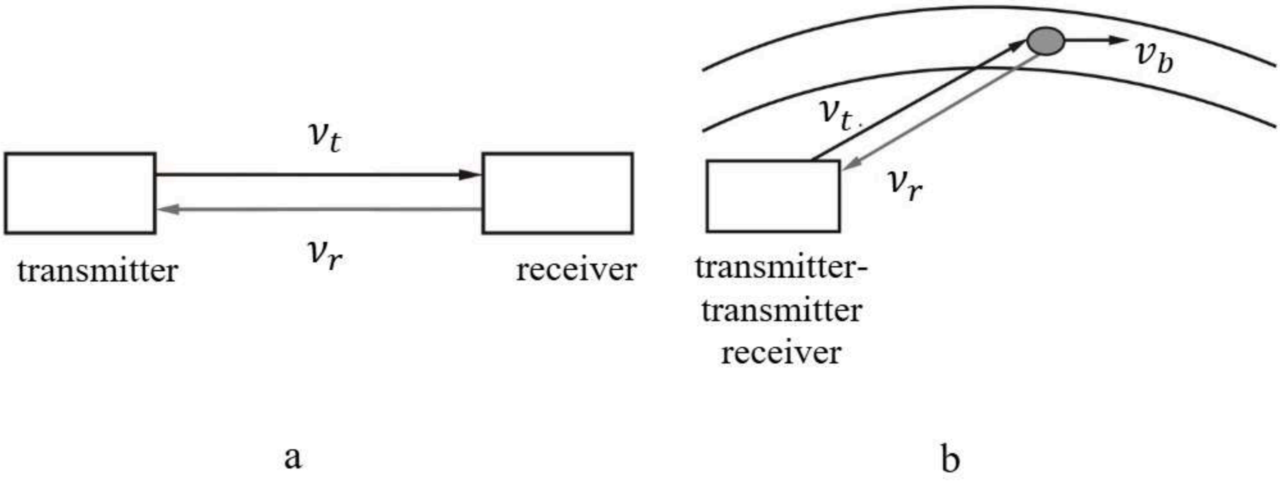


Fig. 3.13. Schematic representation of Doppler effect in a blood vessel: a) – general scheme, b) – in a blood vessel

From the ultrasonic source, a beam of ultrasonic waves with a frequency of $\nu_{\text{trans.}}$. is guided to the blood vessel. On some moving with speed υ_{b} macroscopic blood element, such as a clot of erythrocytes (as long as it is a remote "receiver" moving away from the source), the ultrasound would be perceived at the frequency defined by the relation (3.21g) in the form:

$$\nu' = \left(1 - \frac{\upsilon_{\text{b.}}}{\upsilon}\right) \cdot \nu_{\text{trans.}}, \tag{3.22}$$

where the frequency given on the right is the transmitter frequency $\nu_{\text{trans.}}$ On this element of the blood clot there is a reflection of ultrasonic waves in different directions, including the direction to the receiver. Thus, this macroscopic blood element becomes the "source" of ultrasound (with frequency ν', determined by equation (3.22)), which moves away from the receiver with the speed $\upsilon_{\text{b.}}$. Now let's use equation (3.21b) and get expression for frequency $\nu_{rec.}$ of the received sound signal:

$$\nu_{\text{rec.}} = \frac{1}{\left(1+\frac{\upsilon_{\text{b.}}}{\upsilon}\right)} \cdot \nu' = \frac{1}{\left(1+\frac{\upsilon_{\text{b.}}}{\upsilon}\right)}\left(1 - \frac{\upsilon_{\text{b.}}}{\upsilon}\right) \cdot \nu_{\text{trans.}} \tag{3.23}$$

Here we substitute the frequency ν' from equation (3.23). Equation (3.23) is easily converted to equation

$$\text{x}\nu_{\text{rec.}} = \frac{\upsilon-\upsilon_{\text{b.}}}{\upsilon+\upsilon_{\text{b.}}} \cdot \nu_{\text{trans.}}, \tag{3.24a}$$

from which it follows that $\Delta\nu = \nu_{\text{trans.}} - \nu_{\text{rec.}} = \upsilon_{\text{b.}}(\nu_{\text{rec.}}+\nu_{\text{trans.}}) \cdot \upsilon^{-1}$. But $\Delta\nu \ll \nu_{\text{trans.}}, \nu_{\text{rec.}}$. Let us assume that $\nu_{\text{trans.}} \approx \nu_{\text{rec.}} = \nu$. Then:

$$\Delta\nu = \nu_{\text{trans.}} - \nu_{\text{rec.}} \approx 2\upsilon_{\text{b.}}\nu_{\text{trans.}}/\upsilon. \tag{3.24b}$$

Thus, the addition of these ultrasonic waves with close frequencies leads to the resulting oscillation at the carrier frequency ν, with the amplitude also changing according to the harmonic law with the frequency 2Δν (see oscillation beats), detected by the device. I.e. the value Δν, called frequency shift, is proportional to the velocity of this macroscopic blood element:

$$\nu_{\text{trans.}} - \nu_{\text{rec.}} = D\upsilon_{\text{b.}}, \tag{3.25}$$

where D is the normalization factor.

3.2.4. Sound investigation methods. Auscultation and percussion

Auscultation (in Latin *auscultatio*, listening) is a physical method of medical diagnosis that involves listening to the sounds produced during the functioning of internal organs. It was proposed by R. Laennec in 1819.[*] Auscultation can be direct-carried out by holding the ear to the listened organ, and indirect – with a *stethoscope* or *phonendoscope* (Fig. 3.15) [1][**].

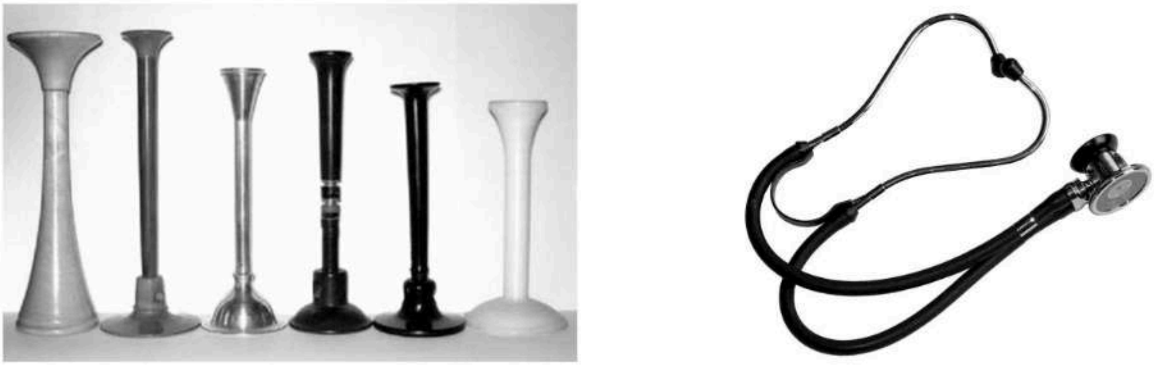
---

[*] ***Rene Laennek,*** (1781–1826), French physician and anatomist, inventor of the stethoscope, a pioneer of clinical and anatomical diagnostics.

[**] The stethoscope (from Greek στηθοσκόπιο: στῆθος – chest + σκοπή – examination) is an elongated tube in the form of a thin hollow cylinder with one end wider than the other and a concave shell for the ear. The doctor places the ear to the wide end of the stethoscope and listens to the patient's internal organs (lungs, heart, bronchi, intestines, etc.) for any sounds or hums. The analog of the stethoscope that was used to auscultate the heart was just a rolled up piece of paper.

Phonendoscope (from Greek Φωνη – sound, ένδον -inside and σκοπέω – observe) is a more modern stethoscope. Its main advantage over the stethoscope is that you can listen to high-pitched sounds due to the tightly stretched membrane, which is sensitive to sound vibrations and amplifies them. The term "phonendoscope" was suggested by ***N. S. Korotkoff.*** Stethophonendoscope is a combined version of stethoscope and phonendoscope. This device allows you to listen to both low and high sounds. It is what modern doctors use. It consists of three main parts: a sound-receiving device – a head (capsule) with a membrane, – a tube that conducts the sound, – tips for the doctor's ears (olives).

The stethophonendoscope is often referred to simply as a stethoscope or phonendoscope with this modern modified version in mind.



a b

Fig. 3.15. Early stethoscopes (a), modern stethoscope (b)

Percussion (*lat. Percussio*) – blow, tapping – one of the main objective physical methods of examination of the patient, consisting in tapping areas of the body and determining the nature of the resulting sound depending on the physical properties of the organs and tissues located under the percussed area (mainly their density, "airiness"[*] and elasticity). The method was invented in 1761 by L. Auenbrugger. The physical description was given by the Czech physician J. Škoda in 1839.

Physical aspects of the method. When tapping an area of the body, sound oscillations of the media are produced. High and low percussion sounds are distinguished depending on the frequency of oscillations (it also depends on the impact force). The height of the sound is directly proportional to the density of the medium (with the same impact force). Thus, when percussing the chest, low sounds are produced over low-density airy lung tissue, and high sounds are produced over dense heart tissue. The loudness of the sound is proportional to the amplitude of vibration, which is higher the greater the percussion force, and usually the percussion sound is louder and longer the less dense the tissues percussed. Low density tissues or organs include those containing large amounts of air. Percussion of the lungs with normal "airiness" gives a clear, i.e. long enough and loud percussion sound. When the "airiness" of the lung tissue decreases

---

[*] There is no strict physical definition. We should understand airiness as a measure of air content in tissues (a measure of its heterogeneity) and organ cavities.

(atelectasis), for example in inflammatory infiltration, as a result of which inflammatory fluid accumulates in the lungs, its density increases, and the percussion sound becomes quiet and short (obtuse). Obtuse percussion is caused by the percussion of dense bones, muscles, fluids in cavities, organs such as the liver, heart, and spleen.

When percussing relatively large cavities filled with air (stomach, or loop of gut, accumulation of air in the pleural cavity), there is a musical sound, in which the main tone dominates. This sound is similar to the sound of beating a drum, so it is called tympanitic (tympanon-drum) or tympanic percussion sound. A characteristic feature of the tympanic sound is the ability to change the pitch of the basic tone with changes in the tension of the cavity walls (air pressure in it). For example, tympanitis disappears with increasing pressure in the pleural cavity with a valve pneumothorax, and the percussive sound is firstly thumb tympanic and then non-tympanic. Thus, when percussing different parts of the body of a healthy person, you can get three basic percussion sounds: clear, obtuse and tympanic. A clear percussion sound occurs with percussion of normal pulmonary tissue. Obtuse percussion is observed when percussing areas under which there are dense media, airless organs (heart, liver, spleen), massive muscle groups (on the thigh – "hip stiffness"). Tympanic sound is produced by percussion of the areas adjacent to the air cavities; in a healthy person, it is detected above the place where the air-filled stomach is adjacent to the chest (the so-called Traube space).

Percussion techniques. Depending on the method of percussion, there are direct and indirect percussion. Direct percussion is performed by striking the surface of the body with the fingertips. In indirect percussion, the finger or hammer blows are applied to another finger or plessimeter – a special plate made of metal, wood, plastic or bone (plēssō, Greek – to strike + metreō – to measure), placed on the body. (Plessimeter and hammer appeared in the first half of the 19th century).

Among the methods of direct percussion we know the methods of Auenbrugger, Obraztsov, Janowski. Auenbrugger covered the place to be percussed with a shirt or put a glove on his hand and tapped the chest with the outstretched fingertips, making slow, not strong strokes. This method is practically not used because it does not give clear and accurate results. V.P. Obraztsov proposed a more efficient method, which is still used quite often. It consists in the following: "Preliminarily the index finger-mallet is placed flat on the back surface of the middle finger, pressing it strongly, then the finger-mallet as if slips off the middle finger and the flesh of the nail phalanx with force strikes on the skin of the tapped area". The left hand spreads the skin folds of the percussed area and limits the sound propagation. F.G. Yanovsky used one-finger percussion, in which percussive strokes are made with minimal force by the pulp of two (terminal and adjacent) phalanges of the right middle finger. Direct percussion is used to determine the borders of the liver, spleen, absolute "obtuse heart", especially in pediatric practice and in emaciated patients.

Methods of mediated percussion include finger tapping on the plessimeter, hammer tapping on the plessimeter, and so-called finger bimanual percussion. The priority of introducing finger bimanual percussion belongs to G.I. Sokolsky, who struck one or two fingers of the left hand with the tips of two or three fingers of the right hand folded together. Gerhardt suggested finger percussion; it was universally accepted. The advantage of this method is that the doctor, along with sound perception, receives with the finger plessimeter a tactile sense of the resistance force of the percussed tissues. During finger percussion, the middle finger of the left hand (serving as a plessimeter) is placed flat against the place under study, the remaining fingers of this hand are spread apart and barely touch or do not touch the body surface. The end phalanx of the middle finger of the right hand, bent at the first joint almost at right angles, is struck (like a hammer) on the middle phalanx of the plessimeter-finger. To obtain a clear sound, uniform, jerky, short strokes directed vertically to the surface of the plethysmograph finger are applied. During percussion, the right hand, bent at right angles at the elbow, only bends and extends at the wrist joint to deliver percussive strokes.

When examining areas of the body whose surface does not allow the plessimeter-finger to be pressed tightly to the skin along its length, percussion is performed according to the Plesh method: the plessimeter-finger is bent at right angles in the first interphalangeal joint and pressed to the skin by the end of the nail phalange, blows are applied with a hammer finger along the main phalange.

Depending on the strength of the percussion, there are strong (loud, deep), weak (quiet, superficial), and medium percussion. Strong percussion is used to identify deep organs and tissues (thickening or cavity in the lung at 5–7 cm from the chest wall). Medium percussion is used to determine relative heart obtuse and liver[*].

Quiet percussion is used to find the boundaries of absolute obtuse heart and liver, lung and spleen, small pleural exudates and superficially located lung masses. The so-called silent (minimal), limiting percussion is made with such weak strokes that the sound is in the threshold of perception of the ear – threshold percussion. It is used to determine more accurately the absolute heart sound pressure; the tapping is done in the direction from the heart to the lungs.

There is also auscultatory percussion. It involves listening to percussion with a stethoscope placed on the opposite side of the chest wall to the organ being percussed (lungs) or above the organ being percussed (liver, stomach, heart) at the point of contact with the abdominal or chest wall.

---

[*] *Obtuse heart* (*lat. obtusio cardiaca*) is the area of the chest wall where the heart area is identified by the blunting of the sound during percussion. Usually determine the boundaries of cardiac obtuseness: right, upper and left (the lower boundary is not defined, because the bottom of the heart is not adjacent to the lungs). In addition, the character of the sound distinguishes between relative and absolute obtuse hearts.
Absolute cardiac obtuseness (*lat. absoluta*) – Cardiac obtuseness detected by quiet (silent) percussion. It corresponds closely to the projection of the part of the heart not covered by the lungs onto the pedicle wall of the chest. Approximately corresponds to the projection of the part of the heart not covered by the lungs on the front wall of the chest. Relative cardiac obtuseness (*lat relativa)* – Cardiac obtuseness detected by percussion along the intercostal spaces with sharp percussive strokes strictly in the sagittal direction. This roughly corresponds to the projection of the entire heart, including the part covered by the lung, onto the front wall of the chest.

Weak percussion strokes (or strokes on the skin with the fingertip) are made on the body from the place of contact with the stethoscope towards the edge of the examined organ. As long as the percussion strokes are made within the organ, the percussion sound is heard clearly: as soon as the percussion goes beyond the examined organ, the sound is sharply muffled or disappears completely. A detailed description of these methods and consideration of questions of their clinical application are beyond the scope of this textbook.

3.3. Ultrasound and fauna

Bats that use echolocation for night orientation (echolocation is a method by which the position of an object is determined by the time delay of the reflected wave return), emit signals of extremely high intensity either with their mouth (the family "leatherback" – Vespertilionidae), or with a parabolic mirror-shaped nasal opening (horseshoe-nosed Rhinolophidae). At a distance of 1÷5 cm from the head of the animal the pressure of ultrasound reaches 60 mbar, i.e. corresponds in the audible frequency range to the pressure of the sound produced by a jackhammer. Bats are able to perceive echoes of their signals at a pressure as low as 0,001 mbar, i.e. 10,000 times lower than the pressure of the emitted signals. Bats are able to avoid obstacles in flight even when ultrasonic signals are overlaid with 20 mbar pressure. The mechanism for this high noise immunity is unknown. When bats localize objects, e.g. vertically stretched threads only 0,005÷0,008 mm in diameter at a distance of 20 cm (half the wingspan), time shift and difference in intensity of emitted and reflected signals play a decisive role. Horseshoe bats can also orient themselves with the help of only one ear (monoaurally) due to the continuously moving auricles. They are even able to compensate for the frequency shift between emitted and reflected signals due to the Doppler effect. By lowering the emitted (echo-locating) frequency during the flight so that the frequency of the reflected ultrasound remains in the area of maximum sensitivity of their "auditory" centers, they can determine the speed of their own displacement.

Nocturnal butterflies of the Bearfly family have an ultrasonic noise generator that "throws off the trail" the bats stalking these insects.

Echolocation is also used to navigate by the birds, the fat goatees, or guajaros. They inhabit the mountain caves of Latin America, from Panama in the northwest to Peru in the south and Suriname in the east. Living in pitch-black darkness, they have nonetheless adapted to flying masterfully through caves. They make soft clicking sounds, which the human ear can also perceive (their frequency is about 7 kHz). Each click lasts for one or two milliseconds. Click sound is reflected from the walls of the dungeon, various protrusions and obstacles, and is perceived by the sensitive hearing of birds. Cetaceans use ultrasonic echo-location in water.

3.4. The application of ultrasound in biology, medicine and other fields. A brief review

The range of used ultrasound frequencies is quite wide: from 0,5 to 15 MHz. The lowest frequencies – up to 1 MHz are used in head examination, 1,5÷3 MHz – in obstetrics and gynecology and in examination of abdominal cavity internal organs, 2÷5 MHz – in cardiology, the highest – from 5 to 15 MHz – in ophthalmology and in examination of bone and joint apparatus. It is important to note that the almost complete reflection of ultrasonic waves at the boundary between the air and soft tissues (99.95%) prevents ultrasound from entering the body through the air. To eliminate this obstacle, the skin in the place of its contact with the ultrasound wave transducer is lubricated with a special contact gel, or in its absence – with vaseline oil or even water.

Sonography. Due to proper distribution of ultrasound in human soft tissues, relative harmlessness in comparison to X-rays and ease of use, ultrasound methods are widely used for visualization of human internal organs. Ultrasound is used to study many internal organs, but only ultrasonic study of the brain and heart is isolated in separate disciplines: echoencephalography and echocardiography. Moreover, there are more studies on echocardiography than on echoencephalography. This is probably due to the fact that the heart is in constant motion and the overwhelming majority of echocardiographic techniques and signs are related to the analysis of its different structures of movement – valves, heart cavity walls and walls of large vessels. Therefore, the information obtained here is much richer than in the study of immobile organs. The issues of Doppleroscopy have been discussed above.

Ultrasonic therapy. In addition to its widespread use for diagnostic purposes, ultrasound is used in medicine (including regenerative medicine) as a treatment method. This is due to the following factors:

- anti-inflammatory, dissolving action;
- analgesic, antispasmodic action;
- cavitational enhancement of skin permeability.

Phonophoresis – is a combined treatment method in which a therapeutic substance (both medications and substances of natural origin) is applied to tissues instead of the usual ultrasound emission gel (used, for example, in ultrasonic studies). It is known that ultrasound leads to deeper penetration of the drug substance in the tissue. In this case, the smaller the frequency of ultrasound, the greater the depth of penetration, as the coefficient of attenuation of the wave is directly proportional to the frequency. For therapeutic purposes, ultrasound with a frequency of 0,8 to 3 MHz is used. In the human body, ultrasound at a frequency of $0{,}8 \div 1$ MHz propagates to a depth of $8 \div 10$ cm, and at a frequency of $2{,}5 \div 3{,}0$ MHz to $1{,}0 \div 3{,}0$ cm.

Ultrasound is absorbed by tissues nonuniformly: the higher the acoustic density, the lower the absorption.

There are three factors that affect the human body during ultrasound therapy:

- mechanical (vibratory micro-massage of cells and tissues);
- thermal (increase of tissue temperature and cell membrane permeability);

- physical and chemical (stimulation of tissue exchange and regeneration processes).

Biological effect of ultrasound depends on its dose (intensity), which can be stimulating, depressing or even destructive for tissues. The most adequate for therapeutic and preventive effects are small doses of ultrasound (up to 1,2 W/cm^2), especially in the pulsed mode. Ultrasound has analgesic, antiseptic (antimicrobial), co-vasodilating, resolving, anti-inflammatory, desensitizing (anti-allergic) effect.

Ultrasound therapy is not used on the brain area, cervical vertebrae, bony protrusions, areas of growing bones, tissues with a pronounced circulatory disorder, on the abdomen during pregnancy, scrotum. With caution, ultrasound is used on the heart and endocrine organs.

Ultrasound in cosmetology. Multifunctional cosmetology devices generating ultrasonic vibrations at a frequency of 1 MHz are used to regenerate skin cells and stimulate their metabolism. With the help of ultrasound is made a micromassage of cells, activating blood microcirculation and lymphatic drainage. As a result, it increases the tone of the skin, subcutaneous tissues and muscles. Ultrasonic massage promotes the release of biological active substances, eliminates muscle spasm, resulting in smoothed wrinkles, tightens tissue facial and body. Ultrasound promotes not only a deeper introduction of cosmetics and preparations, but also removes toxins and cleans cells.

Ultrasound in biology. The ability of ultrasound to destroy cell membranes is used in biological research, for example when it is necessary to separate enzymes from the cell. Ultrasound is also used to disrupt intracellular structures such as mitochondria and chloroplasts to study the relationship between their structure and function. Another application of ultrasound in biology is related to its ability to cause mutations. There are reports that ultrasound even of low intensity can damage DNA molecules. Artificial targeting of mutations plays a major role in plant breeding. The main advantage of ultrasound over other mutagens (X-rays, ultraviolet rays) is that it is extremely easy to work with. A discussion of these special issues is beyond the scope of this book.

Ultrasonic cleaning. Application of ultrasound for mechanical cleaning is based on the appearance of various nonlinear effects in the liquid under its influence. These include cavitation, acoustic flows, sound pressure. The main role is played by cavitation. Its bubbles appearing and collapsing near contaminants destroy them. This effect is known as *cavitation erosion.* The ultrasound used for this purpose has a low frequency and a high power. In the laboratory and production conditions ultrasonic baths filled with solvent (water, alcohol, etc.) are used for washing small parts and dishes. Sometimes with their help even root crops (potatoes, carrots, beets, etc.) are washed from soil particles [1].

3.5. Infrasound

Infrasound (from Latin *Infra* – below, under) is an elastic wave similar to sound but with a frequency lower than that perceived by the human ear. The upper limit of the infrasound frequency range is usually 16÷25 Hz. The lowest boundary of the infrasound range is conventionally defined as 0.001 Hz. Vibrations from tenths and even hundredths of hertz, i.e. with periods of tens of seconds, may be of practical interest. The nature of infrasound oscillations and audible sounds are the same, so infrasound follows the same laws. The same mathematical apparatus is used to describe them as for ordinary audible sound (except for the concepts related to the sound level). Infrasound is weakly absorbed by the medium, so it can travel considerable distances from the source. Infrasound has a number of features associated with the low frequency of oscillation of the elastic medium:

- has larger amplitudes of oscillation than acoustic waves of equal power;
- propagates much farther in the air, as its absorption in the atmosphere is negligible;
- because of its very long wavelength it is diffractive and therefore penetrates easily and evades obstacles that block audible sounds;

- causes the oscillation of large objects, leading under certain conditions to their mechanical resonance.

The mentioned features of infrasound make it difficult to reduce the negative effects of its effects, since conventional methods of noise elimination are ineffective.

Sources of infrasound. Natural. Infrasound occurs during earthquakes, lightning strikes, strong winds (infrasound aerodynamic noise) during storms and hurricanes. Whales and elephants communicate with each other using infrasound.

Technogenic. Technogenic infrasound is generated by oscillation of equipment with large surfaces, powerful turbulent flows of liquids and gases, impact excitation of structures, rotational and reciprocating movement of massive parts of mechanisms. The main technogenic sources of infrasound are heavy machine tools, wind power plants, fans, electric arc furnaces, reciprocating compressors, turbines, vibrating platforms, spillway dams, jet engines, ship engines. In addition, infrasound occurs in surface, underwater and underground explosions.

Infrasound propagation. Infrasound is characterized by low absorption in various media, so infrasound waves in the air, water and in the earth's crust can propagate over very long distances, and it can serve as a harbinger of storms, hurricanes, tsunamis. This phenomenon is of practical use in determining the location of strong explosions or the position of the firing gun. Sounds of explosions containing a large number of infrasound frequencies are used to study the upper layers of the atmosphere, the properties of the aquatic environment.

Physiological effects of infrasound. The physiological effect of infrasound on humans depends only on its spectral, temporal characteristics, power and does not depend on whether a person is in an open space or indoors.

Pathogenic effect of infrasound is to damage the structures of the brain, endocrine glands and internal organs due to the development of tissue hypoxia due to liquor-hemodynamic and microcirculatory disorders. At 180÷190 dB infrasound action leads to rupture of pulmonary alveoli and is usually fatal. Other zones of intense short-term impacts cause a syndrome of severe infrasound discomfort, the limit of tolerance of which is noted already at 154 dB. Researches showed that low-frequency acoustic vibration, including infrasound frequency range from 1 to 100 Hz, during the period from 25 s to 2 min with sound pressure from 145 to 150 dB caused a feeling of the chest wall vibration in the subjects. Dry mouth, visual disturbances, headache, dizziness, nausea, coughing, choking, subcostal restlessness, tinnitus, modulation of speech sounds, pain while swallowing are noted. Some other signs of disturbances in the activity of the body are also observed. It is noted that infra-sound negatively affects eyesight: visual functions deteriorate, visual acuity decreases, the field of vision narrows, accommodation ability weakens, the stability of fixation of the object observed by the eye is broken. Noise at the frequencies of 2÷15 Hz at the intensiveness level of 100 dB results in the increase of tracking errors of indicator pointers. Convulsive twitching of the eyeball, disorders of balance organs function are manifested. Pilots and astronauts exposed to infrasound during training were slower to solve even simple arithmetic problems. There is an assumption that various anomalies in people's condition during bad weather, explained by climatic conditions, are in fact a consequence of infrasound waves. It is believed that the negative impact of infrasound is due to the fact that in the infrasound region are the frequencies of natural vibrations of some organs and human body parts. This causes undesirable resonance phenomena. Let's indicate some natural frequencies of human body vibrations: for the whole body when lying down – 3÷4 Hz, thorax – 5÷8 Hz, abdomen – 3÷4 Hz, eyes – 12÷27 Hz.

Infrasound is particularly harmful to the heart. At a sufficiently high power there are forced vibrations of the heart muscle. At resonance (at frequencies of 6÷7 Hz) their amplitude increases, which can cause bleeding into the heart muscle.

In recent years, infrasound has become more widely used in medical practice. Thus, in ophthalmology infrasound waves with frequencies up to 12 Hz are used in the treatment of myopia. In the treatment of eyelid diseases, infrasound is used for phonophoresis, as well as for cleansing wound surfaces, to improve hemodynamics and regeneration in the eyelids, massage, etc.

MOLECULAR PHYSICS

Human activity is explicitly or implicitly related to the physical phenomena and processes that form the basis of molecular physics. The questions of human life activity are certainly broader and more complex than questions like: can you boil eggs, meat, and other foods in the mountains, can you make tea there, etc. The answer is trivial – of course you can. On the highlands there are villages, monasteries (e.g., Tibetan), training bases for various purposes, weather and geological stations, astronomical observatories, etc., and even large cities[*]. There is another aspect – the activity of living organisms in conditions of extremely low temperatures, in areas of permafrost and snow, etc. Even under these conditions, people normally work and are served by medical personnel of various levels and catering services. People born under such conditions are genetically adapted to them, and the centuries-old experience of survival under such difficult conditions is handed down from generation to generation. Today travel is available to almost all income and age groups, the number of fans of extreme recreation is increasing. That is why knowledge of the rules of conduct in non-standard conditions is extremely important. *Things that previously were the privilege of a small group of well-trained athletes (diving, mountaineering, downhill skiing, slalom, etc.) is now the mainstream of business tourism*.

The life of organisms is determined by temperature to a greater extent than any other factor of the environment due to the fact that all organisms are built of chemical components and all life processes occur on the basis of chemical reactions, subject to the laws of thermodynamics. Temperature determines not only the speed of chemical reactions but is also the reason for the structural reorganization of proteins, phase transformations of fats and changes in water structure. The vital temperature zone

---

[*] The most well-known of them is Mexico City, 2,240 meters above sea level. The highest-altitude city in the world (mining town) is in Peru, La Rinconada (5099 m). Next is Namche Bazaar in Nepal (4150 m) and El Alto in Bolivia (4150 m).

is the interval within which most microorganisms actively carry out their vital functions; it fits, with a few exceptions, into the interval between 0 °C and 50÷60 °C. The lower limit of active life activity of microorganisms is limited, first of all, by the level of droplet-liquid water, whose constant flow in the cell supports the three-dimensional structure of protein molecules and other structural carriers of life and ongoing processes. Therefore, the crystallization of water in washing liquids and cells serves as a critical threshold for their life. However, while the upper threshold of the vital zone, which is determined by thermal coagulation of proteins, is rather narrow, the lower boundary of the vital zone is wider and "blurred" due to many direct and indirect adaptations to retaining a part of water in the liquid state, developed in organisms in the course of evolution. Survival of microorganisms after deep cooling is a well-known fact. (*Fans of winter fishing are well aware of this – frozen fish in warm water come to life*). Storage of biomaterials in laboratories at deep low temperatures of -50÷-60°C and below is well known. The most striking example is the storage and transportation of sperm in Dewar vessels at liquid nitrogen temperature (-195,7 °C) in technologies of artificial insemination widely used in animal husbandry. These and related issues constitute a wide area of research and practical applications of *cryobiology*[*].

Chapters 4–6 in this part of the course address a number of issues

---

[*] Cryobiology (from Greek κρύος – cold, bios – life and logos – science) is a branch of biology that studies the effects of low temperatures on living organisms. In practice, cryobiology studies biological objects or systems at temperatures below normal. Proteins, cells, tissues, organs, or whole organisms can serve as objects. The temperature range is from moderately low to cryogenic (ultra-low). Research principles of cryobiology were laid in the late 19th century by P. I. Bakhmetiev, the outstanding Russian scientist, who studied hypothermia in insects and anabiosis in bats. The results of cryo-biological studies are applied in the storage of plants, cell cultures, embryos for IVF, storage of rare fish eggs, etc. The most important areas of research:

- Study of cold adaptations of microorganisms, plants and animals (hibernation);
- Cryopreservation of cells, tissues, gametes and embryos of animal and human origin for long-term preservation for medical purposes, usually using substances that protect cells during freezing and thawing (cryoprotectors);
- preservation of organs in hypothermic conditions for transplantation;
- lyophilization of pharmaceuticals;
- cryosurgery – surgery using cryogenic gases/liquids to destroy tissues.

that constitute the subject area of molecular physics and its place in the understanding of human life. Molecular physics is a branch of physics that studies the molecular form of motion of matter, the motion of a large ensemble of molecules. A stricter definition: molecular physics is a branch of physics that studies the physical properties of bodies depending on their structure, the forces of interaction between the molecules, atoms, ions that form the body, and the nature of thermal motion.

It follows from this definition that the dynamic method of research, which is used to describe mechanical phenomena, is not applicable in molecular physics, since it does not seem possible to distinguish the motion of a single molecule from the multitude of molecules (there are about 10^{19} in one cubic cm of gas). Therefore, to describe the phenomena of molecular physics, statistical or thermodynamic methods are used, which are actually developed for these purposes.

The statistical method consists in the establishment of regularities between statistical quantities characterizing a set of molecules forming the system. The method is the basis of the Kinetic Molecular Theory of Matter.

The thermodynamic method is the determination of the relationship between the macroscopic parameters of the system (thermodynamics).

This circumstance predetermined the conditional division of this part of the course called "Molecular Physics" into three sections:

- Kinetic Molecular Theory of Matter based on the statistical approach;
- Thermodynamics;
- Description of the properties of physical systems and the processes occurring in them.

SECTION 4. KINETIC MOLECULAR THEORY OF MATTER AND CERTAIN ISSUES OF HUMAN ACTIVITY

4.1. Definitions. Introduction to the Molecular Kinetic Theory of Matter

Systems, states, processes

To a certain extent, the concept of "aggregate state" can be used to characterize matter:

- Solid state, the substance retains its shape;
- liquid state, the substance retains its volume, taking the form of a vessel;
- gaseous state, the substance retains neither shape nor volume.

However, under certain conditions, the same substance can be in solid, liquid and gaseous (in the form of vapor) states. For example, water in the form of ice, liquid and vapor.

A stricter definition of the state of matter is contained in the concept of "phase".

A phase is a macroscopic physically homogeneous part of matter separated from the rest of the system (in mechanics it is a set of considered bodies) by interfaces so that it can be extracted from the system mechanically.

The system may contain several solid phases, two liquid phases, a solid and a liquid phase, but only one gaseous phase.

The simplest system – gas is characterized by a set of three macroscopic parameters: pressure p, volume V and temperature T. The equation establishing the relationship between them is called the equation of state:

$$f(p, V, T) = 0. \tag{4.1}$$

The equilibrium state of a system is the state when macroscopic parameters in all parts of the system are the same. Important for the following discussion of the thermodynamics of living systems is the stationary state, when its parameters do not change with time: $p, V, T \neq f(t)$.

Transition from initial state (1) to final state (2)

$$p_1, V_1, T_1 \rightarrow p_i, V_i, T_i \rightarrow p_2, V_2, T_2, \quad (4.2)$$

where any intermediate states are characterized by equilibrium parameters p_i, V_i, T_i, is called an equilibrium (or quasi-static) process.

If the process can be reversed through the same intermediate equilibrium states, the process is called reversible. This is represented schematically as follows:

$$p_1, V_1, T_1 \leftrightarrow p_i, V_i, T_i \leftrightarrow p_2, V_2, T_2. \quad (4.3)$$

If the system returns to its original position, the process is called a circular (cyclic). If it passes through the same equilibrium intermediate states, the process is called a reversible circular (cyclic). As a rule, reversible processes are represented graphically by continuous lines, while irreversible processes are represented by dotted lines.

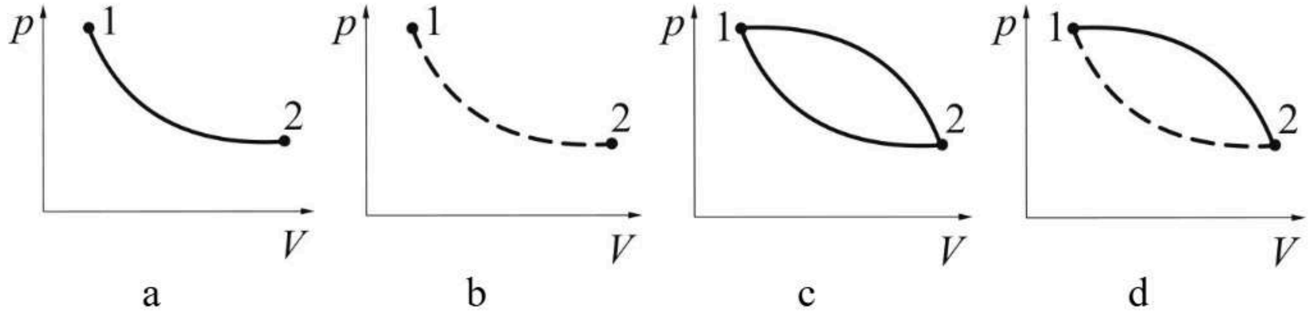


Fig. 4.1. Graphical representation of processes: a) – reversible process, b) – irreversible process, c) – circular reversible process, d) – circular irreversible process

4.1.1. The major concepts of Kinetic Molecular Theory of Matter

The general statements of the Kinetic Molecular Theory of Matter are formulated as follows:

- all bodies are composed of molecules,
- molecules are in continuous chaotic motion,
- There is an interaction between molecules, which together with thermal motion determines the aggregate state: gases, liquids, solids.

An ideal gas, a set of non-interacting particles of a point mass, is used as a model for the statements of the Kinetic Molecular Theory of Matter.

This model corresponds, for example, to atmospheric air at 0 °C and atmospheric pressure. Its characteristics are as follows:

- Loschmidt`s number – $2{,}3 \cdot 10^{19}$ cm^{-3} is the number of molecules in 1 cm^3,
  Avogadro`s number is $6{,}022 \cdot 10^{23}$ $mole^{-1}$ (number of molecules in 1 mole of matter),
- d (diameter of a molecule) ≈ 1,5 Å and slightly larger (e.g., for water vapor H_2O = 3Å) (1 Å = 10^{-8} cm = 10^{-10} m = 0,1 nm),
- υ (speed interval) = $0 \div 10^5$ m/s ,
- z (collision frequency) = $10^9 \div 10^{12}$ s^{-1},

The adequacy of using atmospheric air as an ideal gas model follows from well-known gas laws.

4.1.2. Isoprocesses in an ideal gas

To determine the basic characteristics of a system in the gas phase and to establish the basic laws of change of its state (gas laws), it is common to consider processes occurring at one fixed (constant) parameter.

Isothermal (Iso-T) process (Fig. 4.2). It is described by Boyle-Mariott equation (established by R. Boyle in 1662 and independently by E. Marriott in 1676):

at $$T = const. \quad pV = const. \tag{4.4}$$

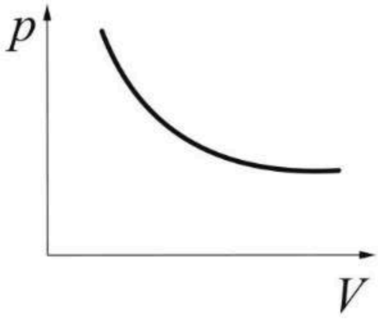


Figure 4.2. Graphical representation of the isothermal process

The isobaric (isobaric, Iso-p) process obeys the Gay-Lussac`s law[**] (Fig. 4.3):

at $$p = const \qquad V = V_0(1 + \alpha t) \tag{4.5}$$

or $$V = V_0 \alpha T, \tag{4.5a}$$

where α=1/273,15 (°C^{-1}) is the coefficient of thermal expansion of the gas, T is the temperature in the Kelvin scale, which is related to the widely used Celsius temperature scale (t) by the relation[*]:

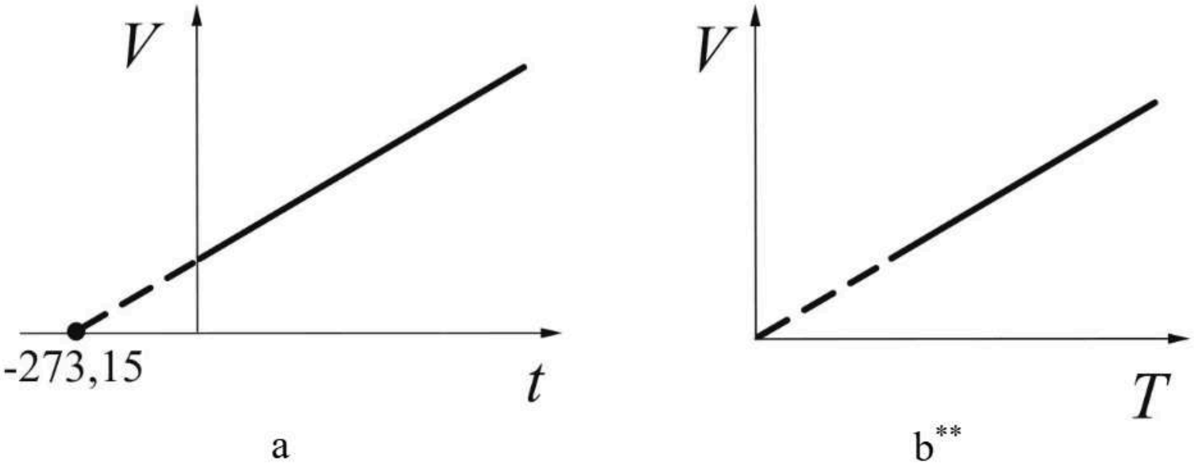
$$T = t + 273{,}15\,. \tag{4.6}$$

[**] ***Joseph Louis Gay-Lussac*** – was a French chemist and physicist, member of the French Academy of Sciences (1806).

[*] As will be shown below, the Kelvin scale is adopted as the absolute thermodynamic temperature scale, in which the unit of absolute temperature is Kelvin (K). The International Practical Temperature Scale is the scale, formerly called the Celsius scale, in which the unit of measure is degrees Celsius (°C). The temperature of 0 K is called absolute zero (-273.15 °C).

Given equation (4.5a), the Gay-Lussac`s law can be written in the form:

$$\frac{V_1}{V_2} = \frac{T_1}{T_2}. \tag{4.7}$$



a b[**]

Fig. 4.3. Graphical representation of the isobaric process in temperature scales: a) – Celsius, b) – Kelvin

The isochoric (isochoric, Iso-V) process is described by Charles's law[*], established in 1687 (refined by Gay-Lussac in 1802) (Fig. 4.4):

[**] At small values of *V* and *T*, the real gas does not correspond to the ideal gas model (see also comments in subsection 5.3.3.).

[*] ***Jacques Alexandre César Charles*** was a French inventor and scientist. He is known as the inventor of the balloon filled with hydrogen (or other gas lighter than air), named after the inventor "charlier" as opposed to "mongolfier". His first balloon went up in Paris from the Champ de Mars on August 27, 1783. The same year, on December 3, he undertook his first ballooning trip together with ***Nicolas-Louis Robert***, an engineer who was one of the balloon's makers.

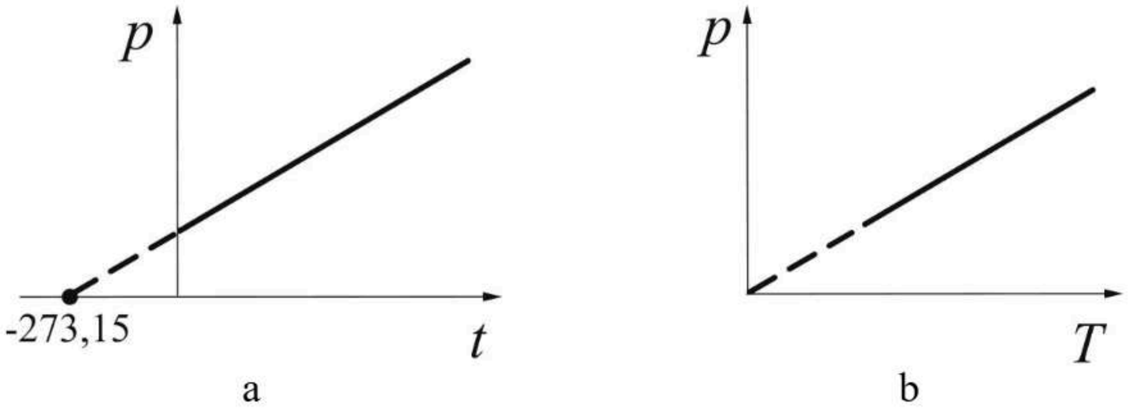


Fig. 4.4. Graphical representation of isochoric process: a) – in Celsius scale, b) – in Kelvin scale

$$\text{at } V = \text{const.} \quad p = p_0(1 + \alpha t), \tag{4.8}$$

$$p = p_0 \alpha T \quad \text{or} \quad \frac{p_1}{p_2} = \frac{T_1}{T_2}. \tag{4.8a}$$

At relatively small values of gas density, changes in its state obey with high accuracy the so-called ideal gas equation of state

$$pV/_T = const. \tag{4.9}$$

According to Avogadro's law, one mole of any gas under normal conditions occupies the same volume V_m and, naturally, the values of the constants in equation (4.9) will be the same for all gases. Then the equation of state of ideal gases for a mole of gas will be written as:

$$pV_m = RT \tag{4.9a}$$

or for an arbitrary mass m of gas:

$$pV = \frac{m}{M} RT. \tag{4.9b}$$

Equation (4.9b) is called the Clapeyron-Mendeleev equation (ideal gas law). Here M is the molecular weight (molar mass), R is the universal gas constant, equal to

$$R = 8{,}3\ \frac{J}{mol \cdot K}\,.$$

Equation (4.9a) was obtained by B. Clapeyron* in 1834 and generalized later by D. Mendeleev (he introduced the universal gas constant *R*).

4.1.3. Maxwell distribution

Due to chaotic motion and collisions, the velocity of molecules is characterized by a set of random numbers, both in direction and in absolute values (theoretically from- ∞ to + ∞). Therefore, it is possible to use methods of mathematical statistics and probability theory to quantitatively characterize such a set of particles.

One of the problems of describing a system in terms of Kinetic Molecular Theory of Matter is formulated as follows: How many molecules dN_υ have a speed in the range of values $\upsilon \div \upsilon + d\upsilon$, or what is the probability $f(\upsilon)$ that the velocity of a molecule will be in this range?

The probability function $f(\upsilon)$ is given as follows:

$$f(\upsilon) = \frac{dN_\upsilon}{N d\upsilon}, \qquad (4.10)$$

where *N* is the total number of molecules, dN_υ is the number of molecules whose speed is in the range $\upsilon \pm d\upsilon$. Since the total number of molecules is

---

* ***B. Clapeyron*** was a French physicist and engineer.

$$N = \int dN_\upsilon = \int\limits_0^\infty N\, f(\upsilon)d\upsilon, \tag{4.10a}$$

then

$$\int\limits_0^\infty f(\upsilon)d\upsilon = 1. \tag{4.10b}$$

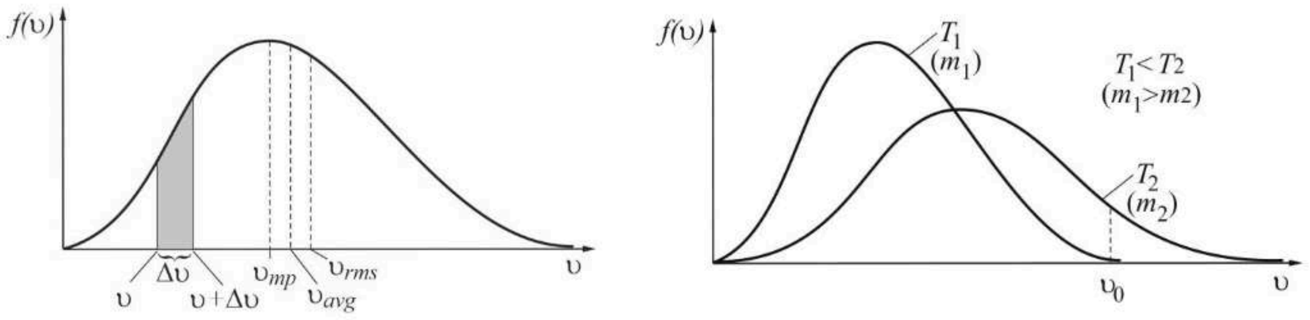
This is the so-called normalization condition of the distribution function, which means that in an infinite set of speed values there will always be a molecule that has a given speed.

Let us omit the derivation of this distribution function, called the Maxwell speed distribution, and give only the final expression:

$$f(\upsilon) = 4\pi\left(\frac{m}{2\pi kT}\right)^{3/2}\upsilon^2\exp\left(-\frac{m\upsilon^2}{2kT}\right). \tag{4.11}$$

The number of molecules in a given range of speed values $\upsilon \div \upsilon + d\upsilon$ is determined by the formula

$$dN = Nf(\upsilon)d\upsilon = 4\pi N\left(\frac{m}{2\pi kT}\right)^{3/2}\upsilon^2\exp\left(-\frac{\mathrm{m}\upsilon^2}{2kT}\right)\mathrm{d}\upsilon. \tag{4.12}$$



a

b

Fig. 4.5. a) – Graphical representation of Maxwell distribution by speed, where the characteristic speed of the molecules is indicated: the most probable speed υ_{mp}, the RMS $\sqrt{\langle\upsilon^2\rangle}$ (= υ_{rms}) and average mean $\upsilon_{\mathrm{avg}} = \langle\upsilon\rangle$; b) – Maxwell distribution for different values of molecule mass and temperature

In Fig. 4.5, where the Maxwell distribution is graphically represented, the characteristic speed of the distribution is marked: the most probable υ_{p}, the root-mean square (RMS) $\langle\upsilon^2\rangle$ and the average speed $\langle\upsilon\rangle$ of molecules – (a) and the view of the Maxwell distribution as a function of molecular mass and temperature – (b).

The most probable velocity υ_{mp} corresponds to the maximum value of the probability function $f(\upsilon)$ and is found from the condition:

$$\frac{\partial f(\upsilon)}{\partial \upsilon} = 0. \tag{4.13}$$

Solving this equation gives the value of the most probable speed

$$\upsilon_{mp} = 1{,}4\sqrt{\frac{kT}{m}}. \tag{4.13a}$$

The mean squared velocity $\langle\upsilon^2\rangle$ and the average mean $\langle\upsilon\rangle$ are defined as

$$\langle\upsilon^2\rangle = \int_0^\infty \upsilon^2 f(\upsilon)d\upsilon \quad \text{and} \quad \langle\upsilon\rangle = \int_0^\infty \upsilon f(\upsilon)d\upsilon\,, \tag{4.13b}$$

respectively. Their numerical values are related to the parameters of the gas components by the following relations:

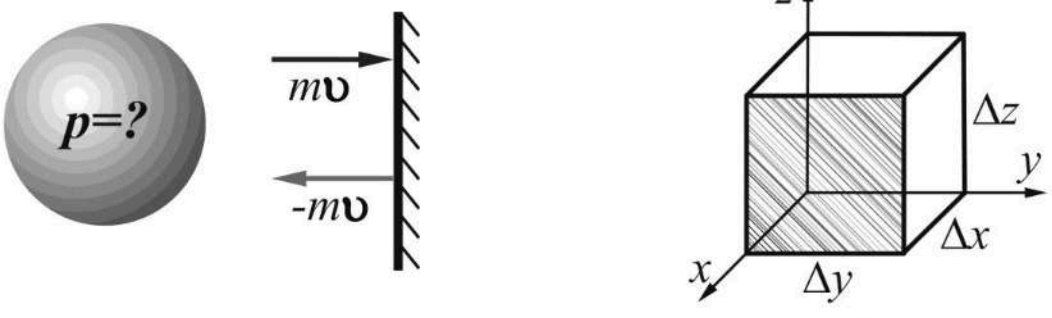
$$\upsilon_{\mathrm{rms}} = \sqrt{\langle\upsilon^2\rangle} = 1{,}73\sqrt{\frac{kT}{m}}, \qquad \upsilon_{\mathrm{avg}} = \langle\upsilon\rangle = 1{,}60\sqrt{\frac{kT}{m}}. \tag{4.13c}$$

Under normal conditions, the average velocity of oxygen molecules is 500 m/s, and that of hydrogen is 2000 m/s.

4.1.4. *The basic equation of kinetic theory of gases*

From the Clapeyron-Mendeleev equation (4.9a) it follows that the pressure $p = nkT$, where n – concentration of gas molecules in the vessel. It would be common sense to predict this kind of dependence by doing a simple experiment. Imagine that you are inflating a soccer ball, and the more air you blow into it, the harder it is for you. The dependence on concentration is obvious: the amount of air increases and the size (volume) of the ball is almost unchanged because the ball is made of a dense, elastic material. Then you put the ball in the sun or in a cool place. In the first case, the ball inflates slightly, becoming more elastic because the walls are more tense due to increased pressure. In the second case, the ball somewhat "deflates", loses its elasticity, i.e., the pressure decreases. There are many variations of such experiments. It is quite obvious that $p{\sim}nT$. The main question is to determine the mechanism of such an empirical dependence of pressure on concentration of molecules and temperature. Let us show that this dependence can be predicted, i.e. we can theoretically describe the experimentally established universal gas law, the meaning of which is contained in the Clapeyron-Mendeleev equation

The theoretical consideration is based on the notion that the pressure in the vessel is associated with the impact of air molecules on the walls of the vessel, as a result of which the momentum of the molecule changes sign after reflection (rebound) from the wall, as shown in Fig. 4.6a.



a

b

Fig. 4.6. a) – Schematic representation of momentum change

of particle $m\upsilon$ during rebound from a vessel wall, b) – illustration of model for calculation of pressure on elementary cube wall-facet with area $\Delta S = \Delta z \cdot \Delta y$ (shaded)

Let us mentally isolate in space an elementary cube whose edges $\Delta x, \Delta y, \Delta z$ are directed along corresponding coordinate axes (Fig. 4.6b). Obviously, molecules move chaotically in all directions. Let us consider only those that move in the direction of the x-axis (let's denote their velocity as υ_x). When they hit the "wall" (a side edge whose area is equal to $\Delta S = \Delta y \cdot \Delta z$) , they bounce back in the opposite direction. The change in the x-component of the momentum of one molecule as a result of the impact is $\Delta K_{0x} = 2m\upsilon_x$. Here m is the mass of the molecule. According to Newton's second law of motion, the change in momentum of a material point equals the momentum of the acting force $(dK = Fdt)$[*]. Here, the force F means the x–component of the force (F_x) interacting with the wall for time dt. However, neither the magnitude of the force F_x nor the time dt are known to us. Therefore, let us assume that the momentum of force $F_x dt$ is equal to the momentum of the average force $\langle F_x \rangle$, acting on the molecule as it moves from "wall" to "wall" and back during time Δt. Then $\Delta K_{0x} = F_x dt = \langle F_x \rangle \Delta t = 2m\upsilon_x$. The change in the x-component of the momentum of all molecules N after the collision with wall ΔS is equal to

$$\Delta K_x = 2m\upsilon_x \cdot N = 2m\upsilon_x \cdot n\Delta x \Delta S = \langle F_x \rangle \Delta t = F_x \frac{2\Delta x}{\upsilon_x}, \qquad (4.14)$$

where n is a molecular concentration, $\langle F_x \rangle$ – middle value of x-components of the total wall impact force of all molecules in the x direction contained in the elementary cube $\Delta x \ \cdot \Delta y \cdot \Delta z$, equal to

---

[*] Since p was previously chosen to denote pressure, to avoid confusion, we denote momentum with the letter "K".

$$\langle F_x \rangle = m\upsilon_x^2 n \Delta S. \tag{4.14a}$$

Now we must consider that this population of molecules obeys the laws of statistics, and the molecules have a scattering of absolute velocity values. Therefore, let us replace the value of υ_x^2 by the RMS value $\langle \upsilon_x^2 \rangle$. Since all components are equal (and the pressure on any wall-edge is the same), then $\langle \upsilon_x^2 \rangle = \frac{1}{3}\langle \upsilon^2 \rangle$. Then the average value of the impact force of a set of molecules is determined by the ratio $\langle F \rangle = \frac{1}{3} nm \langle \upsilon^2 \rangle \Delta S$, and the pressure on any face is

$$p = \frac{\langle F \rangle}{\Delta S} = \frac{1}{3} nm\langle \upsilon^2 \rangle = \frac{2}{3} n \frac{\langle m\upsilon^2 \rangle}{2} = \frac{2}{3} n \langle E_{transl} \rangle, \tag{4.14b}$$

where $\langle E_{transl} \rangle$ is the average value of the kinetic energy of translational motion of the molecules. This is the basic equation of Kinetic Molecular Theory. It is postulated that the average kinetic energy of translational motion of gas molecules is equal to $\langle E_{transl} \rangle = \frac{3}{2} kT$ (see Section 5.1). Then equation 4.14b can be represented as

$$p = nkT, \tag{4.15}$$

where k is Boltzmann constant. Equation (4.15) is transformed into the Clapeyron-Mendeleev equation by multiplying its left and right parts by the mole volume of gas V_m, namely:

$$V_m p = V_m nkT = kN_A T = RT.$$

If a gas consists of molecules of different sort i, then:

$$n = \sum n_i ,$$

where n_i is the concentration of molecules of sort i, and then the total pressure is determined by the ratio

$$p = n_1 kT + n_2 kT + \cdots + n_i kT = \sum p_i \ , \qquad (4.16)$$

which is called Dalton's law. Here $p_i = n_i kT$ is the partial pressure of a given gas, i.e. the pressure exerted by the i-sort of gas.

4.1.5. Barometric formula. Boltzmann distribution

The next task of Kinetic Molecular Theory of Matter is to determine the distribution of molecules in the potential field of forces. This can be done in a generalized way. Let's consider a more illustrative method using an example of the distribution of molecules of atmospheric gas using the equation of state of gases. Initially, we assumed that the Maxwell distribution was not disturbed in the field of gravity. Let us mentally isolate in the air column an elementary infinitesimal volume in the form of a cylinder with height dh and base area dS, as shown in Fig. 4.7.

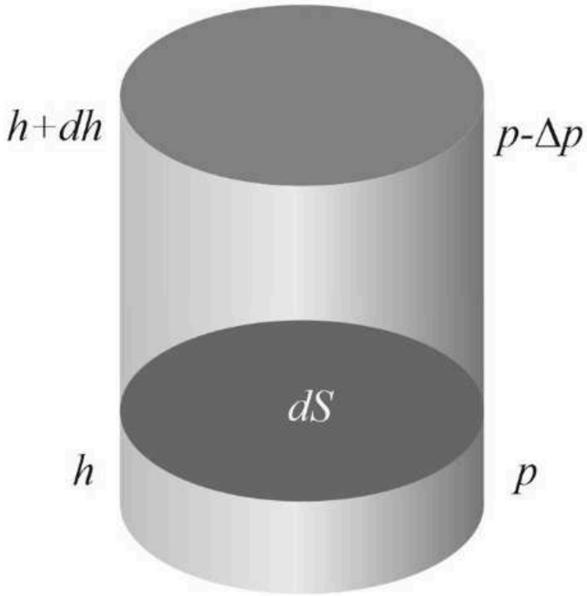


Figure 4.7. To the illustration of the derivation of the barometric formula

Obviously, the pressure p in an arbitrary horizontal section of this cylinder is determined by the overlying layers of air. That is, with increasing height dh the pressure decreases by the value of dp according to the relation

$$dp = -\rho g dh, \tag{4.17}$$

since the weight of the layer dh over the element of section dS at height $h + dh$ is equal to $\rho dSdh$, where ρ is the gas density ($p = {}^{F}/_{S_\perp}$, see Section 2.1).

According to the Clapeyron-Mendeleev equation, the density of a gas is determined by the relation $\rho = \frac{m}{V} = \frac{pM}{RT}$. Then equation (4.17) can be rewritten as:

$$dp = -\frac{pMg}{RT}dh. \tag{4.17a}$$

Assuming that the changes of the gravity acceleration g and temperature with height can be neglected (see Sections 1.2.4 and 4.3, respectively), the solution of this differential equation in separable variables can be represented as:

$$p = C\exp\left(-\frac{Mgh}{RT}\right). \tag{4.17b}$$

The integration constant C is the air pressure at height $h = 0$, i.e. at ground surface level, $p(0)$. Thus, the air pressure at height h is equal to

$$p(h) = p(0)\exp\left(-\frac{Mgh}{RT}\right). \tag{4.18}$$

This is the barometric formula. Let us use the relation $p = nkT$ and obtain an expression for the concentration of air molecules at height h, $n(h)$:

$$n(h) = n(0)\exp\left(-\frac{E_n}{kT}\right), \tag{4.19}$$

where $n(0)$ is the concentration of molecules at height $h = 0$, $E_n = mgh$ is the potential energy of a molecule of mass m_0 in the field of gravity. But this is the distribution of particles in any field of conservative forces, i.e. in any potential field. It, as noted above, can be obtained in a generalized form and is called the Boltzmann distribution in a potential field. Obviously, the barometric formula is its special case. The expression $n \sim \exp\left(-\frac{E_n}{kT}\right)$ is called the Boltzmann factor.

Using Dalton's law, it is easy to obtain an expression for the concentration of molecules of atmospheric air, consisting of different varieties of gas, depending on the altitude h:

$$n_i(h) = n_i(0)e^{-\frac{m_i gh}{kT}} \tag{4.19a}$$

or

$$n_i(h) = n_i(0)e^{-\frac{M_i gh}{RT}}, \tag{4.19b}$$

where m_i and M_i are the molecular and molar masses of gases (components of atmospheric air).

The barometric formula determines the distribution of the components of atmospheric gas, peculiarities of the human body functioning in high altitude or deep diving conditions. Therefore, these issues are highlighted in separate subsections.

4.2. Physical properties of the atmosphere

Let's consider the characteristics of only the layer of the atmosphere closest to the earth – the troposphere, more precisely, the layer up to a height of 10 km. A long stay at this altitude in the open atmosphere is deadly. As already noted, migratory geese ascend to approximately this altitude. The troposphere contains 80% of the mass of the atmospheric air (though by volume it is only about 1% of the total atmosphere) and around 90% of the atmospheric water (it is conventionally assumed that the atmospheric boundary is at a height of 100 km above sea level). It is the main source of winds, clouds, rain and snow. The composition of the atmosphere is shown in Fig. 4.8.

The composition of these gases is practically constant, except for carbon dioxide (CO_2), the content of which is continuously increasing. In addition to the gases shown in Fig. 4.8, the atmosphere contains water in the form of vapor, suspended droplets or ice crystals. The amount of water is not constant and depends on temperature and, to a lesser extent, on air pressure. The Earth's atmosphere also contains small quantities of environmentally harmful components: SO_2, NH_3, CO, HCl, NO, Hg (mercury vapor).

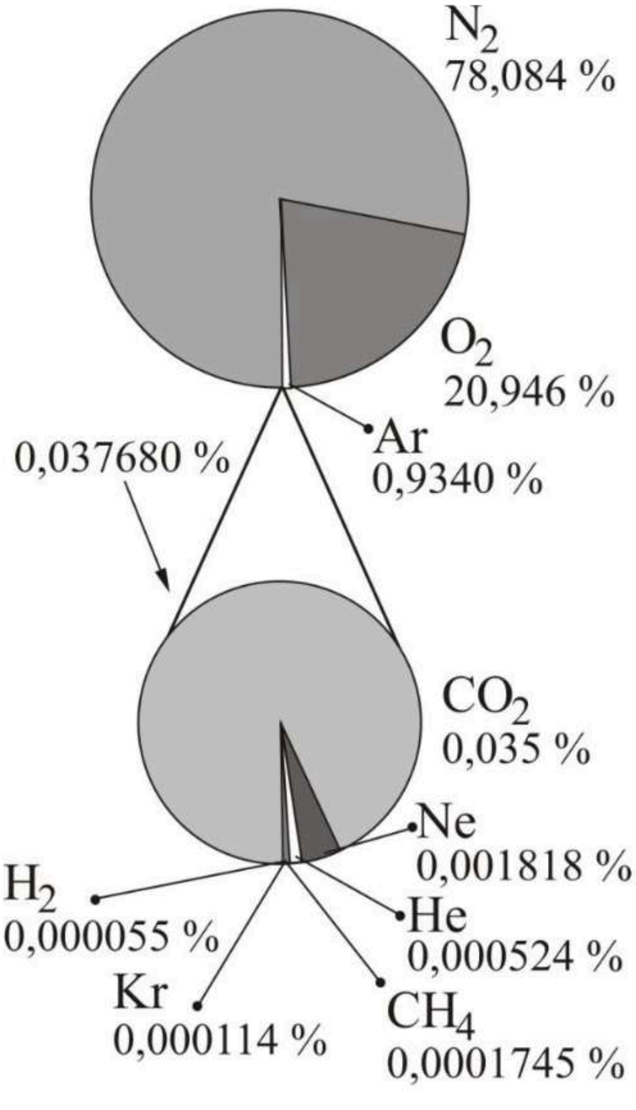


Fig. 4.8. The composition of the Earth's atmosphere

The main characteristic of the atmosphere is temperature. In the near-ground layer, the troposphere, it decreases with altitude by about 0,65 °C for every 100 m. The temperature at a given height h (in km) is approximated by the relations: $t_h = t - 6{,}5\ h$, (°C) or $T_h = 273 + t - 6{,}5\ h$, (K), where t (°C) is the air temperature at the ground surface. Next are air parameters such as pressure, density, humidity, viscosity and compressibility. Air humidity is determined by the amount of water in the air. There is a concept of "relative humidity". This is the ratio of the mass of water vapor in the air to the maximum possible water vapor at a given temperature, and it is expressed in %. The concept of 0% is relative, because absolutely dry air can only be created in a laboratory. On the other hand 100% humidity is a very real thing. This means that the maximum equilibrium concentration of moisture in the air, or saturation, is reached: as the temperature drops further, water vapor condenses on the surface of objects. This temperature is called the dew point. There is a set of reference values called the International Standard Atmosphere (ISA). It shows the change in the state of the air as it rises in altitude (Table 4.1).

ISA tables are used for calibrating pilot-navigational instruments, as well as for navigational and engineering calculations. The basic parameters of air state at zero humidity are taken as basic: pressure $p = 760$ mm Hg (101,3 kPa); temperature $t = +15$ °C (288 K); mass density $\rho = 1{,}225$ kg/m^3.

Table 4.1

Main parameters of atmospheric air[*]

Geometric height, h, m	Temperature T, K	Pressure p, Pa 10^5	Density (ρ), kg/m^3	Sound velocity υ, m/s	Kinematic viscosity, m^2/s
0	**288,150**	**1,01325***	**1,22500**	**340,294**	**1,4607-1**
250	286,525	0,98357	1,19587	339,333	1,4897
500	284,900	0,95461	1,16727	338,370	1,5195
750	283,276	0,92634	1,13921	337,403	1,5500
1000	281,651	0,89876	1,11166	336,435	1,5813
1500	278,402	0,84559	1,05810	334,489	1,6463
2000	275,154	0,79501	1,00655	332,532	1,7147
2500	271,906	0,74691	0,95695	330,563	1,7868
3000	**268,659**	**0,70121****	**0,90925**	**328,584**	**1,8628**
3500	265,413	0,65780	0,86340	326,592	1,9429
4000	262,166	0,61660	0,81934	324,589	2,0275
4500	258,921	0,57752	0,77703	322,573	2,1167
5000	255,676	0,54048	0,73642	320,545	2,2110
5500	252,431	0,50539	0,69746	318,505	2,3107
6000	**249,187**	**0,47217*****	**0,6,6011**	**316,452**	**2,4162**
6500	245,943	0,44075	0,62431	314,385	2,5278
7000	242,700	0,41105	0,59001	312,306	2,6461
7500	239,457	0,38299	0,55719	310,212	2,7714
8000	236,215	0,35651	0,52578	308,105	2,9044
8500	232,974	0,33154	0,49575	305,984	3,0457
9000	229,733	0,30800	0,46706	303,848	3,1957
9500	226,492	0,28584	0,43966	301,697	3,3553
10000	223,252	0,26499	0,41351	299,532	3,5251

* corresponds to an atmospheric pressure of 760 mm Hg.

** corresponds to an atmospheric pressure of 694 mm Hg.

*** corresponds to an atmospheric pressure of 467 mm Hg.

---

[*] The heights at which the "mountain sickness" begins (~3000 m) and at which the permanent human stay is not compatible with life (~6000 m) are highlighted.

4.3 Altitude sickness

Altitude sickness (altitude hypoxia) – a painful condition associated with oxygen starvation due to decreased partial pressure of oxygen in the inhaled air. It occurs when staying at altitudes above 2000–3000 m above sea level, as well as when flying at such altitudes in flight vehicles not equipped with an airtight cabin (e.g. paragliders, hang-gliders, balloons with a leaky gondola). Mountain sickness is a kind of altitude sickness, where besides lack of oxygen such aggravating factors as physical fatigue, cooling, dehydration, ultraviolet radiation, severe weather conditions (e.g. hurricane winds), sharp temperature changes during the day (from +30°C during the day to -20°C at night), etc. play a role. But the main pathological factor of mountain sickness is hypobaric (caused by lower atmospheric pressure) hypoxia (oxygen starvation, more precisely, lack of oxygen in tissues).

Humans are able to adapt to altitude hypoxia, sportsmen specifically train their ability to adapt in order to increase their athletic performance in "usual" conditions. Heights of about 8000 meters are considered to be the maximum possible heights for long-term adaptation (hours to tens of hours) to hypoxia. A longer stay at such altitudes without oxygen breathing apparatus is impossible (lethal outcome). The classification of heights and characteristic physiological changes in a person staying at different heights is as follows:

- Intermediate altitudes (1500÷2500 m). Physiological changes are noticeable, but blood oxygen saturation is within normal limits, >90%; the probability of mountain sickness is low.
- High altitudes (2500÷3500 m). Mountain sickness develops with rapid ascent.
- Very high altitudes (3500÷5800 m), blood oxygen saturation < 90%. Significant hypoxemia (lack of oxygen in the blood) during exercise is observed. Mountain sickness develops often.

- Extreme altitudes (> 5800 m). Hypoxemia increases already at rest, deterioration progresses despite maximal acclimatization. Permanent presence at such altitudes is impossible.

The altitude at which altitude sickness develops varies due to numerous factors, both individual and climatic. The human body functions best at sea level, where the atmospheric pressure is 101,325 kPa or 1013,25 mbar (i.e. 1 atm). The concentration of oxygen (O_2) in the air at sea level is 20,9 % at the partial pressure of oxygen O_2 (p_{O_2}) 21,2 kPa. In healthy people, this is the normal saturation of hemoglobin with oxygen, which binds to red blood cells. When climbing to about 2,100 meters above sea level, the body's saturation with oxyhemoglobin (the protein of hemoglobin bound to oxygen) begins to drop.

Atmospheric pressure decreases exponentially with increasing altitude (see Equation 4.18). The partial pressure p_{O_2} also decreases exponentially with increasing altitude. At the base camp location on Mount Everest (5000 m above sea level), the partial pressure p_{O_2} is about half of the pressure p_{O_2} at sea level, and only one-third at the summit of Everest, at 8848 m. When the partial pressure p_{O_2} drops, the human body responds with adaptive reactions – altitude acclimatization.

Alcohol, even in small doses at an altitude of over 3,000 meters, reduces respiratory rate and increases hypoxia, so the use of alcohol in the high mountain zone should be excluded.

So, the term "mountain acclimatization" refers to a set of specific physiological adaptations that occur during a more or less long uninterrupted stay at altitude. These adaptations reduce the effect of reduced partial pressure of O_2 in the inhaled air on the human organism and increase its efficiency. The main mechanism of natural adaptation to the mountain conditions can be divided into two stages: the first stage is the increase in the transport of O_2 to the tissues of the body and the second stage is the increase in the efficiency of the use of O_2 by cells for aerobic energy generation. The main factors of adaptation to hypobaric hypoxia conditions include:

- increased pulmonary ventilation and accompanying changes in acid-base balance in blood and other tissues;
- increase of pulmonary diffusion capacity;
- increase of erythrocyte and hemoglobin content in blood, changes at tissue level.

The main adaptation changes in the circulatory system are aimed at increasing its oxygen-transport capacity. Acclimatization to altitude is essentially an adaptation to low partial pressure of O_2 and CO_2 in blood and other tissues. High hyperventilation prevents the partial pressure of O_2 in the alveolar air and respectively in the arterial blood from decreasing. However, the degree of arterial blood O_2 partial pressure decrease, observed immediately after the arrival at the altitude, remains constant throughout the acclimatization period of several weeks. During a short-term stay at altitude together with the growth of pulmonary ventilation the partial pressure of carbon dioxide CO_2 in arterial blood continues to decrease. But as a result of prolonged acclimatization to the altitude it increases which is revealed both at rest and especially during muscular work. Let us consider only the third factor related to the medical physics course itself.

The volume of blood plasma during the first few days of staying at altitude decreases compared to the indicators on the plain (sea level). Therefore, the hematocrit index is increased, i.e. the relative concentration of erythrocytes and hemoglobin in the blood is increased. But the number of erythrocytes and hemoglobin in the blood increases not only due to hemoconcentration caused by the loss of part of circulating plasma in the vascular channel, but also due to the body's reaction to hypoxia (erythropoiesis). In the process of further stay at altitude the volume of circulating plasma is restored to its original ("plain") level. In middle mountain conditions it takes several months.

Note that hemoconcentration and erythropoiesis ensure the maintenance of normal O_2 content in arterial blood and therefore play an important role in the body's rapid adaptation to hypoxic conditions.

Further, during the initial days of stay in the mountains, erythropoiesis also intensifies, leading to a true increase in the number of erythrocytes in the blood. It becomes noticeable already on the 3rd – 4th day of staying at the altitude above 3000 m. The number of circulating reticulocytes and large red blood cells increases. The degree of increase in the total number and respectively the concentration of erythrocytes at altitudes up to 4800 m is linearly related to altitude and length of stay in the mountains. Erythropoiesis decreases with increasing altitude up to 6000 m. After several days of being at an altitude of over 7000 meters, the red blood cell count reaches 8,5 million/mm^3. For mountain dwellers, the higher the altitude of the place of permanent residence, the higher it is.

Studies have shown that a decrease in external pressure below 267 mm Hg (this corresponds to an altitude of 8000 m) can lead to the development of "decompression disorders". If external pressure drops to 40 mm Hg, such disorders become irreversible. Reduction of external pressure to dangerous levels is prevented by the use of cabin pressurization systems, special suits, hermetic helmets, etc. It was found that the normal vital activity of the astronaut is provided when the total pressure drops to 500 mm Hg, at the same time the optimal partial pressure of oxygen in the cabin must not exceed 420 mm Hg – higher concentrations turn out to be harmful. The ratio between the total pressure and the partial pressure of oxygen is regulated chemically – by regeneration.

Hemoconcentration at the beginning of altitude acclimatization, and a later true increase in the number of red blood cells in the circulating blood leads to an increase in hematocrit and blood viscosity. This, in its turn, leads to an increase in peripheral vascular resistance and thereby affects hemodynamics. Minor changes in erythrocyte content (hematocrit) have no significant effect on blood viscosity. Only a significant increase in their concentration can have a certain negative effect on blood circulation and heart function.

The oxyhemoglobin dissociation curve shifts to the right during mountain acclimatization, which facilitates the supply of oxygen to tissues*.

This is especially important for functioning muscles. One of the mechanisms of such a shift may be the increased concentration of 2,3-diphosphoglycerate (2,3-DPG) in erythrocytes, which is observed in people constantly living in the mountains. However, even after complete acclimatization at altitude, tissue oxygen supply is difficult, especially during strenuous muscular work, due to reduced partial pressure O_2 in arterial blood [6]. (It is beyond the scope of this textbook to review the issues of changes in the circulatory system and maximum oxygen consumption).

Influence of weather conditions on blood pressure. First of all, blood pressure is influenced by atmospheric pressure, by means of:

- changes in pressure inside the thoracic and abdominal cavities;
- changes in rheological properties and chemical composition of blood viscosity, oxygen saturation*.

Blood pressure is also affected by air temperature. It mainly affects the rheological properties of blood. Air humidity has less influence on blood pressure. There is an average dependence between atmospheric

---

* The oxyhemoglobin dissociation curve is the dependence of hemoglobin oxygen saturation (the ratio of oxyhemoglobin to the amount of total hemoglobin in %) on the partial pressure of oxygen. When the partial pressure is equal to zero, only reduced hemoglobin is in the blood. This means that increasing the partial pressure of oxygen leads to increasing the amount of oxyhemoglobin (it shifts to the right on the dissociation curve). The dissociation of oxyhemoglobin depends on several factors: blood pH (the increase leads to an increase in the partial pressure); the same shift occurs when the concentration of carbon dioxide CO_2 and 2,3-diphosphoglycerate in blood increases, an increase in its temperature.

* From Greek ρεοσ (rheos) – flow and λογότυπο (logos) – word, science. The science about deformation and fluidity of matter. Blood rheology is blood fluidity determined by a combination of functional state of blood forming elements (motility, deformability, aggregation activity of erythrocytes, leukocytes and thrombocytes), blood viscosity (protein and lipid concentration), blood osmolarity (glucose concentration). A key role in the formation of the rheological blood parameters belongs to the blood forming elements, primarily erythrocytes, which make up 98 % of the total volume of blood forming elements [5].

pressure and blood pressure value: atmospheric pressure decrease by 5 mm Hg causes blood pressure increase by 13 mm Hg. [5, 6].

Effects of air travel on the cardiovascular system. During the flight the cabin pressure is approximately equal to the pressure at 2,500 m above sea level. As atmospheric pressure decreases, the amount of oxygen in the cabin decreases – a major risk factor for high blood pressure, hypertensive crisis and heart attack.

Anyway, everyone feels the pressure drops in the aircraft cabin. Rapid decrease of atmospheric pressure and oxygen concentration drop in the passenger cabin is registered at the altitude of 3000 m, and during the long flights plane gains altitude up to 11000 m, which sharply decreases the oxygen supply to the blood. The body's response is the increase of red blood cells content and, as a consequence, the increase of blood viscosity. To maintain blood viscosity in the normal range, appropriate medications are used, for example, the use of tomato juice without salt contributes to this. Even healthy people, especially those suffering from heart disease (coronary heart disease, angina pectoris, myocardial infarction, atherosclerosis and hypertension) may need additional oxygen inhalation during the flight [5, 6].

4.4. The effect of diving underwater (pressure) on a human

The depth of diving without a scuba diver is primarily limited by the oxygen available in the body (about 2,5 liters). The diver is also helped by the fact that water pressure, squeezing blood from the limbs, increases its saturation in the lungs.

When the barometric pressure is significantly elevated, not only the oxygen in the atmospheric air becomes life-threatening, but also the nitrogen contained in it. This gas is highly soluble in nerve tissue, causing first a narcotic and then a toxic effect. Nitrogen narcosis, or "depth intoxication" (nitrogen has long been known as "laughing gas"), usually occurs when a person is breathing atmospheric air at a depth of 30 to 100 meters.

In this state he loses control and death is possible. Therefore, when a diver dives to greater depths he is given a gas mixture where nitrogen is replaced with helium, which is much less soluble in nerve tissue and blood.

There is always some amount of air dissolved in the human body. The oxygen in the air is involved in oxidative processes and is continuously absorbed by the body, while the amount of nitrogen remains unchanged.

When immersed under water and increasing pressure, the amount of neutral gas (nitrogen) dissolved in tissues, entering the breathing mixture, will increase, reaching the saturation limit, which will come when the pressure of dissolved gas in tissues is equal with its partial pressure in the inhaled air or gas mixture (see Dalton's law, equations (4.16)).

Gas saturation and the reverse process – desaturation (hypoxia in medical terminology), are complex processes, as their rate is different for body tissues depending on their blood supply. For example, in muscle tissue with a large number of blood vessels, these processes will proceed faster than in fat tissue with few vessels.

Dissolved gas bubbles do not have time to be transported by blood to the lungs and, increasing in volume as the pressure decreases, can cause blockage of blood vessels, leading to muscle and joint pain and to the development of a dangerous disease called decompression or decompression sickness. The way to prevent this disease was found long time ago when descending to shallow and middle depths. It consists in establishing a special recovery regime (pressure reduction). Diver ascent modes are calculated depending on the diving depth, time of being under pressure and the rate constants of saturation and "desaturation" of various tissues of the body. Note that as the diver's diving time increases, the coefficient of productivity decreases, which is calculated as the ratio of the diver's diving time to the total descent time.

When descending to great depths, which are carried out with the use of artificial gas mixtures for diver breathing, due to high pressures factor of productivity continues to decrease with increasing depth.

Thus, the impact of high pressure during diving, especially in deep water, is primarily reduced to the need for prolonged decompression, which leads to a decrease in the diver's productivity factor to very low values.

Gases contained in the air and other gases that can be used for breathing of divers, with increasing pressure exhibit special properties of impact on human organisms, which are not inherent to them at normal pressure. In this case the effect of a gas is determined by its partial pressure in the inhaled air or gas mixture.

Of the gases considered only oxygen and carbon dioxide are involved in the breathing process, while the other gases, although affect the human body when descending to accessible depths, play the role of fillers, creating a total pressure corresponding to the depth of the dive.

Oxygen. With the increase of partial pressure of oxygen above 15,7÷16,7 kPa (1,6÷1,7 kgf/cm^2) toxic properties start to appear which lead to oxygen poisoning of the organism. Oxygen poisoning can occur at a much lower partial pressure in cases of prolonged breathing of the gas mixture with partial pressure of oxygen over 4,9 kPa (0,5 kgf/cm^2). Symptoms of this disease appear after some time of breathing with a mixture with increased oxygen content, and the safety interval depends on the partial pressure of oxygen.

Thus, when diving in regenerative gear, when a diver breathes pure oxygen, staying at a depth of 10 m is up to 1 hour, and at a depth of 20 m – only 20 minutes.

The lower limit of the oxygen content in the breathing mixture is 1,67 kPa (0,17 kgf/cm^2) (7,88 % at normal pressure). If we slightly deviated from the upper and lower limits in the interest of safety, we would have the partial pressure of oxygen 1,96÷14,7 kPa (0,2÷1,5 kgf/cm^2), which is the oxygen content limit when descending to great depths.

Carbon dioxide. Carbon dioxide can have a toxic effect on the human body even at normal pressure. The content of carbon dioxide in the inhaled air up to 1% does not cause any painful effects, but when its content is up to 3 %, the person has a feeling of heat, shortness of breath and rapid fatigue.

Further increases of carbon dioxide content up to 5÷6 % disturb cardiac activity and a person loses consciousness.

At higher pressures, the degree of carbon dioxide exposure is determined not by its relative content (in %) in the inhaled air (gas mixture), but by its partial pressure, which in this case is usually expressed as a percentage of the normal partial pressure. For example, if a diver goes down to the depth of 50 meters and receives air containing 1 % of carbon dioxide, then the partial pressure of carbon dioxide in recalculation on normal pressure at the depth will be 6 %, that is, its content will be unacceptably high.

This property of carbon dioxide requires a high degree of purification of the gas mixture during deep-water descents. For example, when descending to a depth of 300 m, the gas mixture should not contain more than 0,03 % of carbon dioxide.

Nitrogen. As the pressure increases, the amount of nitrogen dissolved in blood and tissues increases and its narcotic effect begins to appear.

When breathing normal air at a depth of 50÷60 m, when the partial pressure of nitrogen reaches 39,2÷49,0 kPa (3,8÷4,8 atm.) a human shows the phenomena similar to the initial stage of alcohol intoxication: cheerfulness and talkativeness appear, coordination of movements becomes worse. Further pressure increase causes movement disorders, sharp decrease of work capacity, incoordination of speech, visual and auditory hallucinations and finally complete loss of consciousness.

The degree of narcotic effect of nitrogen depends on individual characteristics of people. Some well-trained divers keep their ability to work when diving in air up to the depth of 80÷90 m, while others lose it already at the depth of 50÷60 m.

The property of nitrogen described above makes it unsuitable as a component of a breathing mixture for deep-water descents. It was necessary to search for another gas for this purpose. Helium became such a gas that it found a wide application.

Helium The depth at which the effects of helium become narcotic like nitrogen is not precisely defined. The helium diving limit, or as it is commonly called, the "helium barrier" lies at depths of 600÷1000 meters.

However, the use of helium as a filler is significantly affected by its physical properties: thermal conductivity and density. Helium has a thermal conductivity six times greater than air, which leads to rapid human cooling when breathing helium-containing mixtures, and particularly when in a helium environment. Thus, in order to prevent hypothermia, helium chambers must be kept at temperatures of up to 32°C, whereas 22–23 °C is sufficient when breathing in an air environment.

During underwater descents using helium the situation is more complicated: heat losses increase significantly due to the fact that water has higher thermal conductivity than gases and, washing the diver, intensively cools his body. Moreover, water temperature at great depths is low (0÷5 °C).

Of particular importance when descending to great depths is the ability of helium to distort human speech due to changes in the nature of sound propagation in helium medium – the speech signal in helium medium at a pressure corresponding to a depth of 130 m becomes almost unintelligible.

In order to improve the intelligibility of speech in helium medium there have been numerous experiments on electronic processing of speech signals, which allowed to create correctors of speech, which, however, gave only partial results, and the issue of voice communication with divers during deep water descents has not been finally resolved.

Hydrogen. Unlike other gases used as fillers in breathing mixtures, hydrogen is not inert and, moreover, forms explosive mixtures with oxygen.

It has been established that when the oxygen content in the hydrogen-oxygen mixture is less than 4%, the mixture is not explosive, but there remains the possibility of explosion during its preparation, since explosive concentrations temporarily arise in the gas mixing places.

The above circumstances are the reason why hydrogen is not used during deep-water descents. At the same time, a few experiments have shown the possibility of using hydrogen as a filler in breathing mixtures. Safe systems for the preparation, storage and use of hydrogen-oxygen mixtures are being developed, and they have already given positive results, which suggest that hydrogen can be used during deep-water descents.

Neon. The use of neon is constrained by its high cost. Experiments have shown the possibility of descents using neon-oxygen mixtures to a depth of 200 m. In this case, the inherent high thermal losses and distortion of speech are absent. It is very likely that if inexpensive methods of producing neon can be found, it will be widely used for deep-water descents [5, 6].

SECTION 5. THE FUNDAMENTALS OF THERMODYNAMICS

In the previous subsections, the phenomena interpreting which was given in the framework of the so-called Kinetic Molecular Theory of Matter were considered. According to this theory, the main parameters characterizing a physical system, temperature T, volume V and pressure p are considered as parameters related to the properties of particles (molecules): mass, size (meaning that the volume of space provided for the movement of molecules is smaller than the boundaries of the system – the vessel), and velocity. The methods of probability theory and mathematical statistics are used to establish the relationship between them, to describe the properties of the system as a whole and the processes occurring in it. However, these macroscopic parameters are not sufficient for a complete description of the characteristics of physical systems, the whole variety of processes and phenomena of nature. These problems are solved by methods of thermodynamics, where integral parameters of a system are used out of direct connection with characteristics of particles, of which this system is formed. The thermodynamic method is based on the three principles (laws) of thermodynamics. The relationship between the integral characteristics of a system (substance) and its structure is established within the framework of statistical physics. The thermodynamic method is the basis of classical (equilibrium) thermodynamics, in the framework of which the main law of nature, the law of conservation of energy, is formulated. Human activity as a part of the system "people and environment" is impossible without knowledge of classical thermodynamics. But functions of a human being as a living organism cannot be described within the framework of equilibrium thermodynamics. The life of living organisms is described by methods of nonequilibrium thermodynamics, which are based on representations and principles of classical thermodynamics.

5.1. Thermodynamic state and temperature

In the previous sections, the term "temperature" was widely used as self-evident. Meanwhile, temperature is the most important fundamental characteristic in physics, and its definition in foreign scientific literature is often called the zeroth law of thermodynamics. In the Kinetic Molecular Theory of Matter temperature is related to the kinetic energy of the molecules or particles that make up the physical system, thus establishing an interrelationship between the statistical and thermodynamic ways (methods) of describing phenomena in molecular physics.

It is assumed that, for any kind of molecule, the average kinetic energy of its translational motion is related to temperature by the relation

$$\langle E_{trans.} \rangle = {}^{3}/_{2}\, kT, \tag{5.1}$$

where, as mentioned earlier, *k* is Boltzmann constant, and the energy of any element of the gas volume is the same throughout the system. Such a state is called thermodynamic, and its corresponding temperature *T* is called thermodynamic temperature. Obviously, the local temperature (temperature of infinitesimal elements of gas volume) will not change only when molecules change their places. In detail: the place vacated by a molecule leaving this local region of space is immediately occupied by another molecule (this is the so-called principle of detailed equilibrium widely used to describe kinetic processes in conditions of thermodynamic equilibrium). A formal scheme of such a detailed exchange is shown in Fig. 5.1.

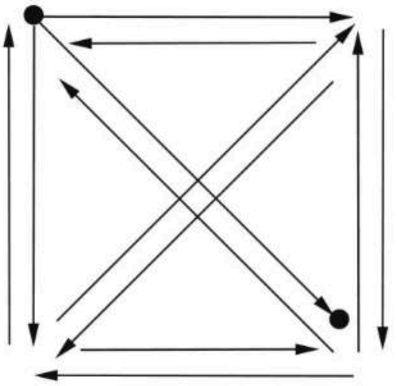


Figure 5.1. The illustration of the principle of detailed equilibrium

The proposed (postulated) quantitative relationship between temperature and kinetic energy (equation 5.1) is based on Taking into account only translational motion of molecules. As it was mentioned in section 4, for free gas the coordinates (x, y, z) and velocities of molecules (υ_x, υ_y, υ_z) are equivalent. Then for the RMS values of velocity of molecules:

$$\langle\upsilon^2\rangle = \langle\upsilon_x^2\rangle + \langle\upsilon_y^2\rangle + \langle\upsilon_z^2\rangle, \tag{5.2}$$

RMS values of velocity x, y, z-components in which are equal:

$$\langle\upsilon_x^2\rangle = \langle\upsilon_y^2\rangle = \langle\upsilon_z^2\rangle = \frac{1}{3}\langle\upsilon^2\rangle. \tag{5.3}$$

Using relations (5.2) and (5.3) for the average energy of translational motion, we obtain the expression:

$$\langle E_{trans.}\rangle = \langle\frac{m\upsilon^2}{2}\rangle = \frac{3}{2}kT. \tag{5.4}$$

So far, we have to consider that the molecule has only kinetic energy of translational motion in three possible directions of motion along three coordinate axes, i.e. the molecule has three degrees of freedom – coordinates x, y, z and their corresponding velocity components (υ_x, υ_y, υ_z).

Consequently, the energy per degree of freedom in translational motion is $kT/2$.

But the total number of independent coordinates for determining the position of the body in space – the total number of degrees of freedom i is also determined by the orientation of the axis of rotation of the molecule and the direction of rotation (rotational motion corresponds to rotational energy levels), and by the character of oscillations of atoms in the molecule (oscillational energy levels). It is assumed that the energy per each degree of freedom is the same – this is the so-called law (theorem) of equal distribution of energy over degrees of freedom, and then the total average thermal energy of the molecule is defined by the formula:

$$\langle E_{total} \rangle = \frac{i}{2} kT. \tag{5.5}$$

Figures 5.2 and 5.3 illustrate the determination of the number of degrees of freedom of a molecule in the simplest cases.

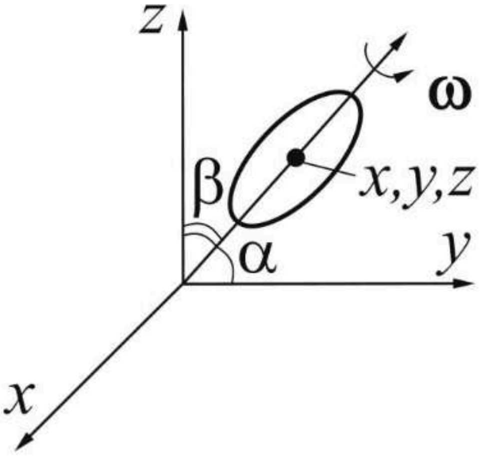
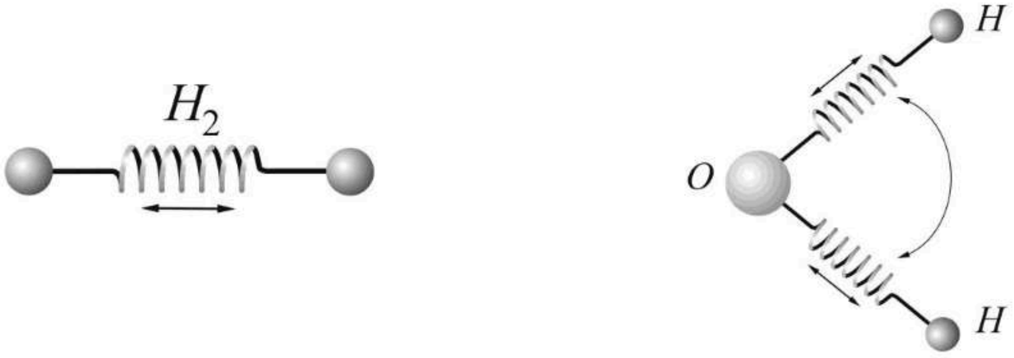


Fig. 5.2. For the determination of degrees of freedom: three degrees of freedom of translational motion are determined by coordinates of center of inertia of masses – *x*, *y*, *z;* three degrees of freedom of rotational motion are defined by two guiding angles of axis of rotation α, β, and the direction of the angular velocity vector **ω**

In the case of a linear molecule, such as a hydrogen molecule, the rotation around its own axis of rotation passing through the hydrogen atoms can be neglected (the bulk form degenerates into a line), and the number of degrees of freedom becomes 5. But two degrees of freedom are added due to oscillations of hydrogen atoms along the chemical bond that occur at high temperatures (it is simulated by a spring, Fig. 5.3a). Figure 5.3b shows a mechanical model of a triatomic water molecule for which the elastic oscillations caused by the deformation of the valence angle must also be taken into account.



a b

Fig. 5.3. Oscillatory degrees of freedom: a) – for a hydrogen molecule, b) – for a water molecule

In the general case, the number of degrees of freedom of a molecule is equal to

$$i = n_{trans.} + n_{rot.} + 2n_{osc.}. \tag{5.6}$$

Note that when the oscillatory motion of molecules is taken into account, the number of degrees of freedom doubles, since both the kinetic and potential energies of oscillations of the molecule are taken into account.

5.2. First law of thermodynamics. Work in isoprocesses

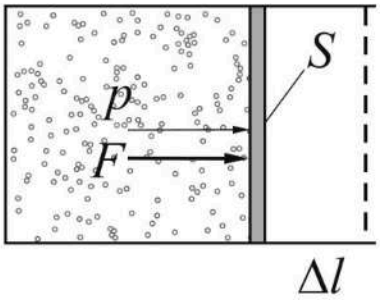
Internal energy

In a real gas, internal energy consists of the total kinetic energy of non-interacting molecules U_{total}, the potential energy of intermolecular interaction $U_{inter.}$ and the intramolecular energy $U_{intra.}$:

$$U = U_{total} + U_{inter.} + U_{intra.}. \tag{5.7}$$

When studying processes occurring in physical systems at the molecular level, intramolecular energy remains unchanged. The motion of the system as a whole is also not taken into the account. Thus, the change in internal energy is determined by the change in the energy of thermal motion and the potential energy of the interaction of molecules: $dU = d\ (U_{inter.} + U_{kin.})$.

Work done by a gas



The work of gas is defined as follows (see Section 1.5):

$$\Delta A = F\Delta l = pS\Delta l = p\Delta V, \tag{5.8}$$

where *S* is the piston area, *p* is the pressure under the piston. Or in differential form:

$$\delta A = pdV\,. \tag{5.8a}$$

Heat

Heat Q is energy in a specific form – the form of chaotic molecular motion. The totality of microscopic processes (i.e. capturing not the whole body at once, but only certain groups of molecules) leading to the transfer of energy from body to body – from one part to another is called heat transfer. The heat transferred to the system is used to change the internal energy and work (and vice versa, when work is performed on the system, the internal energy is changed and heat is released):

$$\Delta Q = \Delta U + \Delta A \Rightarrow \delta Q = dU + \delta A. \quad (5.9)$$

This is the formulation of the first law of thermodynamics, which in differential form is written in the form:

$$\delta Q = dU + pdV. \quad (5.10)$$

Mathematical operators – symbols denoting infinitesimal increment (change) δ and *d,* have different meanings. The heat transferred (as will be shown later) and the work done (as was shown earlier) generally depend on how it is carried out. It is therefore customary to use the symbol δ. The change in internal energy depends only on the initial and final states of the system, so it is denoted as dU. In integral form, the first law of thermodynamics is written as follows:

$$Q = U_2 - U_1 + A, \quad (5.10a)$$

where Q is the heat transferred to the system, as a result of which the internal energy of the system changes $(U_2 - U_1)$ and work *A* is done.

Heat capacity

Heat capacity is a physical characteristic of a substance, determined by the amount of heat required to change the temperature of a mole, or an arbitrary mass *m* of a substance by one degree. A distinction is made between heat capacities at constant volume or pressure (i.e., corresponding to iso-processes), molar and specific. Below is a summary of notations (number of moles $\nu = {}^{m}/_{M}$, where m is the mass of the substance, M is the molar mass):

– molar heat capacity at constant volume C_V, V=*const*:

$$C_V = \left(\frac{\delta Q}{dT}\right)_V; \tag{5.11a}$$

– molar heat capacity at constant pressure C_p, p = *const*:

$$C_p = \left(\frac{\delta Q}{dT}\right)_p; \tag{5.11b}$$

– specific heat capacity at V = *const*, C_V:

$$v = \frac{1}{M} C_V = \left(\frac{\delta Q}{mdT}\right)_V; \tag{5.11c}$$

– specific heat capacity at p = *const*, c_p:

$$c_p = \left(\frac{\delta Q}{mdT}\right)_p = \frac{1}{M} C_p. \tag{5.11d}$$

For a mole of ideal gas $U_{int.} = 0$, $U = \frac{i}{2} RT$, $(R = kN_A)$. From the definition of heat capacities (5.11a – 5.11d) it follows that

$$C_V = \frac{i}{2} R, \tag{5.12}$$

$$C_p = \frac{i}{2} R + p\left(\frac{dV}{dT}\right)_p = \frac{i}{2} R + R = \frac{i+2}{2} R, \tag{5.13}$$

$C_p - C_V = R$ – Meyer's formula (law). (5.14)

Work done in the isoprocesses.

Here are the basic formulas for determining the work done in isoprocesses during expansion of an arbitrary mass of gas. In general case work is calculated by formula

$$A = \int_{1}^{2} \delta A = \int_{V_1}^{V_2} p dV, \qquad (5.15)$$

where V_1 – initial, V_2 – final gas volumes.

It is clear that the work of the gas in the isobaric ($p = \text{const}$) process is:

$$A = p \int_{V_1}^{V_2} dV = p(V_2 - V_1), \qquad (5.15a)$$

in an isothermal ($T = \text{const}$) process:

$$A = \int_{V_1}^{V_2} pd = \int_{V_1}^{V_2} \frac{m}{M} RT \frac{dV}{V} = \frac{m}{M} RT \ln \frac{V_2}{V_1} . \qquad (5.15b)$$

Obviously, in the isochoric process (iso – *V*) no work is done because $dV = 0$.

Adiabatic process

Adiabatic process (see Fig. 5.4) is a process without heat exchange with external environment ($\delta Q = 0$). Then the first law of thermodynamics will take the form:

$$d\,U + pdV = 0. \qquad (5.16)$$

Since the internal energy of a mole of gas is $U = \frac{i}{2}RT$ and $dU = \frac{i}{2}RdT$, equation (5.16) can be written as:

$$\frac{i}{2}RdT + pdV_m = 0 \tag{5.16a}$$

or, using the Clapeyron-Mendeleev equation (4.9b),

$$\frac{i}{2}RdT + RTdV_m/V_m = 0. \tag{5.16b}$$

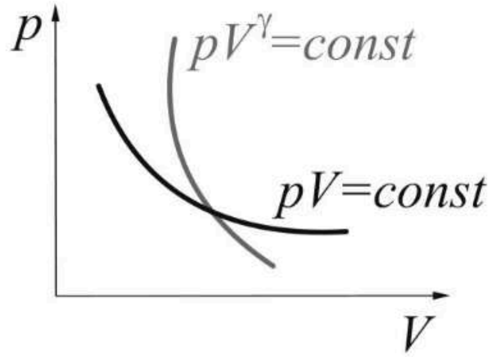


Figure 5.4. Graphical representation of isothermal and adiabatic processes

The solution of this differential equation in divided variables is the Poisson equation

$$TV^{\gamma-1} = \text{const.} \tag{5.16c}$$

Again, using the Clapeyron-Mendeleev equation, we obtain the adiabatic equation in a more common form:

$$pV^{\gamma} = \text{const.} \tag{5.16d}$$

Here $\gamma = {C_p}/{C_V}$ is the so-called adiabatic index. The elementary work of a mole of gas performed in the adiabatic process is

$$\delta A = pdV = -dU = -\frac{i}{2}RdT = -C_V dT. \tag{5.17}$$

The total work of a mole of gas when passing from state 1 to state 2 is defined by the expression:

$$A_{12} = -\int_{T_1}^{T_2} C_V dT = -C_V = C_V(T_1 - T_2) \tag{5.17a}$$

or for an arbitrary mass of gas *m*:

$$A_{12} = \frac{m}{M} C_V(T_1 - T_2). \tag{5.17c}$$

In the general case, processes at arbitrary values of the heat capacity C (polytropic processes) are described by the polytropic equation

$$pV^n = const, \tag{5.17d}$$

where *n* is the so-called polytropic index:

$$n = {(C - C_p)}/{(C - C_V)}. \tag{5.17e}$$

5.3. The second and third laws of thermodynamics

5.3.1. Construction principles of heat machines

Let's consider a heat machine – a periodically acting engine, which performs work due to heat received from the outside. Graphically, the work cycle of heat machines is shown in Fig. 5.5. When the process is directed clockwise, the system receives heat Q^+ from outside (from the heater at temperature T_h) and changes from state 1 to state 2. Then the system returns to the initial state, giving part of the heat Q^- to the external environment (refrigerator T_r, negative "-" value of heat). In an ideal machine, this is a reversible cycle in which heat transfer from the heater to the refrigerator and back occurs without heat loss (i.e. there is no heat exchange at points 1 and 2, $\delta Q = 0$). Let us use the equation describing the first law of thermodynamics $\delta Q = dU + pdV$. Integration over closed loop ($\oint \delta Q = \oint pdV$) leads to the equation

$$Q^+ + Q^- = A, \tag{5.18}$$

because $\oint dU = 0$ (the system returned to the initial state, i.e. to the state with the same value of internal energy).

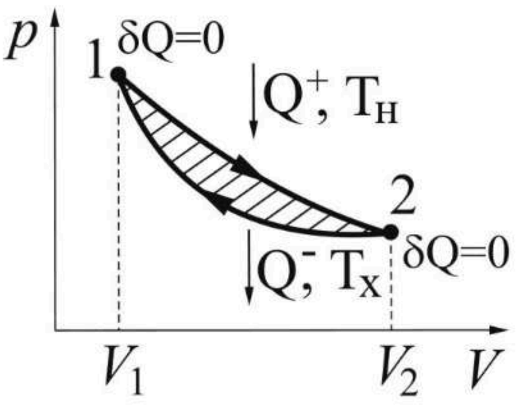


Fig. 5.5. Graphical representation of the operation of a heat machine

To determine the efficiency of a heat machine, a value called the efficiency factor η is used and is defined as the ratio of the produced work *A* (graphically, it is the shaded area on the diagram *p*(*V*), shown in Fig. 5.5.) to the heat received Q^+:

$$\eta = \frac{A}{Q^+} = \frac{Q^+ + Q^-}{Q^+} = 1 + \frac{Q^-}{Q^+}. \tag{5.19}$$

In a process that occurs in the opposite direction, the system works like a refrigeration machine (refrigerator).

Carnot heat machine: An ideal reversible machine works according to the Carnot cycle, which includes two adiabats and two isotherms: (1–2) at T_1 (= T_h) and (3–4) at T_2 (= T_r) (Fig. 5.6). In an isothermal process, according to the first principle of thermodynamics, all heat is converted into work. The adiabats (2-3 and 4-1) provide contact of the working body with the reservoirs without heat loss. Thus, heat losses are minimized in a Carnot heat machine, i.e. the efficiency of such a machine is maximal.

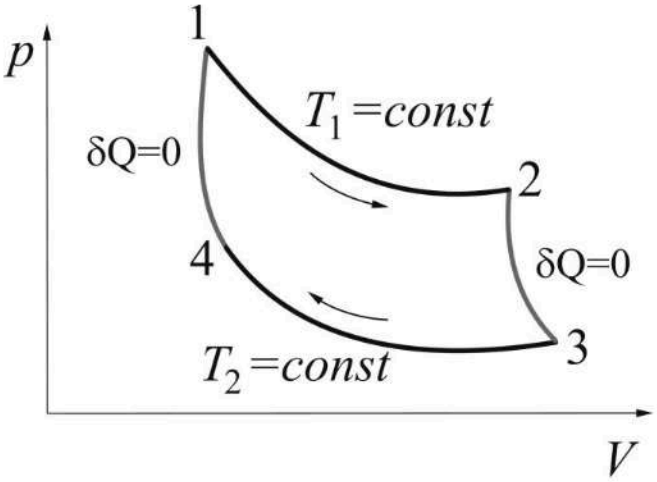


Fig. 5.6. Graphical representation of the Carnot cycle

Calculation of the efficiency of the Carnot cycle gives the value

$$\eta = \frac{Q^+ + Q^-}{Q^+} = \frac{T_1 - T_2}{T_1} = 1 - \frac{T_2}{T_1}, \qquad (5.20)$$

where Q^+ is the amount of heat received by the working body (gas) from the heater; Q^- is the amount of heat transferred by the working body to the cooler; T_1 is the heater temperature; T_2 is the cooler temperature.

5.3.2. The Second Law of Thermodynamics. Entropy

The laws of thermodynamics have dispelled the centuries-old dream of man to build a perpetual motion machine. In the history of physics we can find many examples of devices, which, at first glance, are such. It would seem that according to the first law of thermodynamics all received heat can be converted into work. But returning to the initial state 1 ($dU = 0$), by isotherm T_1 the machine will not do any work, while returning to the initial state by isotherm T_2 involves giving away heat. Thus, it is impossible to build a periodic machine that would do more work than the heat supplied to it from outside (a perpetual motion machine of the first kind). This statement represents the 1st law of thermodynamics.

From the consideration of the Carnot cycle it follows that two heat reservoirs and a working body are necessary to build a heat machine (it provides the transfer of heat from one reservoir to the other). This is the so-called Carnot principle of construction of heat machines, constituting the essence of the second principle of thermodynamics, which was formulated by other founders of thermodynamics, as the impossibility of building perpetual engines of the second type.

Clausius postulate (statement): It is impossible to have a cyclic process, the only result of which would be the transfer of heat from a less heated body to a more heated body (It is impossible to construct a cyclical machine whose sole effect is to transfer energy continuously by heat from one object to another object at a higher temperature without the input of energy by work). (See below).

Kelvin's principle: It is impossible to have a cyclic process whose only result is the production of work and the exchange of heat with a single heat reservoir. The conversion of a certain amount of heat must be accompanied by a transfer of heat to the refrigerator. (It is impossible to construct a heat engine that, operating in a cycle, produces no effect other than the input of energy by heat from a reservoir and the performance of an equal amount of work. Kelvin–Planck form of the second law of thermodynamics)**.** All of these statements are essentially identical.

The principles of heat machines are summarized in the form of the 1st and 2nd Carnot Theorems.

Irreversible processes. Clausius inequality. Entropy. Let us proceed to the quantitative formulation of the second law of thermodynamics, which is based on the Clausius statement that the efficiency factor of an irreversible thermal machine η_i is always less than the efficiency factor of a reversible Carnot machine η_0:

$$\eta_i < \eta_0. \tag{5.21}$$

In the general case:

$$\eta \le \eta_0, \text{ or } \eta \le 1 - \frac{T_2}{T_1} \qquad \text{or} \qquad \frac{Q^+ + Q^-}{Q^+} \le \frac{T_1 - T_2}{T_1}. \tag{5.21a}$$

After simple transformations we obtain the so-called Clausius inequality

$$\frac{Q^+}{T_1} + \frac{Q^-}{T_2} \le 0. \tag{5.22}$$

(That is, in an isolated system, the sum of the reduced heat $\frac{Q}{T}$ for a cyclic process is less than zero (for an irreversible process) or equal to 0 (for a reversible process). In generalized form it is written as follows:

$$\oint \frac{\delta Q}{T} \le 0. \tag{5.22a}$$

Consider a reversible cyclic process for which $\oint \frac{\delta Q}{T} = 0$. Then the infinitesimal change in the reduced heat $\frac{\delta Q}{T}$ can be replaced by the differential of the state function S ($\frac{\delta Q}{T} = dS$ or $\delta Q = TdS$), called entropy. So, in a cyclic reversible process occurring in a closed isolated system, $\oint dS = 0$.

This can be shown in another way. Let us divide the closed loop into two reversible branches I and II (Fig. 5.7):

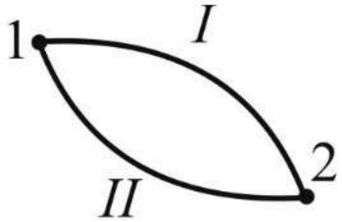


Figure 5.7. Image of a cyclic reversible process consisting of two reversible branches I and II

$$\oint \frac{\delta Q}{T} = \int_{1_I}^{2} \frac{\delta Q}{T} + \int_{2_{II}}^{1} \frac{\delta Q}{T} = 0 \quad \text{and} \quad \int_{1_I}^{2} \frac{\delta Q}{T} = - \int_{2_{II}}^{1} \frac{\delta Q}{T} = \int_{1_{II}}^{2} \frac{\delta Q}{T} \qquad (5.22b).$$

It follows from the condition (5.22b) that the reduced heat $\frac{Q}{T}$ does not depend on the way of transition from one state to another, i.e. it is a function of the state. Then the change in entropy during the transition from state 1 to state 2 is determined by the relation

$$\int_1^2 dS = S_2 - S_1. \qquad (5.22c)$$

It is obvious that $\oint dS = 0$.

Now let's imagine a cyclic process including irreversible section «I» (Fig. 5.8), in which the system first irreversibly passes from state 1 to state 2 and then reversibly returns to the initial state, i.e. in general the cyclic process is irreversible.

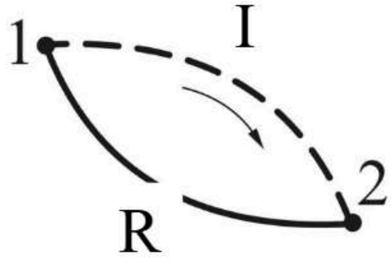


Figure 5.8. The representation of the cyclic process including reversible "R" and irreversible "I" regions

Let us use the Clausius inequality for an irreversible cyclic process (Eq. 5.22a):

$$\oint \frac{\delta Q}{T} < 0 \qquad \text{or} \qquad \int\limits_{\substack{1\\I}}^{2} \frac{\delta Q}{T} + \int\limits_{\substack{2\\R}}^{1} \frac{\delta Q}{T} < 0. \tag{5.22d}$$

The first integral relates to an irreversible process, so taking into account relation (5.22c), equation (5.22d) will be written in the form:

$$S_1 - S_2 + \int\limits_{\substack{1\\I}}^{2} \frac{\delta Q}{T} < 0 \quad \text{or} \quad S_2 - S_1 > \int\limits_{\substack{1\\I}}^{2} \frac{\delta Q}{T}\ . \tag{5.22e}$$

For an isolated system $\delta Q = 0$ and then:

$$S_2 - S_1 > 0. \tag{5.23}$$

So, in isolated systems entropy either does not change (in reversible processes) or increases (in irreversible processes). This statement is called the law of non-decreasing (increasing) entropy or the second law of thermodynamics, which in differential form is written as follows:

$$dS \geq 0. \tag{5.24}$$

Note that all natural processes are irreversible. This fundamental law means that all processes in nature proceed with increasing entropy, reaching a maximum value when the system is in equilibrium. That is, the second law of thermodynamics indicates the direction of the process.

5.3.3. Statistical representation of entropy. Nernst heat theorem, third law of thermodynamics

The state of the system as a set of molecules can be set by macroscopic parameters: volume, temperature, pressure, internal energy, entropy, etc. However, this so-called macroscopic state can also be set in another way – through the characteristics of all its molecules (for example, in the model of ideal gas through their coordinates and velocities (or momentums). Such a state is called a microscopic state. The number of different micro-states corresponding to a macro-state is called thermodynamic probability or statistical weight. The thermodynamic probability is fundamentally different from the familiar concept of "probability". Let's explain this with a simple example. Take a cube with numbered sides 1,2,3,4,5,6 (for example, a backgammon dice). The probability that after tossing one of these numbers will fall is 1/6. This is the so-called frequency probability. If the faces are unlabeled, a cube-shaped body has 6 identical ways to be placed on the game board, a tetrahedron has 4, and a ball has an infinite number. Accordingly, the thermodynamic probability (the term "statistical weight" is more appropriate here) is 6, 4 and infinity (determining the statistical weight with the symmetry properties is a common technique in statistical physics).

Now let us turn to the model of a molecular system. Imagine a flat square box, the bottom of which is divided by squares with low walls, which simulates a flat coordinate grid. Place the numbered balls in it (we number them so that it is possible to trace the movement of a particular ball, because it takes time to move it). We choose the depth and size of the cell so that the balls can easily roll over and, after they stop moving, slightly touch each other in the square. Let us now place in the box as many balls as there are at a distance of about 30-40 squares from each other. Thus, we have constructed a mechanical model of ideal gas. Naturally, the more squares and balls there are, the more accurately the model reflects the properties of this statistical system. Obviously, all squares are of equal value and the probability of placing a ball in any of the squares is the same. Each set of squares with balls (they are marked!) is one of the possible micro-states of the macrosystem, which are equally probable. This statement is called the postulate of equal probability of microstates. Let us now place a box with balls in a shaker and, after each stirring procedure, record their positions in the cells. The number of such placements (thermodynamic probability) simulating a microstate will naturally depend on the total number of balls. Let us now carry out a simple experiment. Let's divide the square into two equal parts. We will put 4 numbered balls inside the box, put a transparent lid on the box, and then we will repeat the mixing of balls and count after each procedure of mixing the number of balls on this or that side of the dividing strip. The result will coincide with the calculation predicted on the basis of the laws of statistics.

Event 1. All the balls will be placed in the left half of the box: there is only one way of this arrangement, which corresponds to the thermodynamic probability (denoted by Γ) equal to 1.

Event 2. There are three balls on the left and one on the right (3:1): there are 4 such combinations (the balls are labeled!) $\Gamma = 4$.

Event 3, two balls on each side: 6 ways, $\Gamma = 6$.

Event 4, one ball on the left, three on the right, $\Gamma = 4$.

Event 5, all the balls are in the right half of the box, $\Gamma = 1$.

The total thermodynamic probability Γ is 16.

Total combinations with the placement of 4 labeled balls is 16 (2^4). The probability of the first event is 1/16, the second 1/4, the third 1/6, the fourth 1/4, and the fifth 1/16. These are frequency probabilities of events. Their sum is 1. This means that some of these events will necessarily happen.

Obviously, the more balls (events), the lower the probability of the event when all the balls are on the same half. And the greater the probability of the event when both sides of the dividing strip will have approximately the same number of balls, i.e. the number of such combinations (thermodynamic probability) will be the greatest. Let us continue our reasoning. The initial arrangement of balls on one side of the box – the only arrangement – is disturbed by mixing, and it cannot be reached spontaneously. Hence, it is a nonequilibrium state from which the system moves to an equilibrium state to which the maximum number of ways of its realization corresponds (the maximum value of the thermodynamic probability). Now let's establish a correlation with the second law of thermodynamics.

In the equilibrium state $\Gamma = \Gamma_{max}$. But according to the second law of thermodynamics, entropy in irreversible processes occurring in isolated systems increases and reaches its maximum value, i.e. $S \to S_{\max}$. That is, there is a logical connection between the value of Γ and entropy S, established by the Boltzmann formula:

$$S = k\ln\Gamma\,. \tag{5.25}$$

Further, at the hypothetical stopping of molecular motion, which would correspond to temperature $T = 0$, there is only one way of realization of the state, i.e. $\Gamma = 1$. Then it is fair to say that when temperature tends to absolute zero* ($T \to 0$), entropy tends to be zero:

---

[*] Absolute zero is unattainable in practice. Its existence and position on the temperature scale follows from the extrapolation of observed physical phenomena (see Section 4.1.2), according to which at absolute zero the energy of thermal motion of molecules and atoms of matter must be equal to zero. This means that the chaotic motion of particles stops and they form an ordered structure (liquid helium is an exception). However, from the point of view of quantum physics, even at absolute zero there are zero-point fluctuations, which are caused by quantum properties of particles and the physical vacuum surrounding them.
Absolute zero (zero) is a minimal temperature, which a physical body in the Universe can have in the limit. It serves as the beginning of the Kelvin absolute temperature scale. The reference temperature is the temperature corresponding on the phase diagram (see Section 6.4) to the triple point of water (273.16 K, 0.01 °C).

$$\lim_{T\to 0} S = 0. \tag{5.26}$$

This statement is the third law of thermodynamics, also called Nernst's theorem. It establishes limits for thermal processes. Since for quantitative calculations the difference of the state function S is more important (see equation 5.22c), we can take as a starting point the state to which $S = 0$ corresponds. Then the entropy at some temperature T is

$$S = \int_0^T dS = S_T\,. \tag{5.26a}$$

Let's note that in the thermodynamic sense the second law of thermodynamics is absolute. But it is not statistically absolute because it is not always fulfilled: there are processes that are conditioned exclusively by manifestations of the statistical properties of the system. These are numerous physical processes resulting from fluctuations of physical parameters of the system – their deviations from mean values which correspond to equilibrium values of macroscopic parameters. For example, deviation of particle concentration (i.e. density) from equilibrium value.

5.3.4. Thermodynamic functions and potentials

So, to the already listed thermodynamic state functions of the system *p, V, T, U* one more function is added – entropy *S.* In principle, there are many of them ‑ any function of the state function is also a thermodynamic function. But it is important to distinguish from them the so-called thermodynamic potentials – *definite thermodynamic functions of independent macroscopic parameters, fully defining the thermodynamic state of the system.* In this case, knowing thermodynamic potentials as functions of any complete set of system parameters, by differentiation and integration it is possible to calculate not only other parameters of the system, but also any macroscopic characteristics of the system and processes proceeding in them. Among the above thermodynamic functions only the internal energy is a thermodynamic potential. Note that entropy is an additive quantity: if a system is divided into two parts and each part corresponds to entropy S_1, S_2, *respectively,* then the entropy of the whole system is equal to the sum of entropies of its parts $S = S_1 + S_2$.

To describe the thermodynamic features of living systems, we need another thermodynamic potential – *free energy*.

First, we obtain the relationship between system parameters and internal energy. It follows from the first law of thermodynamics that $\delta Q = dU + pdV$, and from the second $\delta Q = TdS$ (for isothermal process). Combining them into one relation, we obtain the so-called thermodynamic identity:

$$TdS = dU + pdV, \tag{5.27}$$

since in reversible processes both parts of this equality are fulfilled identically[*]. It follows that:

$$dU = -pdV + TdS, \tag{5.27a}$$

and then:

$$-p = \left(\frac{\partial U}{\partial V}\right)_S, \qquad T = \left(\frac{\partial U}{\partial S}\right)_V, \tag{5.28}$$ [**]

---

[*] An identity (in mathematics) is an equality that is satisfied over the entire set of values of its constituent variables.

[**] Thermodynamic potentials are functions of state. Therefore, the increment of any of the potentials is equal to the full differential of the function by which it is expressed. The full differential of the function $f(x, y)$ of the variables x and y is defined by the expression

$$df = \left(\frac{\partial f}{\partial x}\right)_y dx + \left(\frac{\partial f}{\partial y}\right)_x dy.$$

Therefore, if in the course of transformations we obtain for an increment of some quantity f an expression of the form

$$df = X(\xi, \eta)d\xi + Y(\xi, \eta)d\eta,$$

it can be said that this quantity is a function of the parameters ξ and η, with the functions $X(\xi, \eta)$ and $Y(\xi, \eta)$ being partial derivatives of the function

$$\left(\frac{\partial f}{\partial \xi}\right)_\eta = X(\xi, \eta), \quad \left(\frac{\partial f}{\partial \eta}\right)_\xi = Y(\xi, \eta).$$

i.e. to determine the parameters T and p we choose the thermodynamic potential – internal energy as a function of volume and entropy $U = U(S, V)$. The minimum of the internal energy corresponds to the equilibrium state of the system. Let us write expression (5.27) in another form:

$$pdV = -dU + TdS. \tag{5.29}$$

Since in an isothermal process $TdS = d(TS)$, then

$$pdV = -dU + d(TS) = -d(U - TS) = -dF = \delta A \tag{5.29a}$$

or

$$\delta A = -dF. \tag{5.29b}$$

Here F is the so-called Helmholtz free energy (or Lewis free energy, as it is also called), which is the thermodynamic potential defined as

$$F = U - TS. \tag{5.30}$$

In integral form, i.e. for transition from initial state 1 to final state 2, equation (5.30 a) is written as

$$A = F_1 - F_2. \tag{5.31}$$

Thus, in a reversible isothermal process, the work is done due to the loss of free energy F, which plays the role of potential energy. However, equation 5.29b is satisfied only in a static (quasi-static) process. This is a model representation. In reality, the process of isothermal expansion is irreversible because the work done is less. And in general case

$\delta A \leq dF$ (5.31a) or $A \leq F_1 - F_2$. (5.31b)

(The inequality sign refers to irreversible processes). This means that the loss of free energy determines the maximum work that the system can do in an isothermal process. Using equations 5.27 and 5.29, the differential of the state function, the free energy given by parameters *S* and *V*, as a function *F*(*SV*) can be represented in the form:

$$dF = -SdT - pdV. \tag{5.32a}$$

The system parameters – pressure and entropy at constant values of temperature and volume are equal, respectively:

$$-p = \left(\frac{\partial F}{\partial V}\right)_T; \ -S = \left(\frac{\partial F}{\partial T}\right)_V. \tag{5.32b}$$

Thus, the place of molecular physics in natural science is determined by the following:

1. The concept of temperature (the zeroth law of thermodynamics) was introduced.
2. The law of conservation of energy in general form – the 1st law of thermodynamics – is formulated.
3. The direction of processes is established – the 2nd law of thermodynamics. It is absolute thermodynamically and not absolute statistically.
4. The 3rd law of thermodynamics indicates the limitation of processes.
5. In addition to macroscopic parameters *p, V, T* thermodynamic potentials – thermodynamic functions *U, S, F* (and others) of independent macroscopic parameters, completely defining the thermodynamic state of the system, are introduced.

SECTION 6. PHYSICAL SYSTEMS AND PHENOMENA (PROCESSES)

6.1. Real gases. Van der Waals equation of state. Liquefaction of gases

In real gases molecules interact with each other. It is natural to assume that the interaction is caused by attraction and repulsion forces, which determine the type of the potential interaction curve (Fig. 6.1). The attraction forces, the so-called van der Waals forces, are electromagnetic in nature. They are manifested at distances of the order of 10^{-8} cm between the centers of molecules and are described by the function

$$F(r) \sim -r^{-7}. \qquad (6.1)$$

There are three types of Van der Waals forces (arranged in descending order of absolute value): dipole-dipole interaction of electric dipole moments of molecules (so-called orientation forces); dipole-induced dipole interaction (induction forces), i.e. interaction between polar molecules and non-polar molecules where the electric dipole moment is induced by the electric field of polar molecules, dispersion interaction (dispersion forces, caused by electromagnetic interaction of electrons rotating around the atomic nuclei in electrically neutral molecules). The repulsive forces are of quantum nature (Fig. 6.1) and are described by the formula

$$F(r) \sim r^{n}, \qquad (6.2)$$

where $n \geq 9$. They appear at a distance of less than 10^{-8} cm.

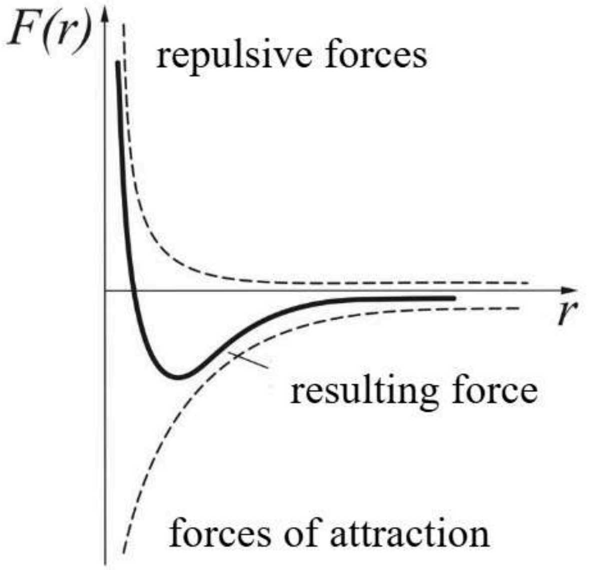


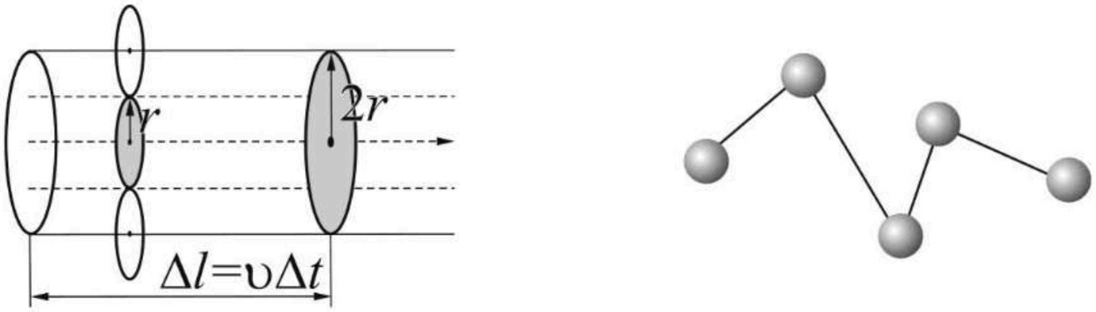
Fig. 6.1. Schematic representation of the dependence of intermolecular interaction forces on the distance between molecules

What experimental evidence suggests that there are interaction forces between molecules?

1. Dependence of the average free path length on temperature. The free path of a molecule $\langle\lambda\rangle$ is the average distance that a molecule travels between two successive collisions, one of the kinetic characteristics of gas in the Kinetic Theory of Gases. For an ideal gas, it does not depend on temperature and is defined by the expression

$$\langle\lambda\rangle = \frac{1}{n 4\pi r^2 \sqrt{2}}, \tag{6.3}$$

where n is the concentration of molecules, r is the radius of the molecule. In the model of collision of rigid spheres this relation can be obtained from simple considerations (explanations in Fig. 6.2).



a b

Fig. 6.2. The determination of the mean free path length of molecules in the model of collision of rigid spheres

Let the molecule move in a straight line with average velocity $\langle\upsilon\rangle$ (Fig. 6.2a). During the time Δt it travels an average path $\langle\Delta l\rangle = \langle\upsilon\rangle\Delta t$ and collides with all molecules whose centers are no further away than the surface of a cylinder of radius $2r$. The cross-sectional area of this cylinder (shaded) is $4\pi r^2$, where r is the radius of a rigid sphere with cross-sectional area πr^2 (also shaded). The number of such molecules is equal to $\langle\Delta l\rangle \cdot 4\pi r^2 \cdot n = \langle\upsilon\rangle\Delta t 4\pi r^2 n$.

The average number of collisions ν per unit time is determined by the relation:

$$\nu = 4\pi r^2\langle\upsilon\rangle n. \qquad (6.3a)$$

On average, the path traveled by a molecule per unit time is equal to its average velocity $\langle\upsilon\rangle$, so the average free path $\langle\lambda\rangle$ (sometimes denoted λ), would be

$$\langle\lambda\rangle \equiv \lambda = \frac{\langle\upsilon\rangle}{\nu}. \qquad (6.3b)$$

However, chaotically moving molecules change direction after each collision. This means that when determining the average free path length, it is necessary to consider the average relative velocity $\langle\upsilon_{rel}\rangle = \langle\upsilon\rangle\sqrt{2}$. Taking this circumstance into account, we obtain the relation (6.3) for the average free path length of a molecule.

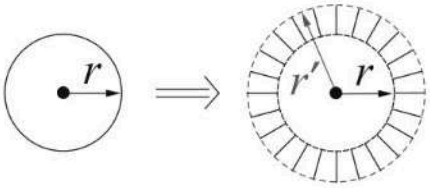
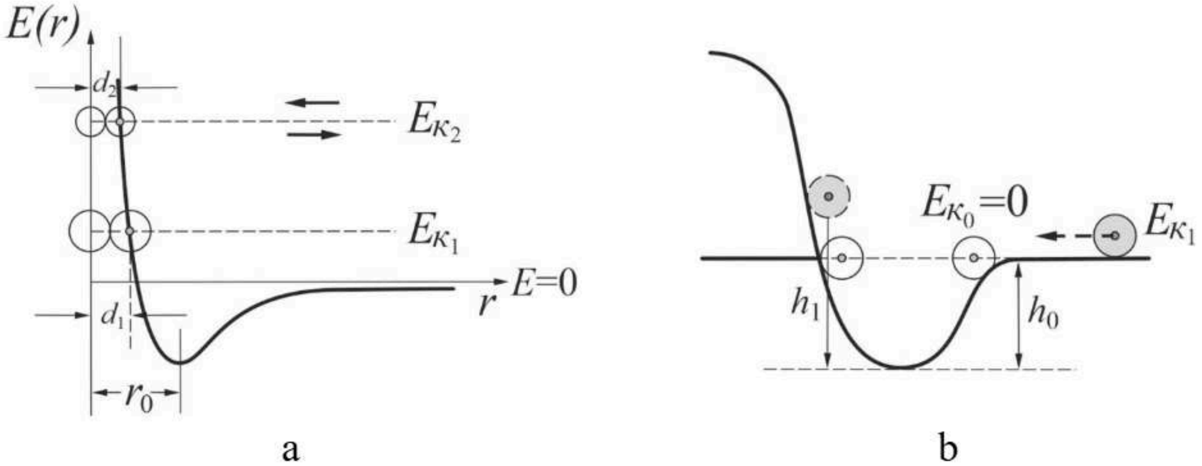


Fig. 6.3. Schematic representation of a cross section of a molecule

In the model of rigid spheres, there is no dependence on temperature. But in real gas $\langle\lambda\rangle = f(T) \sim \frac{1}{T}$, i.e. it decreases with increasing temperature. This is due to the fact that in reality the radius of the molecule is slightly larger than in the rigid sphere model. It is necessary to consider some molecule with radius r' (Fig. 6.3), which determines the minimum distance that two molecules can come close to each other, after which they fly apart in different directions if their energy is not enough to overcome the repulsive forces that would lead to the formation of a chemical bond between them (a new molecule). This minimum distance determines the so-called effective collision cross section $\sigma = 4\pi(r')^2$.

The greater the energy of the molecules (i.e., T), the closer the molecules come to each other. The effective collision cross section σ decreases, i.e. the average distance $\langle\lambda\rangle$ before the collision increases. This is clearly illustrated in Fig. 6.4, which shows schematically the dependence of the potential energy of intermolecular interaction on the distance between the conditional centers of the molecules.



a

b

Fig. 6.4. a) – Schematic representation of the potential energy of interaction between molecules; b) – mechanical model explaining the degree of convergence of interacting molecules

Let's imagine that one molecule is placed at the origin of the coordinates and the other molecule is moved along the *r*-axis (direction). When $r > r_0$ the other molecule is in the zone of attraction forces and tends to get closer with increasing speed. At $r < r_0$ it enters the repulsive force zone and rapidly loses its velocity to 0 at the maximum permissible distance $d_1 = 2r'$. The greater its kinetic energy reserve (i.e., the higher the system temperature), the higher on the potential curve the molecule can move or, in other words, get closer to a smaller distance d_1.

This can be explained with mechanical models from sports, such as the one shown in Fig. 6.4b. Imagine that you have to ski down the slope of a ravine and climb to the opposite side of the ravine without any additional conditions. Let's assume that the snow on the slopes of the ravine is treaded and the skis are lubricated so much that friction losses are minimal. Obviously, you will ascend the opposite slope by inertia if you accelerate or start descending with the highest possible initial speed, which can be set in various ways (for example, by pushing off with ski poles). This is equivalent to the motion of a molecule (shaded) from infinity with some initial kinetic energy E_{k_1}. *At the end of the descent, the kinetic energy will be the sum of the initial* E_{k_1} *and the acquired potential energy* mgh_0, *determined by the depth of the ravine (the depth of the potential well* $E(r_0)$ *at* r_0 *is an imitation of the gravitational force). If there is no loss of energy, the total energy, which at the beginning of acceleration was equal to* E_{k_1}, *remains unchanged, and the skier will ascend to the reverse slope (imitation of repulsive forces, the slope can be imagined higher) to the height* h_1, *determined by the condition* $mgh_1 = E_{k_1}$

To what limit can molecules come closer together as temperature increases while maintaining their molecular structure? Up to a distance comparable to twice the covalent radius of the molecule ($2r_c$). As the kinetic energy of the molecules (temperature) increases further, it is sufficient to overcome the repulsive energy (the exchange interaction of molecules), which will lead to the formation of a chemical bond, i.e., the formation of a new molecule, which will be characterized already by a different value of the covalent radius (size) and van der Waals radius. This is how the formation of a molecule by the collision mechanism occurs. Obviously, $r_c < r_0$.

In accordance with the above, the so-called van der Waals radius r_0 is the radius of the sphere of effective intermolecular interaction.

For air molecules: $r_0 \sim 2{,}5\cdot10^{-10}$ m, $\lambda \sim 10^{-7}$ m.

At higher pressures, there is a deviation of pV, which follows from the equation $pV_m = RT$. For example, for nitrogen at $p = 100$ atm, the deviation is 7%. This means that certain corrections must be made in the Clapeyron-Mendeleev equation. This was most successfully done by Van der Waals, whose name was given to the equation of the state of real gases.

Van der Waals equation . Let's consider the simplest model representation of the equation of the state of real gases. This is the Clapeyron-Mendeleev equation $pV_m = RT$, in which it is necessary to take into account the presence of attraction forces of molecules, which is equivalent to an additional pressure p_i $(p \Rightarrow p + p_i)$. And, since the molecule is sized, the volume of space to move the molecules decreases by the amount V', $(V_m \Rightarrow V_m - V_m{}')$. Now the equation of state will take the form:

$$(p + p_i)(V_m - V_m{}') = RT. \tag{6.4}$$

In the approximate estimation of the correction for volume $V_m{}'$ we can proceed from the principle of pairwise interaction (Fig. 6.5), according to which only a pair of molecules participates in one act of collision.

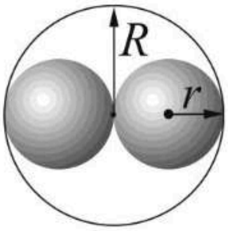


Fig. 6.5. The explanation of the principle of pairwise interaction
r is the radius of a molecule, *R* is the radius of a sphere in which only two molecules can be in it at the same time

The volume of one molecule $V_0 = \frac{4}{3}\pi r^3$. As can be seen from Fig. 6.5, two interacting molecules have a volume $V(R) = \frac{4}{3}\pi(2r)^3$, $(R = 2r)$. As a result, the correction for the volume is of the order of the value of the quadrupled volume of one molecule $V' = 4V_0$. The Van der Waals equation for a mole of gas with the above[*] corrections takes the following form:

$$\left(p + \frac{a^2}{V_m}\right)(V_m - b) = RT, \qquad (6.5)$$

where *a, b* are Van der Waals corrections – individual for each gas. They are determined from experiments on liquefaction of gases.

Gas liquefaction. The gas liquefaction experiment is shown schematically in Fig. 6.6.

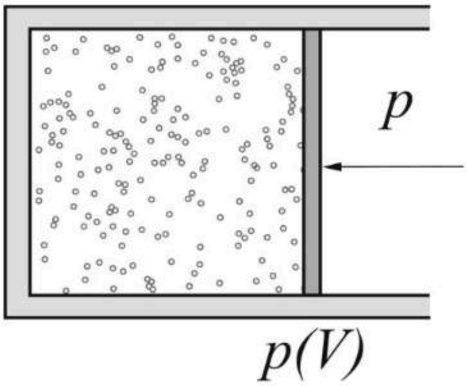


Fig. 6.6. Schematic representation of the gas liquefaction experiment

There is gas in the cylinder under the piston, which is compressed by moving the piston with great force. In this way, the relationship between pressure, the volume of gas under the piston (the piston position can be recorded) at different temperatures is established, i.e. isothermal

---

[*] The van der Waals equation is a model representation of the equation of state of a real gas. A gas whose state is satisfactorily described by this equation is called a van der Waals gas. Other model representations (equations) of real gas are also known.

dependences $p(V)$ are obtained. Under certain conditions, liquid droplets form inside the cylinder, i.e. a two-phase gas-liquid system is formed, in which dynamic equilibrium is established at pressure p_s, called saturated (saturating) vapor pressure. As the piston continues to move, the pressure p_s remains unchanged until the entire volume under the piston is filled with liquid. Subsequent compression requires considerably more force, since the compressibility of the liquid is much lower than that of the gas. Such dependences $p = p(V)$ at various values of $T_1 ... T_i$ (real gas isotherms) are shown in Fig. 6.7. They form a family of isotherms, which are separated by the so-called critical isotherm T_{cr}. At $T > T_{cr}$ the isotherms correspond to the gas phase, at $T < T_{cr}$ there is a two-phase state. The state in which the liquid and gaseous phases become indistinguishable (their densities become the same) is called the critical state. The critical state is characterized by a set of parameters p_{cr}, T_{cr}, V_{cr}, which are related to the corrections a and b by the following relations[*]:

$$V_{cr} = 3b, \qquad p_{cr} = {}^{a}\!/_{27b^2}, \quad T_{cr} = {}^{8a}\!/_{27Rb}. \qquad (6.5a)$$

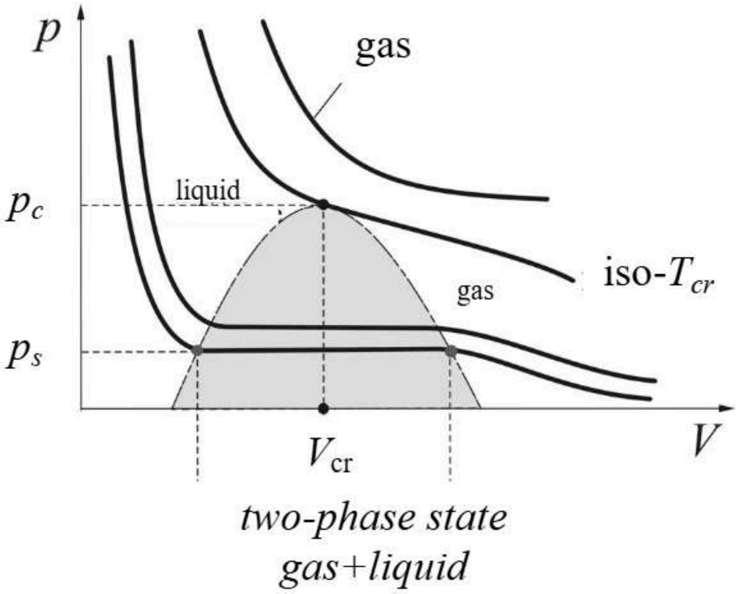


Fig. 6.7. Schematic representation of real gas isotherms. The two-phase state region is marked by a gray background

---

[*] The above theoretical relations are somewhat different from the experimental.

Note that saturating vapor pressure p_s does not depend on the volume of the system $p_s \neq f(V)$, but depends on temperature, i.e. $p_s = f(T)$.

For water T_{cr} = 647,2 K (374,1 ºC), p_{cr} = 218 atm. The densities of water and saturating steam in the critical state are the same and equal to 329 kg/cm^3.

6.2. Structure and general properties of liquids. Surface phenomena

6.2.1. The simplest model of a liquid. Basic properties of a liquid

Frenkel's model. Liquid, unlike gases, can be stored in an open vessel. It follows that there are interaction forces between the molecules forming a liquid sufficient to hold them together, i.e. the interaction energy of the molecules is comparable to the energy of their thermal motion. Consequently, the thermal motion of liquid molecules obeys different laws, since they are no longer able to move freely. Let us consider the model of a simple liquid as a set of molecules of spherical shape, between which there are no specific interactions, but only the van der Waals interaction. This model was proposed by J. Frenkel.

Assume that there is a molecule in an arbitrary region of space with the coordinates x, y, z. It vibrates about an equilibrium position with a certain frequency and stays in this position for time τ_0, called the sedentary time. The transition of this molecule to a new equilibrium position with coordinates x', y', z' can only happen by overcoming the potential barrier E_a, as depicted in Fig. 6.8.

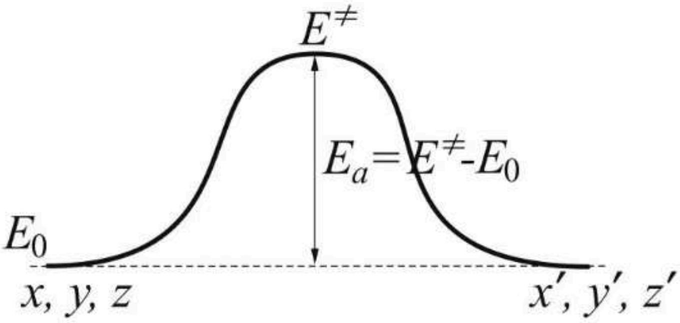


Figure 6.8. The illustration of thermal motion in the Frenkel's model

Any molecule with kinetic energy equal to $E^{\neq}$ due to thermal motion can climb to the top of this barrier (usually denoted by $E^{\neq}$), and then can either return to the initial position or move to a new equilibrium position with the same probability. Now let us use the Boltzmann distribution, according to which the number of molecules in the potential field E_p is proportional to $\exp\left(-{}^{E_p}/_{RT}\right)$. That is, the number of molecules at the stable equilibrium positions $(x,\ y,\ z,\ x',\ y', z')$ is proportional to $\exp\left(-{}^{E_0}/_{RT}\right)$, and at the top of the barrier $\exp\left(-{}^{E^{\neq}}/_{RT}\right)$. Molecules are in dynamic equilibrium and the equilibrium constant K between activated molecules (i.e. those at the top of the barrier-transition state) and unactivated ones in a stable equilibrium position is determined by the relation:

$$K=\frac{N^{\neq}}{N}=A\exp\left({}^{-E_a}/_{RT}\right)=k, \tag{6.6}$$

where E_a is the activation energy of thermal motion of molecules ($E_a=E^{\neq}-E_0$), A is the frequency factor determining the probability that a molecule possessing energy E_a, will pass this potential barrier. According to the absolute velocity theory for monomolecular reactions of the first (pseudo first) order (transition state theory), the equilibrium constant K is the reaction (process) rate constant k (in this case $k={}^{1}/_{\tau_0}$) , i.e. equation (6.6) can be written in the form:

$$k = \tau_0^{-1} = A(\tau)\exp\left({}^{-E_a}/_{RT}\right). \tag{6.6a}$$

Equation (6.6a) is a kinetic equation similar to the empirical Arrhenius equation, in which the pre-exponential multiplier A means the collision frequency of the molecules.

In turn, τ_0 is proportional to viscosity η. Thus, the temperature dependence of viscosity can be represented as:

$$\eta = B(\tau)\exp\left({}^{E_a}/_{RT}\right). \tag{6.6b}$$

From the experimental relationship $\eta = f\left(\frac{1}{T}\right)$ the activation energy of thermal motion of the fluid is determined. Note that the coefficients $A(\tau)$ and $B(\tau)$ have a temperature dependence, but it is small and is often neglected.

The presence of the above-mentioned local stable equilibrium positions can be represented graphically (Fig. 6.9) as a distribution function $G(r)$.

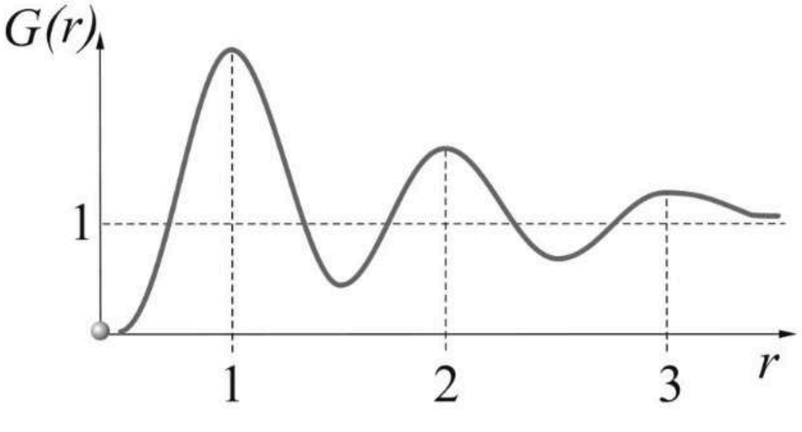


Fig. 6.9. Schematic representation of the distribution function *G(r)*, 1, 2, 3 – coordination spheres in the Frenkel model

Experiments show that there are three such regions where the probability of detecting a molecule is greater than 1. They are called coordinate spheres. Thus, unlike gases, in liquids there is a so-called near-order. It is obvious that at infinite distance $G = 1$, i.e. there is always another molecule in the surrounding of the molecule under consideration.

Basic properties of the fluid:

1. *Compressibility*. It is characterized by the compressibility coefficient, which is determined at a constant temperature as follows:

$$\chi = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T, \tag{6.7}$$

where χ is the coefficient of isothermal compressibility, V is the initial volume of liquid.

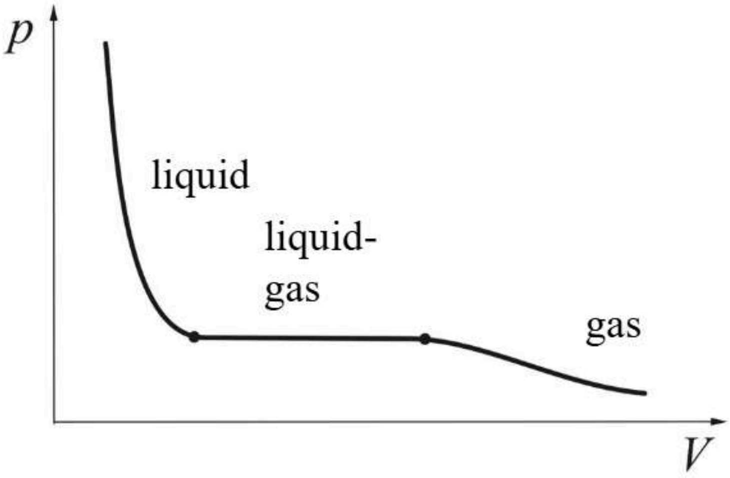


Fig. 6.10. General view of the real gas isotherm

As already noted (see section 2.) and as it follows from the general view of the experimental isotherm (Fig. 6.10), the compressibility of liquid is much less than that of gas: $\chi_l \ll \chi_g$. The order of these values is as follows:

$$\chi_l \approx 10^{-4} \div 10^{-5}\ \text{atm}^{-1}, \qquad \chi_g \approx 1\ \text{atm}^{-1}.$$

It has been experimentally established that in liquids the compressibility is related to the *a* and *b* corrections in the van der Waals equation for the gas whose molecules form the liquid:

$$\chi_l = \frac{A}{V_m\left(p + \frac{a}{V_m^2}\right)}. \tag{6.7a}$$

This ratio is valid within a certain pressure range. The determination of χ is called piezometry.

Thermal expansion of liquids is characterized by the coefficient of thermal expansion α, which is determined at constant pressure as follows:

$$\alpha = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_p. \tag{6.7b}$$

The coefficient of thermal expansion of liquids α is proportional to temperature, $\alpha_l \approx \alpha(T)$, and in gas is proportional to the inverse of temperature, $\alpha_g \approx \frac{1}{T}$.

The procedure for determining is called dilatometry.

The heat capacity of liquids, as well as in gases, is characterized by C_p, C_V values, but the Meyer relation is not fulfilled, $C_p - C_V \neq R$.
For liquid metals $C_p \approx C_V \approx 3R$ (as for solids).

As a rule, liquids have only one liquid phase, but there is an exception – liquid helium (obtained from gaseous helium at $T = 4{,}21$ K), which has the properties of a classical liquid (He I), with a finite viscosity value in the range 4,24 K – 2,17 K; at temperatures below 2,17 K (more precisely 2,172 K) it is a quantum non-viscous ($\eta = 0$) superfluid (He II).

6.2.2. Surface phenomena

For a molecule inside the liquid (see Fig. 6.11), the attraction forces from other molecules are zero in total. For molecules near the liquid-air boundary, the interaction zone between the liquid molecules is not completely filled, so they are subjected to a force directed inside the liquid. Transition of molecules from liquid depth to surface is connected with overcoming of this force. This applies equally to molecules tending to leave the liquid in the process of evaporation. This work is done at the expense of the kinetic energy of the thermal motion of the molecules. Thus, the molecules in the surface layer (and thus the entire layer) have additional potential energy. As a consequence, the liquid tends to shrink its surface. The liquid borders the vessel, and on each element of this border (contour) there is a force uniformly distributed throughout the contour, called the surface tension force. It is directed along a tangent to the liquid surface perpendicular to the contour line of this elementary section of the surface boundary. A quantitative measure of surface tension is the value σ, called the surface tension coefficient:

$$\sigma = {}^{F}/_{l} \quad \text{(N/m)}, \qquad (6.8)$$

where l is the contour boundary, F is the surface tension force.

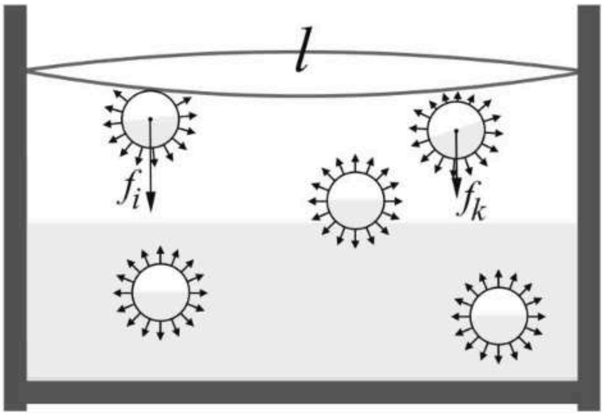


Fig. 6.11. The explanation of the surface tension forces

The surface profile of a liquid depends on the ratio of forces of intermolecular interaction at the boundary of contacting media: vessel walls-liquid (F_{sl}), vessel walls-atmospheric gas (F_{sg}), liquid-atmospheric gas (F_{gl}). In accordance with this, liquids are called wetting and non-wetting. Fig. 6.12 shows profiles of drops of such liquids on the horizontal surface of a solid body. The directions of forces acting at the boundary of the contacting media are also shown there.

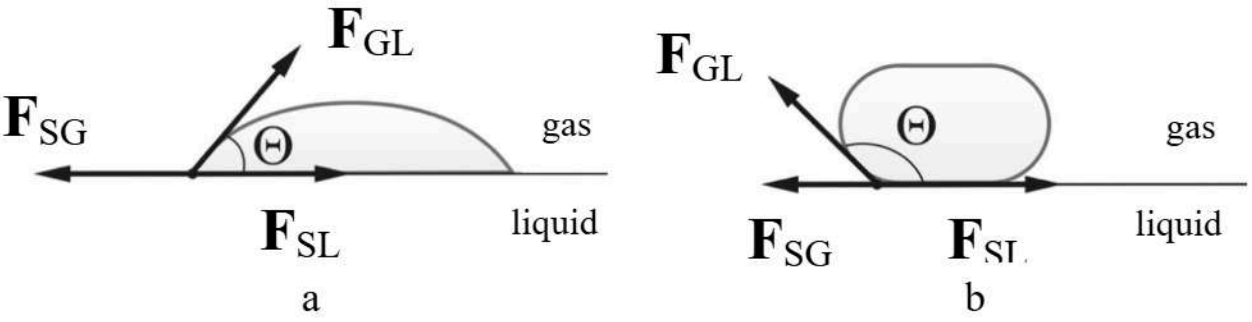


Fig. 6.12. The examples of profiles: a) – wetting liquid,
b) – non-wetting liquid

A quantitative characteristic of the surface layer profile is the edge angle Θ defined by the relation:

$$\cos\Theta = \frac{\sigma_{SG} - \sigma_{SL}}{\sigma_{GL}}, \tag{6.8a}$$

where σ_{SL}, σ_{GL}, σ_{SG} are the corresponding surface tension coefficients at the boundaries of media: "solid-liquid", "gas-liquid" and "solid-gas", connected by inequality $\sigma_{SL} > (\sigma_{GL} + \sigma_{SG})$. An important consequence of the surface tension is capillary phenomena, which are manifested by the fact that, for example, a liquid in a thin tube lowered into a vessel with liquid may rise to a certain height if the liquid is wetting, or conversely, fall below the liquid in the vessel outside the tube if the liquid is non-wetting (Fig. 6.13a and 6.13b, respectively).

In both cases, the absolute value of the change in the liquid level inside the tubes with respect to the liquid level outside the tubes is the same and is determined by the ratios:

$$h = \frac{2\sigma}{\rho g R} \quad \text{or} \quad h = \frac{2\sigma\cos\Theta}{\rho g r}. \tag{6.9}$$

The manifestation of capillary properties is very diverse. The most typical examples in medical practice are gauze swabs drying the surface of wounds, drainage of human internal organs with them after surgery, blood sampling using thin tubes, etc.

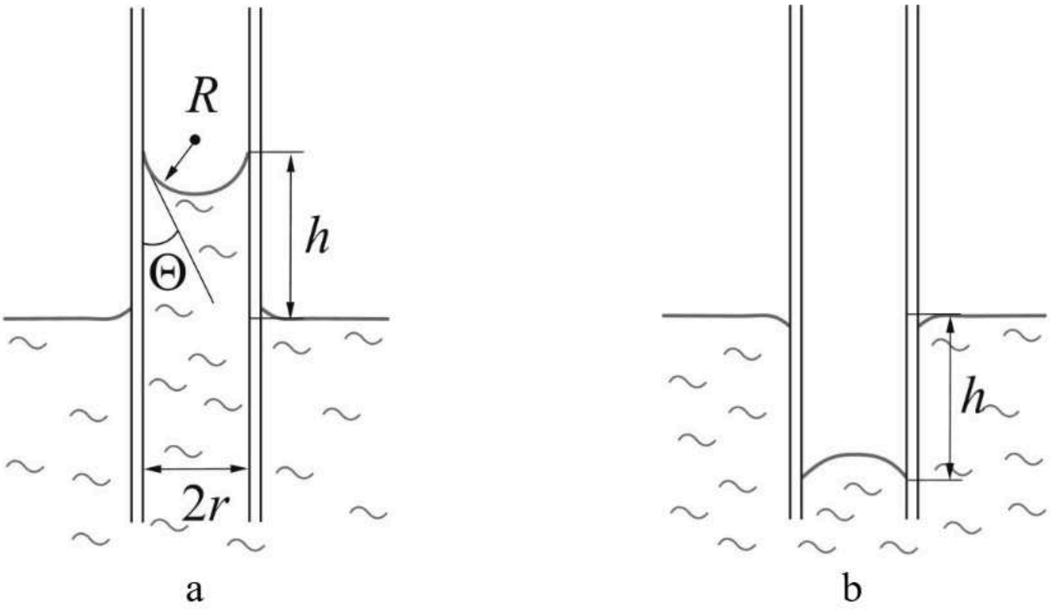


Fig. 6.13. The illustration of capillary phenomena: a) – in case of wetting liquid, b) – non-wetting liquid

6.3. Crystalline state

Note that in physics, the terms "crystalline state" and "solid phase" (solid body) are synonymous. A solid body retains its volume and shape, hence the interaction energy of the particles is much greater than their thermal motion energy, i.e. $E_{\text{int}} >> kT(RT)$. It is known that all liquids change to the crystalline phase when the temperature drops and/or pressure increases. However, there are so-called anomalous liquids (water, antimony, bismuth), where liquid-solid transition has its own peculiarities. For example, the density of water in interval 0÷4 decreases as the temperature decreases, not increases. This property of water is extremely important for living nature. When the ambient temperature decreases below 0 °C under the ice layer forming on the surface of the reservoir, there is an intense exchange of warm and cold layers of water. As the temperature decreases, the ice thickness increases to a certain value. If a body of water does not freeze to the bottom, the life of microorganisms and living beings is preserved in it (including the bottom layer of the ground). The only problem is the lack of oxygen in this environment. It is solved in a simple way: if it is necessary to preserve the life of inhabitants of shallow bodies of water, ice-holes are made in the ice

Let us highlight the main features of the solid phase (crystalline state):

- anisotropy of physical properties;
- molar heat capacity at room temperature and above is equal to $C_{p,V} = 3R$ (Dulong -Petit law);
- transition from the liquid phase to the solid phase takes place discontinuously.

The manifestation of such properties is a consequence of the ordering of the structure of matter, which extends to the entire substance as a whole. This means that the mutual arrangement of matter particles is preserved in all directions, forming a so-called crystal lattice, in the nodes of which these particles are located (Fig. 6.14). The curve of potential interaction of adjacent particles has approximately the same form as in liquids and gases. Obviously, such potential curves can be plotted along all three coordinate axes, then we obtain a set of structural parameters a_1, a_2, a_3, called periods of the crystal lattice: they correspond to distances between periodically arranged potential wells on these interaction curves. Of course, this is a schematic representation of the lattice structure.

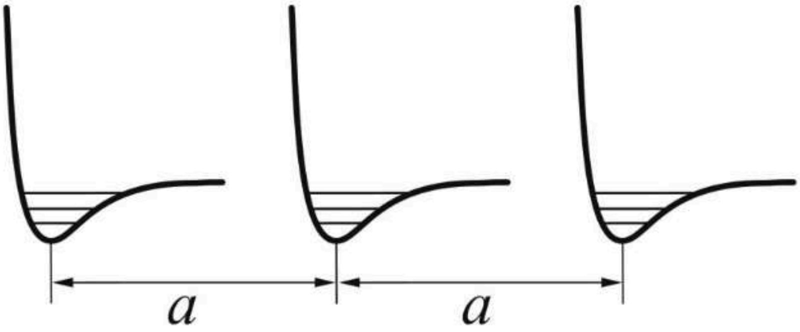


Fig. 6.14. The family of potential interaction curves between adjacent elements of matter located in the nodes of crystal lattice. Such a family can be represented in other directions of three-dimensional space

Crystals (solid bodies) are classified according to the following features:

- symmetry properties (crystalline classes);
- chemical properties of elements located in nodes of crystal lattice (atomic, ionic, molecular crystals, metal structure), these are so called physical types of crystals.

Any crystal lattice can be constructed of geometric figures (structural elements) by moving them along coordinate axes. Schematically, it can be represented as a mosaic picture (Fig. 6.15).

The notions of elementary cell and primitive cell are used. An elementary cell is a structural element of a crystal lattice, the symmetry of which coincides with the symmetry of the whole crystal lattice. In Fig. 6.15 it is a hexagon made of two types of equilateral triangles with different colors. Rhombic shapes 1, 2 are made of a minimal set of selected shapes – they are primitive cells. I.e. a primitive cell is an elementary cell with a minimal set of chemical elements characterizing the substance.

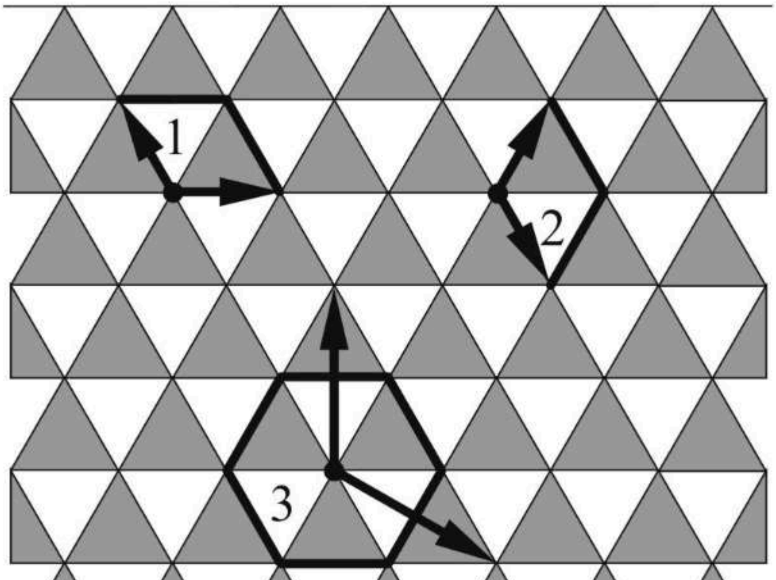
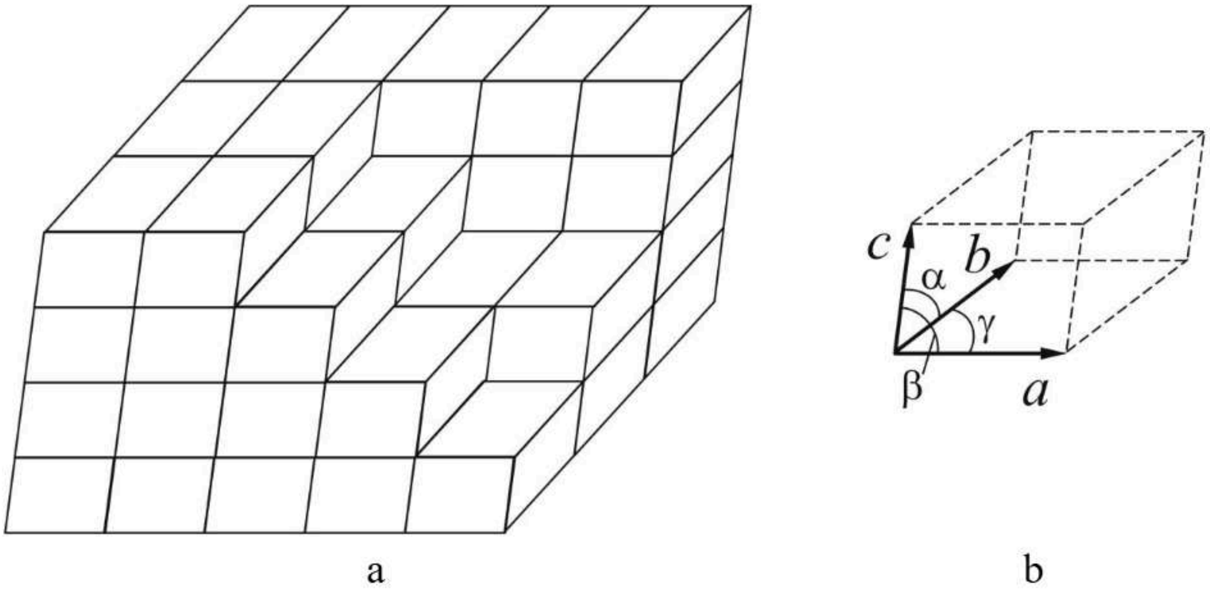


Fig. 6.15. The illustration of elementary (figure 3) and primitive (1, 2) crystal lattice cells

The strict construction of the crystal lattice is based on the so-called Bravet cell (Fig. 6.16b).



a b

Fig. 6.16. a) – Bravais lattice, b) – primitive Bravais cell (a, b, c – periods of crystal lattice)

A more detailed discussion of these issues is beyond the scope of this course.

6.4. Thermal properties of crystals

Let’s consider the basic thermal properties of crystals.

Phase transitions of the 1st order "liquid -solid" (Fig. 6.17). Let us carry out the following experiment. We will cool water in a vessel. As a measure of influence we choose the time during which cooling will take place (left graph in Fig. 6.17).

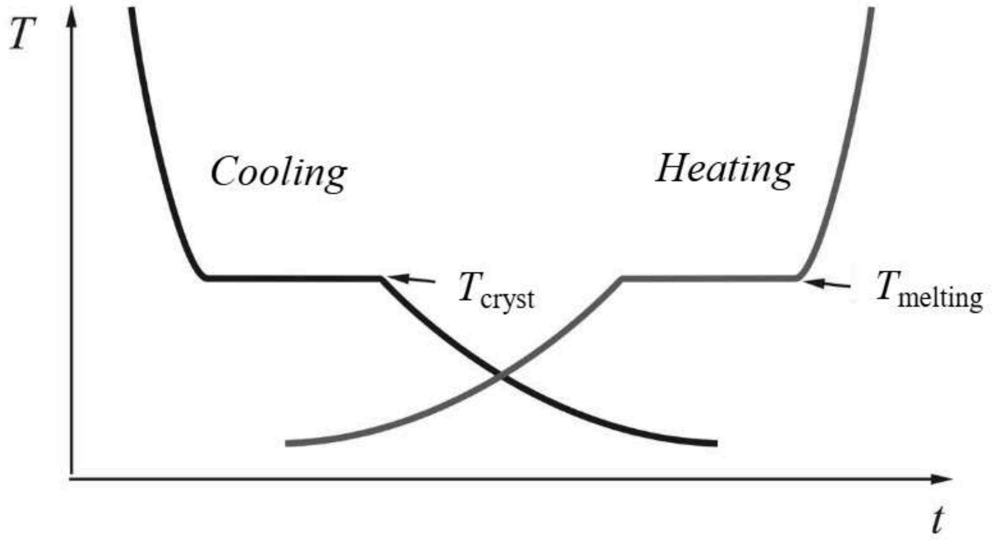


Fig. 6.17. Schematic representation of substance transition from the liquid state to the solid state and back in the form of dependence of temperature change on time as it cools or heats up

Naturally, water temperature will decrease. As soon as the ice chip(s) appear, the water temperature will be about 0 °C. With further cooling, the size of the ice block(s) will increase. In this case both the temperature of the formed piece of ice and the temperature of water will be the same. This is the so-called two-phase state in which the "liquid-solid" transition takes place at a constant temperature while cooling continues. After all the water turns into ice, the temperature of the ice will start to decrease, because the crystallization process, characterized by the crystallization temperature, is complete. Now let's lower the ice into a vessel with water (right graph). The water will begin to cool, and the temperature of the ice will increase, and at a temperature called the melting point, the ice will begin to melt (fuse). Then the proportion of ice will decrease and the total proportion of water will increase. Once again we have a two-phase system, in which the melting process takes place (transition from the crystalline state to the liquid state). During this process, the temperature of water and ice remains unchanged also near 0 °C

This is the so-called phase transition of the first order. Its characteristics are:

- the transition is performed by a jump $T_{\text{cryst}} \approx T_{\text{melting}}$[*];
- during the transition both phases exist simultaneously;
- there is a latent heat of phase transition L.

Liquid-gas transitions (boiling, evaporation, condensation, gas liquefaction process) also refer to phase transitions of the first order.

Phase diagram. Triple point. Quantitatively, the phase transition is described by the Clapeyron-Clausius equation

$$\frac{dp}{dT} = \frac{L}{T(V_2 - V_1)}, \qquad (6.10)$$

where L is the latent heat of the phase transition of the first order, $(V_2 - V_1)$ is the difference in the specific volumes of the two phases in dynamic equilibrium.

Fig. 6.18 shows schematically the so-called phase diagram – a curve of dependence $p(T)$, separating the areas of existence of two phases. It follows from relation (6.10) that the temperature (point) of the phase transition (marked as T^* on the T axis) depends on the value of pressure at which it occurs. This can be observed in everyday life and practice. Atmospheric pressure is lower in the highlands than in the plains, so water boils at $T <$ +100° C. Clinics sterilize instruments at temperatures above +100° C in autoclaves, sturdy containers with lockable lids that can withstand fairly high pressures.

---

[*] The melting and crystallization temperatures are usually somewhat different, due to a number of factors.

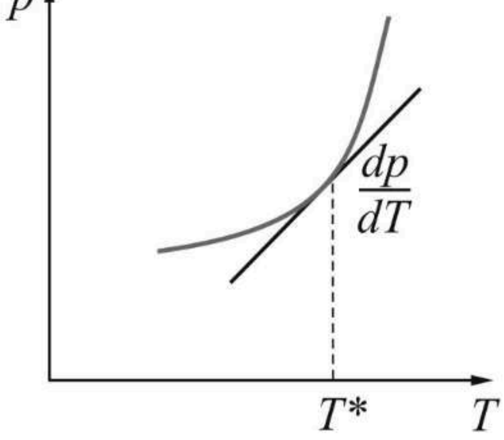


Fig. 6.18. Schematic representation of the phase diagram

At certain values of pressure and temperature, three phases of the same substance (for example, water in the form of ice, water itself and its vapor) can be in dynamic equilibrium simultaneously. This is the so-called triple point with the parameters p_{tr} and T_{tr} (Fig. 6.19).

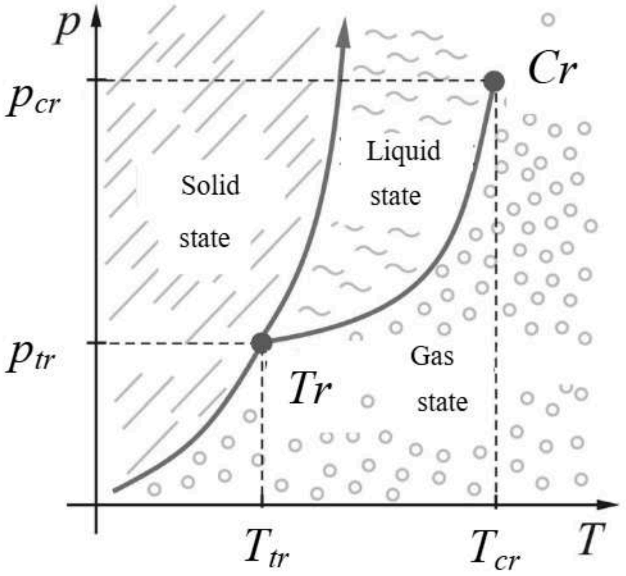


Fig. 6.19. Illustration of phase diagrams for substances that can be simultaneously in three phase states

The phase diagram for the transition between liquid and solid states breaks off at the critical point (Cr). The phase diagram for the solid-liquid transition at high values of p and T is not limited. It follows from the above phase diagrams that a direct transition from the solid phase to the gaseous phase (and vice versa), without passing the liquid phase, is possible. This phenomenon is called dry sublimation. It is well known that meat, fish and some food products dehydrate when stored for a long time in the open air in freezing conditions. This phenomenon is used in food dehydration technology for long-term storage. It is well known that frosting in the cold quickly disappears. Until recently, wet laundry and other items were dried in the cold. Dry ice – carbon dioxide CO_2 (p_{tr} = 5,11 atm., T_{tr} = 216,5 K (–56,6 °C) – is often used to store substances. Sublimation temperature at normal pressure is – 78,5 °C.

Thermal expansion is determined by the ratio:

$$\Delta l = l_0 \alpha \Delta T, \tag{6.11a}$$

where α is the coefficient of thermal expansion, equal to

$$\alpha = \frac{1}{l_0}\frac{\Delta l}{\Delta T}. \tag{6.11b}$$

Here Δl is the absolute, $\frac{\Delta l}{l_0}$ is the is the relative increase in one of the dimensions (geometric parameter), such as length l, (l_0 is its original value). In the general case, thermal expansion along arbitrary directions is different (anisotropic) and is characterized by the values $\alpha_1, \alpha_2, \alpha_3$, which correspond to the directions of coordinate axes x, y, z. Only for crystals of cubic symmetry $\alpha_1 = \alpha_2 = \alpha_3$.

Heat capacity. According to the law (rule) of Dulong-Petit, as already noted, the molar heat capacities C_p and C_V of any crystalline body at room temperature and above are the same and equal to $3R$ (except diamond, for which a heat capacity equal to $3R$ sets at 250 K). In the general case, the temperature dependence has the form shown in Fig. 6.20.

The temperature at which this dependence reaches a plateau is called the Debye temperature Θ.

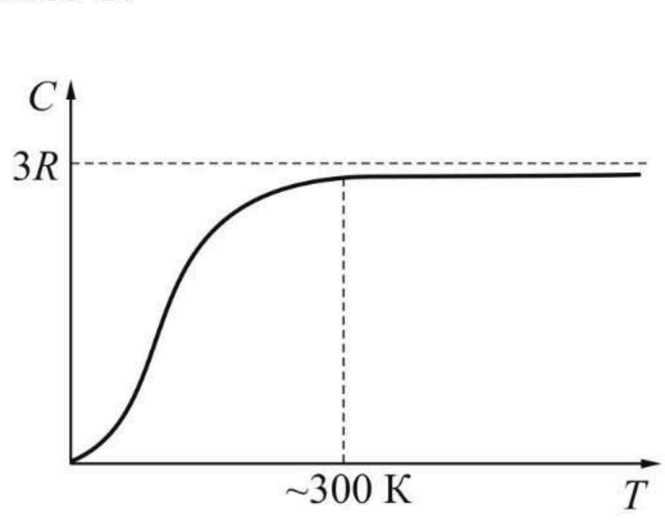


Fig. 6.20. General view of the heat capacity of solids on temperature

This experimental temperature dependence of the heat capacity of a solid cannot be interpreted in the framework of classical physics, according to which the crystal lattice node (a chemical element located in the crystal lattice node) is a harmonic oscillator. The energy of the harmonic oscillator is $3kT$, or in terms of mole of matter $3RT$, which according to formulas (5.11a) and (5.12) gives the value of heat capacity $3R$. Moreover, it is inconsistent and does not even describe the law (rule) of Dulong-Petit, which is fulfilled for all crystalline bodies, both conductors and dielectrics. Quantum theory eliminates these contradictions and fully describes the experimentally observed temperature dependence of heat capacity. Consideration of this issue is beyond the scope of this course. Let us note only that, according to this theory (Einstein-Debye model), the vibrational energy of a lattice node is a set of discrete levels corresponding to certain frequencies of its vibrations $E_{ni} = nh\nu_i$. The total energy of the lattice is determined by a double summation: by the index n from 1 to infinity and by the index i from 1 to m. That is, it is assumed that the frequency spectrum is limited to some frequency v_m characteristic of each crystal substance. Energy level discreteness does not appear at room temperatures and above ($hv < kT$). But it becomes significant below a certain boundary temperature, called the Debye temperature Θ, determined from the condition $h\nu_m = k\Theta$, at which the curve of the temperature dependence of the heat capacity as the temperature rises reaches a "plateau" to which the value of the molar heat capacity $3R$ corresponds.

The nature of thermal expansion and heat capacity are the same. This feature of crystals is reflected in the Grüneisen law, according to which the ratio of the coefficient of thermal expansion α to the molar heat capacity C is independent of temperature:

$$\frac{\alpha}{C} = const. \tag{6.12}$$

6.5. Transport phenomena

Any system, being taken out of equilibrium, tends to return to an equilibrium state. This process is called relaxation. As has been repeatedly noted, a system is characterized by a set of parameters, such as the number of particles in a certain volume, pressure and temperature. Hence, relaxation means that the value of one or another parameter of the system with time comes to an equilibrium value. The process of equalization of concentrations is called diffusion, equalization of temperature – thermal conductivity, equalization of flow velocities in different layers of liquid or gas – viscous (internal) friction. These processes are called transport processes because they are caused either by the transfer of particles (mass) – diffusion, or the transport of energy – thermal conduction, or the transfer of momentum – internal friction. The field of knowledge where the basic regularities of transport phenomena are established is called physical kinetics. There are different approaches. In the model case, all phenomena can be described within the framework of the Kinetic Molecular Theory of Matter, using as the main parameter the mean free path of molecules $\langle\lambda\rangle$. This approach is used to describe transport processes in both gases, liquids and solids. A detailed consideration of these questions can be found in the textbooks of the course of general physics.

We will limit ourselves only to defining the main characteristics and relations on the basis of the transport phenomena in gases.

Mass transfer (*M*) – *diffusion* (equalization of concentration of particles or density) in gas, liquid, solid (Fig. 6.21). The density j of the flux of the component q of the substance transported per unit time t through unit area S is determined by the relation (Fick's law)

$$j = -D\frac{dq}{dx}, \tag{6.13}$$

where $\frac{dq}{dx}$ is the gradient of the component q in the transfer direction x. For diffusion, Fick's law is written in the form:

$$j = -D\frac{d\rho}{dx} = -\mathrm{D}\frac{mdn}{dx}, \tag{6.13a}$$

where ρ is the gas density, n is the concentration of molecules of mass m, D is the diffusion coefficient, $\frac{d\rho}{dx}$ is the density gradient in the particle (mass) transport direction.

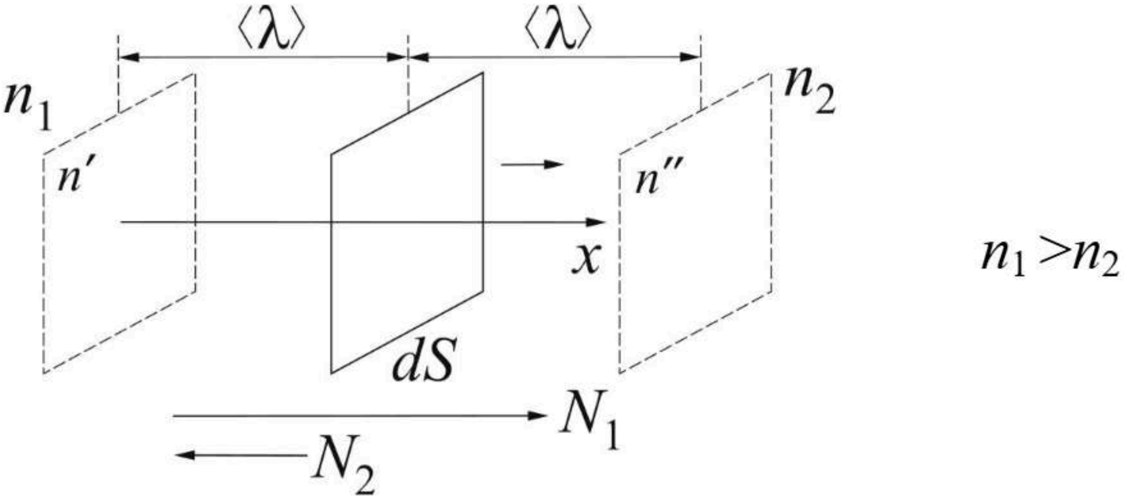


Fig. 6.21. Schematic representation of diffusion to calculate steady-state diffusion coefficient

The transfer of the number of particles *dN* is described by the equation

$$dN = -D \frac{dn}{dx} dSdt, \tag{6.13b}$$

and the transfer of an element of mass *dM* – by the equation

$$dM = -D \frac{d\rho}{dx} dSdt. \tag{6.13c}$$

The dimensionality of the diffusion coefficient is defined by the relation $[D] = [L]^2 \cdot [t]^{-1}$. Measurement units are $m^2 \cdot s^{-1}$ (SI); $cm^2 \cdot s^{-1}$ (CGS).

Let us estimate the value D in the framework of Kinetic Theory of Gases for stationary diffusion (the most important type of diffusion for living organisms). This means that the difference in concentration of molecules in arbitrarily chosen regions on both sides of the elementary site dS does not change with time, i.e. $\Delta n = n_1 - n_2$ (let's assume for certain that $n_1 > n_2$ in accordance with Figure 6.21). Hence, the stationary diffusion flux of particles along direction x, (N_1) and in opposite direction, (N_2) is determined as $j = N_1 - N_2 = {}^1/_6\,(n' - n'')$.

The concentration gradient is defined as:

$$\frac{n' - n''}{2\langle\lambda\rangle} = \frac{dn}{dx}, \qquad \text{hence:} \qquad n' - n'' = -2\langle\lambda\rangle \frac{dn}{dx}.$$

According to Fick's law

$$j = -\frac{1}{3}\langle\lambda\rangle\langle\upsilon\rangle \frac{dn}{dx}. \tag{6.14}$$

Thus, for the steady-state diffusion coefficient we obtain the value

$$D = \frac{1}{3}\langle\lambda\rangle\langle\upsilon\rangle, \tag{6.14a}$$

which is confirmed by experimental data.

Transfer of energy (heat Q) – heat conduction (Fig. 6.22). The energy flux density q is defined as

$$q = -\chi \frac{dT}{dx}, \tag{6.15}$$

where χ is the heat transfer coefficient, $\frac{dT}{dx}$ is the temperature gradient along the heat transfer direction.

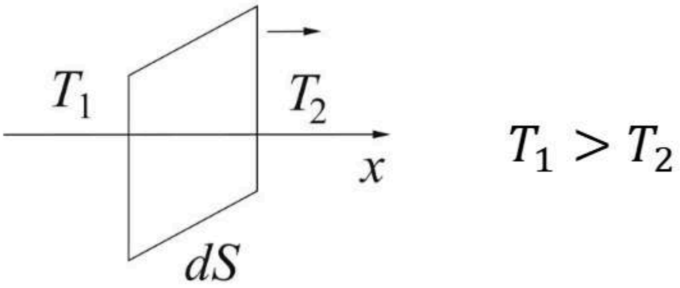


Fig. 6.22. Schematic representation of thermal conductivity

The elementary amount of heat transferred dQ is determined by the equation

$$dQ = -\chi \frac{dT}{dx} dS dt. \tag{6.15a}$$

The steady-state heat-transfer coefficient obtained within the framework of Kinetic Molecular Theory of Matter is determined by the relation

$$\chi = \frac{1}{3} \rho \langle \upsilon \rangle \langle \lambda \rangle C_V. \tag{6.15b}$$

Internal friction is caused by the transfer of particle momentum from one layer of a moving gas or liquid stream to the neighboring one (Fig. 6.23).

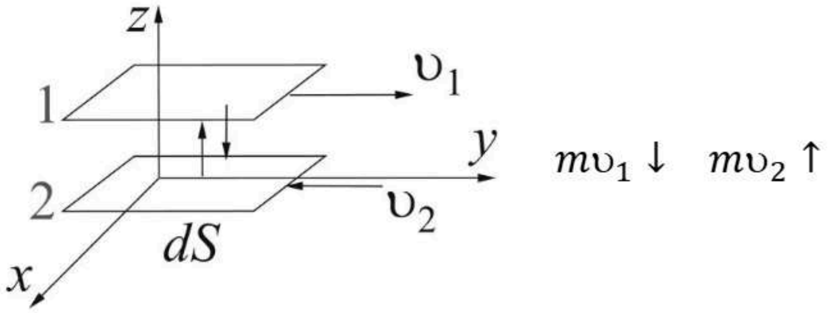


Fig. 6.23. Schematic representation of momentum transfer in a gas (liquid) flow

The mechanism of interaction between the layers can be represented as follows. If, for example, layer 1 moves faster than layer 2 relative to the reference frame (Fig. 6.23), molecules entering layer 2 from layer 1, having a smaller momentum $m\upsilon_2$, "slow down" the upper layer. Molecules coming from layer 1 into layer 2, having a larger value of momentum $m\upsilon_1$, accelerate this layer ("drag it along"). Quantitative consideration within Kinetic Molecular Theory of Matter leads to an expression for internal friction forces that agrees with Newton's formula experimentally obtained for liquids (see section 2, equation 2.8). The viscosity coefficient η for gases is defined by the relation

$$\eta = \frac{1}{3}\rho\langle\upsilon\rangle\langle\lambda\rangle. \qquad (6.16)$$

Transport phenomena play an exceptional role in nature. They provide stationary regimes in a variety of processes that ensure the stability of the functioning of dynamic systems, including living organisms. These issues will be discussed at the end of the book.

6.6. Solutions. Osmosis

Osmosis (from Greek ὄσμος – push, pressure) is the spontaneous transition of a substance through a semipermeable partition (membrane) separating two solutions of different concentrations or solution and solvent (Fig. 6.24). In this case there is an alignment of concentrations on both sides of the partition, i.e., an equilibrium in the system is established. The mechanisms of this phenomenon are complex. There are different points of view. In the literature the diffusion mechanism of the process is more often mentioned and is linked to the size of the substance molecules and pores in the membrane. There is another point of view, according to which the process proceeds by preliminary dissolution of solvent molecules in the membrane and subsequent recoil (transition) of solvent molecules into the solution. The most important case is the transition of solvent molecules through a semipermeable membrane that does not allow the solute molecules to pass through. This will cause the concentration of the substance in the solution to decrease. That is, if there are solutions with different concentrations on different sides of the partition, there will be a flow of the substance toward the solution with the lower concentration. An increase in the number of solvent molecules will result in an increase in the liquid column on one side and therefore an increase in hydrostatic pressure. As a result of concentration equalization, the process will stop. The resulting pressure difference between the two sides of the partition is called osmotic pressure.

For highly dilute solutions, the osmotic pressure is determined by the Vant-Goff law:

$$\Pi = {}^{NkT}\!/_{V} = {}^{\mu RT}\!/_{V}, \qquad (6.17)$$

where N and μ are the numbers of molecules and moles of the substance, respectively, i.e. the pressure does not depend on the nature of the dissolved substance, but only on its concentration.

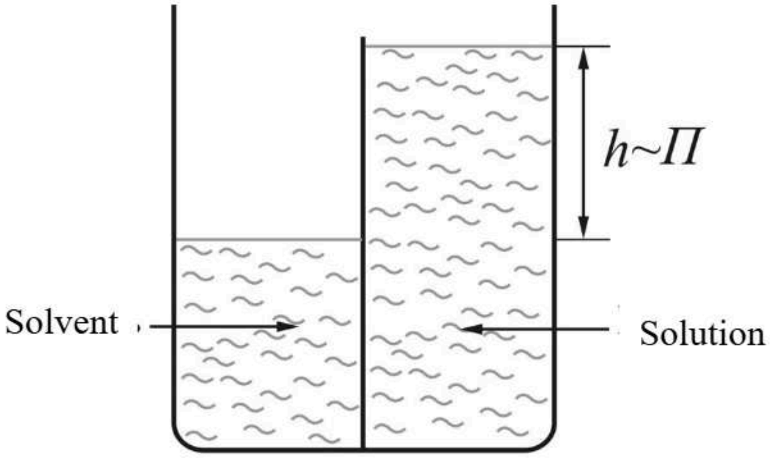


Fig. 6.24. Schematic representation of osmosis

The membrane can be permeable not only to the solvent, but also to some dissolved substances. This makes dialysis possible. Dialysis is the purification of colloidal solutions and substances of high molecular weight substances from low molecular weight compounds dissolved in them using a semi-permeable membrane. In dialysis, molecules of dissolved low molecular weight substances pass through the membrane, while colloidal particles remain behind it because they cannot pass through the membrane. The simplest dialyzer is a collodion (semi-permeable material) bag containing the liquid to be dialyzed. The bag is immersed in a solvent (e.g., water). Gradually the concentration of the dialysis agent in the dialyzable liquid and in the solvent becomes the same. By changing the solvent, it is possible to achieve almost complete purification of unwanted impurities. The dialysis rate is usually extremely low (measured in weeks). The dialysis process is accelerated by increasing the membrane area and temperature, continuously changing the solvent. In medical terms it is *hemodialysis* – the process of purifying blood using a semi-permeable porous membrane in an artificial kidney machine.

Dialysis is used to purify colloidal solutions of electrolytes and low molecular weight non-electrolytes. In the industry dialysis is used for the purification of various substances, for example, in the production of artificial fibers, in the manufacture of medicines.

Note that the membrane surrounding a normal blood cell is permeable only to water molecules, oxygen, some ions (potassium, calcium, chlorine), some nutrients dissolved in blood and products of cellular activity. For large protein molecules dissolved in the cell it is impermeable. Therefore, proteins, so important in biological processes, stay inside the cell.

Osmosis plays a huge role in the life support of organisms of different levels and plants. Consideration of these issues can be found in the specialized literature. Let us give only some examples. The use of blood substitutes and physiological solutions in clinics is well known. As a rule, they should be *isotonic* in relation to body fluids (isotonic, isoosmotic are solutions with the same values of osmotic pressure).

Osmosis is involved in nutrient transport in tree trunks, where capillary transport is unable to perform this function. Osmosis underlies cellular plasmolysis. *Plasmolysis* is artificially induced lag (separation) of cytoplasm from the cell membrane (*deplasmolysis* – disappearance of plasmolysis). Non-poisonous substances weakly penetrating through cytoplasm into the vacuole are used as plasmolytics – substances whose solutions cause plasmolysis. Plasmolysis can be induced by immersing the cell in a salt or sugar solution whose concentration is higher than the concentration of cell sap (hypertonic solution). If the cytoplasmic membrane were permeable, there would be an equalization of cellular fluid and hypertonic solution concentrations by moving water and dissolved substances from the cell to the solution and back. However, it has the property of semipermeability and does not allow substances dissolved in water to pass inside. On the contrary, only water will be sucked out of the cell by the hypertonic solution, i.e. move through the semipermeable cytoplasmic membrane. The volume of the vacuole will decrease. The cytoplasmic membrane due to elasticity follows the contracting vacuole and lags behind the cell membrane first in the corners, then in many places with the formation of concave surfaces (concave plasmolysis), and finally, the protoplast takes a rounded shape (convex plasmolysis). When a plasmolysed cell is immersed in water or hypotonic solution, deplasmolysis is observed.

Since ancient times, humans have used the effect of osmosis in the process of salting food, during which plasmolysis of pathogen cells[*] occurred.

To increase the volume of the vacuole (so that it expands the cell walls – turgor pressure), plant cells store sucrose. By increasing or decreasing sucrose concentration in the cytoplasm, cells can control osmosis. This increases the elasticity of the plant as a whole. Many plant movements (e.g. movement of whiskers of peas and other "climbing plants") are connected with changes in turgor pressure. Freshwater protozoa also have vacuoles, but the task of protozoan vacuoles is only to pump out excess water from the cytoplasm to maintain a constant concentration of dissolved substances.

Osmosis is widely used in laboratory techniques: for determination of molar characteristics of polymers, concentration of solutions and the study of various biological structures. The physics of the phenomenon is described fully enough in [5, 6].

---

[*] Pathogen is a term used to describe an infectious agent (pathogenic bacteria, viruses, fungi, protozoa, helminthes) that has the ability to cause an infectious process in animals, plants and microorganisms. It is synonymous with the more widely accepted term "causative agent". Recently, the term pathogen has been extended to include non-infectious etiological factors. Pathogenesis is a complex of interconnected staged damaging, protective-adaptive and reparative (restorative) reactions, expressed in more or less deep damage and restoration of structures and functions of organs and systems, as well as changes in the behavior of the organism as a whole.

6.7. Pharmacokinetic issues

The medication taken passes from the place of insertion (e.g., gastrointestinal tract, muscle) into the blood, which carries it through the body and delivers it to the various tissues of the organs and systems. This process is referred to by the term *absorption*[*]. The rate and completeness of absorption characterize the bioavailability of the drug, determine the time of action and its strength. Naturally, when injected intravenously and intra-arterially, the drug substance enters the bloodstream immediately and completely, and its bioavailability is 100%.

During absorption, the drug must pass through the cell membranes of the skin, mucous membranes, capillary walls, cellular and subcellular structures. All mechanisms of absorption can be divided into four main types (by physical nature): *diffusion*, *filtration* (passage of molecules through pores under pressure), *active transport* (transfer with energy expenditure), *osmosis* and *pinocytosis*[**]. The same membrane transport mechanisms are involved in the distribution of drugs in the body and their

---

[*] Means the absorption of a substance by the entire volume of the absorbing substance; there is the concept of *adsorption* – surface absorption.

[**] Pinocytosis (from the Greek πίνω – drinking, absorbing and κύτος – receptacle, here – cell) – 1) The absorption by the cell surface of liquid containing substances. 2) The process of absorption and intracellular destruction of macromolecules.
One of the main mechanisms of penetration of high molecular weight compounds into the cell, in particular proteins and carbohydrate-protein complexes.

The phenomenon of pinocytosis was discovered by American scientist W. Lewis in 1931.

During pinocytosis, short, thin outgrowths appear on the cell's plasma membrane surrounding a liquid droplet. This section of the plasma membrane bulges out and then sloughs away inside the cell in the form of a bubble. The formation of pinocytosis vesicles up to 2 μm in diameter was traced using phase-contrast microscopy and microfilming. The electron microscope distinguished vesicles 0,07 to 0,1 μm in diameter (*micropinocytosis*). Pinocytosis vesicles can move inside the cell, fuse with each other and with intracellular membrane structures. Pinocytosis is most active in amoebae, intestinal epithelial cells, renal tubules, vascular endothelium, and growing oocytes. Pinocytosis activity depends on the physiological state of the cell and the co-state of the environment. Active inducers of pinocytosis are γ-globulin, gelatin, and some salts.

excretion. Some drugs taken orally are absorbed by simple diffusion in the stomach, but most of them are absorbed in the small intestine, which has a large surface area (about 200 m^2, if you "spread" all the epithelial villi[*]) and an intensive blood supply. The stomach is the first stop on the way of drugs taken by mouth. Here drugs can be destroyed by interaction with food or digestive juices, in particular with hydrochloric acid[**]. To avoid this, they are placed in special acid-resistant shells that dissolve only in the alkaline environment of the small intestine. Gastric acid degradation is not the only reason why chewing or crushing a tablet is undesirable. There are drugs from which the active ingredient is not released immediately, but gradually (slowly, over a long period of time, or calculated in time – "batchwise"). If the integrity of the delivery system (capsule or tablet shell) of these drugs is violated, the active substance is released immediately. In this case, the concentration of the drug in the patient's body may be much higher than the therapeutic dose or even toxic.

The stomach absorbs drugs with acidic properties: salicylic acid, acetylsalicylic acid, sleeping pills of the barbituric acid derivatives group (barbiturates), which have sedative, sleeping or anticonvulsant effects, and others.

Drugs are also absorbed by diffusion from the rectum during rectal injection.

Filtration through membrane pores is much less common, since the diameter of these pores is small and only small molecules can pass through them. Capillary walls are the most permeable for drugs, and least of all – the skin, the upper layer of which consists mainly of keratinized cells.

---

[*] Epithelium (Latin epithelium), from the Greek ἐπι- – over- and θηλή – mammary gland nipple), or epithelial tissue is a layer of cells lining the surface (epidermis) and body cavities, as well as the mucous membranes of internal organs, food tract, respiratory system and urogenital tract. In addition, it forms most of the body's glands. Epithelium cells lie on thin basal membrane, they lack blood vessels, and their nutrition is carried out at the expense of underlying connective tissue

[**] There are so-called prodrugs – chemically modified forms of the drug (ester, salt, ester salt, etc.), which are transformed into the drug itself as a result of metabolic processes in the biological environment.

But the intensity of absorption through the skin can be increased. It is known that nourishing creams and masks are applied to specially prepared skin (removing excess keratinized cells, cleansing pores, improving blood circulation is achieved, for example, with the help of a water bath), and enhance the analgesic effect of muscle inflammation (myositis) is achieved by local massage, rubbing ointments and solutions into the sore spot.

Absorption of the drug when used sublingually (under the tongue) is faster and more intense than from the gastrointestinal tract. Doctors advise that in situations where quick help is needed (for example, to relieve pain during renal or hepatic colic), crush a tablet of No-Spa (drotaverine) and keep it in the mouth, without swallowing, along with a sip of hot water. Hot water causes dilation of blood vessels in the mouth, and the antispasmodic effect of the drug in this case comes very quickly, almost as after intramuscular injection[*].

Of course, it would be very convenient if all drugs could be taken internally. However, so far this has not been achieved. Some substances (e.g., insulin) are completely destroyed by enzymes in the gastrointestinal tract, while others (benzylpenicillins) are destroyed by the presence of hydrochloric acid in the stomach. These medications are used by injection. The same method is used if emergency treatment is needed.

If the drug is to have an effect only at the injection site, it is administered externally in the form of ointments, lotions, gargles, etc. Some drugs taken in small doses (e.g. nitroglycerin) can also be absorbed through the skin if they are used in special dosage forms such as transdermal ("percutaneous") therapy systems.

For gaseous and volatile medications, the main method of application is inhalation. When this occurs, absorption occurs in the lungs, which have a large surface area and an abundant blood supply. The same way aerosols are absorbed [5, 6].

---

[*] Attention: In case of abdominal pain, taking any analgesics or antispasmodics is strictly forbidden until the diagnosis is made!

In lieu of conclusion

The solution to any physical problem is based on the modeling principle. At the first stage, a model is chosen, with the help of which it is possible to give a quantitative description of the system or process. At the second stage, it is necessary to select methods of solution within the framework of this model that are adequate to the formulated problem. When considering mechanical phenomena, this is usually done relatively simply. For example, the falling body without a parachute and with a parachute. In the first case it is a free fall in the model representation, the law of conservation of mechanical energy is fulfilled. The force of friction in the motion of the body is neglected. But friction, in fact, there is. As a result, for example, a falling spacecraft burns up in the dense layers of the atmosphere in full compliance with the law of conservation of total energy. Falling with a parachute. The model is obvious: the motion of the body under the action of gravity and the force of air resistance, which in the first approximation can be assumed constant. And it is possible to calculate its value, at which the body in the final phase of the flight will fall at a constant speed, at which the destruction of the body (in contact with the ground) will not occur. Or calculating the motion of spacecraft, in which it is necessary to include the gravitational fields created by other celestial bodies as they move through space. This means that the limits of an isolated system and the properties of the reference frame in which we apply Newton's laws will change. In thermodynamics, things are much more complicated. Therefore, let's consider the classification of systems (models) on the example of solving specific problems.

Isolated systems: do not exchange either energy or substance with the external environment. The study of such systems led to the creation of classical thermodynamics. Its main provisions were considered in Section 5.

Adiabatically isolated: there is no substance exchange with the external environment, but energy exchange through work is possible. Such a system is used to obtain low temperatures on the basis of the Joule-Thomson effect. The principle of operation of the system is as follows. A porous baffle is placed in a thermally insulated tube and a gas such as nitrogen is pumped through it under high pressure (throttles). The air pressure is created by a pump (work is done). Throttling through this partition, the gas exits under lower pressure, i.e. it expands. Since there is no heat exchange, all the work is used to change the internal energy of the gas, which leads to a change in kinetic energy (temperature) as a consequence of potential energy change in the interaction of molecules during gas expansion (pressure decrease). At certain temperatures this pressure drop leads to a temperature drop. This principle of operation of adiabatically isolated system is implemented in Linde machine, which produces liquid cooling agents – cryoliquids, for example, nitrogen, oxygen, helium, hydrogen. Another method of obtaining low temperatures (liquid refrigerants) is based on application of Joule-Thomson effect during adiabatic gas expansion (an example of such machine is a domestic refrigerator). Here the compressor compresses gas (freon) which heats up and the heat is dissipated to atmosphere through radiator grid, then the gas expands adiabatically in special chamber, after cooling it is compressed again and so on. For industrial applications, machines are used to produce liquid refrigerants using both methods of temperature reduction (this principle was proposed by Claude in 1902). Adiabatic demagnetization is used to obtain lower temperatures.

Closed systems have energy exchange with the external environment, but no substance exchange. An example of such a system is a sealed flask (retort) in which Lomonosov and Lavoisier conducted experiments on annealing of metal (i.e. under conditions of heat exchange) – experiments that allowed them to establish the law of conservation of matter. Closed thermodynamic systems tend to move to a thermodynamic equilibrium state, which is characterized by the cessation of all macroscopic processes in the system. Without external influences, the system can remain in this state for any length of time.

Partially open systems exchange both energy and substance with the environment, but not all components of the substance take part in the material exchange. A typical example of such a system is a vessel divided by a semi-permeable partition that causes solvent molecules to pass through the partition from the solution in one part of the vessel to the other part of the vessel – the phenomenon of osmosis (Subsection 6.6).

And finally, *open systems* in which both energy and matter are exchanged with the external environment. These are humans and the animal world. An ingenious insight during the observation of human activity in unusually hot conditions led R. Mayer to the discovery of the basic law of nature – the first principle of thermodynamics.

The field of knowledge dealing with the study of the mechanisms of energy transformation in the processes of vital activity of organisms, the thermo-dynamics of living systems, is called bioenergetics (biological energetics). This is a very complex interdisciplinary field of science at the interface of physics, chemistry and biology, both in terms of research methods (physical and chemical) and in terms of objectives and research objects (biological). Thus, it is part of both molecular biology, biophysics, and biochemistry. Naturally, they are joined by medicine through biomedicine, since the ultimate goal of all these studies is to ensure the effective functioning of the living organism. In the context of this textbook, we will consider only the general principles of the functioning of living systems.

What are the fundamental differences between classical thermodynamics (physical thermodynamics and chemical thermodynamics of systems with a variable number of components) and the thermodynamics of living systems?

The key concepts of classical thermodynamics – equilibrium states and equilibrium (quasi-equilibrium) processes – are not acceptable for describing processes under conditions of constant exchange of matter and energy with the external environment. It is necessary to use other representations: stationary states and stationary (quasi-stationary) processes.

This means that the second principle of thermodynamics, the law of increasing (non-increasing) entropy, which describes natural processes in nature (a system removed from the equilibrium position tends to the equilibrium state), should be reformulated.

The steady-state mode of functioning of a living organism means that constant flows of energy and matter in various forms must be maintained. In classical thermodynamics such a transfer of energy and substance is called a transport phenomenon (Subsections 6.5, 6.6).

The function of the living organism is to create and maintain a gradient of several parameters: the concentration of matter (concentration-transition gradient), the gradient of electrochemical potential (electrochemical gradient) and the gradient of automatism in the cardiac conduction system.

The concentration gradient responsible for diffusion in gases and liquids has already been discussed (Subsection 6.5).

The electrochemical gradient or electrochemical potential gradient is a combination of the concentration gradient and the membrane potential, which determines the direction of ion movement across the membrane. It consists of two components: the concentration gradient due to the difference in concentration of the dissolved substance on both sides of the membrane and the electric gradient due to the difference in the magnitude of the charges on the opposite sides of the membrane. Thus, the electrochemical gradient arises due to different concentrations of ions on opposite sides of the water-permeable membrane. The ions are transported through the membrane from the region of higher concentration to the region of lower concentration by diffusion. The ions also carry an electric charge which forms an electric potential on the membrane (membrane potential). If the charge distribution on both sides of the membrane is unequal, the difference in the electric potential generates a force that causes ionic diffusion until the charges on both sides are balanced out. This issue will be discussed in more detail in the next part of the course.

Gradient of automatism. Cardiac automatism is the ability of the heart cells to self-excitation without any external influences. An isolated heart, when supplied with a nutrient solution, is capable of contracting outside the body for a long time. Potentially all elements of the cardiac conduction system (Section 2.3.3) are capable of generating an automatic rhythm to varying degrees. The sinoatrial node is the pacemaker of the heart. Interacting with the extracardiac nerves, it determines the heart rate 60 to 80 per minute. If the node is damaged, the atrioventricular node (40÷50 heartbeats per min), followed by the His bundle (30÷40) and Purkinje fibers (20) act as rhythm driver. The decrease in the frequency of excitation generation by the cardiac conduction system in the direction from the atria to the apex of the heart is called an automatism *gradient*.

An illustrative example of the stationary mode is the operation of any transport system, for example, a pipeline. As applied to a human it is the introduction of a substance into the body and excretion of metabolic products out of it. Normally it is a stationary process with constant concentration gradients. Thus, the natural reaction of an organism is to restore the stationary state of an open system disturbed in any way.


I. Prigogine
(1917–2003)

The physical basis for such regulation is the Prigogine* theorem, which is a formulation of the second law of thermodynamics as applied to open systems. According to Prigogine's theorem: "The minimum entropy production corresponds to the stationary state of the system (under conditions preventing the achievement of an equilibrium state)" – the minimum increase in entropy in a physical system per unit time as a result of non-equilibrium processes occurring in it.

---

* ***I. Prigogine*** was a Belgian physicist and physicochemist of Russian origin. One of two laureates of the Nobel Prize in Chemistry (1977) of Russian origin (the first was ***Nikolai Nikolaevich Semenov***), Viscount of Belgium. His main works are devoted to nonequilibrium thermodynamics and statistical mechanics of irreversible processes.

The second principal difference is that processes in cells occur in the absence of changes in temperature, pressure and volume. For this reason, the direct transfer of heat to work in the body is impossible and heat production is a non-returnable loss of energy. Therefore, in the course of evolution, organisms have developed a number of specific mechanisms for direct conversion of one form of energy into another, bypassing its transition into heat. In the organism only a small part of the released energy is transformed into heat and is lost. Most of it is transformed into the form of free chemical energy of certain compounds, in which it is extremely mobile, i.e. it can transform into other forms at constant temperature, in particular, to perform work.

Now the adaptive reaction of an organism to various external influences, as a rule, leading to deterioration of the general state of an organism at the initial stage and not always passing without leaving a trace, becomes clear. Let's add to them the organism reaction to sharp changes of environment temperature (it is considered that temperature difference by 10 degrees or more can cause immune system disturbance), crossing of time zones, sharp infringement of diet, overdose of medicines and many other things.

The main and practically the only source of energy for life on Earth is the energy of the Sun's radiation[*], part of which is absorbed by the pigments of plants and some bacteria and in the process of photosynthesis[**] is accumulated by autotrophic organisms[***] in the form of chemical energy: partly in the form of ATP – adenosine triphosphate (due to photosynthetic

---

[*] The second source of energy is the world's Ocean, which is heated by the heat generated by the processes taking place inside the Earth..

[**] Photosynthesis (from the Greek φῶς – light and σύνθεσις – combination, folding, bonding, synthesis) is the process of conversion of light energy into chemical bonds of organic substances in the light by photoautotrophs with the participation of photosynthetic pigments (chlorophyll in plants, bacteriochlorophyll and bacteriorhodopsin in bacteria). In modern plant physiology, photosynthesis is more often understood as a photoautotrophic function – a set of processes of absorption, conversion and use of energy of light quanta in various endergonic reactions (where work is performed over the system), including the conversion of carbon dioxide into organic substances.

[***] Autotrophs (from Greek αὐτός – self + τροφή – food) are organisms that synthesize organic matter from inorganic.

phosphorylation processes), partly in the form of energy of some specific compounds (reduced nicotinamide-adenindinucleotides), which are the most important intermediate energy accumulators. The whole further process of synthesis of carbohydrates, then of lipids, proteins and other cell components is performed in the cycle of dark enzymatic reactions (complex of enzymatic reactions, during which the reduction of absorbed carbon dioxide occurs due to the products formed during the light phase) due to the energy of the above mentioned compounds.

Detailed consideration of these issues is beyond the scope of this textbook. They are discussed in the course "Bioenergetics".

References

1. Harms V. Physic fur Mediziner und Pharmazeuten. Mit 131 Testfragen und 289 Abbildungen / V. Harms. – Harms Verlag,In't Holt 37, 24214, Lindhoft, 2016– 256 p.
2. Davidovits P. Physics in Biology and Medicine 4th edition / P. Davidovits –San Diego, United States: Elsevier Science Publishing Co Inc, 2013. – 352 p.
3. Hobbie R.K. Intermediate Physics for Medicine and Biology / R.K. Hobbie, B.J. Roth – Cham, Switzerland: Springer International Publishing AG, 2015. – 629 p.
4. Knudson D. Fundamentals of Biomechanics / D. Knudson – New York, United States: Springer-Verlag New York Inc., 2007. – 354 p.
5. Brown B.H. Medical Physics and Biomedical Engineering / B.H Brown, R.H Smallwood, D.C. Barber, P.V Lawford, D.R Hose – London, United Kingdom: Taylor & Francis Ltd., 1999. – 768 p.
6. Herman I.P. Physics of the Human Body / I.P. Herman. – Cham, Switzerland: Springer International Publishing AG, 2016. – 953 p.
7. Glaser R. Biophysics: An Introduction / R. Glaser. – Berlin, Germany: Springer-Verlag Berlin and Heidelberg GmbH & Co., 2012. – 407 p.

Подписано в печать….

Подписано в печать 04.10.2022.
Формат 60x84 $^{1/16}$. Бумага офсетная.
Гарнитура “Таймс”. Печать цифровая.
Усл. печ. л. 17,5. Тираж 100 экз. Заказ № 107.

Отпечатано с готового оригинал-макета в ИП Кашапов Л.И..
г. Казань, ул. Ю. Фучика, 50. Тел. +7 917 919 95 99.

Table of corrections to the book: A.V. Aganov, K.S. Usachev
Medical Physics. Part 1. Mechanics, Molecular Physics. Edition 3rd

Page	Paragraph number (above or below)	Incorrect variant	Correct variant
62	1 above	The so-called **rules** of dimensionality …	The so-called rule and formulas of dimensionality…
77	1 above	…elastin (a fibrous protein just like collagen ~~**and the main tissue, the matrix**~~).	… elastin (a fibrous protein just like collagen)
78	2 above	Sections along (shaded) and across the vessel axis ~~**(highlighted in gray)**~~ of the cylinder are shown in Fig. 1.34.	Sections along (shaded) and across the vessel axis of the cylinder are shown in Fig. 1.34.
85	1 above	The work A_{12} to move a body of mass m … $A_{s-f} = \int_s^f m\mathbf{g}d\mathbf{r} = m\int_s^f g\cos\alpha\, dr = -mg\int_s^f dh = -mg(h_2 - h_1)$. (1.58a)	The work $A_{s\text{-}f}$ to move a body of mass m … $A_{s-f} = \int_s^f m\mathbf{g}d\mathbf{r} = m\int_s^f g\cos\alpha\, dr = -mg\int_s^f dh = mg(h_2 - h_1)$. (1.58a)
89	1 above	~~$A_{12} = E_{k_2} - E_{k_1}$, $A_{21} = E_{p_2} - E_{p_1}$.~~ Since $A_{12} = -A_{21}$,	$A_{12} = -(E_{P_2} - E_{P_1})$. So we can write
127	2 below	3 seconds	0.3 seconds
159	2 above	…ultrasound is reflected from the air layer, Z_2).	…ultrasound is reflected from the air layer, $Z_2 << Z_1$).
205	1 below	There is an average **dependence** between atmospheric pressure and blood pressure ~~**value:**~~ atmospheric pressure ~~**decrease by 5 mm Hg**~~ causes blood pressure increase ~~**by 13 mm Hg**~~. [5, 6].	There is an average relationship between atmospheric pressure and blood pressure, namely: a decrease in atmospheric pressure causes an increase in blood pressure. [5, 6].
239	1 above	But in real gas $\langle\lambda\rangle = f(T) \sim \frac{1}{T}$, i.e. it **decreases** with increasing temperature.	But in real gas $\langle\lambda\rangle = f(T) \sim T$, i.e. it increases with increasing temperature.

***Albert Aganov** is Professor at Kazan Federal University (KFU), Honored Scientist of the Republic of Tatarstan and the Russian Federation, Dean of the Faculty of Physics and Director of the Institute of Physics at KFU (till 2015), Head of the Medical Physics Department. Albert Aganov is the author (co-author) of more than 400 publications, including monographs and textbooks.*

***Konstantin Usachev** is Professor at Medical Physics Department at the Kazan Federal University (KFU), Head of the Laboratory for Structural Analysis of Biomacromolecules at Kazan Scientific Center of Russian Academy of Science. Winner of the E. K. Zavoisky Youth Prize among young scientists. Konstantin Usachev is the Winner of the All-Russian competition Golden Names of High School 2019.*