S.V. Eliseeva, Yu.N. Zubkov

Physics, Mathematics

A COURSE OF LECTURES
for foreign medical students of the first year:
educational book

Ulyanovsk 2023

Ministry of Science and Higher Education of the Russian Federation

Federal State Budget-Financed Educational Institution of Higher Education
«Ulyanovsk State University»
Faculty of High Technology Physics and Engineering
Department of Physical Methods in Applied Researches

S.V. Eliseeva, Yu.N. Zubkov

Physics, Mathematics

A COURSE OF LECTURES
for foreign medical students of the first year:
educational book

Ulyanovsk 2023

УДК 53(076.5)
ББК 22.3
Е51

Печатается по решению Ученого совета
инженерно-физического факультета высоких технологий
Ульяновского государственного университета
(протокол № 9 от 18 апреля 2023 г.)

Рецензенты:
доцент кафедры естественнонаучных дисциплин Ульяновского института гражданской авиации имени Главного маршала авиации Б.П. Бугаева,
кандидат физико-математических наук, доцент *В. В. Ефимов*;

профессор кафедры радиофизики и электроники ФГБОУ ВО Ульяновского государственного университета, доктор физико-математических наук,
профессор *О. Н. Гадомский*

Eliseeva S.V.

Е51 Physics, Mathematics: A course of lectures for foreign medical students of the first year: educational book / S.V. Eliseeva, Yu.N. Zubkov. – Ulyanovsk : Ulyanovsk State University, 2023. – 392 p.

The course of lectures is prepared in conformity with the typical syllabus for medical faculty. This issue contains all necessary data (text, formulas and diagrams) for study of medical and biological physics and mathematics for the specialties of the Faculty of Medicine 31.05.01 - General Medicine, 31.08.19. - Pediatrics and 33.05.01 - Pharmacy.

УДК 53(076.5)
ББК 22.3

Contents

Introduction

Biophysics is discipline that studies physical mechanisms and processes which are the basis of vital activity of biological objects. It is associated with the application of the physical principles, methods and instrumentation to living organisms or systems.

Course of „Physics in Biology and Medicine" includes biophysics of human and elements of medical physics. In general biophysics is an interdisciplinary branch of science which develops at the border of physics, chemistry, biology and medical sciences. It employs physical methods and techniques to study phenomena taking place in living organisms at all levels of their organization, from the micro to the macro scale, from molecules to cells and whole organisms. Medical physics means application of physical phenomena and physical methods and techniques (e.g. ultrasounds, light, laser, ionizing radiation, nuclear resonance) in medical diagnosis and therapy.

Lecture No 1

Kinematics

Mechanics is the branch of physics describing physical properties of bodies and rules of their mechanic motion. Mechanics, science concerned with the motion of bodies under the action of forces, including the special case in which a body remains at rest. Of first concern in the problem of motion are the forces that bodies exert on one another. This leads to the study of such topics as gravitation, electricity, and magnetism, according to the nature of the forces involved. Given the forces, one can seek the manner in which bodies move under the action of forces; this is the subject matter of mechanics proper.

Historically, mechanics was among the first of the exact sciences to be developed. Its internal beauty as a mathematical discipline and its early remarkable success in accounting in quantitative detail for the motions of the Moon, the Earth, and other planetary bodies had enormous influence on philosophical thought and provided impetus for the systematic development of science into the 20th century.

Mechanics is basically divided in to two parts Static's and Dynamics, Dynamics is further divided in kinematics and kinetics

Statics: It deals with study of behavior of body at rest under the action of various forces, that are in equilibrium.

Dynamics: Dynamics is study of object in motion

Kinematics: It deals with motion of body without considering the forces acting on it.

Kinetics: It deals with motion of the body considering the forces acting on it.

Kinematics, as a field of study, is often referred to as the "geometry of motion" and is occasionally seen as a branch of mathematics.A kinematics problem begins by describing the geometry of the system and declaring the initial conditions of any known values of position, velocity and/or acceleration of points within the system. Then, using arguments from geometry, the position, velocity and acceleration of any unknown parts of the system can be determined.

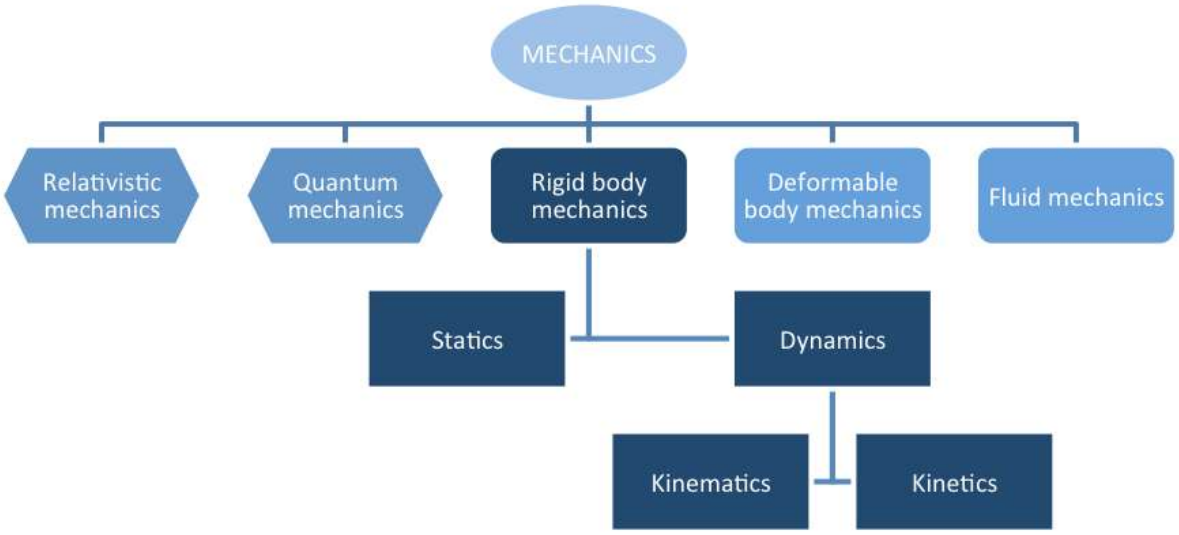

Fig. 1.1. The Branches of Mechanics.

1) Translational motion is the motion by which a body shifts from one point in space to another. One example of translational motion is the the motion of a bullet fired from a gun. An object has a rectilinear motion when it moves along a straight line or straight direction

2) Rotational motion deals only with rigid bodies. A rigid body is an object that retains its overall shape, meaning that the particles that make up the rigid body remain in the same position relative to one another. A wheel and rotor of a motor are common examples of rigid bodies that commonly appear in questions involving rotational motion. Angular motion is a common type of rotational motion.

1.1. Translational motion

Defining Kinematics:

1) *Kinematics* is the study of the motion of points, objects, and groups of objects without considering the causes of its motion.

2) *Kinematics* is the branch of classical mechanics that describes the motion of points, objects and systems of groups of objects, without reference to the causes of motion (i.e., forces). The study of kinematics is often referred to as the "geometry of motion."

Objects are in motion all around us. Everything from a tennis match to a space-probe flyby of the planet Neptune involves motion. When you are resting, your heart moves blood through your veins. Even in inanimate objects there is

continuous motion in the vibrations of atoms and molecules. Interesting questions about motion can arise: how long will it take for a space probe to travel to Mars? Where will a football land if thrown at a certain angle? An understanding of motion, however, is also key to understanding other concepts in physics. An understanding of acceleration, for example, is crucial to the study of force.

To describe motion, kinematics studies the trajectories of points, lines and other geometric objects, as well as their differential properties (such as velocity and acceleration). Kinematics is used in astrophysics to describe the motion of celestial bodies and systems; and in mechanical engineering, robotics and biomechanics to describe the motion of systems composed of joined parts (such as an engine, a robotic arm, or the skeleton of the human body).

A formal study of physics begins with kinematics. The word "kinematics" comes from a Greek word "kinesis" meaning motion, and is related to other English words such as "cinema" (movies) and "kinesiology" (the study of human motion). Kinematic analysis is the process of measuring the kinematic quantities used to describe motion. The study of kinematics can be abstracted into purely mathematical expressions, which can be used to calculate various aspects of motion such as velocity, acceleration, displacement, time, and trajectory.

Displacement is the change in position of an object relative to its reference frame. For example, if a car moves from a house to a grocery store, its displacement is the relative distance of the grocery store to the reference frame, or the house. The word "displacement" implies that an object has moved or has been displaced. Displacement is the change in position of an object and can be represented mathematically as follows:

$$\Delta x = x_f - x_0$$

where Δx is displacement, x_f is the final position, and x_0 is the initial position.

Introduction to Scalars and Vectors

A **vector** is any quantity that has both magnitude and direction, whereas a **scalar** has only magnitude. What is the difference between distance and displacement? Whereas displacement is defined by both direction and magnitude, distance is defined by magnitude alone. Displacement is an example of a vector quantity. Distance is an example of a scalar quantity. In mathematics, physics, and engineering, a vector is a geometric object that has a magnitude (or length) and direction and can be added to other vectors according to vector algebra. The

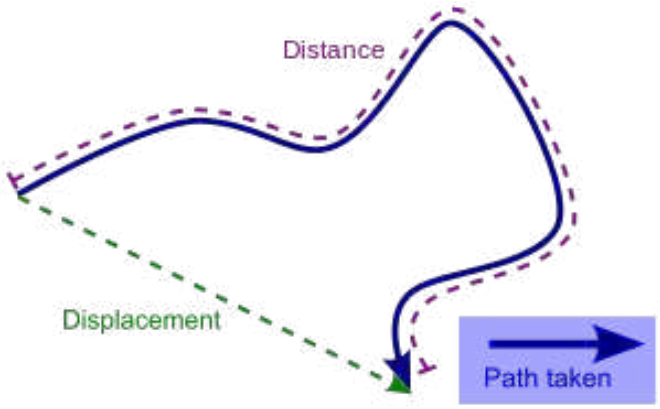

Fig. 1.2. The distance travelled is always greater than or equal to the displacement.

direction of a vector in one-dimensional motion is given simply by a plus (+) or minus (−) sign.

In everyday usage, the terms "speed" and "velocity" are used interchangeably. In physics, however, they are distinct quantities. Speed is a scalar quantity and has only magnitude. Velocity, on the other hand, is a vector quantity and so has both magnitude and direction. This distinction becomes more apparent when we calculate average speed and velocity.

Average speed is calculated as the distance traveled over the total time of travel. In contrast, average velocity is defined as the change in position (or displacement) over the total time of travel.

$$v_{avg} = \frac{d}{\Delta t},$$

where v_{avg} is average speed, d is distance, and Δt is change in time.

Average velocity is displacement (change in position) divided by the time of travel,

$$\overline{v} = \frac{\Delta x}{\Delta t} = \frac{x_f - x_0}{t_f - t_0},$$

where $\overline{v}$ is the *average* (indicated by the bar over the v) velocity, Δx is the change in position (or displacement), and x_f and x_0 are the final and beginning positions at times t_f and t_0, respectively. If the starting time t_0 is taken to be zero, then the average velocity is simply

$$\overline{v} = \frac{\Delta x}{t}$$

The SI unit for velocity is meters per second, or m/s, but many other units (such as km/h, m/h and cm/s) are commonly used.

Instantaneous velocity is the velocity of an object at a single point in time and space as calculated by the slope of the tangent line.

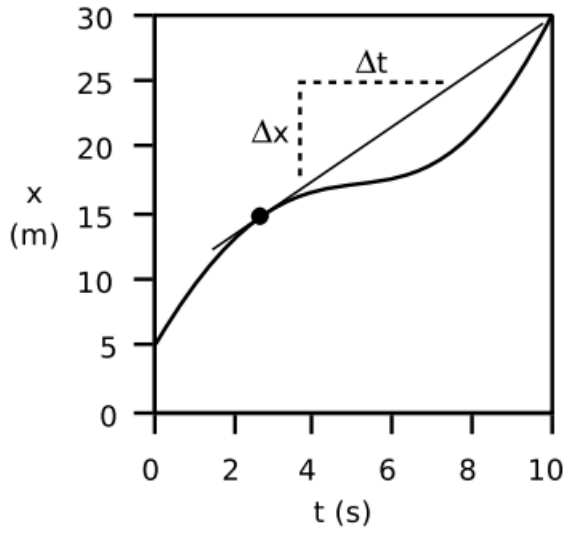

Fig. 1.3. The tangent line through the relevant point on its x versus t graph.

Typically, motion is not with constant velocity nor speed. While driving in a car, for example, we continuously speed up and slow down. A graphical representation of our motion in terms of distance vs. time, therefore, would be more variable or "curvy" rather than a straight line, indicating motion with a constant velocity as shown below. (We limit our discussion to one dimensional motion. It should be straightforward to generalize to three dimensional cases.)

To calculate the speed of an object from a graph representing constant velocity, all that is needed is to find the slope of the line; this would indicate the change in distance over the change in time. However, changing velocity it is not as straightforward.

Since our velocity is constantly changing, we can estimate velocity in different ways. One way is to look at our instantaneous velocity, represented by one point on our curvy line of motion graphed with distance vs. time. In order to determine our velocity at any given moment, we must determine the slope at that point. To do this, we find a line that represents our velocity in that moment, shown graphically in. That line would be the line tangent to the curve at that point. If we extend this line, we can easily calculate the displacement of distance over time and determine our velocity at that given point. The velocity of an object at any

given moment is the slope of the tangent line through the relevant point on its x vs. t graph.

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

Determining instantaneous velocity: The velocity at any given moment is defined as the slope of the tangent line through the relevant point on the graph. **Instantaneous speed** equals to the magnitude of the instantaneous velocity.

When a particle's velocity changes, the particle is said to undergo **acceleration** (or to accelerate). For motion along an axis, the **average acceleration** a_{avg} over a time interval Δt is

$$a_{avg} = \frac{v_2 - v_1}{t_2 - t_1} = \frac{\Delta v}{\Delta t},$$

where the particle has velocity v_1 at time t_1 and then velocity v_2 at time t_2. The **instantaneous acceleration** (or simply **acceleration**) is

$$a = \frac{dv}{dt}.$$

In words, the acceleration of a particle at any instant is the rate at which its velocity is changing at that instant. Graphically, the acceleration at any point is the slope of the curve of $v(t)$ at that point. In words, the acceleration of a particle at any instant is the second derivative of its position $x(t)$ with respect to time.

$$a = \frac{dv}{dt} = \frac{d}{dt}(\frac{dx}{dt}) = \frac{d^2x}{dt^2}.$$

A common unit of acceleration is the meter per second per second: $m/(s \cdot s)$ or m/s^2. Acceleration has both magnitude and direction (it is yet another vector quantity). Its algebraic sign represents its direction on an axis just as for displacement and velocity; that is, acceleration with a positive value is in the positive direction of an axis, and acceleration with a negative value is in the negative direction.

1.1.1. Components of the Acceleration Vector

When you divide the acceleration vector into its components relative to a plane curve, you express it in terms of the acceleration tangent to the curve and

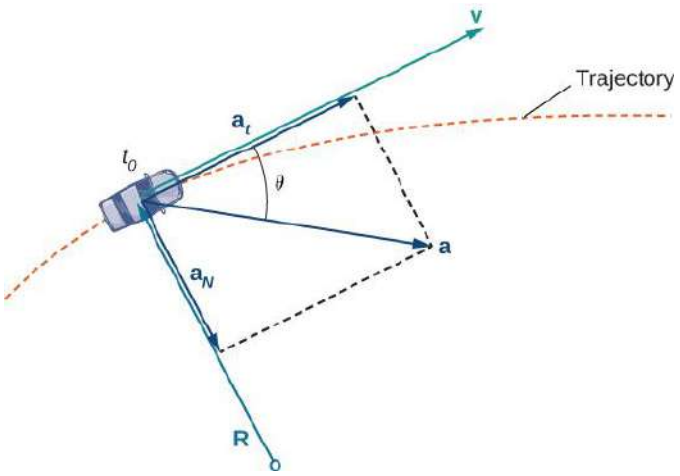

Fig. 1.4. Normal and Tangential components of acceleration.

the acceleration normal (perpendicular) to the curve.

$$|a| = \sqrt{a_N^2 + a_t^2}, \quad a_N = \frac{v^2}{R}, \quad a_t = \frac{dv}{dt}.$$

1.1.2. Free-Fall Acceleration

If you tossed an object either up or down and could somehow eliminate the effects of air on its flight, you would find that the object accelerates downward at a certain constant rate. That rate is called the free-fall acceleration, and its magnitude is represented by g. The acceleration is independent of the object's characteristics, such as mass, density, or shape; it is the same for all objects.

An important example of straightline motion with constant acceleration is that of an object rising or falling freely near Earth's surface. The constant acceleration equations describe this motion, but we make two changes in notation:

(1) we refer the motion to the vertical y axis with $+y$ vertically *up*;

(2) we replace a with g, where $-g$ is the magnitude of the free-fall acceleration. Near Earth's surface, $g = 9.8\ m/s^2$.

Large accelerations are sometimes expressed in terms of g units, with

$$1\ g = 9.8\ \ m/s^2 \quad (g\ \ unit).$$

1.1.3. Constant Acceleration: A Special Case

In many types of motion, the acceleration is either constant or approximately so. For example, you might accelerate a car at an approximately constant rate when a traffic light turns from red to green. Later when you brake the car to a

stop, the acceleration (or deceleration in common language) might also be approximately constant. Such cases are so common that a special set of equations has been derived for dealing with them. These equations are valid only for constant acceleration (or situations in which you can approximate the acceleration as being constant).

When the acceleration is constant, the average acceleration and instantaneous acceleration are equal

$$a = a_{avg} = \frac{v - v_0}{t - 0}.$$

Here v_0 is the velocity at time $t = 0$ and v is the velocity at any later time t. We can recast this equation as

$$\boxed{v = v_0 + at}, \quad \text{(This formula is missing } \Delta x\text{)}. \tag{1.1}$$

As a check, note that this equation reduces to $v = v_0$ for $t = 0$, as it must. Doing so yields $dv/dt = a$, which is the definition of a. In a similar manner, we can rewrite equation (with a few changes in notation) as

$$v_{avg} = \frac{x - x_0}{t - 0}$$

and then as

$$x = x_0 + v_{avg}t, \tag{1.2}$$

in which x_0 is the position of the particle at $t = 0$ and v_{avg} is the average velocity between $t = 0$ and a later time t.

For the linear velocity function in (1.1), the *average* velocity over any time interval (say, from $t = 0$ to a later time t) is the average of the velocity at the beginning of the interval ($= v_0$) and the velocity at the end of the interval ($= v$). For the interval from $t = 0$ to the later time t then, the average velocity is

$$v_{avg} = \frac{1}{2}(v_0 + v). \tag{1.3}$$

Substituting the right side of (1.1) for v yields, after a little rearrangement,

$$v_{avg} = v_0 + \frac{1}{2}at. \tag{1.4}$$

Finally, substituting (1.4) into (1.2) yields

$$\boxed{x - x_0 = v_0 t + \frac{1}{2}at^2}, \quad \text{(This formula is missing } v\text{)}. \tag{1.5}$$

As a check, note that putting $t = 0$ yields $x = x_0$, as it must. As a further check, taking the derivative of (1.5) yields (1.1), again as it must. Equations (1.1) and (1.5) are the basic equations for constant acceleration.

However, we can derive other equations that might prove useful in certain specific situations. First, note that as many as five quantities can possibly be involved in any problem about constant acceleration – namely, $x - x_0$, v, t, a, and v_0. Usually, one of these quantities is *not* involved in the problem, *either as a given or as an unknown.* We are then presented with three of the remaining quantities and asked to find the fourth.

Equations (1.1) and (1.5) each contain four of these quantities, but not the same four. In (1.1), the "missing ingredient" is the displacement $x-x_0$. In (1.5), it is the velocity v. These two equations can also be combined in three ways to yield three additional equations, each of which involves a different "missing variable." First, we can eliminate t to obtain

$$\boxed{v^2 = v_0^2 + 2a(x - x_0)}, \quad \text{(This formula is missing } t\text{)}. \tag{1.6}$$

This equation is useful if we do not know t and are not required to find it. Second, we can eliminate the acceleration a between (1.1) and (1.5) to produce an equation in which a does not appear:

$$\boxed{x - x_0 = \frac{1}{2}(v_0 + v)t}, \quad \text{(This formula is missing } a\text{)}. \tag{1.7}$$

Finally, we can eliminate v_0, obtaining

$$\boxed{x - x_0 = vt - \frac{1}{2}at^2}, \quad \text{(This formula is missing } v_0\text{)}. \tag{1.8}$$

Note the subtle difference between this equation and (1.5). One involves the initial velocity v_0; the other involves the velocity v at time t.

1.2. Angular motion

1.2.1 Angular Position In mathematics, the angle of rotation (or **angular position**) is a measurement of the amount (i.e., the angle) that a figure is rotated about a fixed point which we take as the **zero angular position**. (often the center of a circle). From geometry, we know that u is given by

$$\theta = \frac{s}{r} \quad \text{(radian measure)}$$

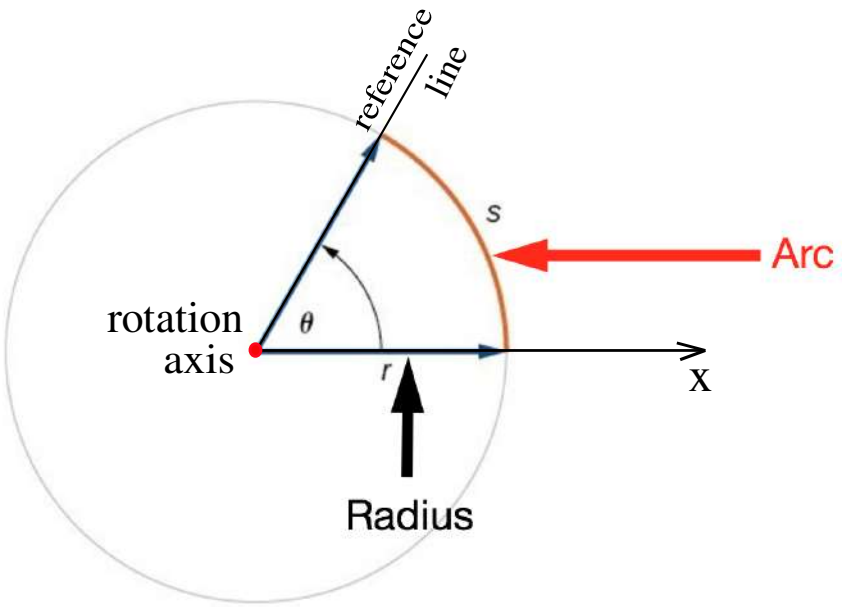

Fig. 1.5. The position of the reference line.

Here s is the length of a circular arc that extends from the x axis (the zero angular position) to the reference line, and r is the radius of the circle.

An angle defined in this way is measured in **radians** (rad) rather than in revolutions (rev) or degrees. The radian, being the ratio of two lengths, is a pure number and thus has no dimension. Because the circumference of a circle of radius r is $2\pi r$, there are 2π radians in a complete circle:

$$1\ rev = 360^o = \frac{2\pi r}{r} = 2\pi\ \ rad, \quad 1\ rad = 57.3^o = 0.159\ rev.$$

We do *not* reset θ to zero with each complete rotation of the reference line about the rotation axis. If the reference line completes two revolutions from the zero angular position, then the angular position θ of the line is $\theta = 4\pi\ rad$.

For pure translation along an x axis, we can know all there is to know about a moving body if we know $x(t)$, its position as a function of time. Similarly, for

pure rotation, we can know all there is to know about a rotating body if we know $\theta(t)$, the angular position of the body's reference line as a function of time.

1.2.2. Angular Displacement

The changing the angular position of the reference line from θ_1 to θ_2, the body undergoes an **angular displacement** $\Delta\theta$ given by

$$\Delta\theta = \theta_2 - \theta_1.$$

This definition of angular displacement holds not only for the rigid body as a whole but also for *every particle within that body.*

If a body is in translational motion along an x axis, its displacement Δx is either positive or negative, depending on whether the body is moving in the positive or negative direction of the axis. Similarly, the angular displacement $\Delta\theta$ of a rotating body is either positive or negative, according to the following rule:

> An angular displacement in the counterclockwise direction is positive, and one in the clockwise direction is negative.

The phrase "clocks are negative" can help you remember this rule (they certainly are negative when their alarms sound off early in the morning).

1.2.3. Angular Velocity

Suppose that our rotating body is at angular position θ_1 at time t_1 and at angular position θ_2 at time t_2. We define the **average angular velocity** of the body in the time interval Δt from t_1 to t_2 to be

$$\omega_{avg} = \frac{\theta_2 - \theta_1}{t_2 - t_1} = \frac{\Delta\theta}{\Delta t},$$

where $\Delta\theta$ is the angular displacement during Δt (ω is the lowercase omega).

The **(instantaneous) angular velocity** ω, with which we shall be most concerned, is the limit of the ratio as Δt approaches zero. Thus,

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

If we know $\theta(t)$, we can find the angular velocity ω by differentiation. The magnitude of an angular velocity is called the **angular speed**, which is also represented with ω.

1.2.4. Angular Acceleration

If the angular velocity of a rotating body is not constant, then the body has an angular acceleration. Let ω_2 and ω_1 be its angular velocities at times t_2 and t_1, respectively. The **average angular acceleration** of the rotating body in the interval from t_1 to t_2 is defined as

$$\varepsilon_{avg} = \frac{\omega_2 - \omega_1}{t_2 - t_1} = \frac{\Delta\omega}{\Delta t},$$

in which $\Delta\omega$ is the change in the angular velocity that occurs during the time interval Δt. The **(instantaneous) angular acceleration** ε, with which we shall be most concerned, is the limit of this quantity as Δt approaches zero. Thus,

$$\varepsilon = \lim_{\Delta t \to 0} \frac{\Delta\omega}{\Delta t} = \frac{d\omega}{dt}.$$

The unit of angular acceleration is commonly the radian per second-squared (rad/s^2) or the revolution per second-squared (rev/s^2).

1.2.5. Angular Quantities Are Vectors

We can describe the position, velocity, and acceleration of a single particle by means of vectors. If the particle is confined to a straight line, however, we do not really need vector notation. Such a particle has only two directions available to it, and we can indicate these directions with plus and minus signs.

In the same way, a rigid body rotating about a fixed axis can rotate only clockwise or counterclockwise as seen along the axis, and again we can select between the two directions by means of plus and minus signs. The question arises: "Can we treat the angular displacement, velocity, and acceleration of a rotating body as vectors?" The answer is a qualified "yes" .

Consider the angular velocity. We can represent angular velocity as a vector $\vec{\omega}$ pointing along the axis of rotation. Then we establish a direction for the vector $\vec{\omega}$ by using a **right-hand rule**. Curl your right hand about the rotating record, your fingers pointing *in the direction of rotation*. Your extended thumb will then point in the direction of the angular velocity vector. If the record were to rotate in the opposite sense, the right-hand rule would tell you that the angular velocity vector then points in the opposite direction.

Now for the caution: Angular displacements (unless they are very small) cannot be treated as vectors. Why not? We can certainly give them both magnitude

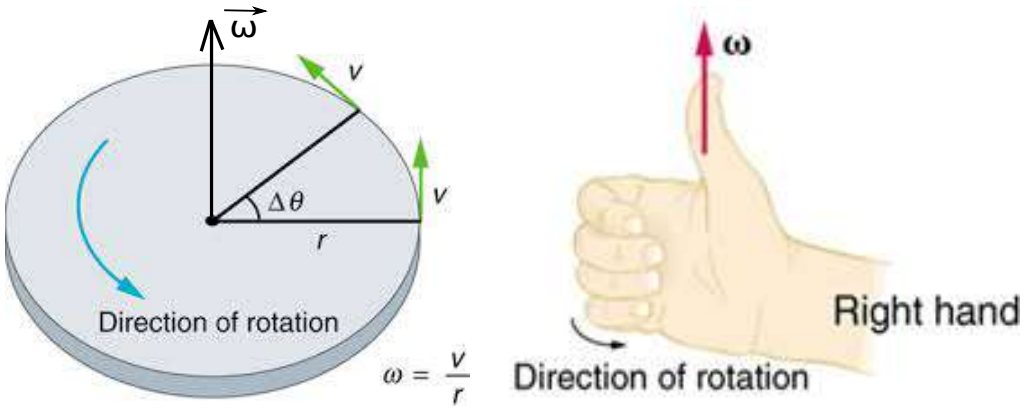

Fig. 1.6. The right-hand rule.

and direction, as we did for the angular velocity vector. However, to be represented as a vector, a quantity must also obey the rules of vector addition, one of which says that if you add two vectors, the order in which you add them does not matter. Angular displacements fail this test.

1.3. Relating the Linear and Angular Variables

Rotation with Constant Angular Acceleration In pure translation, motion with a constant linear acceleration (for example, that of a falling body) is an important special case. In pure rotation, the case of constant angular acceleration is also important, and a parallel set of equations holds for this case also. We shall not derive them here, but simply write them from the corresponding linear equations, substituting equivalent angular quantities for the linear ones (see table.1.1).

Table 1.1

Equivalent angular quantities to the linear

Linear Equation	Missing Variable	Angular Equation
$v = v_0 + at$	$\Delta x = x - x_0 \sim \quad \Delta\theta = \theta - \theta_0$	$\omega = \omega_0 + \varepsilon t$
$x - x_0 = v_0 t + \frac{1}{2}at^2$	$v \sim \quad \omega$	$\theta - \theta_0 = \omega_0 t + \frac{1}{2}\varepsilon t^2$
$v^2 = v_0^2 + 2a(x - x_0)$	$t \sim \quad t$	$\omega^2 = \omega_0^2 + 2\varepsilon(\theta - \theta_0)$
$x - x_0 = \frac{1}{2}(v_0 + v)t$	$a \sim \quad \varepsilon$	$\theta - \theta_0 = \frac{1}{2}(\omega_0 + \omega)t$
$x - x_0 = vt - \frac{1}{2}at^2$	$v_0 \sim \quad \omega_0$	$\theta - \theta_0 = \omega t - \frac{1}{2}\varepsilon t^2$

When a rigid body, such as a merry-go-round, rotates around an axis, each particle in the body moves in its own circle around that axis. Since the body is

rigid, all the particles make one revolution in the same amount of time; that is, they all have the same angular speed ω.

However, the farther a particle is from the axis, the greater the circumference of its circle is, and so the faster its linear speed v must be. You can notice this on a merry-go-round. You turn with the same angular speed ω regardless of your distance from the center, but your linear speed v increases noticeably if you move to the outside edge of the merry-go-round.

We often need to relate the linear variables s, v, and a for a particular point in a rotating body to the angular variables θ, ω, and ε for that body. The two sets of variables are related by r, the *perpendicular distance* of the point from the rotation axis. This perpendicular distance is the distance between the point and the rotation axis, measured along a perpendicular to the axis. It is also the radius r of the circle traveled by the point around the axis of rotation.

The Position. If a reference line on a rigid body rotates through an angle θ, a point within the body at a position r from the rotation axis moves a distance s along a circular arc

$$s = \theta r \quad \text{(radian measure)}.$$

This is the first of our linear – angular relations. *Caution:* The angle θ here must be measured in radians.

The Speed. Differentiating a distance s with respect to time – with r held constant – leads to

$$\frac{ds}{dt} = \frac{d\theta}{dt} r.$$

However, ds/dt is the linear speed (the magnitude of the linear velocity) of the point in question, and $d\theta/dt$ is the angular speed ω of the rotating body. So

$$v = \omega r \quad \text{(radian measure)}.$$

Caution: The angular speed ω must be expressed in radian measure. This equation tells us that since all points within the rigid body have the same angular speed ω, points with greater radius r have greater linear speed v. If the angular speed ω of the rigid body is constant, then this equation tells us that the linear speed v of any point within it is also constant. Thus, each point within the body undergoes uniform circular motion. The period of revolution T for the motion of

each point and for the rigid body itself is

$$T = \frac{2\pi r}{v}$$

This equation tells us that the time for one revolution is the distance $2\pi r$ traveled in one revolution divided by the speed at which that distance is traveled.

$$T = \frac{2\pi}{\omega} \quad \text{(radian measure)}.$$

This equivalent equation says that the time for one revolution is the angular distance 2π rad traveled in one revolution divided by the angular speed (or rate) at which that angle is traveled.

The Acceleration. Differentiating a linear speed v with respect to time – with r held constant – leads to

$$\frac{dv}{dt} = \frac{d\omega}{dt} r.$$

Here we run up against a complication. Expression dv/dt represents only the part of the linear acceleration that is responsible for changes in the *magnitude* v of the linear velocity $\vec{v}$. Like $\vec{v}$, that part of the linear acceleration is tangent to the path of the point in question. We call it the *tangential component* a_t of the linear acceleration of the point, and we write

$$a_t = \varepsilon r \quad \text{(radian measure)}.$$

where $\varepsilon = d\omega/dt$. *Caution:* The angular acceleration ε must be expressed in radian measure.

In addition a particle (or point) moving in a circular path has a *radial component* of linear acceleration, $a_r = v^2/r$ (directed radially inward) that is responsible for changes in the *direction* of the linear velocity $\vec{v}$. By substithat taking into account the expression for v, we can write this component as

$$a_r = \frac{v^2}{r} = \omega^2 r \quad \text{(radian measure)}.$$

Thus the linear acceleration of a point on a rotating rigid body has, in general, two components. The radially inward component a_r is present whenever the angular velocity of the body is not zero. The tangential component a_t is present whenever the angular acceleration is not zero.

Test questions

1. What movement is called translational? rotational?

2. What is a displacement vector? Is the absolute value of the displacement vector always equal to the distance traveled by the point?

3. Give definitions of vectors of average speed and average acceleration, instantaneous speed and instantaneous acceleration. What are their directions?

4. What does the tangential acceleration component characterize? normal component of acceleration? What are their modules?

5. Is it possible to move in which there is no normal acceleration? tangential acceleration? Give examples.

6. What is called angular velocity? angular acceleration? How are their directions determined?

7. What is the relationship between linear and angular quantities?

Lecture No 2

Dynamics. Newton's laws.

We have seen that part of physics is a study of motion, including accelerations, which are changes in velocities. Physics is also a study of what can cause an object to accelerate. That cause is a **force**, which is, loosely speaking, a push or pull on the object. The force is said to act on the object to change its velocity. When a car slams into a telephone pole, a force on the car from the pole causes the car to stop. Science, engineering, legal, and medical journals are filled with articles about forces on objects, including people.

Newtonian Mechanics The relation between a force and the acceleration it causes was first understood by Isaac Newton (1642 – 1727) and is the subject of this lecture. The study of that relation, as Newton presented it, is called **Newtonian mechanics**. We shall focus on its three primary laws of motion. Newtonian mechanics does not apply to all situations. If the speeds of the interacting bodies are very large — an appreciable fraction of the speed of light — we must replace Newtonian mechanics with **Einstein's special theory of relativity**, which holds at any speed, including those near the speed of light. If the interacting bodies are on the scale of atomic structure (for example, they might be electrons in an atom), we must replace Newtonian mechanics with **quantum mechanics**. Physicists now view Newtonian mechanics as a special case of these two more comprehensive theories. Still, it is a very important special case because it applies to the motion of objects ranging in size from the very small (almost on the scale of atomic structure) to astronomical (galaxies and clusters of galaxies).

2.1. Newton's First Law

Before Newton formulated his mechanics, it was thought that some influence, a "force," was needed to keep a body moving at constant velocity. Similarly, a body was thought to be in its "natural state" when it was at rest. For a body

to move with constant velocity, it seemingly had to be propelled in some way, by a push or a pull. Otherwise, it would "naturally" stop moving. These ideas were reasonable. If you send a puck sliding across a wooden floor, it does indeed slow and then stop. If you want to make it move across the floor with constant velocity, you have to continuously pull or push it. Send a puck sliding over the ice of a skating rink, however, and it goes a lot farther. You can imagine longer and more slippery surfaces, over which the puck would slide farther and farther. In the limit you can think of a long, extremely slippery surface (said to be a **frictionless surface**), over which the puck would hardly slow. (We can in fact come close to this situation by sending a puck sliding over a horizontal air table, across which it moves on a film of air.) From these observations, we can conclude that a body will keep moving with constant velocity if no force acts on it. That leads us to the first of Newton's three laws of motion:

> **Newton's First Law:** If no force acts on a body, the body's velocity cannot change; that is, the body cannot accelerate.

In other words, if the body is at rest, it stays at rest. If it is moving, it continues to move with the same velocity (same magnitude *and* same direction).

Force

We now wish to define the unit of force. We know that a force can cause the acceleration of a body. Thus, we shall define the unit of force in terms of the acceleration that a force gives to a standard reference body, which we take to be the standard kilogram. This body has been assigned, exactly and by definition, a mass of 1 kg.

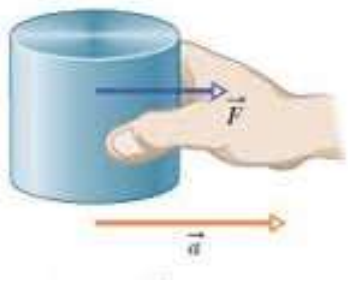

Fig. 2.1. A force $\vec{F}$ on the standard kilogram gives that body an acceleration .

We put the standard body on a horizontal frictionless table and pull the body to the right (Fig.2.1) so that, by trial and error, it eventually experiences a measured acceleration of 1 m/s^2. We then declare, as a matter of definition, that the force we are exerting on the standard body has a magnitude of 1 newton (abbreviated N).

We can exert a 2 N force on our standard body by pulling it so that its measured acceleration is 2 m/s^2, and so on. Thus in general, if our standard body of 1 kg mass has an acceleration of magnitude a, we know that a force F must be acting on it and

that the magnitude of the force (in newtons) is equal to the magnitude of the acceleration (in meters per second per second).

Thus, a force is measured by the acceleration it produces. However, acceleration is a vector quantity, with both magnitude and direction. Is force also a vector quantity? We can easily assign a direction to a force (just assign the direction of the acceleration), but that is not sufficient. We must prove by experiment that forces are vector quantities. Actually, that has been done: forces are indeed vector quantities; they have magnitudes and directions, and they combine according to the vector rules.

This means that when two or more forces act on a body, we can find their **net force**, or **resultant force**, by adding the individual forces vectorially. A single force that has the magnitude and direction of the net force has the same effect on the body as all the individual forces together. This fact is called the **principle of superposition for forces**. The world would be quite strange if, for example, you and a friend were to pull on the standard body in the same direction, each with a force of 1 N, and yet somehow the net pull was 14 N.

Here, forces are most often represented with a vector symbol such as $\vec{F}$, and a net force is represented with the vector symbol $\vec{F}_{net}$. As with other vectors, a force or a net force can have components along coordinate axes.When forces act only along a single axis, they are single-component forces. Then we can drop the overhead arrows on the force symbols and just use signs to indicate the directions of the forces along that axis. Instead of the wording used abowe, the more proper statement of Newton's First Law is in terms of a *net* force:

> **Newton's First Law:** If no *net* force acts on a body ($\vec{F}_{net} = 0$), the body's velocity cannot change; that is, the body cannot accelerate.

There may be multiple forces acting on a body, but if their net force is zero, the body cannot accelerate.

Inertial Reference Frames

Newton's first law is not true in all reference frames, but we can always find reference frames in which it (as well as the rest of Newtonian mechanics) is true. Such special frames are referred to as **inertial reference frames**, or simply **inertial frames**.

An inertial reference frame is one in which Newton's laws hold.

For example, we can assume that the ground is an inertial frame provided we can neglect Earth's astronomical motions (such as its rotation).

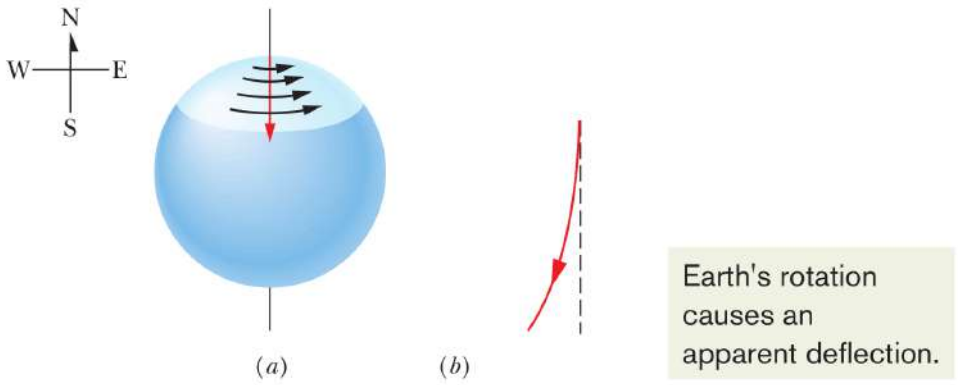

Fig. 2.2. (a) The path of a puck sliding from the north pole as seen from a stationary point in space. Earth rotates to the east. (b) The path of the puck as seen from the ground.

That assumption works well if, say, a puck is sent sliding along a *short* strip of frictionless ice – we would find that the puck's motion obeys Newton's laws. However, suppose the puck is sent sliding along a *long* ice strip extending from the north pole (Fig.2.2a). If we view the puck from a stationary frame in space, the puck moves south along a simple straight line because Earth's rotation around the north pole merely slides the ice beneath the puck. However, if we view the puck from a point on the ground so that we rotate with Earth, the puck's path is not a simple straight line. Because the eastward speed of the ground beneath the puck is greater the farther south the puck slides, from our groundbased view the puck appears to be deflected westward (Fig.2.2b). However, this apparent deflection is caused not by a force as required by Newton's laws but by the fact that we see the puck from a rotating frame. In this situation, the ground is a **noninertial frame**.

If measurements are made in, say, an elevator that is accelerating relative to the ground, then the measurements are being made in a noninertial frame and the results can be surprising.

Mass

Everyday experience tells us that a given force produces different magnitudes

of acceleration for different bodies. Put a baseball and a bowling ball on the floor and give both the same sharp kick. Even if you don't actually do this, you know the result: The baseball receives a noticeably larger acceleration than the bowling ball. The two accelerations differ because the mass of the baseball differs from the mass of the bowling ball – but what, exactly, is mass?

We can explain how to measure mass by imagining a series of experiments in an inertial frame. In the first experiment we exert a force on a standard body, whose mass m_0 is defined to be 1.0 kg. Suppose that the standard body accelerates at 1.0 m/s^2. We can then say the force on that body is 1.0 N.

We next apply that same force (we would need some way of being certain it is the same force) to a second body, body X, whose mass is not known. Suppose we find that this body X accelerates at 0.25 m/s^2. We know that a *less massive* baseball receives a *greater acceleration* than a more massive bowling ball when the same force (kick) is applied to both. Let us then make the following conjecture: The ratio of the masses of two bodies is equal to the inverse of the ratio of their accelerations when the same force is applied to both. For body X and the standard body, this tells us that

$$\frac{m_X}{m_0} = \frac{a_0}{a_X}.$$

Solving for m_X yields

$$m_X = m_0 \frac{a_0}{a_X} = (1.0\ kg)\frac{1.0\ m/s^2}{0.25\ m/s^2} = 4.0\ kg.$$

Our conjecture will be useful, of course, only if it continues to hold when we change the applied force to other values. For example, if we apply an 8.0 N force to the standard body, we obtain an acceleration of 8.0 m/s^2. When the 8.0 N force is applied to body X, we obtain an acceleration of 2.0 m/s^2. Our conjecture then gives us

$$m_X = m_0 \frac{a_0}{a_X} = (1.0\ kg)\frac{8.0\ m/s^2}{2.0\ m/s^2} = 4.0\ kg,$$

consistent with our first experiment. Many experiments yielding similar results indicate that our conjecture provides a consistent and reliable means of assigning a mass to any given body.

Our measurement experiments indicate that mass is an *intrinsic* characteristic of a body – that is, a characteristic that automatically comes with the existence of

the body. They also indicate that mass is a scalar quantity. However, the nagging question remains:What, exactly, is mass?

Since the word *mass* is used in everyday English, we should have some intuitive understanding of it, maybe something that we can physically sense. Is it a body's size, weight, or density? The answer is no, although those characteristics are sometimes confused with mass. We can say only that *the mass of a body is the characteristic that relates a force on the body to the resulting acceleration.* Mass has no more familiar definition; you can have a physical sensation of mass only when you try to accelerate a body, as in the kicking of a baseball or a bowling ball.

2.2. Newton's Second Law

All the definitions, experiments, and observations we have discussed so far can be summarized in one neat statement:

> **Newton's Second Law:** The net force on a body is equal to the product of the body's mass and its acceleration.

In equation form,

$$\vec{F}_{net} = m\vec{a} \quad \text{(Newton's second law)}. \tag{2.1}$$

This equation is simple, but we must use it cautiously. First, we must be certain about which body we are applying it to. Then $\vec{F}_{net}$ must be the vector sum of *all* the forces that act on *that* body. Only forces that act on *that* body are to be included in the vector sum, not forces acting on other bodies that might be involved in the given situation. Every time you work a force problem, your first step is to clearly state the body to which you are applying Newton's law.

Equation (2.1) tells us that if the net force on a body is zero, the body's acceleration $\vec{a} = 0$. If the body is at rest, it stays at rest; if it is moving, it continues to move at constant velocity. In such cases, any forces on the body *balance* one another, and both the forces and the body are said to be in *equilibrium.* Commonly, the forces are also said to *cancel* one another, but the term "cancel" is tricky. It does *not* mean that the forces cease to exist (canceling forces is not like canceling dinner reservations). The forces still act on the body.

For SI units, (2.1) tells us that

$$1\ N = (1\ kg)(1\ m/s^2) = 1\ kg \cdot m/s^2.$$

To solve problems with Newton's second law, we often draw a **free-body diagram** in which the only body shown is the one for which we are summing forces. A sketch of the body itself is preferred by some teachers but, to save space in these chapters, we shall usually represent the body with a dot. Also, each force on the body is drawn as a vector arrow with its tail on the body. A coordinate system is usually included, and the acceleration of the body is sometimes shown with a vector arrow (labeled as an acceleration).

2.2.1. Some Particular Forces

The Gravitational Force

A **gravitational force** $\vec{F}_g$ on a body is a certain type of pull that is directed toward a second body. Above we do not discuss the nature of this force and usually consider situations in which the second body is Earth. Thus, when we speak of *the* gravitational force $\vec{F}_g$ on a body, we usually mean a force that pulls on it directly toward the center of Earth – that is, directly down toward the ground. We shall assume that the ground is an inertial frame.

Suppose a body of mass m is in free fall with the free-fall acceleration of magnitude g. Then, if we neglect the effects of the air, the only force acting on the body is the gravitational force $\vec{F}_g$. We can relate this downward force and downward acceleration with Newton's second law ($\vec{F} = m\vec{a}$). We place a vertical y axis along the body's path, with the positive direction upward. For this axis, Newton's second law can be written in the form $F_{net,y} = ma_y$, which, in our situation, becomes

$$-F_g = m(-g)$$

or In equation form,

$$F_g = mg. \tag{2.2}$$

In words, the magnitude of the gravitational force is equal to the product mg.

This same gravitational force, with the same magnitude, still acts on the body even when the body is not in free fall but is, say, at rest on a pool table or moving across the table. (For the gravitational force to disappear, Earth would have to disappear.)

We can write Newton's second law for the gravitational force in these vector forms:

$$\vec{F}_g = -F_g\hat{j} = -mg\hat{j} = m\vec{g},$$

where $\hat{j}$ is the unit vector that points upward along a y axis, directly away from the ground, and $\vec{g}$ is the free-fall acceleration (written as a vector), directed downward.

Weight

The **weight** P of a body is the magnitude of the net force required to prevent the body from falling freely, as measured by someone on the ground. For example, to keep a ball at rest in your hand while you stand on the ground, you must provide an upward force to balance the gravitational force on the ball from Earth. Suppose the magnitude of the gravitational force is 2.0 N. Then the magnitude of your upward force must be 2.0 N, and thus the weight P of the ball is 2.0 N. We also say that the ball *weighs* 2.0 N and speak about the ball *weighing* 2.0 N.

Now let us generalize the situation. Consider a body that has an acceleration $\vec{a}$ of zero relative to the ground, which we again assume to be an inertial frame. Two forces act on the body: a downward gravitational force $\vec{F}_g$ and a balancing upward force of magnitude P. We can write Newton's second law for a vertical y axis, with the positive direction upward, as

$$F_{net,y} = ma_y.$$

In our situation, this becomes

$$P - F_g = m(0)$$

or $P = F_g$ (weight, with ground as inertial frame).

This equation tells us (assuming the ground is an inertial frame) that

> The weight P of a body is equal to the magnitude F_g of the gravitational force on the body.

Substituting mg for F_g from (2.2), we find

$$P = mg \quad \text{(weight)}. \tag{2.3}$$

which relates a body's weight to its mass.

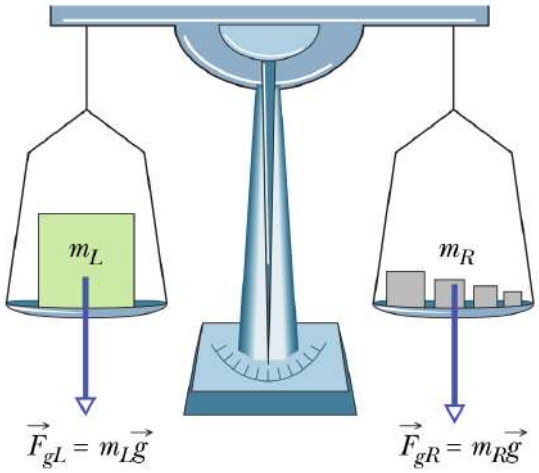

Fig. 2.3. An equal-arm balance. When the device is in balance, the gravitational force $\vec{F}_{gL}$ on the body being weighed (on the left pan) and the total gravitational force $\vec{F}_{gR}$ on the reference bodies (on the right pan) are equal. Thus, the mass m_L of the body being weighed is equal to the total mass m_R of the reference bodies.

To *weigh* a body means to measure its weight. One way to do this is to place the body on one of the pans of an equal-arm balance (Fig.2.3) and then place reference bodies (whose masses are known) on the other pan until we strike a balance (so that the gravitational forces on the two sides match). The masses on the pans then match, and we know the mass of the body. If we know the value of g for the location of the balance, we can also find the weight of the body with (2.3).

We can also weigh a body with a spring scale (Fig.2.4). The body stretches a spring, moving a pointer along a scale that has been calibrated and marked in either mass or weight units. If the scale is marked in mass units, it is accurate only where the value of g is the same as where the scale was calibrated.

The weight of a body must be measured when the body is not accelerating vertically relative to the ground. For example, you can measure your weight on a scale in your bathroom or on a fast train. However, if you repeat the measurement with the scale in an accelerating elevator, the reading differs from your weight because of the acceleration. Such a measurement is called an *apparent weight.*

Caution: A body's weight is not its mass. Weight is the magnitude of a force and is related to mass by (2.3). If you move a body to a point where the value of g is different, the body's mass (an intrinsic property) is not different but the weight is. For example, the weight of a bowling ball having a mass of 7.2 kg is

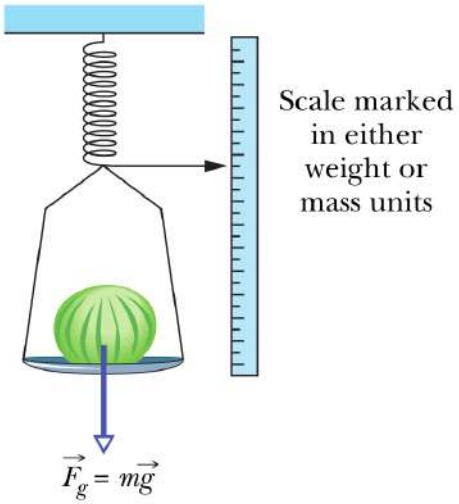

Fig. 2.4. A spring scale. The reading is proportional to the *weight* of the object on the pan, and the scale gives that weight if marked in weight units. If, instead, it is marked in mass units, the reading is the object's weight only if the value of g at the location where the scale is being used is the same as the value of g at the location where the scale was calibrated.

71 N on Earth but only 12 N on the Moon. The mass is the same on Earth and Moon, but the free-fall acceleration on the Moon is only 1.6 m/s^2.

The Normal Force

If you stand on a mattress, Earth pulls you downward, but you remain stationary. The reason is that the mattress, because it deforms downward due to you, pushes up on you. Similarly, if you stand on a floor, it deforms (it is compressed, bent, or buckled ever so slightly) and pushes up on you. Even a seemingly rigid concrete floor does this (if it is not sitting directly on the ground, enough people on the floor could break it).

The push on you from the mattress or floor is a **normal force** $\vec{F}_N$. The name comes from the mathematical term *normal*, meaning perpendicular: The force on you from, say, the floor is perpendicular to the floor.

When a body presses against a surface, the surface (even a seemingly rigid one) deforms and pushes on the body with a normal force $\vec{F}_N$ that is perpendicular to the surface.

Fig.2.5a shows an example. A block of mass m presses down on a table, deforming it somewhat because of the gravitational force $\vec{F}_g$ on the block. The table pushes up on the block with normal force $\vec{F}_N$. The free-body diagram for the block is given in Fig.2.5b. Forces $\vec{F}_g$ and $\vec{F}_N$ are the only two forces on the block and they are both vertical. Thus, for the block we can write Newton's

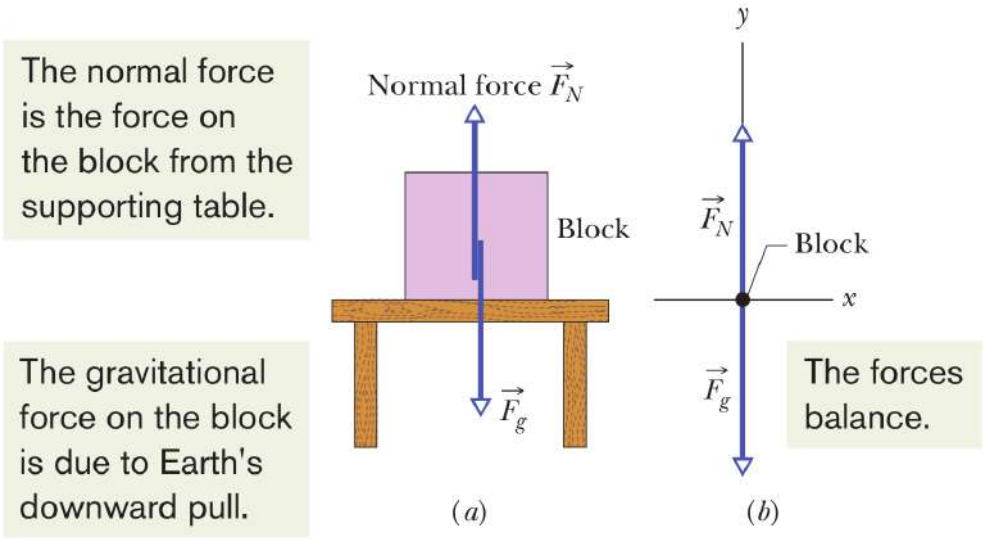

Fig. 2.5. (a) A block resting on a table experiences a normal force $\vec{F}_N$ perpendicular to the tabletop. (b) The free-body diagram for the block.

second law for a positive-upward y axis ($F_{net,y} = ma_y$) as

$$F_N - F_g = ma_y.$$

From (2.2), we substitute mg for F_g, finding

$$F_N - mg = ma_y.$$

Then the magnitude of the normal force is

$$F_N = mg + ma_y = m(g + a_y)$$

for any vertical acceleration a_y of the table and block (they might be in an accelerating elevator). If the table and block are not accelerating relative to the ground, then $a_y = 0$ and

$$\boxed{F_N = mg}.$$

Friction

If we either slide or attempt to slide a body over a surface, the motion is resisted by a bonding between the body and the surface. The resistance is considered to be a single force $\vec{f}$, called either the **frictional force** or simply **friction**. This force is directed along the surface, opposite the direction of the intended motion (Fig.2.6). Sometimes, to simplify a situation, friction is assumed to be negligible (the surface is *frictionless*).

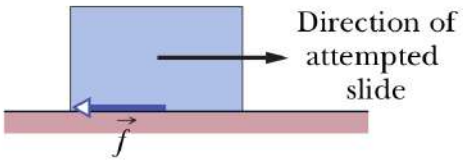

Fig. 2.6. A frictional force $\vec{f}$ opposes the attempted slide of a body over a surface.

Tension

When a cord (or a rope, cable, or other such object) is attached to a body and pulled taut, the cord pulls on the body with a force $\vec{T}$ directed away from the body and along the cord (Fig.2.7a). The force is often called a *tension force* because the cord is said to be in a state of *tension* (or to be *under tension*), which means that it is being pulled taut. The *tension in the cord* is the magnitude T of the force on the body. For example, if the force on the body from the cord has magnitude $T = 50\ N$, the tension in the cord is $50\ N$

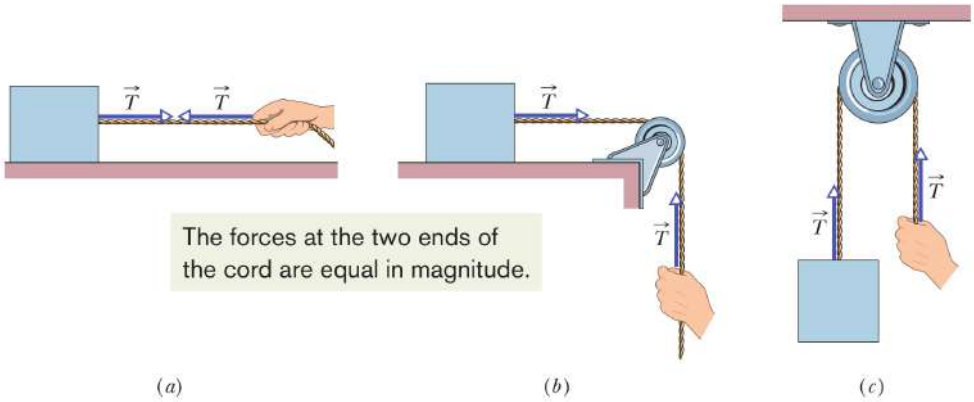

Fig. 2.7. (a) The cord, pulled taut, is under tension. If its mass is negligible, the cord pulls on the body and the hand with force $\vec{T}$, even if the cord runs around a massless, frictionless pulley as in (b) and (c).

A cord is often said to be *massless* (meaning its mass is negligible compared to the body's mass) and *unstretchable*. The cord then exists only as a connection between two bodies. It pulls on both bodies with the same force magnitude T, even if the bodies and the cord are accelerating and even if the cord runs around a *massless, frictionless pulley* (Figs.2.7b and c). Such a pulley has negligible mass compared to the bodies and negligible friction on its axle opposing its rotation. If the cord wraps halfway around a pulley, as in Fig.2.7c, the net force on the pulley from the cord has the magnitude $2T$.

2.3. Newton's Third Law

Two bodies are said to *interact* when they push or pull on each other – that is, when a force acts on each body due to the other body.

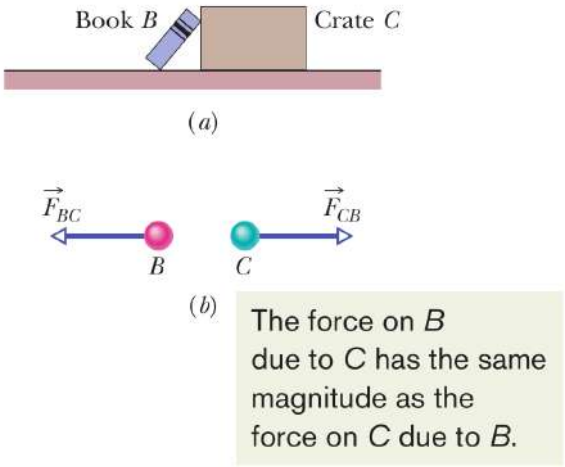

Fig. 2.8. (a) Book B leans against crate C. (b) Forces $\vec{F}_{BC}$ (the force on the book from the crate) and $\vec{F}_{CB}$ (the force on the crate from the book) have the same magnitude and are opposite in direction.

For example, suppose you position a book B so it leans against a crate C (Fig.2.8a). Then the book and crate interact: There is a horizontal force $\vec{F}_{BC}$ on the book from the crate (or due to the crate) and a horizontal force $\vec{F}_{CB}$ on the crate from the book (or due to the book). This pair of forces is shown in Fig.2.8b. Newton's third law states that

> **Newton's Third Law:** When two bodies interact, the forces on the bodies from each other are always equal in magnitude and opposite in direction.

For the book and crate, we can write this law as the scalar relation

$$F_{BC} = F_{CB} \quad \text{(equal magnitudes)}$$

or as the vector relation

$$\vec{F}_{BC} = -\vec{F}_{CB} \quad \text{(equal magnitudes and opposite directions)}$$

where the minus sign means that these two forces are in opposite directions. We can call the forces between two interacting bodies a **third-law force pair**.

When any two bodies interact in any situation, a third-law force pair is present. The book and crate in Fig.2.8a are stationary, but the third law would still hold if they were moving and even if they were accelerating.

2.6. Energy

The term *energy* is so broad that a clear definition is difficult to write. Technically, energy is a scalar quantity associated with the state (or condition) of one or more objects.

Kinetic energy E_K is energy associated with the *state of motion* of an object. The faster the object moves, the greater is its kinetic energy. When the object is stationary, its kinetic energy is zero.

For an object of mass m whose speed v is well below the speed of light,

$$E_K = \frac{1}{2}mv^2 \quad \text{(kinetic energy)}. \tag{2.4}$$

The SI unit of kinetic energy (and every other type of energy) is the **joule** (J), named for James Prescott Joule, an English scientist of the 1800s.

$$1\ joule = 1\ J = 1\ kg \cdot m^2/s^2.$$

Work A is energy transferred to or from an object via a force acting on the object. Energy transferred to the object is positive work, and from the object, negative work.

Work Done by a Constant Force

The work done on a particle by a constant force $\vec{F}$ during displacement $\vec{d}$ is

$$A = Fd\cos\phi = \vec{F} \cdot \vec{d} \quad \text{(work, constant force)}. \tag{2.5}$$

in which ϕ is the constant angle between the directions of $\vec{F}$ and $\vec{d}$. Only the component of $\vec{F}$ that is along the displacement $\vec{d}$ can do work on the object. When two or more forces act on an object, their **net work** is the sum of the individual works done by the forces, which is also equal to the work that would be done on the object by the net force $\vec{F}_{net}$ of those forces.

Work and Kinetic Energy

For a particle, a change ΔE_K in the kinetic energy equals the net work A done on the particle:

$$\Delta E_K = E_{Kf} - E_{Ki} = A \quad \text{(work – kinetic energy theorem)}$$

in which E_{Ki} is the initial kinetic energy of the particle and E_{Kf} is the kinetic energy after the work is done.

$$E_{Kf} = E_{Ki} + A. \tag{2.6}$$

Potential Energy

A **potential energy** is energy that is associated with the configuration of a system in which a conservative force acts. When the conservative force does work A on a particle within the system, the change ΔU in the potential energy of the system is

$$\Delta U = -A.$$

If the particle moves from point x_i to point x_f, the change in the potential energy of the system is

$$\Delta U = -\int_{x_i}^{x_f} F(x)dx.$$

The potential energy associated with a system consisting of Earth and a nearby particle is **gravitational potential energy**. If the particle moves from height y_i to height y_f, the change in the gravitational potential energy of the particle – Earth system is

$$\Delta U = mg(y_f - y_i) = mg\Delta y.$$

If the **reference point** of the particle is set as $y_i = 0$ and the corresponding gravitational potential energy of the system is set as $U_i = 0$, then the gravitational potential energy U when the particle is at any height y is

$$U(y) = mgy.$$

Conservation of Energy

The **total energy** E of a system (the sum of its mechanical energy and its internal energies, including thermal energy) can change only by amounts of energy that are transferred to or from the system. This experimental fact is known as

the **law of conservation of energy**. If work A is done on the system, then

$$A = \Delta E = \Delta E_{mec} + \Delta E_{th} + \Delta E_{int},$$

Where the mechanical energy E_{mec} of a system is the sum of its kinetic energy E_K and potential energy U. If the system is isolated ($A = 0$), this gives

$$\Delta E_{mec} + \Delta E_{th} + \Delta E_{int} = 0$$

and $E_{mec,2} = E_{mec,1} - \Delta E_{th} - \Delta E_{int}$, where the subscripts 1 and 2 refer to two different instants.

2.3.1. Linear Momentum

Center of mass and center of gravity

The **center of mass** is the virtual point about which the mass of the body is evenly distributed. The **center of gravity** is the virtual point at which the gravitational force acts on a body and is the point where the weight force can be said to be acting. We can assume that the center of mass and the center of gravity coincide as humans operate within a constant gravitational field when on the Earth. However, in space, where the influence of gravity has been removed, the center of gravity no longer exists, while the center of mass remains.

The **center of mass** of a system of n particles is defined to be the point whose coordinates are given by

$$\vec{r}_{com} = \frac{1}{M} \sum_{i=1}^{n} m_i \vec{r}_i,$$

where M is the total mass of the system.

Linear Momentum and Newton's Second Law

For a single particle, we define a quantity $\vec{p}$ called its **linear momentum** as

$$\vec{p} = m\vec{v},$$

and can write Newton's second law in terms of this momentum:

$$\vec{F}_{net} = \frac{d\vec{p}}{dt}.$$

Conservation of Linear Momentum

If a system is isolated so that no net *external* force acts on it, the linear momentum $\vec{P}$ of the system remains constant:

$$\vec{P} = constant \quad \text{(closed, isolated system)}.$$

Test questions

1. What is the difference between the concepts of energy and work?

2. What is the relationship between force and potential energy?

3. What caused the change in potential energy?

4. Is the condition of closedness of the system necessary to fulfill the law of conservation of mechanical energy?

5. What is the law of conservation of mechanical energy? For which systems is it performed?

6. What is the physical essence of the law of conservation and transformation energy? Why is it a fundamental law of nature?

Lecture No 3

Angular motion

3.1. Rotational Kinetic Energy and Rotational Inertia

The kinetic energy E_K of a rigid body rotating about a fixed axis is given by

$$E_K = \frac{1}{2} I\omega^2 \quad \text{(radian measure)},$$

in which I is the **rotational inertia** (or **moment of inertia**) of the body, defined as

$$I = \sum m_i r_i^2$$

for a system of discrete particles and defined as

$$I = \int r^2 dm$$

for a body with continuously distributed mass. The r and r_i in these expressions represent the perpendicular distance from the axis of rotation to each mass element in the body, and the integration is carried out over the entire body so as to include every mass element.

Fig.3.1 gives the results of such integration for nine common body shapes and the indicated axes of rotation.

3.1.1. The Parallel-Axis Theorem

The *parallel-axis theorem* relates the rotational inertia I of a body about any axis to that of the same body about a parallel axis through the center of mass:

$$I = I_{com} + Ma^2$$

Here a is the perpendicular distance between the two axes, and I_{com} is the rotational inertia of the body about the axis through the com. We can describe a

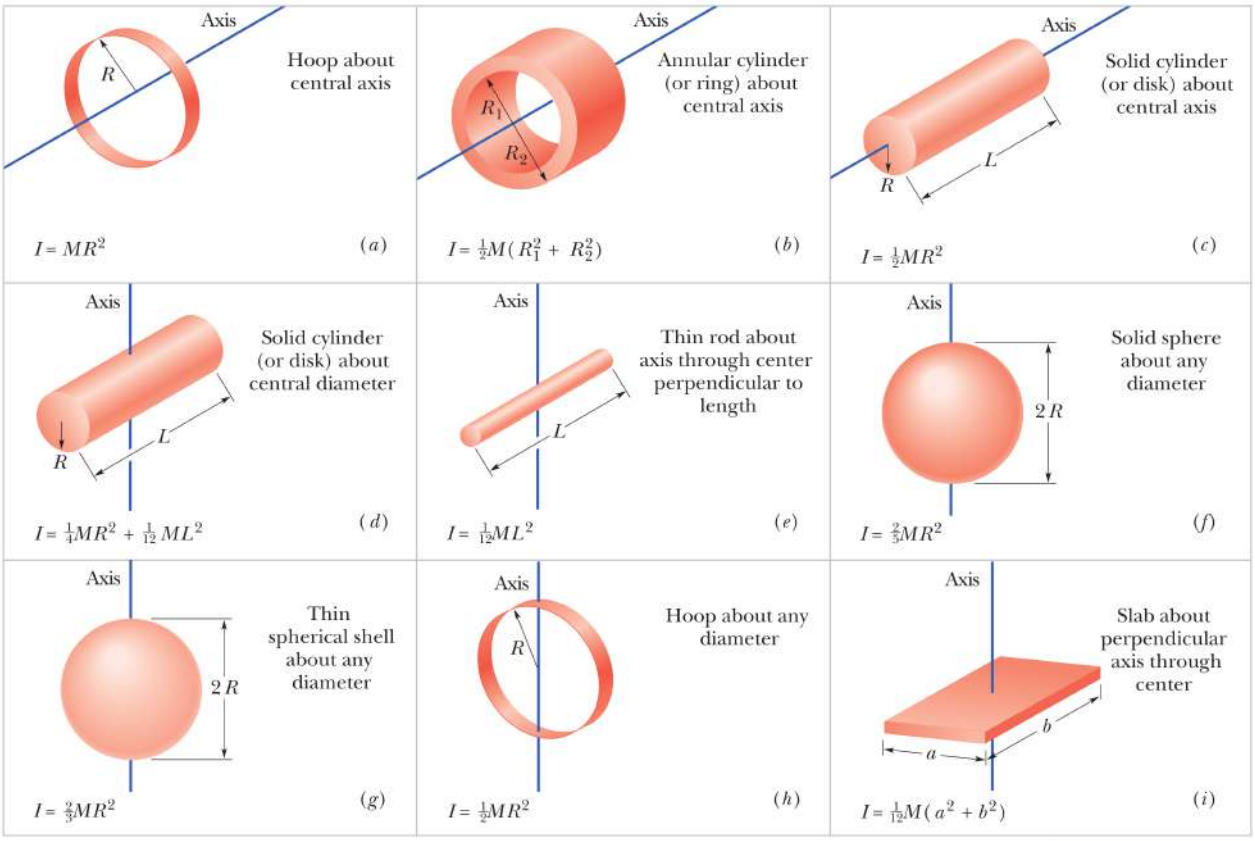

Fig. 3.1. Some Rotational Inertias.

as being the distance the actual rotation axis has been shifted from the rotation axis through the com.

3.1.2. Torque

Torque is a turning or twisting action on a body about a rotation axis due to a force $\vec{F}$. If $\vec{F}$ is exerted at a point given by the position vector $\vec{r}$ relative to the axis, then the magnitude of the torque is

$$\tau = rF_t = r_\perp F = rF\sin\phi,$$

where F_t is the component of $\vec{F}$ perpendicular to $\vec{r}$ and ϕ is the angle between $\vec{r}$ and $\vec{F}$. The quantity $r_\perp$ is the perpendicular distance between the rotation axis and an extended line running through the $\vec{F}$ vector. This line is called the **line of action** of $\vec{F}$, and $r_\perp$ is called the **moment arm** of $\vec{F}$. Similarly, r is the moment arm of F_t.

The SI unit of torque is the newton-meter $(N \cdot m)$. A torque τ is positive if it tends to rotate a body at rest counterclockwise and negative if it tends to rotate the body clockwise.

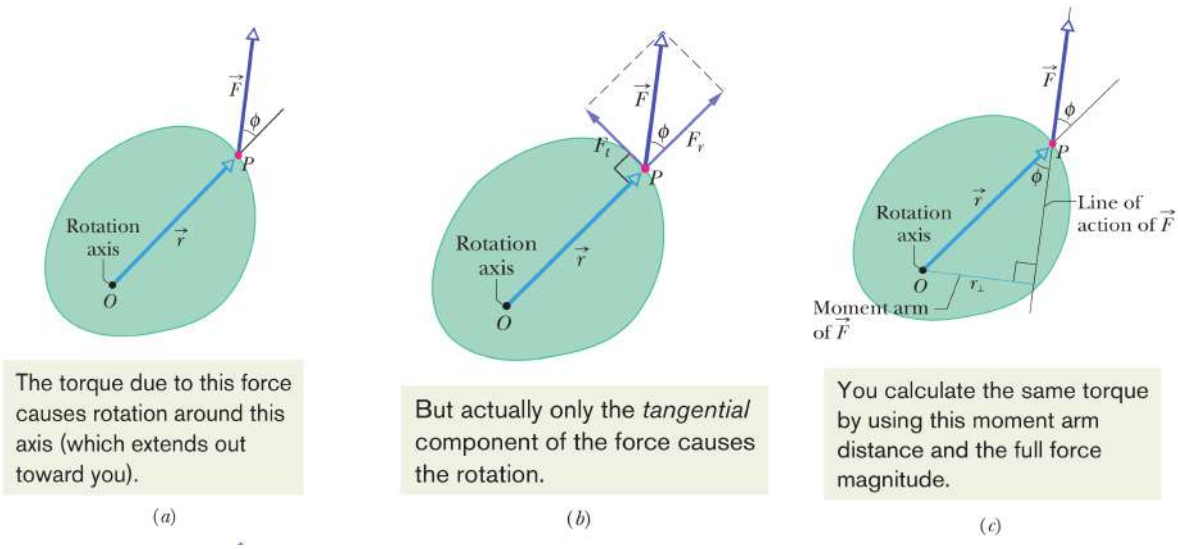

Fig. 3.2. (a) A force $\vec{F}$ acts on a rigid body, with a rotation axis perpendicular to the page. The torque can be found with (a) angle ϕ, (b) tangential force component F_t, or (c) moment arm $r_\perp$.

3.1.3. Newton's Second Law in Angular Form

The rotational analog of Newton's second law is

$$\tau_{net} = I\varepsilon,$$

where τ_{net} net is the net torque acting on a particle or rigid body, I is the rotational inertia of the particle or body about the rotation axis, and ε is the resulting angular acceleration about that axis.

3.2. Work and Rotational Kinetic Energy

The equations used for calculating work and power in rotational motion correspond to equations used for translational motion and are

$$A = \int_{\theta_i}^{\theta_f} \tau d\theta$$

$$P = \frac{dA}{dt} = \tau\omega.$$

When τ is constant

$$A = \tau(\theta_f - \theta_i).$$

The form of the work – kinetic energy theorem used for rotating bodies is

$$\Delta E_K = E_{K_f} - E_{K_i} = \frac{1}{2} I\omega_f^2 - \frac{1}{2} I\omega_i^2 = A.$$

3.2.1. Torque as a Vector

In three dimensions, *torque* $\vec{\tau}$ is a vector quantity defined relative to a fixed point (usually an origin); it is

$$\vec{\tau} = \vec{r} \times \vec{F},$$

where $\vec{F}$ is a force applied to a particle and $\vec{r}$ is a position vector locating the particle relative to the fixed point. The magnitude of $\vec{\tau}$ is given by

$$\tau = rF \sin\phi = rF_\perp = r_\perp F,$$

where ϕ is the angle between $\vec{F}$ and $\vec{r}$, $F_\perp$ is the component of $\vec{F}$ perpendicular to $\vec{r}$, and $r_\perp$ is the moment arm of $\vec{F}$. The direction of $\vec{\tau}$ is given by the right-hand rule.

3.3. Angular Momentum of a Particle

The *angular momentum* $\vec{\ell}$ of a particle with linear momentum $\vec{p}$, mass m, and linear velocity $\vec{v}$ is a vector quantity defined relative to a fixed point (usually an origin) as

$$\vec{\ell} = \vec{r} \times \vec{p} = m(\vec{r} \times \vec{v}).$$

The magnitude of $\vec{\ell}$ is given by

$$\ell = rmv \sin\phi = rp_\perp = rmv_\perp = r_\perp p = r_\perp mv,$$

where ϕ is the angle between $\vec{r}$ and $\vec{p}$, $p_\perp$ and $v_\perp$ are the components of $\vec{p}$ and $\vec{v}$ perpendicular to $\vec{r}$, and $r_\perp$ is the perpendicular distance between the fixed point and the extension of $\vec{p}$. The direction of $\vec{\ell}$ is given by the right-hand rule for cross products.

3.3.1. Newton's Second Law in Angular Form

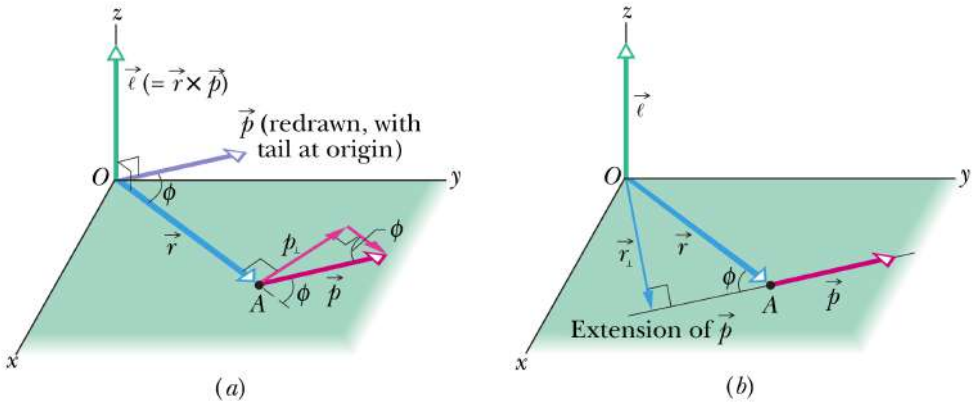

Fig. 3.3. Defining angular momentum. A particle passing through point A has linear momentum $\vec{p}(= m\vec{v})$, with the vector $\vec{p}$ lying in an xy plane. The particle has angular momentum $\vec{\ell}(= \vec{r}\times\vec{p})$ with respect to the origin O. By the right-hand rule, the angular momentum vector points in the positive direction of z. (a) The magnitude of $\vec{\ell}$ is given by $\ell = rp_\perp = rmv_\perp$. (b) The magnitude of $\vec{\ell}$ is also given by $\ell = r_\perp p = r_\perp mv$.

Newton's second law for a particle can be written in angular form as

$$\vec{\tau}_{net} = \frac{d\vec{\ell}}{dt},$$

where $\vec{\tau}_{net}$ is the net torque acting on the particle and ℓ is the angular momentum of the particle.

3.3.2. Angular Momentum of a System of Particles

The angular momentum $\vec{L}$ of a system of particles is the vector sum of the angular momenta of the individual particles:

$$\vec{L} = \vec{\ell_1} + \vec{\ell_2} + \cdots + \vec{\ell_n} = \sum_{i=1}^{n} \vec{\ell_i}.$$

The time rate of change of this angular momentum is equal to the net external torque on the system (the vector sum of the torques due to interactions of the particles of the system with particles external to the system):

$$\vec{\tau}_{net} = \frac{d\vec{L}}{dt} \quad \text{(system of particles)}.$$

Angular Momentum of a Rigid Body. For a rigid body rotating about a fixed axis, the component of its angular momentum parallel to the rotation axis

is

$$L = I\omega \quad \text{(rigid body, fixed axis).}$$

3.3.3. Conservation of Angular Momentum

The angular momentum $\vec{L}$ of a system remains constant if the net external torque acting on the system is zero:

$$\vec{L} = a\ constant \quad \text{(isolated system)}$$

or $\vec{L}_i = \vec{L}_f$ (isolated system).
This is the **law of conservation of angular momentum.**

The spinning volunteer

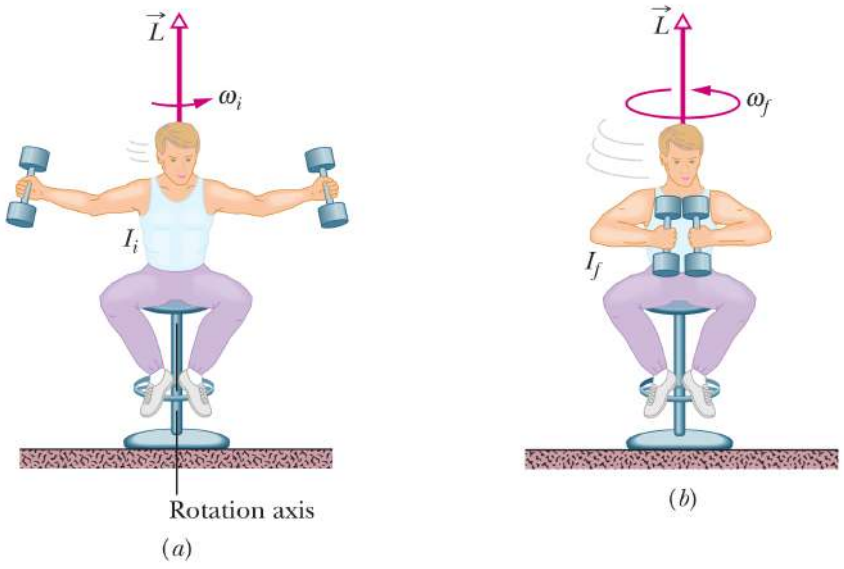

Fig. 3.4. (a) The student has a relatively large rotational inertia about the rotation axis and a relatively small angular speed. (b) By decreasing his rotational inertia, the student automatically increases his angular speed. The angular momentum $\vec{L}$ of the rotating system remains unchanged.

Suppose that the initially rigid body somehow redistributes its mass relative to that rotation axis, changing its rotational inertia about that axis. We write this conservation law as

$$I_i\omega_i = I_f\omega_f. \tag{3.1}$$

Here the subscripts refer to the values of the rotational inertia I and angular speed ω before and after the redistribution of mass.

Fig.3.4 shows a student seated on a stool that can rotate freely about a vertical axis. The student, who has been set into rotation at a modest initial angular speed ω_i, holds two dumbbells in his outstretched hands. His angular momentum vector $\vec{L}$ lies along the vertical rotation axis, pointing upward.

The instructor now asks the student to pull in his arms; this action reduces his rotational inertia from its initial value I_i to a smaller value I_f because he moves mass closer to the rotation axis. His rate of rotation increases markedly, from ω_i to ω_f. The student can then slow down by extending his arms once more, moving the dumbbells outward.

No net external torque acts on the system consisting of the student, stool, and dumbbells. Thus, the angular momentum of that system about the rotation axis must remain constant, no matter how the student maneuvers the dumbbells. In Fig.3.4a, the student's angular speed ω_i is relatively low and his rotational inertia I_i is relatively high. According to (3.1), his angular speed in Fig.3.4b must be greater to compensate for the decreased I_f.

Table 3.1

Some Corresponding Relations for Translational and Rotational Motion

Pure Translation (Fixed Direction)	Pure Rotation (Fixed Axis)
Position x	Angular position θ
Velocity $\vec{v} = \dfrac{dx}{dt}$	Angular velocity $\vec{\omega} = \dfrac{d\theta}{dt}$
Acceleration $\vec{a} = \dfrac{d\vec{v}}{dt} = \dfrac{d^2x}{dt^2}$	Angular acceleration $\vec{\varepsilon} = \dfrac{d\vec{\omega}}{dt} = \dfrac{d^2\theta}{dt^2}$
Force $\vec{F}$	Torque (moment of force) $\vec{M} = [r \times F]$
Linear momentum $\vec{p} = m\vec{v}$	Angular momentum $\vec{L} = [r \times p] = J\vec{\omega}$
Basic dynamics equation	
$\vec{F} = m\vec{a}$	$\vec{M} = J\vec{\varepsilon}$
$\vec{F} = \dfrac{d\vec{p}}{dt}$	$\vec{M} = \dfrac{d\vec{L}}{dt}$
Work $dA = F_s dS$	Work $dA = M_z d\varphi$
Kinetic energy $E_k = \dfrac{mv^2}{2}$	Kinetic energy $E_k = \dfrac{J\omega^2}{2}$

3.4. Musculoskeletal System

The muscular and skeletal systems provide support to the body and allow for movement. The bones of the skeleton protect the body's internal organs and support the weight of the body. The muscles of the muscular system contract and pull on the bones, allowing for movements as diverse as standing, walking, running, and grasping items.

Injury or disease affecting the musculoskeletal system can be very debilitating. The most common musculoskeletal diseases worldwide are caused by malnutrition, which can negatively affect development and maintenance of bones and muscles. Other diseases affect the joints, such as arthritis, which can make movement difficult and, in advanced cases, completely impair mobility.

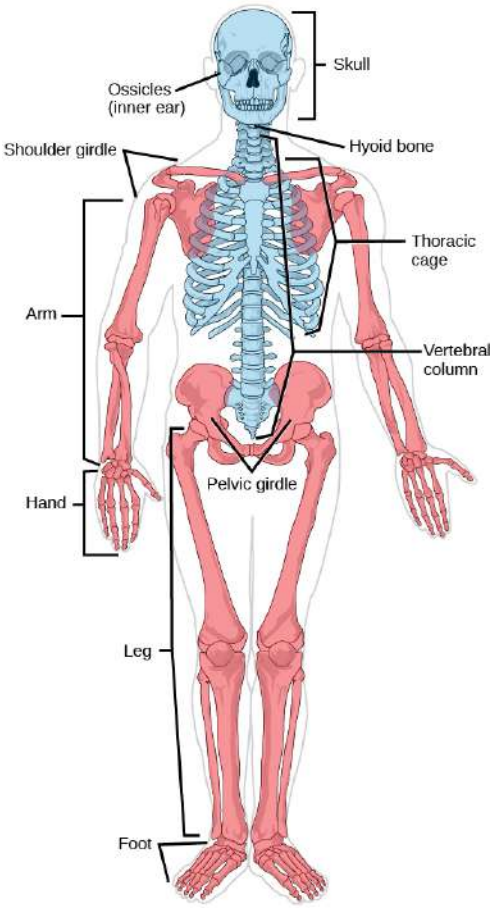

Fig. 3.5. The axial skeleton, shown in blue, consists of the bones of the skull, ossicles of the middle ear, hyoid bone, vertebral column, and thoracic cage. The appendicular skeleton, shown in red, consists of the bones of the pectoral limbs, pectoral girdle, pelvic limb, and pelvic girdle.

Progress in the science of prosthesis design has resulted in the development of artificial joints, with joint replacement surgery in the hips and knees being the most common. Replacement joints for shoulders, elbows, and fingers are also available.

3.4.1 Skeletal System

The human skeleton is an endoskeleton that consists of 206 bones in the adult. An endoskeleton develops within the body rather than outside like the exoskeleton of insects. The skeleton has five main functions: providing support to the body, storing minerals and lipids, producing blood cells, protecting internal organs, and allowing for movement. The skeletal system in vertebrates is divided into the axial skeleton (which consists of the skull, vertebral column, and rib cage), and the appendicular skeleton (which consists of limb bones, the pectoral or shoulder girdle, and the pelvic girdle).

The axial skeleton forms the central axis of the body and includes the bones of the skull, ossicles of the middle ear, hyoid bone of the throat, vertebral column, and the thoracic cage (rib cage) (Figure 3.5.).

The bones of the skull support the structures of the face and protect the brain. The skull consists of cranial bones and facial bones. The cranial bones form the cranial cavity, which encloses the brain and serves as an attachment site for muscles of the head and neck. In the adult they are tightly jointed with connective tissue and adjoining bones do not move.

The auditory ossicles of the middle ear transmit sounds from the air as vibrations to the fluid-filled cochlea. The auditory ossicles consist of two malleus (hammer) bones, two incus (anvil) bones, and two stapes (stirrups), one on each side. Facial bones provide cavities for the sense organs (eyes, mouth, and nose), and serve as attachment points for facial muscles.

The hyoid bone lies below the mandible in the front of the neck. It acts as a movable base for the tongue and is connected to muscles of the jaw, larynx, and tongue. The mandible forms a joint with the base of the skull. The mandible controls the opening to the mouth and hence, the airway and gut.

The vertebral column, or spinal column, surrounds and protects the spinal cord, supports the head, and acts as an attachment point for ribs and muscles of the back and neck. It consists of 26 bones: the 24 vertebrae, the sacrum, and the coccyx. Each vertebral body has a large hole in the center through which the spinal cord passes down to the level of the first lumbar vertebra. Below this level, the hole contains spinal nerves which exit between the vertebrae. There is a notch on each side of the hole through which the spinal nerves, can exit from the spinal cord to serve different regions of the body. The vertebral column is approximately 70 cm (28 in) in adults and is curved, which can be seen from a side view.

Intervertebral discs composed of fibrous cartilage lie between adjacent vertebrae from the second cervical vertebra to the sacrum. Each disc helps form a slightly moveable joint and acts as a cushion to absorb shocks from movements such as walking and running.

The thoracic cage, also known as the rib cage consists of the ribs, sternum, thoracic vertebrae, and costal cartilages. The thoracic cage encloses and protects the organs of the thoracic cavity including the heart and lungs. It also provides support for the shoulder girdles and upper limbs and serves as the attachment point for the diaphragm, muscles of the back, chest, neck, and shoulders. Changes

in the volume of the thorax enable breathing. The sternum, or breastbone, is a long flat bone located at the anterior of the chest. Like the skull, it is formed from many bones in the embryo, which fuse in the adult. The ribs are 12 pairs of long curved bones that attach to the thoracic vertebrae and curve toward the front of the body, forming the ribcage. Costal cartilages connect the anterior ends of most ribs to the sternum.

The appendicular skeleton is composed of the bones of the upper and lower limbs. It also includes the pectoral, or shoulder girdle, which attaches the upper limbs to the body, and the pelvic girdle, which attaches the lower limbs to the body (Figure 3.5).

The pectoral girdle bones transfer force generated by muscles acting on the upper limb to the thorax. It consists of the clavicles (or collarbones) in the anterior, and the scapulae (or shoulder blades) in the posterior.

The upper limb contains bones of the arm (shoulder to elbow), the forearm, and the hand. The humerus is the largest and longest bone of the upper limb. It forms a joint with the shoulder and with the forearm at the elbow. The forearm extends from the elbow to the wrist and consists of two bones. The hand includes the bones of the wrist, the palm, and the bones of the fingers.

The pelvic girdle attaches to the lower limbs of the axial skeleton. Since it is responsible for bearing the weight of the body and for locomotion, the pelvic girdle is securely attached to the axial skeleton by strong ligaments. It also has deep sockets with robust ligaments that securely attach to the femur. The pelvic girdle is mainly composed of two large hip bones. The hip bones join together in the anterior of the body at a joint called the pubic symphysis and with the bones of the sacrum at the posterior of the body.

The lower limb consists of the thigh, the leg, and the foot. The bones of the lower limbs are thicker and stronger than the bones of the upper limbs to support the entire weight of the body and the forces from locomotion. The femur, or thighbone, is the longest, heaviest, and strongest bone in the body. The femur and pelvis form the hip joint. At its other end, the femur, along with the shinbone and kneecap, form the knee joint.

3.4.2 Joints and Skeletal Movement

The point at which two or more bones meet is called a joint, or articulation.

Joints are responsible for movement, such as the movement of limbs, and stability, such as the stability found in the bones of the skull.

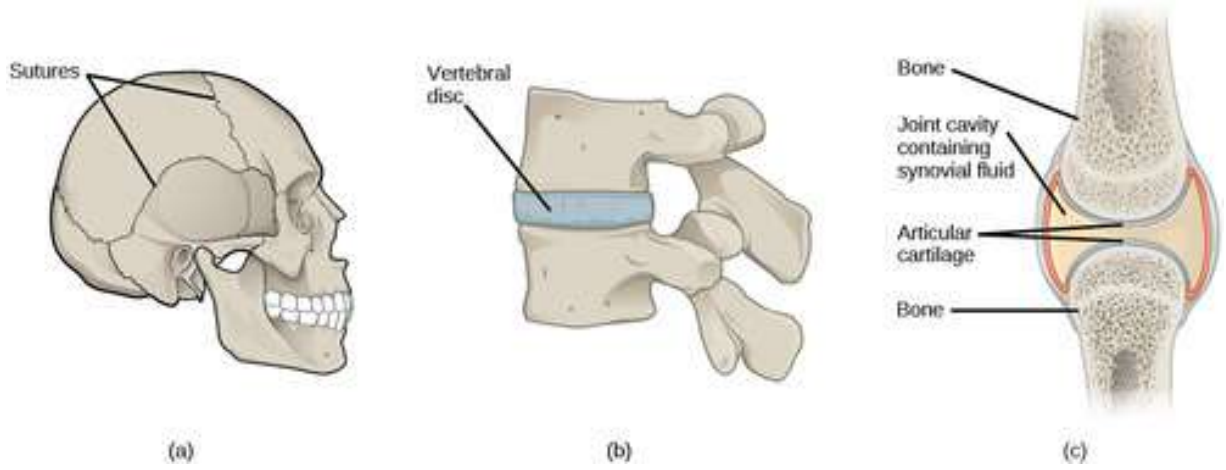

Fig. 3.6. (a) Sutures are fibrous joints found only in the skull. (b) Cartilaginous joints are bones connected by cartilage, such as between vertebrae. (c) Synovial joints are the only joints that have a space or "synovial cavity" in the joint.

There are two ways to classify joints: based on their structure or based on their function. The structural classification divides joints into fibrous, cartilaginous, and synovial joints depending on the material composing the joint and the presence or absence of a cavity in the joint. The bones of fibrous joints are held together by fibrous connective tissue. There is no cavity, or space, present between the bones, so most fibrous joints do not move at all, or are only capable of minor movements. The joints between the bones in the skull and between the teeth and the bone of their sockets are examples of fibrous joints (Figure 3.6a).

Cartilaginous joints are joints in which the bones are connected by cartilage (Figure 3.6b). An example is found at the joints between vertebrae, the so-called "disks" of the backbone. Cartilaginous joints allow for very little movement.

Synovial joints are the only joints that have a space between the adjoining bones (Figure 3.6c). This space is referred to as the joint cavity and is filled with fluid. The fluid lubricates the joint, reducing friction between the bones and allowing for greater movement. The ends of the bones are covered with cartilage and the entire joint is surrounded by a capsule. Synovial joints are capable of the greatest movement of the joint types. Knees, elbows, and shoulders are examples of synovial joints.

The wide range of movement allowed by synovial joints produces different types of movements. Angular movements are produced when the angle between the bones of a joint changes. Flexion, or bending, occurs when the angle between

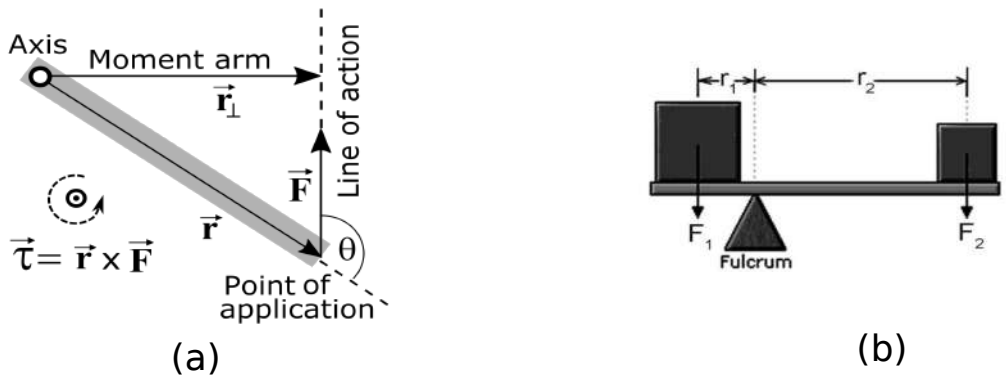

Fig. 3.7. (a) Moment of force (torque). (b) Moments of forces acting on lever (first class).

the bones decreases. Moving the forearm upward at the elbow is an example of flexion. Extension is the opposite of flexion in that the angle between the bones of a joint increases. Rotational movement is the movement of a bone as it rotates around its own longitudinal axis. Movement of the head as in saying "no" is an example of rotation.

3.4.3 Three types of levers

The lever names a solid body having a nonmotile axis of rotation, on which the forces react, aspiring to turn it around of this axis. In dependence on a relative positioning of the affixed forces and fulcrum (pivot – points of a rest) all levers are divided on three types.

Main idea at studying of levers is moment of force. Moment of force $\vec{\tau}$ is determined in relation to some point O: it is vector product of vectorial force F on radius-vector r from point O to point of force application: $\vec{\tau} = \vec{r} \times \vec{F}$ (see fig.3.7, a). Equilibrium condition for lever is equality to zero of vector sum of moments of all applied forces. For example, lever on fig.3.7 (b) is in balance if $F_1 r_1 = F_2 r_2$.

Rule of the lever: the lever is in equilibrium, if the algebraic sum of the moments of forces (product of force on its lever) is equal to zero. A corollary: magnitude of force is inversely proportional to length of a lever. The rule of the lever plays the important role at analysis of the forces leaped by muscles, and reviewing of equilibrium, both separate components a locomotorium, and all system as a whole.

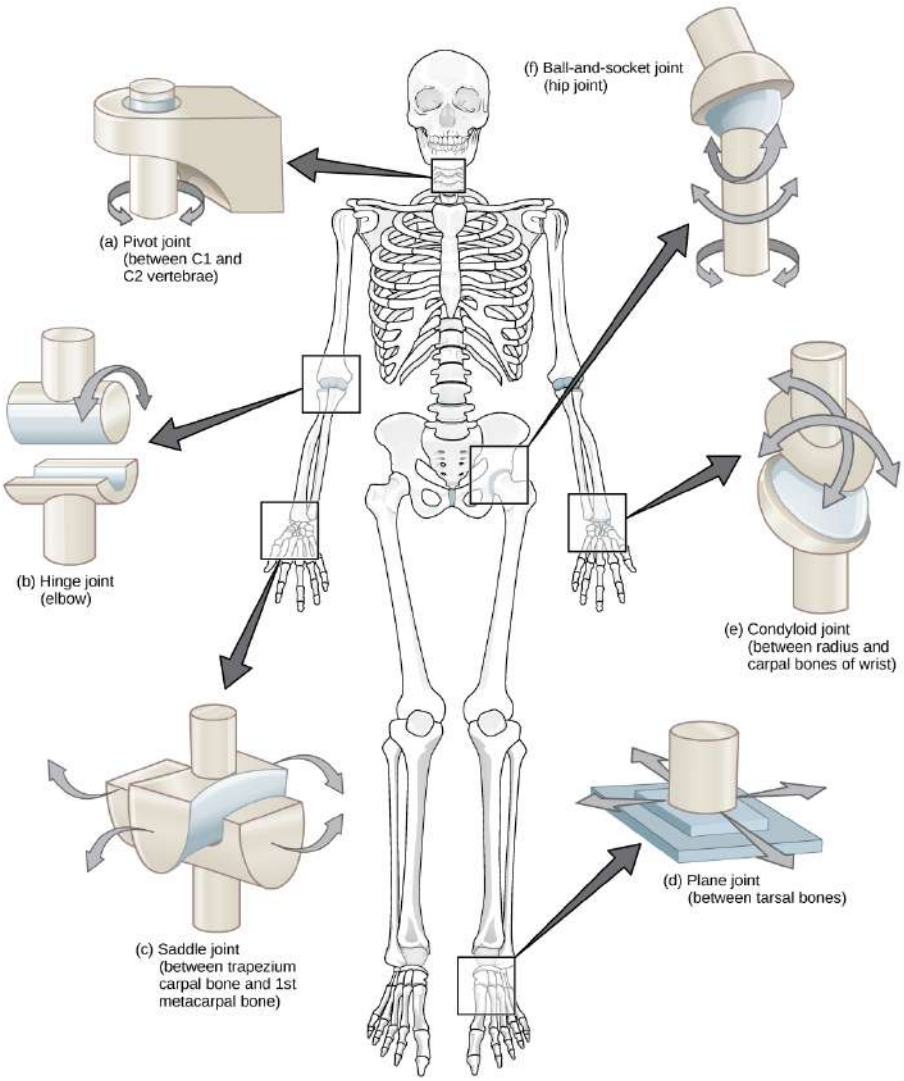

Fig. 3.8. Kinds of joints in arm kinematics' chain.

Levers in the organism. From the mechanical point of view an skeleton is a lever-swivel system, that is kept in equilibrium and reduced in locomotion by skeletal muscles. This schemes illustrate why wrong method of movement execution can carry to trauma as result of to high loading on the bones and joints, even if load is relatively slight, but distribution of loading is such that long lever is formed.

3.4.4 Muscles

Muscles allow for movement such as walking, and they also facilitate bodily processes such as respiration and digestion. The body contains three types of muscle tissue: skeletal muscle, cardiac muscle, and smooth muscle (Figure 3.10).

Skeletal muscle tissue forms skeletal muscles, which attach to bones and sometimes the skin and control locomotion and any other movement that can be consciously controlled. Because it can be controlled intentionally, skeletal muscle is

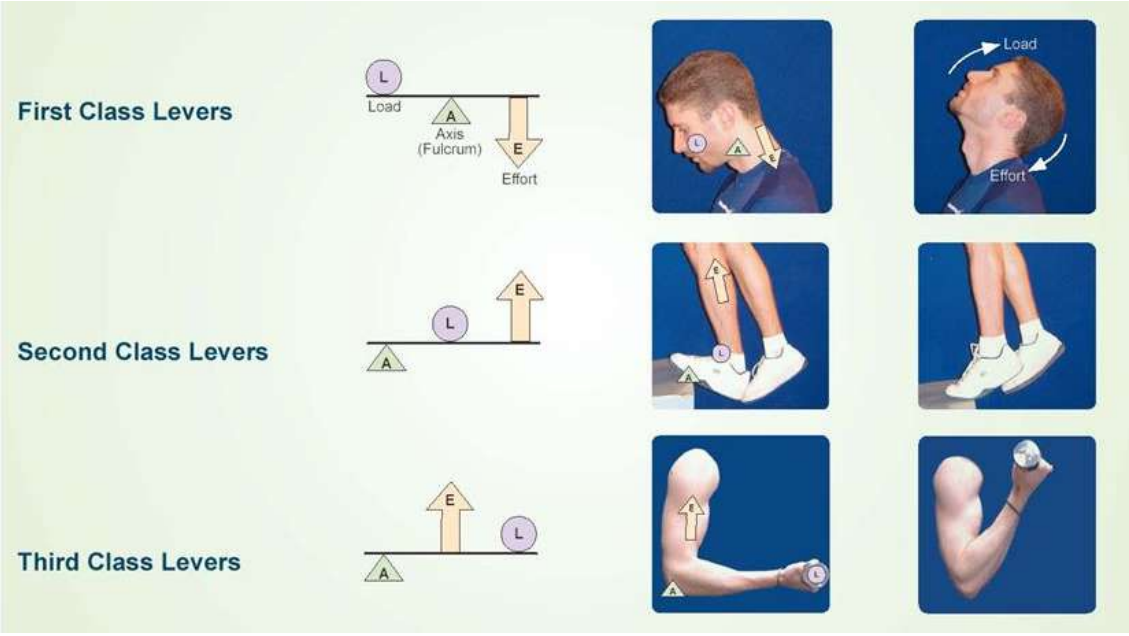

Fig. 3.9. Classes of levers.

also called voluntary muscle. When viewed under a microscope, skeletal muscle tissue has a striped or striated appearance. This appearance results from the arrangement of the proteins inside the cell that are responsible for contraction. The cells of skeletal muscle are long and tapered and have multiple nuclei on the periphery of each cell.

Smooth muscle tissue occurs in the walls of hollow organs such as the intestines, stomach, and urinary bladder, and around passages such as in the respiratory tract and blood vessels. Smooth muscle has no striations, is not under voluntary control, and is called involuntary muscle. Smooth muscle cells have a single nucleus.

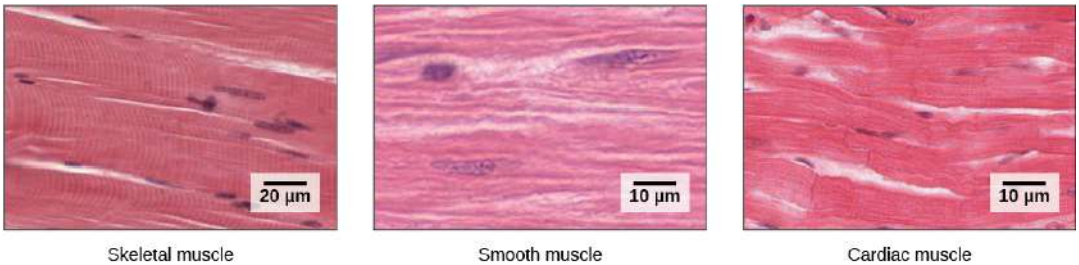

Fig. 3.10. The body contains three types of muscle tissue: skeletal muscle, smooth muscle, and cardiac muscle. Notice that skeletal muscle cells are long and cylindrical, they have multiple nuclei, and the small, dark nuclei are pushed to the periphery of the cell. Smooth muscle cells are short, tapered at each end, and have only one nucleus each. Cardiac muscle cells are also cylindrical, but short. The cytoplasm may branch, and they have one or two nuclei in the center of the cell.

Cardiac muscle tissue is only found in the heart. The contractions of cardiac muscle tissue pump blood throughout the body and maintain blood pressure. Like skeletal muscle, cardiac muscle is striated, but unlike skeletal muscle, cardiac muscle cannot be consciously controlled and is called involuntary muscle. The cells of cardiac muscle tissue are connected to each other through intercalated disks and usually have just one nucleus per cell.

Skeletal Muscle Fiber Structure and Function

On fig.3.11. different levels of skeletal muscle organization are represented. Further features of a microstructure of muscles, providing them contractive activity, re analyzed and the physicochemical processes descending during it process are considered.

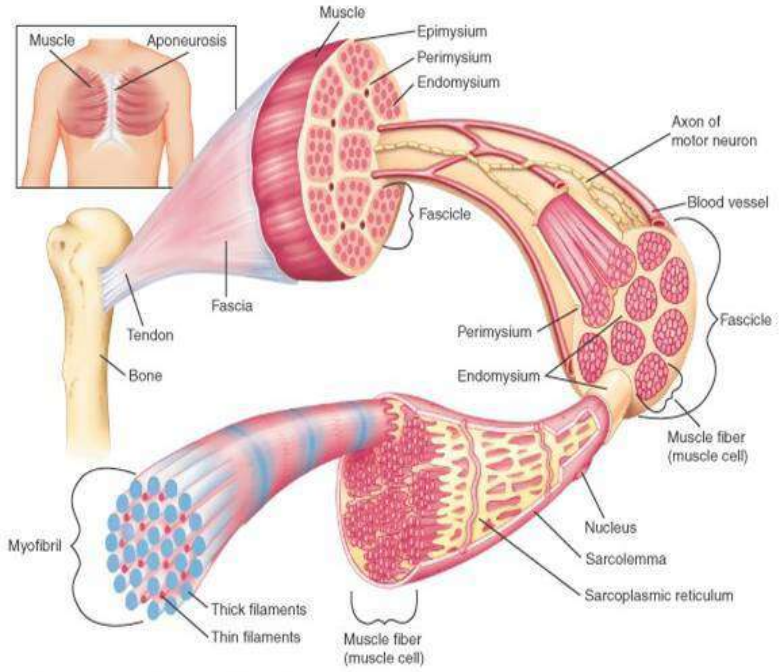

Fig. 3.11. A skeletal muscle structure.

Each skeletal muscle fiber is a skeletal muscle cell. Within each muscle fiber are myofibrils, long cylindrical structures that lie parallel to the muscle fiber. Myofibrils run the entire length of the muscle fiber. They attach to the plasma membrane, called the sarcolemma, at their ends, so that as myofibrils shorten, the entire muscle cell contracts (Figure 3.12).

The striated appearance of skeletal muscle tissue is a result of repeating bands of the proteins actin and myosin that occur along the length of myofibrils.

Myofibrils are composed of smaller structures called myofilaments. There are two main types of myofilaments: thick filaments and thin filaments. Thick filaments are composed of the protein myosin. The primary component of thin

filaments is the protein actin.

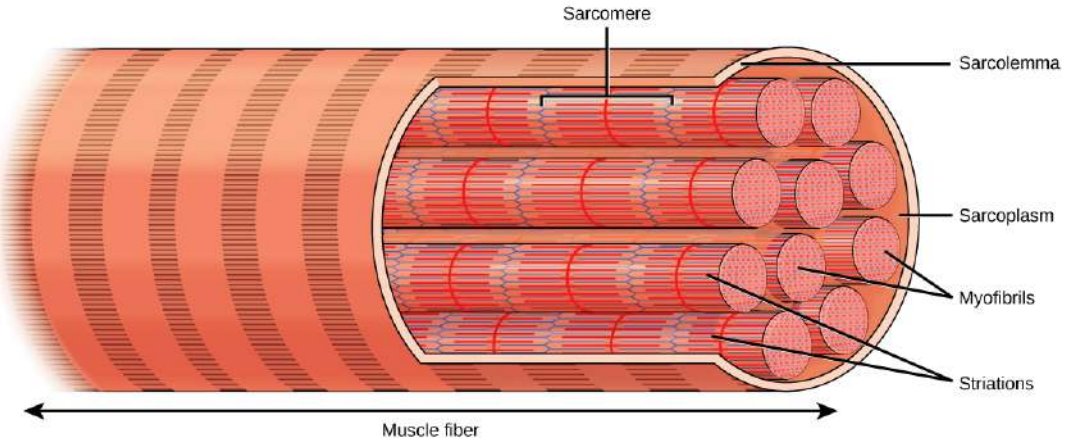

Fig. 3.12. A skeletal muscle fiber is surrounded by a plasma membrane called the sarcolemma, with a cytoplasm called the sarcoplasm. A muscle fiber is composed of many fibrils packaged into orderly units. The orderly arrangement of the proteins in each unit, shown as red and blue lines, gives the cell its striated appearance.

The thick and thin filaments alternate with each other in a structure called a sarcomere. The sarcomere is the unit of contraction in a muscle cell. Contraction is stimulated by an electrochemical signal from a nerve cell associated with the muscle fiber. For a muscle cell to contract, the sarcomere must shorten. However, thick and thin filaments do not shorten. Instead, they slide by one another, causing the sarcomere to shorten while the filaments remain the same length. The sliding is accomplished when a molecular extension of myosin, called the myosin head, temporarily binds to an actin filament next to it and through a change in conformation, bends, dragging the two filaments in opposite directions. The myosin head then releases its actin filament, relaxes, and then repeats the process, dragging the two filaments further along each other. The combined activity of many binding sites and repeated movements within the sarcomere causes it to contract. The coordinated contractions of many sarcomeres in a myofibril leads to contraction of the entire muscle cell and ultimately the muscle itself. The movement of the myosin head requires ATP, which provides the energy for the contraction.

Test questions

1. What is called the moment of force (torque) relative to a fixed point ki? How is the direction of the moment of force determined?

2. How is rotational work determined?

3. What is the moment of inertia of a material point? solid state? What is the role of the moment of inertia in rotational motion and how to calculate it?

4. Formulate and explain The Parallel-Axis Theorem.

5. What is the angular momentum of a material point? Solid State?

6. Derive and formulate the equation of the dynamics of the rotational movement.

7. What is the physical essence of the law of conservation of angular momentum?

8. What is called angular velocity? angular acceleration? How are their directions determined?

9. What is the relationship between linear and angular quantities?

10. Explain the basic elements of musculoskeletal mechanics apparatus.

Lecture No 4

Mechanical oscillations

Periodicity is peculiar to many processes taking place in biological systems. It may be observed in functional activity of the heart, lungs, stomach. Some processes in live organisms may be considered as oscillatory: oscillation of vascular walls at propagation of pulse waves, oscillation of blood pressure in vessels, air volume in lungs, oscillation of tympanic membranes, vocal ligaments, biopotential values in different points of human body.

4.1. Free Undamped Oscillations

Simple Harmonic Motion. Free undamped oscillations are those that appear in a system due to the action of interval forces only as a result of some initial deviation of the system from the state of stable equilibrium. Let's consider a spring pendulum. When a material point of mass m shifts on distance x relative to equilibrium position it starts to be influenced by elastic force resulting from spring deformation. This force is directed oppositely to the displacement. Write Hooke's law as:

$$\boxed{F_{el} = -kx} \quad (\textbf{Hooke's law}), \tag{4.1}$$

where k is coefficient of elasticity (or stiffness).

In accordance with Newton's law this force will accelerate the material point:

$$\sum_{i=1}^{n} \vec{F_i} = m\vec{a}. \tag{4.2}$$

where a is the acceleration, $a = x''$ with respect to time.

After comparing the right parts of equalities (4.1) and (4.2) we will receive:

$$-kx = mx'',$$

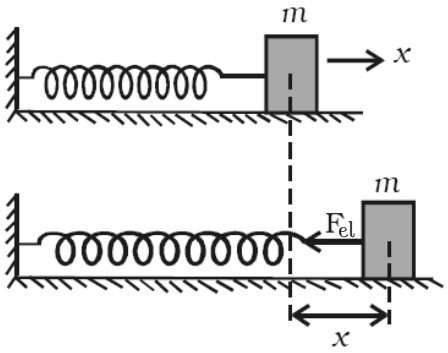

Fig. 4.1. Spring pendulum.

$$mx'' + kx = 0. \tag{4.3}$$

Divide equation (4.3) by mass m and the ratio $\frac{k}{m}$ may be denoted by square of some value ω_0: $\frac{k}{m} = \omega_0^2$ Then equation (4.3) will look like:

$$\boxed{x'' + \omega_0^2 x = 0}. \tag{4.4}$$

Equation (4.4) – **differential equation of undamped oscillations**.

The solution for equation (4.4) will look like:

$$\boxed{x(t) = A \cdot \cos(\omega_0 t + \varphi_0)} \quad \text{or} \quad \boxed{x(t) = A \cdot \sin(\omega_0 t + \varphi_0)}.$$

Table 4.1

Characteristics of undamped harmonic oscillations

x is a displacement (or shift, bias);	ω_0 is the cyclic (or circular, angular) frequency;
A is the amplitude;	a is the acceleration;
T is the period;	φ is the phase;
f is the linear frequency;	φ_0 is the initial phase;
v is the velocity (or speed);	E is the energy.

The formulae:

1. The formula of frequency is $\boxed{f = \frac{N}{t}}$

N is the number of complete (or full) oscillations,

t is time of complete oscillations (N).

2. The formula of period is $\boxed{T = \frac{t}{N}}$, $\boxed{T = \frac{1}{f}}$.

3. The formula of cyclic (or circular) frequency is $\boxed{\omega_0 = 2\pi f}$, $\boxed{\omega_0 = \frac{2\pi}{T}}$

4. The formula of velocity is

$$\boxed{v = x^{'} = -A\omega_0 \cdot \sin(\omega_0 t + \varphi_0)} \quad \text{or} \quad \boxed{v = A\omega_0 \cdot \cos(\omega_0 t + \varphi_0)}.$$

5. The formula of acceleration is

$$\boxed{a = x^{''} = -A\omega_0^2 \cdot \cos(\omega_0 t + \varphi_0)} \quad \text{or} \quad \boxed{a = -A\omega_0^2 \cdot \sin(\omega_0 t + \varphi_0)}.$$

$\varphi = \omega_0 t + \varphi_0$ is the phase undamped harmonic oscillations.

6. The formula of the total (or full) energy of harmonic oscillations of oscillator is the sum of its kinetic energy and potential energy:

$$E = \frac{mv^2}{2} + \frac{kx^2}{2} = \frac{mA^2\omega_0^2}{2}\cos^2(\omega_0 t + \varphi_0) + \frac{mA^2\omega_0^2}{2}\sin^2(\omega_0 t + \varphi_0) =$$

$$= \frac{mA^2\omega_0^2}{2}(\cos^2 x + \sin^2 x) = \frac{mA^2\omega_0^2}{2} = \frac{kA^2}{2}$$

where $k = m\omega_0^2$.

$$\boxed{E_{total} = \frac{mA^2\omega_0^2}{2} = \frac{kA^2}{2}}$$

4.2. Free Damped Harmonic Oscillations

Assume that the friction or resistance exists in the considered systems where the force of friction (resistance) is in proportion to speed: $F_f = -rv$, r – coefficient of friction (resistance).

$v = x^{'}$ – velocity.

$$F_f = -rx^{'}.$$

Write in this case the equation of motion (2^{nd} Newton's law):

$$\sum_{i=1}^{n} \vec{F_i} = m \cdot \vec{a} \qquad \vec{F}_{el} + \vec{F}_f = m\vec{a}$$

$$-kx - rx' = mx''$$

$$mx'' + rx' + kx = 0. \tag{4.5}$$

Denoting $\frac{k}{m} = \omega_0^2$, $\frac{r}{m} = 2\beta$ we will receive **differential equation of damped harmonic oscillations (4.6):**

$$\boxed{x'' + 2\beta x' + \omega_0^2 \cdot x = 0}, \tag{4.6}$$

β – damping coefficient.

The solution for equation (4.6) will look like:

$$\boxed{x = A_0 e^{-\beta t} \sin(\omega t + \varphi_0)}, \tag{4.7}$$

where A_0 is the oscillation amplitude at the initial moment of time. $\boxed{\omega = \sqrt{\omega_0^2 - \beta^2}}$ is the formula of the cyclic frequency of damped oscillations.

The oscillation amplitude decreases to the exponential law:

$$\boxed{A_t = f(t)}, \qquad \boxed{A_t = A_0 e^{-\beta t}}$$

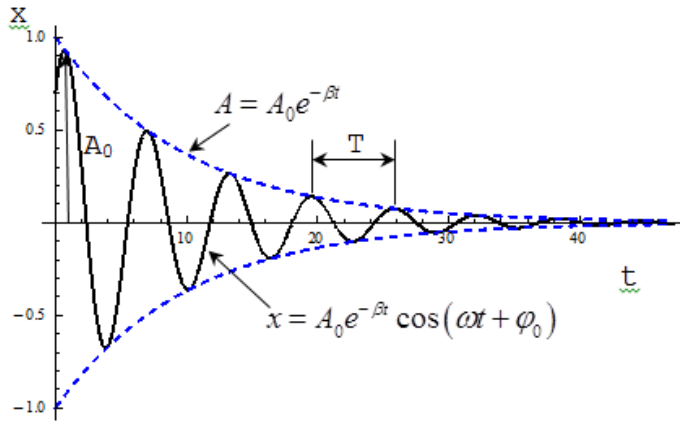

Fig. 4.2. The graph of damped oscillations.

Characteristics of damped oscillations

1) **period** of damped oscillations is $\boxed{T = \frac{2\pi}{\sqrt{\omega_0^2 - \beta^2}}}$;

2) **cyclic frequency** of damped oscillations is $\boxed{\omega = \sqrt{\omega_0^2 - \beta^2}}$;

3) **logarithmic decrement** of damping is

$$\boxed{\delta = \ln \frac{A_t}{A_{t+T}}} \quad \text{or} \quad \boxed{\delta = \beta T}$$

$$\delta = \ln \frac{A_t}{A_{t+T}} = \ln \frac{A_0 e^{-\beta t}}{A_0 e^{-\beta(t+T)}} = \ln \frac{A_0 e^{-\beta t}}{A_0 e^{-\beta t} e^{-\beta T}} = \beta T \ln e = \beta T$$

4.3. Forced Oscillations

Let us assume that a material point with mass m in addition to elastic force and force of frictions is influenced by external driving force which changes in accordance with periodical law:

$$F_e = F_0 \cdot \sin \omega t$$

where F_0 is the amplitude, ω is cyclic frequency of driving force. In this case equation of motion will look like:

$$\sum_{i=1}^{n} \vec{F}_i = m\vec{a}$$

$$\vec{F}_{el} + \vec{F}_f + \vec{F}_e = m\vec{a}$$

$$mx'' = -kx - rx' + F_0 \sin \omega t$$

Denoting $\frac{k}{m} = \omega_0^2$; $\frac{r}{m} = 2\beta$; $\frac{F_0}{m} = f_0$ we will receive **differential equation of forced oscillations (4.8):**

$$\boxed{x'' + 2\beta x' + \omega_0^2 x = f_0 \cdot \sin(\omega t)}. \tag{4.8}$$

The general solution for differential equation will be

$$\boxed{x = A \cdot \cos(\omega t - \alpha)},$$

where A is the amplitude of forced oscillations equal to

$$A = \frac{f_0}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\beta^2\omega^2}}, \tag{4.9}$$

$$\alpha = \arctan \frac{2\omega\beta}{\omega_0^2 - \omega^2}.$$

4.4. Resonance

The phenomenon of a sharp increase (reaching maximum amplitude) of forced oscillations of a system when the cyclic frequency of the driving force approaches the value ω_{res} res is called **resonance**.

Resonance phenomenon is observed at the frequency $\omega_0 = \omega_{res}$ where the amplitude of forced oscillations A reaches maximum value.

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

The maximum value A_{max} of the amplitude corresponds to the resonance cyclic frequency:

$$A_{res} = \frac{f_0}{2\beta\sqrt{\omega_0^2 - \beta^2}}$$

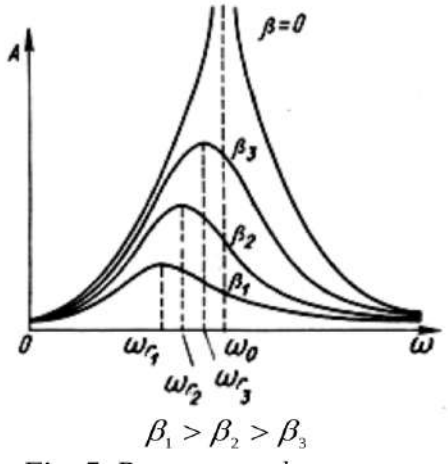

Fig. 4.3. Resonance phenomenon $\beta_1 > \beta_2 > \beta_3$.

Infrasonic vibrations of great amplitude may be hazardous for human being as some processes in human organism take place within the slot of infrasonic frequencies. For example, α-rhythms of the brain have frequency slot $9 - 13\ Hz$

and, therefore, action of infrasonic waves may cause harmful resonance in human brain.

At the nuclear level, the process called nuclear magnetic resonance (NMR) uses resonant behavior of photons to probe the structure of matter and to produce image used in medical diagnosis. Magnetic resonance imaging of a human brain involves the resonant absorption of energy by protons in the brain tissue.

4.5. Autooscillations

Systems, which automatically regulate supply of energy from external source, are called autooscillating, and periodical processes which occur in them, are called autooscillations. Amplitude and frequency of autooscillations depend on the properties of the system. The scheme of autooscillating system consisting of four compulsory elements is represented in figure.

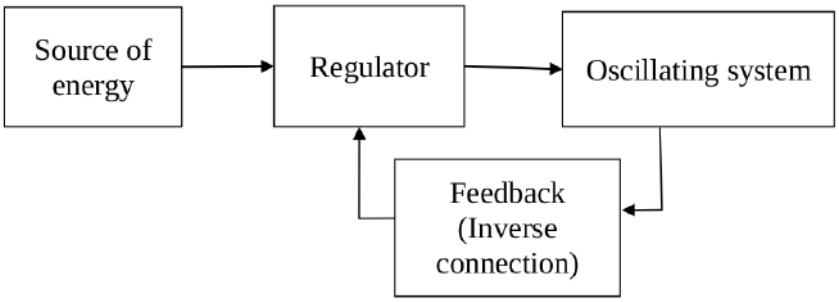

Fig. 4.4. Autooscillating system.

Examples of the autoscillatory systems are following:

1. *Clock* (pendulum – oscillating system, lifted balance weight or spring is the source of energy, anchor-regulator of energy entry from the source into oscillating system connected with oscillating system by feedback).

2. *Generator of the electromagnetic oscillations.*

3. *Heart, lungs are biological autooscillatory systems.*

Autooscillations may be of different forms: it may be oscillations close to harmonic (pendulum clock, oscillations in LC-generators), or pulse oscillations of different forms – rectangular, exponential, saw-like.

Vibration therapy (VT) has been proposed as an option to improve physical performance and reduce the negative effects of ageing on bone, muscles and tendons. Several discrepancies exist on the type of applications, frequency and

magnitude. These differences reflex on the contradictory clinical results in literature.

Test questions

1. What are called vibrations? Give examples of mechanical and non-mechanical vibrations.

2. Define frequency, period, circular frequency and phase fluctuations and indicate the relationship between them.

3. What are free (own) oscillations? Under what conditions are these oscillations undamped? Get the differential equation of free oscillations. Explain the physical meaning of its solutions.

4. Write down the equation of damped oscillations and explain its solution. What characterizes the logarithmic damping decrement?

5. What is resonance? What are the conditions for resonance? Give examples of the phenomenon of mechanical resonance.

6. Give examples of vibrations in biological systems and living organisms.

Lecture No 5

Mechanical waves

The process of propagation of oscillatory movement in an elastic medium is called **mechanical wave**.

5.1. The types of mechanical waves

There are two types of wave: *transverse and longitudinal.*

1. Transverse waves

*A **transverse wave** is a travelling wave in which the particles of the disturbed medium move perpendicular to the direction of wave motion.*

Transverse waves are like waves along a string or a rope. In the fig.5.1 the energy is transferring from left to right, but the particles of the rope are only moving up or down. In transverse waves medium's particles perform oscillations in direction perpendicular to the direction of propagation.

Transverse waves can be produced in solids and on a surface of liquids. In the transverse wave all the particles of medium execute simple harmonic oscillations about their mean positions. The position of maximal displacement in the upward direction is called the "crest" and the position of maximum displacement in downward direction is called the "trough". The distance between two successive crests or troughs is called the wavelength of the transverse wave.

Fig. 5.1. The transverse wave.

2. Longitudinal (or compressional) waves

*A **longitudinal wave** is a travelling wave in which the particles of the disturbed medium move parallel to the direction of wave motion.*

In this type of waves particles are oscillating along the direction of wave propagation. Longitudinal waves can propagate through solids, fluids and gases. In the longitudinal wave all the particles of a medium execute also simple harmonic oscillations about their mean positions. Longitudinal waves are alternation of the regions of rarefaction and compression of the medium where these waves propagate. The distance between two successive compressions or two successive rarefactions is called wavelength of the longitudinal wave.

A spring like a "slinky" can be used to demonstrate a longitudinal wave. Here the particles oscillate about fixed points from left to right, and the energy is transferred from left to right. Sound waves are longitudinal waves (Fig. 5.2).

Fig. 5.2. The longitudinal wave.

5.2. The characteristics of mechanical waves

Amplitude (A) is the absolute value (module) of the maximum displacement of this quantity from its equilibrium value. Unit of amplitude in SI system is meter (m).

Period (T) is the time required for one full oscillation. Unit of period in SI system is second (s).

$$\boxed{T = \frac{t}{n}}$$

Frequency (f) *(also frequency denoted as ν)* is the number of full oscillations by particles of medium per unit time. There is an inverse relation between a wave's frequency and its period. Unit of frequency in SI system is the hertz ($Hz = 1/s = s^{-1}$).

$$\boxed{f = \frac{n}{t}}$$

Phase of oscillating particle denotes its position and direction of motion at any moment of time. Unit of phase in SI system is the radian (rad).

$$\varphi = 2\pi\Big(\frac{t}{T} - \frac{r}{\lambda}\Big)$$

Velocity of a wave (v) is the speed which a wave front travels in a medium. Unit of speed in SI system is meter per second (m/s).

Wave front is surface, which the wave has reached in some time point. In the case wave front is a plane $x = const$, that is why the wave is called plane. Wave surface form is determined by configuration of oscillation source and medium properties. In isotropic medium a spherical wave, which has sphere as surface, propagated from the point source.

Wavelength (λ) is the shortest distance between the points of a wave whose oscillation phases differ by 2π (or the distance which wave passes per period of oscillations). Unit of wavelength is meter (m).

$$\lambda = v \cdot T = \frac{v}{\nu}.$$

Energy flow is the quantity of energy transferred by the wave through surface S per unit time:

$$\boxed{\Phi = \frac{\Delta E}{\Delta t}}, \qquad [\Phi] = \frac{J}{s} = W.$$

Wave intensity is the energy flow transferred by the wave through a unit area of the surface perpendicular to the of wave propagation:

$$\boxed{I = \frac{\Phi}{S} = \frac{\Delta E}{\Delta t \cdot \Delta S}}, \qquad S \text{ is area,}$$

$$[I] = \frac{W}{m^2} \quad \text{(watts per square meter)}$$

Volume density of energy:

$$\boxed{\varepsilon = \frac{\Delta E}{\Delta V}}; \qquad [\varepsilon = \frac{J}{m^3}].$$

5.3. Wave equation

The displacement of particles caused by the plane wave can be described by the general equation of wave motion: $S = f(x;t)$.

Assume that wave process is propagated in positive direction of axe OX and the source of oscillations is in the plane perpendicular to the propagation direction, and oscillates according to the law (Fig. 5.3):

$$S = A\cos\omega t$$

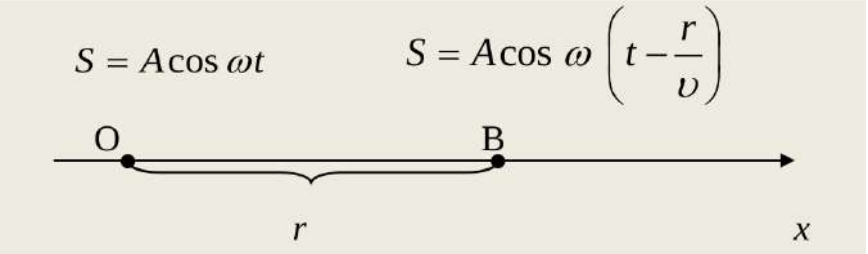

Fig. 5.3. The displacement.

Let v be the speed of propagation of wave process in medium. In the time interval $\tau = \dfrac{r}{v}$ wave process will reach point B, which is at the distance r from the source of oscillations, and will cause oscillations of this point in time interval τ according to the law:

$$S = A\cos\omega\Big(t - \frac{r}{v}\Big).$$

Considering the correlation between λ, v, T, equation may represented in the form:

$$S = A\cos\frac{2\pi}{T}\Big(t - \frac{r}{v}\Big), \qquad \omega = \frac{2\pi}{T}$$

$$S = A\cos 2\pi\Big(\frac{t}{T} - \frac{r}{vT}\Big), \qquad \lambda = vT$$

Wave equation: $\boxed{S = A\cos 2\pi\Big(\dfrac{t}{T} - \dfrac{r}{\lambda}\Big)}$

5.4. Umov vector

1. $I = \dfrac{\Delta E}{\Delta t \cdot S} \cdot \dfrac{v}{v} = \dfrac{\Delta E \cdot v}{V} = \varepsilon \cdot v.$

Wave intensity I is vector quantity, since speed v is vector, it is called

Umov vector: $\boxed{\vec{I} = \varepsilon \cdot \vec{v}}$

Umov vector is numerically equal to density of energy flow and coincides with the direction of the speed vector of wave propagation.

2. Represe Umov vector in other form. The full mechanical energy of harmonic oscillations of one particle is equal to:

$$E = \frac{mv^2}{2} + \frac{kx^2}{2} = \frac{mA^2\omega^2}{2}.$$

$$m = \rho V, \qquad E = \frac{\rho V A^2\omega^2}{2}$$

$$I = \frac{\Delta E}{\delta t \cdot S}, \qquad I = \frac{\rho V A^2\omega^2}{2\Delta t S},$$

$$V = l \cdot S, \qquad I = \frac{\rho l S A^2\omega^2}{2\Delta t S} = \frac{\rho l A^2\omega^2}{2\Delta t}$$

$$v = \frac{l}{\Delta t}, \qquad \boxed{\vec{I} = \frac{\rho A^2\omega^2}{2} \cdot \vec{v}},$$

where ρ is density of medium.

5.5. The Doppler effect

The Doppler effect is the change in frequency and wavelength of a wave when the observer and the source of the wave move relative to each other.

Let us consider a source of waves with a constant frequency and amplitude.

1) The observer and the source are fixed:

$$\lambda = vT = \frac{v}{f}, \qquad f = \frac{v}{\lambda}$$

2) The observer moves to the source. The source is fixed:

$$f^{'} = \frac{v + v_0}{\lambda} f$$

$f^{'} > f$ – the observer moves towards the source;

$f^{'} < f$ – the observer moves away from the source.

3) If the source of waves follows a wave, that the wavelength is smaller.

$$\lambda^{'} = \lambda - v_s T = \frac{v}{f} - \frac{v_s}{f} = \frac{v - v_s}{f}$$

$$f'' = \frac{v}{\lambda'} = \frac{v}{v - v_s} f$$

$f'' > f$ – the source moves towards the observer;

$f'' < f$ – the source moves away from the observer.

4) The general equation for the Doppler effect is

$$\boxed{f''' = \frac{v \pm v_0}{v \mp v_s} f.}$$

Here f is the frequency of waves emitted by the source and f''' is the frequency of waves perceived by the observer. v is the velocity of waves in the immovable medium, v_0 and v_s are the speeds of the observer and the source of waves. The "upper" symbols (+ and −) concern the case when relative motion between the source and observer is moving them to meet. The "down" symbols concern the case when they are moving away from each other. The change of frequency of the waves due to Doppler effect is named the Doppler shift. This phenomenon is used to measure the speed of various moving bodies including the erythrocytes in the blood vessels.

Test questions

1. How to explain the propagation of vibrations in an elastic medium? What is a wave?

2. What is called a transverse wave? longitudinal? When do they arise?

3. What is called wavelength? What is the relationship between wavelength, speed and period?

4. What are the main energy characteristics of the wave?

5. What is the Doppler effect? How to determine the frequency of sound perceived by the receiver if the sound source and receiver are moving?

6. What are the main diagnostic methods in medicine based on the Doppler effect?

Lecture No 6

Acoustics. Sound waves.

6.1. The nature of sound

Acoustics is the field of physics. The subject matter of acoustics is the study of physical nature of sound, mechanisms of its generation, propagation and practical application.

Sound is the mechanical waves with frequency in the interval from 16 Hz to 20 kHz, which propagate in elastic medium. The particles of a medium do not travel forward. They only oscillate about their mean positions in the direction of wave propagation. But the energy of a sound is transmitted from one particle of medium to another. The sound is transferred by this sequence of compression and rarefactions.

The wavelength (λ) is the average distance from one compression to the next, or from one rarefaction to the next.

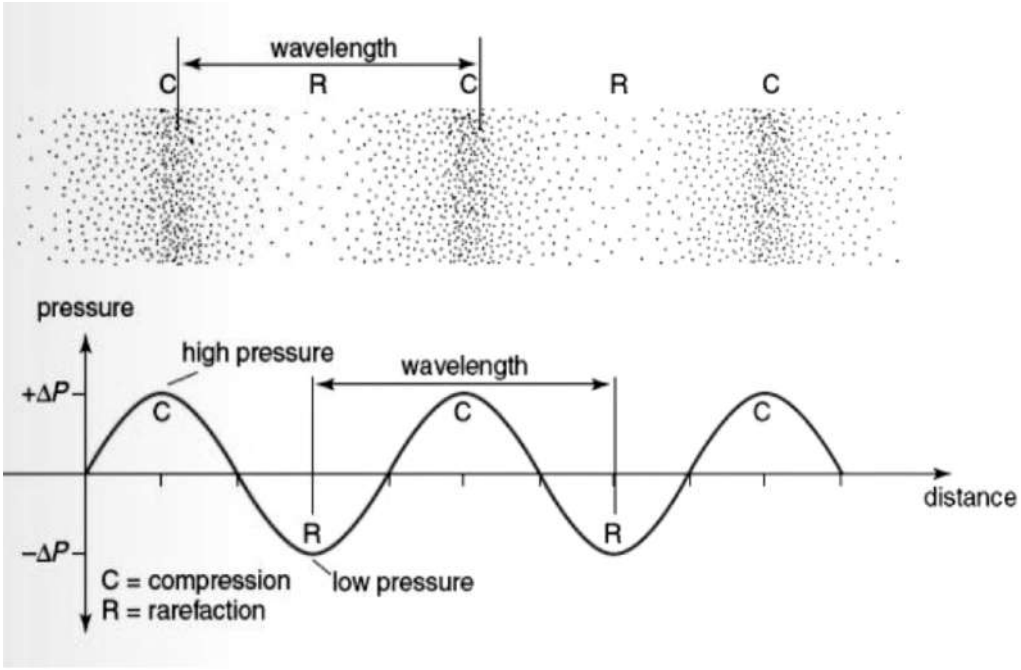

Fig. 6.1. Sound wave

Sound wave is an oscillatory motion of the particles that travels through an elastic medium as a wave. Sound waves lie within the range of sensitivity of the human ear (these waves also are called audible waves) and have a frequency ranging between approximately 16 Hz to 20000 Hz.

A speaker produces a sound wave by oscillating a cone, causing vibrations of air molecules. In fig.6.1, a speaker vibrates at a constant frequency and amplitude, producing vibrations in the surrounding air molecules. As the speaker oscillates back and forth, it transfers energy to the air, mostly as thermal energy. But a small part of the speaker's energy goes into compressing and expanding the surrounding air, creating slightly higher and lower local pressures. These compressions (high-pressure regions) and rarefactions (low-pressure regions) move out as longitudinal pressure waves having the same frequency as the speaker—they are the disturbance that is a sound wave. (Sound waves in air and most fluids are longitudinal, because fluids have almost no shear strength. In solids, sound waves can be both transverse and longitudinal.)

Fig.6.2(a) shows the compressions and rarefactions, and also shows a graph of gauge pressure versus distance from a speaker. As the speaker moves in the positive x-direction, it pushes air molecules, displacing them from their equilibrium positions. As the speaker moves in the negative x-direction, the air molecules move back toward their equilibrium positions due to a restoring force. The air molecules oscillate in simple harmonic motion about their equilibrium positions, as shown in part (b). Note that sound waves in air are longitudinal, and in the figure, the wave propagates in the positive x-direction and the molecules oscillate parallel to the direction in which the wave propagates.

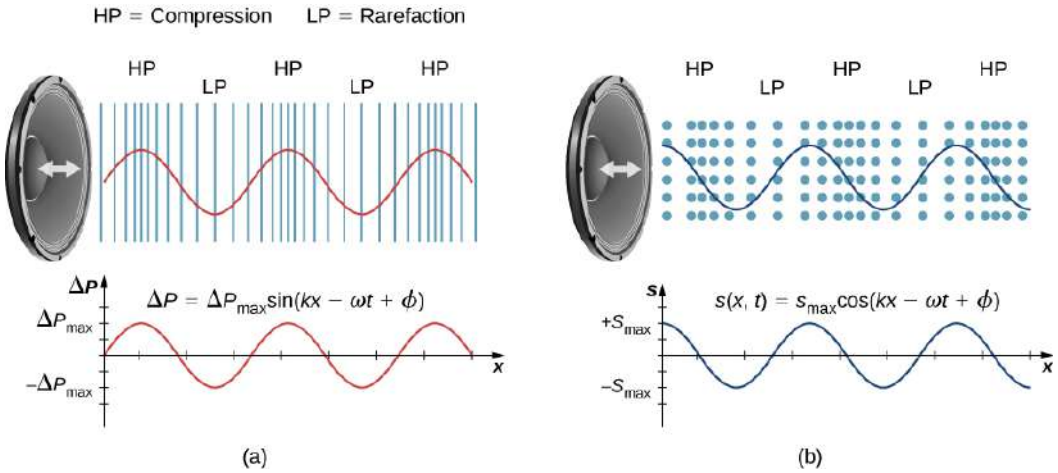

Fig. 6.2. Models Describing Sound

Sound can be modeled as a pressure wave by considering the change in pressure

from average pressure,

$$\Delta P = \Delta P_{max} \sin(kx \mp \omega t + \varphi)$$

This equation is similar to the periodic wave equations seen in waves, where ΔP is the change in pressure, ΔP_{max} is the maximum change in pressure, $k = 2\pi/\lambda$ is the wave number, $\omega = 2\pi/T = 2\pi f$ is the angular frequency, and φ is the initial phase. The wave speed can be determined from $v = \omega/k = \lambda/T$. Sound waves can also be modeled in terms of the displacement of the air molecules. The displacement of the air molecules can be modeled using a cosine function:

$$s(x,t) = s_{max} \cos(kx \mp \omega t + \varphi)$$

In this equation, s is the displacement and s_{max} is the maximum displacement.

Ultrasound waves are longitudinal waves with frequencies over 20000 Hz, which is about the upper limit of human hearing.

Infrasonic waves are longitudinal waves with frequencies below the audible range (i.e., less than approximately 16 Hz).

Table 6.1

Acoustic scale

Type of sound	Infrasound	Audible sound	Ultrasound	Hypersound
Frequency	$f < 16$ Hz	$16 < f < 2 \cdot 10^4$ Hz	$2 \cdot 10^4 < f < 10^9$ Hz	$f > 10^9$ Hz

All sounds fall into tones, or musical sounds, noises and sonic booms.

Tones are those resulting from oscillations of a source with constant amplitude and frequency.

Pure (simple) tones is harmonic oscillations of sources. Frequency is the main characteristic of a pure tone.

For example: sound of tuning fork (fig.6.3).

Complex (compound) tones is resulting from non-harmonic oscillations. A complex tone can be split into simple tones.

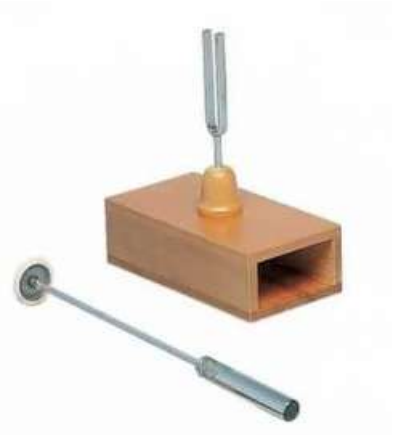

Fig. 6.3. Tuning fork

For example: human voice, sounds of musical instruments.

Complex tone may be decomposed into pure tones. A pure tone included into a complex one and which has the lowest frequency f_0 is called **fundamental** (or

main). Pure tones included into complex tone and those which have frequencies devisable by frequency of a fundamental tone ($f = nf_0$, where $n = 1, 2, ...$) are called overtones. Complex tone has a line acoustic spectrum (fig.6.4, a).

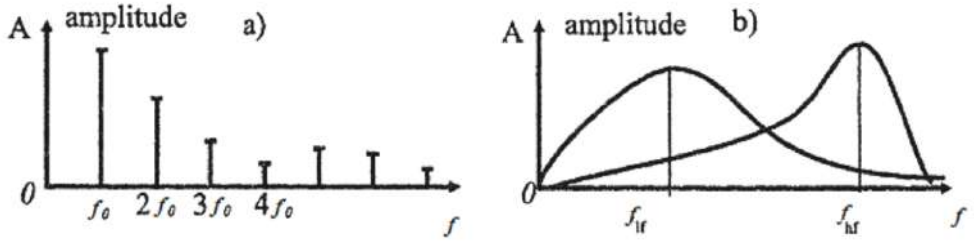

Fig. 6.4. Types of sound spectrums

Noise is a complex aperiodic sound with continuous spectrum. Noises may differ by spectrums. For example, low-frequency noises and high- frequency noises have different amplitudes in correspondent zones of spectrum (fig.6.4, b). Its amplitude and period of oscillations change irregularly.

The examples of noises are rustles, the consonants of human voice, claps etc.

6.2. The use of sound in medicine.

1. Audiometry is the method for measurement of hearing acuity. The individual curve of a patient's hearing threshold at different frequencies is determined by means of an audiometer; the respective curve is called an audiogram. Comparing a patient's audiogram with a normal one makes it possible to diagnose ear diseases. The audiometer is a sound generator with telephones, which allows independent control of output signal frequency and intensity.

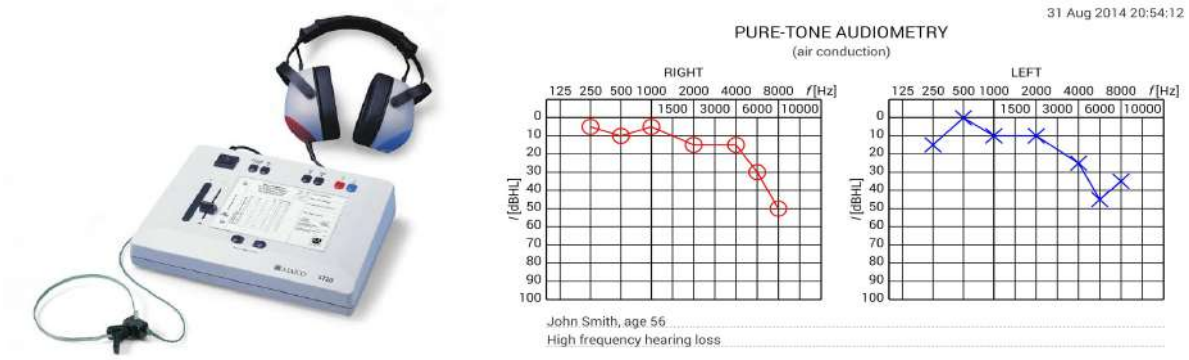

Fig. 6.5. Audiometr and audiogram.

2. Auscultation is the act of listening to sounds arising within organs (as

the lungs or heart) as an aid to diagnosis and treatment. One uses a stethoscope or phonendoscope.

Fig. 6.6. Authentic Doctor's Stethoscope.

3. Percussion is the act or technique of tapping the surface of a body part to learn the condition of the parts beneath by the resulting sound. Percussion sounds are extracted by tapping with either a hammer on a special plate, called a pleximeter, or simply by the fingers of one hand on the phalanxes of the other hand applied to the patient's body.

4. Phonocardiography is the method consists in graphical registration of heart sounds and noises with their subsequent clinical interpretation. The method of phonocardiography (PCG) is used to diagnose heart functioning. Phonocardiography is realized by means of a phonocardiograph.

6.3. Ultrasound (US), sources of US. Features of distribution of ultrasonic waves.

Ultrasound is called sound oscillations which frequency borrows the range from 20 kHz up to 10^{10} Hz. The top limit is accepted completely conditionally from such reasons, that the length of wave in substance and tissues for such frequency appears is commensurable with intermolecular distances in view of that speed of distribution of US in water and tissues is identical.

The greatest distribution, both in technics and in medical practice piezoelectric radiators of US have received. As piezoelectric radiators use crystals of quartz or the synthetic ceramic, lead zirconate titanate etc. Piezoeffect (direct) is called phenomenon of occurrence on surfaces of the mentioned crystal plates of opposite on mark charges under action of mechanical deformations. After removal of deformation charges disappear.

There is also inverse piezoeffect which has found application and in medical practice for reception of high-frequency US. If on silvered sides of surface of plate of

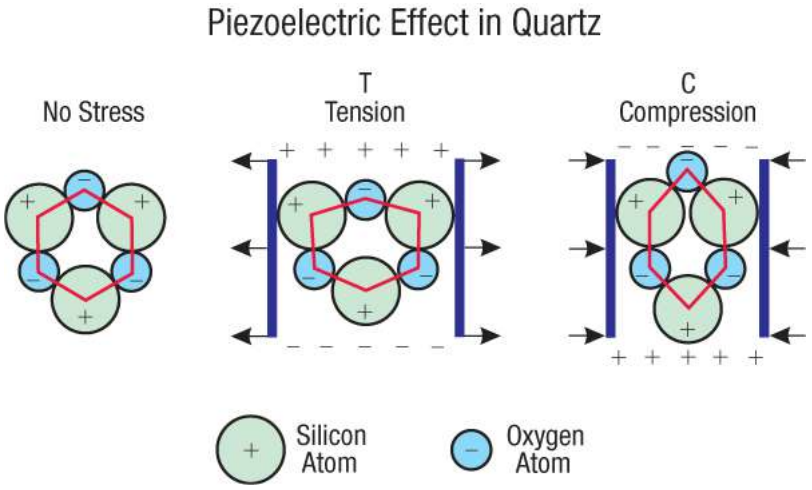

Fig. 6.7. Piezoeffect.

the piezoelement to submit alternating voltage from generator, the quartz plate will come in fluctuation in a step of alternating voltage of the generator. The amplitude of oscillations will be maximal when own frequency of quartz plate ν_0 coincides with frequency of generator ν_g, i.e. there will be resonance ($\nu_0 = \nu_g$). The detector of US can be created on the basis of direct piezoelectric effect. Thus under influence of US-waves there is a deformation of a crystal that results in occurrence of alternating voltage which can be measured or fixed on the screen of an electronic oscillograph after its preliminary amplification. A piezoelectric transducer can operate as US generator (radiator) and US detector. A typical transducer used in medical applications is shown in fig.6.8.

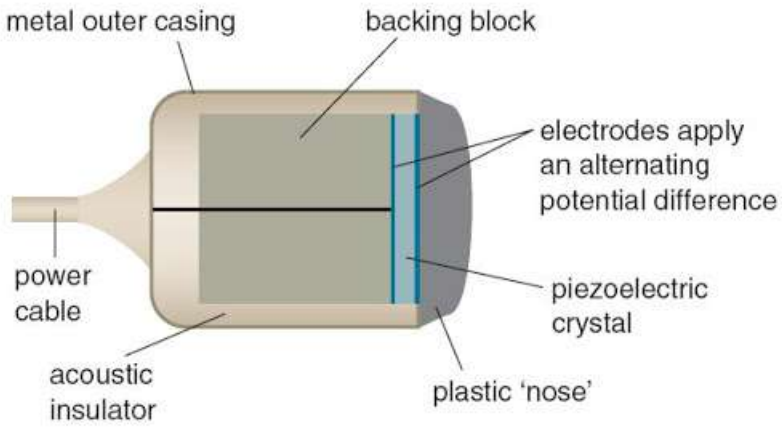

Fig. 6.8. A typical transducer used in medical applications.

The ultrasound we can get with help of the devices based on the phenomenon of magnitostrictia (for reception of low frequencies). Magnitostrictia is the phenomenon, which will consist in change of length of the ferromagnetic core placed in a high-frequency magnetic field. End faces of this core will radiate *low-frequency*

US. Except for the specified sources of US there are *mechanical sources* in which mechanical energy will be transformed to energy of US oscillations.

By the nature US as well as sound is the mechanical wave extending in the elastic medium. Speeds of propagation of sound and ultrasonic waves are approximately identical. However the length of wave of US is much less, than for sound. It allows focusing easily US oscillations. Ultrasonic wave has the much greater intensity than sound, owing to the big frequency it can achieve several Watt on square centimeter (Wt/cm^2), and at focusing it is possible to receive US with intensity of 50 Wt/cm^2 and more.

Propagation of US in the medium differs (due to small length of a wave) other feature: liquids and firm bodies represent good conductors of US, air and gas are bad conductors. So, in water US fades in 1000 times more poorly than in air. At propagation of US in the non-uniform environment arises its reflection and refraction. Reflection of US on border of two mediums depends on parity of their wave resistance. If US in the medium with wave resistance $\omega_1 = \rho_1 v_1$ falls perpendicularly on a flat surface of the second medium with $\omega_2 = \rho_2 v_2$ part of energy will pass through the boundary surface and part will be reflected. The coefficient of reflection will be equal to zero, if $\rho_1 v_1 = \rho_2 v_2$, i.e. US energy will not be reflected from border of the unit of surfaces and will pass from one medium to another lost-free. For the borders of «air - liquid», «liquid - air», «firm body - air» and on the contrary the coefficient of reflection will be equal almost to 100%. It speaks that air has very small acoustic (wave) resistance.

Therefore in all cases of connection of a radiator of US with the irradiated medium, for example, with a body of the person, it is necessary to watch that between radiator and a tissue there *was no even a minimal air layer* (wave resistance of biological mediums in 3000 times is more of wave resistance of air). To exclude air layer, a surface of US-radiator covered by a layer of oil or it is rendered by a thin layer on a surface of a body.

At propagation of US in the medium arises sound pressure which changes, accepting positive value in the field of compression and negative in the area following it. So, for example, at intensity of ultrasound of 2 Wt/cm^2 in tissues of the person pressure is created in the field of compression of $+2,6$ $atm.$, which in the following area passes in value of $-2,6$ $atm.$. Compression and low preassure, created by ultrasound result in formation of breaks of continuous liquid with formation of microscopic cavities (**cavitation**). If this process occurs in a liquid bubbles are filled by pairs of liquid or the gases dissolved of it. Then on a

place of a cavity the site of compression of substance is formed, the cavity slams quickly, allocated a significant amount of energy in small volume that results in *destruction of microstructures of substance.*

6.4. Ultrasound.

Ultrasound are mechanical waves with frequencies from 20 kHz to 10^9 Hz.

Ultrasound is generated by using such physical phenomena as **the inverse piezoelectric effect (piezoeffect) and magnetostriction**.

Piezoelectric effect was discovered by **Pierre Curie** in 1880.

The essence of the inverse piezoeffect is that piezoelectrics (for example, quartz and certain types of ceramic) are subject to strain (expand or compress) when placed in an electric field. Therefore, placing a piezoelectric crystal in an alternating electric field of ultrasound frequency makes the crystal surface oscillate with this frequency and excite ultrasound waves in the environment.

Magnetostriction is changing of dimensions (expansion or compression) of ferrites in an alternating magnetic field. Placing ferrites into an alternating magnetic field of ultrasound frequency causes oscillation of the ferrite surface with this frequency, i.e. excitation of ultrasound waves in the environment.

6.4.1. Aplication of ultrasound in medicine.

The most widespread application of ultrasound is in diagnostics. Such methods are called ultrasonographic detection (USD). USD methods are based on reflection of ultrasound waves from external and internal surfaces of different human organs. It is possible as short-wave ultrasounds cannot turn round obstacle with sizes more than wavelength, partly reflect from boundaries of mediums (soft tissues of various density, bones, body cavities) information regarding the internal organs.

$$f = 1 - 20\ MHz, \qquad I = 0.001 \div 0.05 \frac{W}{cm^2}$$

λ is wavelength of ultrasound; δ is linear dimension of the obstacles.

$$\lambda = \frac{v}{f} = \frac{1500\ m/s}{10^6\ Hz} = 1.5\ mm$$

a) If δ comparable to λ this is called the phenomenon of *diffraction.*

Diffraction is a rounding of the wave barriers.

b) If $\delta >> \lambda$ then there is the shadow of ultrasound, reflection and absorption of ultrasonic waves (ultrasound-echolocation).

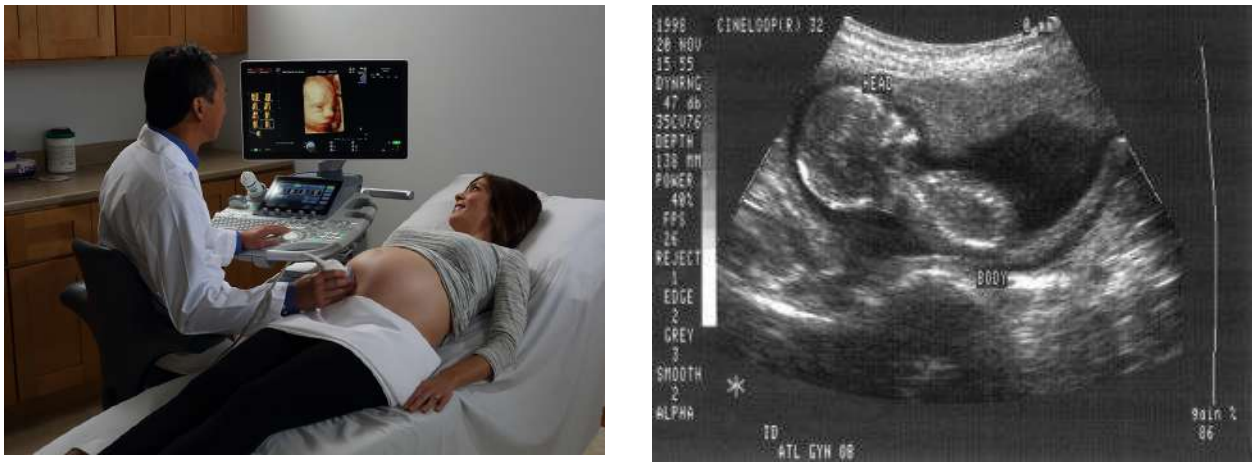

Fig. 6.9. Ultrasonographic detection.

Short ultrasound impulses produced echo-signals with different response times, which are formed at different boundaries. A computer processes the ultrasound signals received to display an image of the reflecting surfaces on a monitor.

USD is applied for detecting tumours and other pathological changes in the organs; to detect stones in the urinary system and in the gall bladder; to measure the dimensions of organs or their parts (cardiac chambers, and the renal pelvis).

c) The absorption. In passing from one medium to another ultrasound intensity varies according to the Rayleih formula: $\dfrac{I_2}{I_1} = \dfrac{4(\chi_1/\chi_2)}{(1+(\chi_1/\chi_2))^2}$; where $\chi = \rho v$ is **wave impedance**.

Wave impedance of biological media 3000 times more than air. The surface of the body covered a layer of oil.

Passing through the biological tissue ultrasound intensity decreases by the law:

$$\boxed{I = I_0 \cdot e^{\alpha d}}$$

d is thickness of the tissue; α is the monochromatic absorption coefficient.

Three types of scanning are common in the ultrasound diagnostics

1) A–scan ("A" – *amplitude*).

A-scans can be used in order to measure distances. A transducer emits an ultrasonic pulse and the time taken for the pulse to bounce off an object and come back is graphed in order to determine how far away the object is. A-scans only give one-dimensional information and therefore are not useful for imaging.

Ultrasound beam is directed along a line through some part of a body, for instance through the brain. Ultrasound investigation of brain is named echoencephalography. Ultrasound waves traverse the structures of brain and reflect from them. The mechanical waves are transformed into electric pulses of different amplitude on a display. The horizontal motion of the display beam is adjusted in time to represent both the original ultrasound pulse and its echo. The position of electric pulses on the display indicates the distance to the normal and pathological structures of brain and their location.

2) B–scan ("B" – brightness).

B-scans can be used to take an image of a cross-section through the body. The transducer is swept across the area and the time taken for pulses to return is used to determine distances, which are plotted as a series of dots on the image. B-Scans will give two-dimensional information about the cross-section.

This technique gives the possibility to observe two-dimensional picture of interior organs. It uses a movable ultrasound transducer. The electric voltage changes resulting from ultrasound pulses and their echo are added to the accelerating voltage in the electron gun of electronic tube. When the echo is present, a bright spot on the display appears.

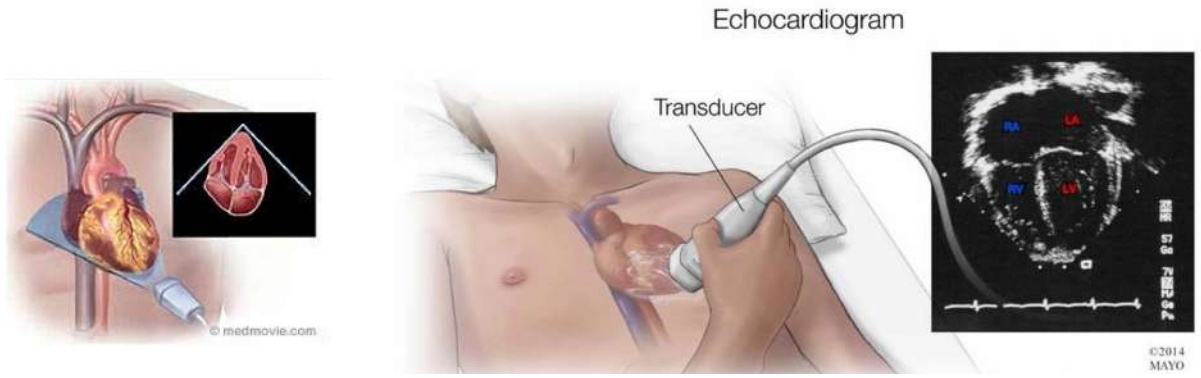

Fig. 6.10. Echocardiography.

3) M–scan ("M" – motion). This method is useful in investigation of motion, for instance the motion of the heart wall or the heart valves (echocardiography). Ultrasound wave is directed to the moving object and reflects from its surface. The horizontal motion of the display beam represents time intervals. The vertical deflection of the display beam reflects the amplitude of object movements.

4) **Ultrasound Doppler technique** (echodopplerography). The use of the Doppler Effect during USD allows determining the blood flow velocity, to examine the functioning of moving parts of organs (for example, the cardiac valve).

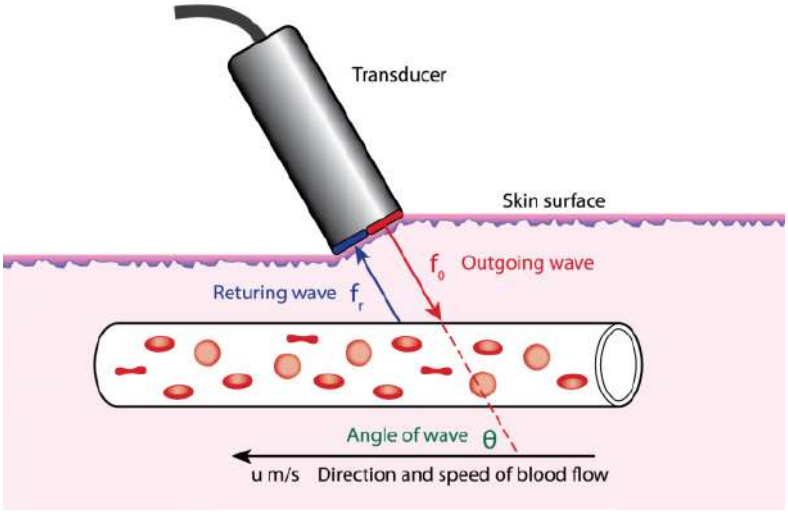

Fig. 6.11. Echodopplerography.

A high-frequency ultrasound beam is directed along the blood vessel. Ultrasound is reflected from blood cells that act as a moving source of ultrasound. They cause a Doppler shift (this is the frequency difference between emitted and reflected ultrasound that depends on the velocity of blood flow). The ultrasound receiver measures the amount of Doppler shift. Information arriving from the receiver is computed and visualized to obtain a picture of blood vessel with the indication of the blood flow rate.

Before considering the therapeutic application of ultrasound, we will indicate the **effects that appear when ultrasound waves pass through biological tissues**:

- microvibrations at cellular and subcellular levels;
- destruction of macromolecules;
- reconstruction and damage of membranes, leading to changes in their permeability;
- heat release;
- destruction of cells and microorganism;
- formation of chemically highly active ions and free radicals.

The primary mechanisms of ultrasound therapy are mechanical and heat action on the tissue. The high frequency corresponds to the high intensity of ultrasound:

$$I = \frac{\rho A^2 \omega^2}{2} v, \qquad I \sim f^2 (\omega = 2\pi f);$$

$$f = 10^4 - 10^9 \ Hz \qquad I \sim 10^8 - 10^{18} \frac{W}{cm^2}$$

In surgery, ultrasound is applied as

- an *"ultrasound scalpel"*, i.e. for dissection of both soft and bone tissues;
- for ultrasound welding of bone tissues (*ultrasound osteosynthesis*).

In nephrology, focused ultrasound is used for destruction of kidney stones (*lithotripsy*) in the urinary system and in the gall bladder, *in stomatology* one for removal of dental calculus, drilling dental channels, etc

Fig. 6.12. Lithotripsy.

The destructive action of ultrasound on microorganisms is used **for sterilization** of various medium.

Cavitation is the appearance of microvoids in material medium (for example, in fluid) under the action of pressure fluctuation. $I \geq 0.3\frac{W}{cm^2}$.

In pharmaceutics cavitation processes caused by action of ultrasonic wave of great intensity are used for dispersion of hard and fluid materials with the purpose of producing medical powders, emulsions, suspensions and aerosols as well as for homogenization of tissues when extracting biologically active substances (enzymes, toxins, vitamins and others) therefrom.

Low-intensity ultrasound is used **in stomatology, ophthalmology** and other fields of medicine for micromassage of tissue structures. In so doing, permeability of cell membranes and energization of the processes of tissue metabolism increase.

Ultrasound can produce certain physical alterations in biological tissues, which depend on its intensity level. Ultrasound of moderate intensity $I \leq 1\frac{W}{cm^2}$ is used as one of the physiotherapeutic methods of treatment.

Ultrasound waves perform "micromassage" of tissues and slightly heat them activating the blood circulation and metabolism.

Ultrasound increases the permeability of biological membranes. It gives the possibility to introduce some medicines through the intact skin (**phonophoresis**).

6.5. Infrasound.

Acoustic waves with frequencies less than 16 Hz are called **infrasound waves**, or **infrasound**. Weak absorption by different medium is typical for infrasound; hence, it propagates over great distances, travels around obstacles and penetrates into rooms.

Sources of infrasound in nature include plants and trees, earthquakes and tsunami, hurricanes, tornado and storms, eruption of volcanoes, avalanches, thunderstorms, floods and waterfalls. Industrial society offers such sources of infrasound waves as cars, aircrafts, motorcycles, air heating and cooling systems, agricultural mechanisms. Heart beatings, lung oscillations, activity of bowel, vibration of vocal system are accompanied with the generation of infrasound.

Human organism is sensitive to infrasound. People feel the state of fear, anxiety, uneasiness, extreme sorrow, nausea, imbalance and spatial disorientation. Infrasound has an adverse effect on the human body, causing tiredness, headache, drowse, irritation, and a sense of fear. It is supposed that the phenomenon of resonance is the cause of many of the said physiological effects since the natural oscillation frequencies of the human body and its parts are in the range of 3 to 13 Hz.

High-intensity industrial noises and vibrations are harmful for human beings. The intensity level of these sounds is measured by means of special devices called noise dosimeters.

6.6. The human speech apparatus and its work.

Unless there is a special problem, speaking our mother tongue is something we do so effortlessly and unconsciously that we are unaware not only of the extremely complex cognitive processes that underlie the act of speaking, but also of the incredibly precise mechanics involved in articulating our words correctly.

The human vocal apparatus (fig.6.13) is like two kinds of musical instruments at once: a **wind instrument** and a **string instrument**. This apparatus includes a source of wind (the lungs), components that vibrate (the vocal cords in the larynx), and a series of resonant chambers (the pharynx, the mouth, and the nasal cavities). Here is how all these components work together when you speak.

The first component of this apparatus is the lungs that provide the necessary air and that can thus be described as the **"generator"**. When you are speak-

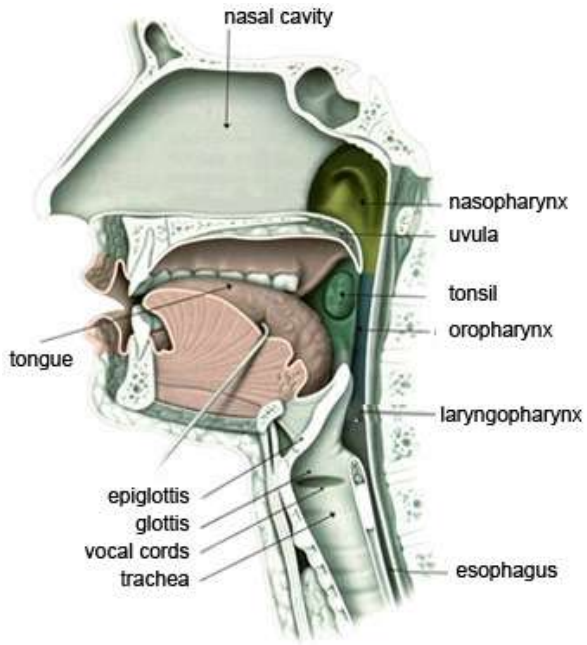

Fig. 6.13. The human vocal apparatus.

ing, your inhalations become faster and shorter and you breathe more with your mouth, whereas otherwise you inhale only with your nose. When you exhale while speaking, you increase the volume and pressure of your airstream to vibrate the vocal cords in your larynx.

The larynx consists of a set of muscles and pieces of cartilage, with varying degrees of mobility, that can be raised or lowered like a gate to protect your bronchi and lungs from food and other foreign bodies. When you swallow food, your larynx rises, while the epiglottis, a flap of cartilage at the entry to the larynx, closes down over it to block the upper airways and let the food move down your esophagus safely into your stomach.

When you speak, the air expelled from your lungs moves up through the trachea to the larynx, where it passes over the vocal cords. These cords are a matched pair of muscles and ligaments, pearly white in colour, 20 to 25 millimetres long, and coated with mucus. They constitute the second component of your vocal apparatus: **the "vibrator"**.

The vocal cords are attached horizontally from the thyroid cartilage (the "Adam's apple" in men) at the front to the arytenoid cartilages at the rear. By moving these cartilages as you speak, you alter the length and position of your vocal cords. When you start to say something, the arytenoid cartilages press the vocal cords against each other, thus closing the opening between them (known as the glottis).

Under the pressure of the air being exhaled, the vocal cords separate, then

close again immediately, causing the air pressure beneath the glottis to increase again. By opening and closing the glottis rapidly during phonation, the vocal cords thus release the air from the lungs in a vibrating stream. When you speak a sentence, you modify the vibration frequency of your vocal cords many times to produce the acoustic vibrations (sounds) that are the raw materials for the words themselves.

For these sounds to be transformed into words, they must then be shaped by the rest of the vocal apparatus. The first step in this process occurs in the pharyngeal cavity, where the respiratory and digestive systems meet. The pharynx and the other cavities with which it communicates (the nasal cavities, mouth, and larynx) act as a **"resonator"** that alters the sounds issuing from your vocal cords, amplifying some frequencies while attenuating others.

The transformation of the sounds from the larynx is then completed by the position of the soft palate, tongue, teeth, lips, and other parts of the mouth, which act as **"modulators"** for this sound. While the larynx produces the vibrations without which you would have no voice, it is these other parts of your vocal apparatus that make your voice so flexible and versatile. They do so in different ways. Your he soft palate either blocks the passage to the upper nasal cavities or leaves it open so that the vibrating airstream can enter them. Your jaws open or close to change the size of the oral cavity. Your tongue changes shape and position to alter this cavity further. Your tongue and the lips obstruct the airflow through the teeth to varying extents. The lips also alter their shape-open, closed, pursed, stretched, and so on-to shape the sound further.

To produce the vowel sound “ee” (as in “teen”), for example, you must move your tongue toward the front of your palate, which widens the pharyngeal cavity while raising the larynx slightly. To produce the sound “ah” (as in “far”), you must lower your jaw and your tongue. To pronounce consonants, you must make various movements of the tongue and lips. For example, to pronounce an “F”or an “S”, you move your tongue and lips so as to slow the outgoing airstream. To pronounce a “B”, “P”, or “T”, you stop the airstream and then release it, with varying degrees of sharpness. To produce a “V” or a “J”, you make the airstream vibrate, and so on.

6.6.1 Is the human vocal apparatus essential for speech?

Scientists long believed that the main reason that other primates had never succeeded in mastering human language despite all the efforts that had been made

to teach them was that the particular anatomy of their vocal apparatus prevented them from doing so. In apes, as in human infants, the larynx is positioned very high in the neck, which would prevent it from producing all the sounds of human language. But this position does have certain advantages: for example, both apes and babies can breathe through their noses while continuing to eat.

In contrast, in adult humans, the low position of the larynx means that the pathways to the stomach and the lungs intersect, thus increasing the risks of choking. It therefore seems that the advantage that this descended larynx provides is a vocal communication system that makes this risk of choking worthwhile.

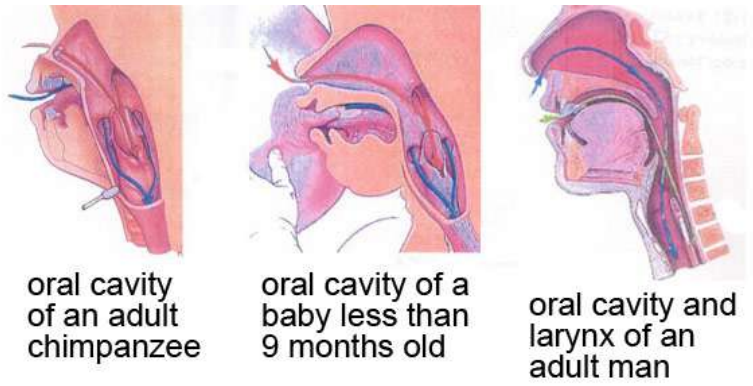

Fig. 6.14. The location of the larynx in humans and apes.

Modelling and simulation studies have shown, however, that the limited phonatory capabilities of the high-positioned larynx in primates and babies represent only a relatively minor handicap in terms of language. For that matter, the high position of the larynx in human babies does not prevent them from imitating the adult vowel sounds "ee", "ah", and "oo" from as early as 4 months of age, and from producing their first words 8 months later, when the larynx is still very high and the pharyngeal cavity is still very small. The reason that apes and younger babies cannot speak would therefore seem to be not that their larynx is too high, but rather that they lack the cognitive abilities needed to master language.

Test questions

1. Which mechanical waves are called sound waves? Infrasound? Ultrasound?

2. Give the definition of physical characteristics of a sound: tone, noise, sonic shock.

3. What does the concept of "sound intensity" express and what determines this value?

4. Explain the meaning of concepts: sound pressure, threshold of audibility, threshold of pain.

5. Explain the meaning of concepts: tone height, timbre, loudness.

6. Formulate the law of Weber - Fechner.

7. What is ultrasound? What physical phenomena are used in sources and receivers of ultrasound?

8. What do you know the effects of ultrasonic waves on biological objects?

9. Explain the essence of ultrasound diagnostic methods used in industry and medicine.

10. What is ultrasound therapy and surgery?

11. What are the features of the propagation of infrasound waves?

12. What is the effect of infrasound waves on humans?

13. What are physical bases of the human's speech apparatus?

Lecture No 7

Physics of hearing

7.1. The Ear

This is not a biology course, so we will not dwell upon *all* of the structure visible in this picture, but rather will concentrate on the parts relevant to the physics. Fig. 7.1 shows a cross-section of the human ear, our basic transduction device for sound.

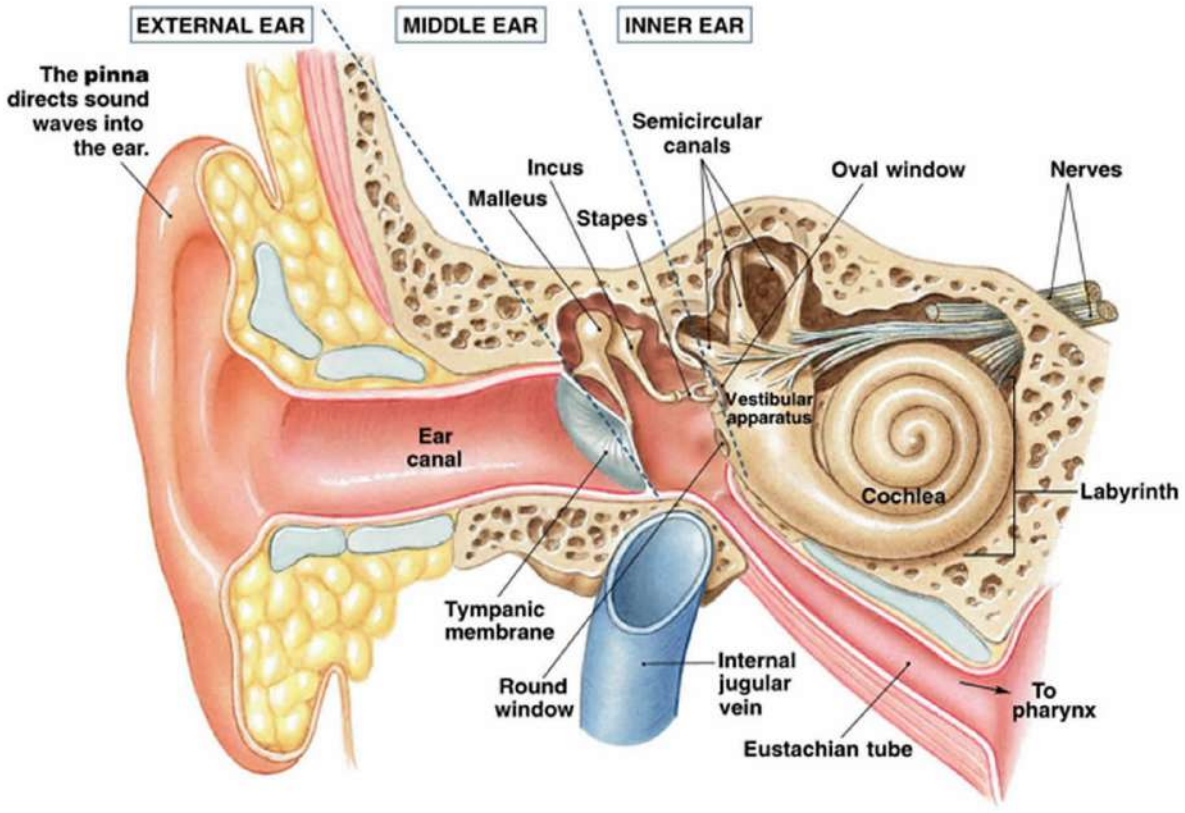

Fig. 7.1. The anatomy of the human ear.

Let's start with the **outer ear** (**external ear**). This structure collects sound waves from a larger area than the ear canal per se and reflects them down to the ear canal. You can easily experiment with the kind of amplification that results from this by cupping your hands and holding them immediately behind

your ears. You should be able to hear both a qualitative change in the frequencies you are hearing and an effective amplification of the sounds from in *front* of you at the expense of sounds originating *behind* you. Many animals have larger outer ears oriented primarily towards the front, and have muscles that permit them to further alter the direction of most favorable sound collection without turning their heads. Human ears are more nearly omnidirection.

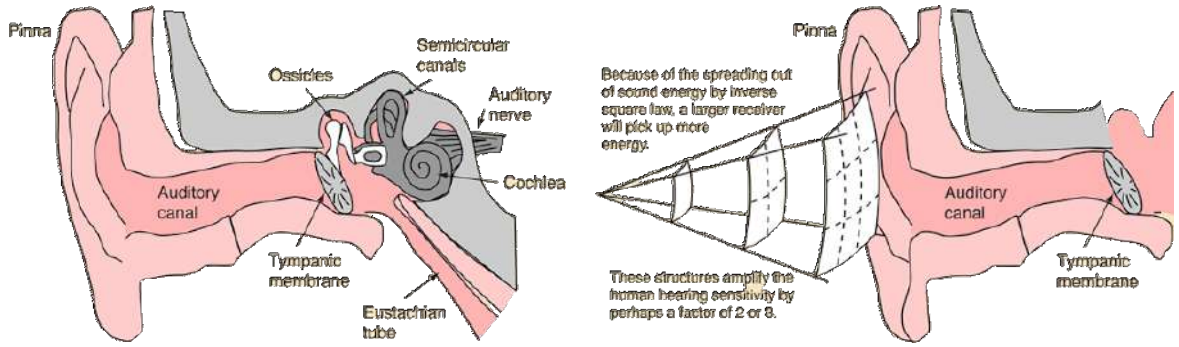

Fig. 7.2. The auditory canal

Sound energy spreads out from its sources. For a point source of sound, it spreads out according to the inverse square law. For a given sound intensity, a larger ear captures more of the wave and hence more sound energy. The outer ear structures act as part of the ear's preamplifier to enhance the sensitivity of hearing.

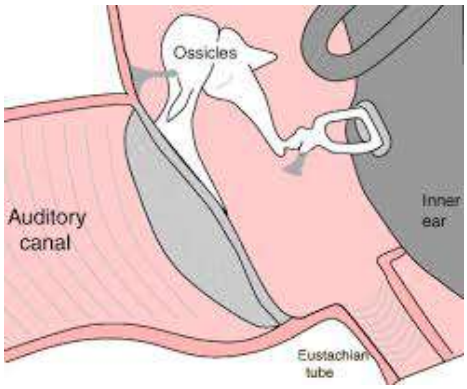

Fig. 7.3. The tympanic membrane

The **auditory canal** (ear canal in the figure above) acts like a resonant cavity to effectively amplify frequencies in the 2.5 kHz range and tune energy deliver to the tympanic membrane or eardrum. This membrane is a strong, resilient, tightly stretched structure that can vibrate in response to driving sound waves. It is connected to a collection of small bones (the ossicles) that conduct sound from the eardrum to the inner ear and that constitute the **middle ear** in the figure above. The common name of the ossicles are: hammer, anvil and stirrup, the latter so named because its shape strongly resembles that of the stirrup on a horse saddle. The anvil effectively amplifies oscillations by use of the principle of leverage, as a fulcrum attachment causes the stirrup end to vibrate through a much larger amplitude than the hammer end. The stirrup is directly connected to the **oval window**, the gateway into the inner

ear.

The tympanic membrane or "eardrum" receives vibrations traveling up the auditory canal and transfers them through the tiny ossicles to the oval window, the port into the inner ear.

The eardrum is some fifteen times larger than the oval window of the inner ear, giving an amplification of about fifteen compared to a case where the sound pressure interacted with the oval window alone.

The tympanic membrane is very thin, about 0.1 mm, but it is resilient and strong. It is made up of three layers: the outer layer of skin, a layer of fibrous connective tissue, and a layer of mucous membrane.

The middle ear is connected to the **eustachian tube** to your throat, permitting pressure inside your middle ear to equalize with ambient air pressure outside. If you pinch your nose, close your mouth, and try to breath out hard, you can actually blow air out through your ears although this is unpleasant and can be dangerous. This is one way your ears equilbrate by "popping" when you ride a car up a hillside or fly in an airplane. If/when this does not happen, pressure differences across the tympanic membrane reduce its response to ambient sounds reducing auditory acuity.

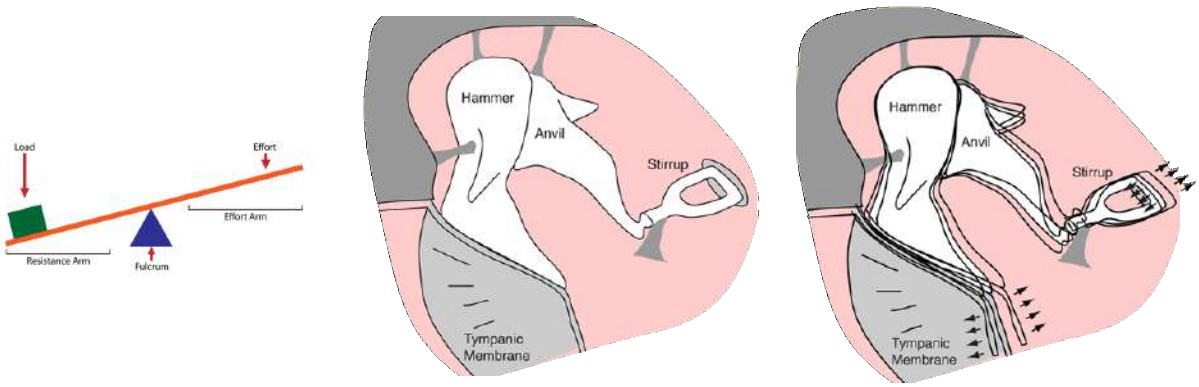

Fig. 7.4. The Ossicles.

The three tiniest bones in the body form the coupling between the vibration of the eardrum and the forces exerted on the oval window of the inner ear. Formally named the malleus, incus, and stapes, they are commonly referred to in English as the hammer, anvil, and stirrup. With a long enough lever, you can lift a big rock with a small applied force on the other end of the lever. The amplification of force can be changed by shifting the pivot point.

The ossicles can be thought of as a compound lever which achieves a multiplication of force. This lever action is thought to achieve an amplification by a

factor of about three under optimum conditions, but can be adjusted by muscle action to actually attenuate the sound signal for protection against loud sounds. The vibration of the eardrum is transmitted to the oval window of the inner ear by means of the ossicles, which achieve an amplification by lever action. The lever is adjustable under muscle action and may actually attenuate loud sounds for protection of the ear.

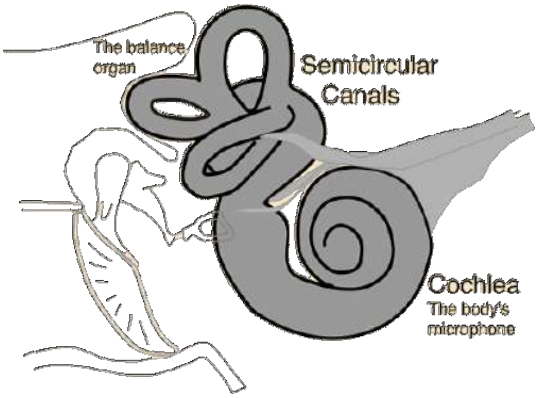

Fig. 7.5. The Inner Ear

The small bone called the **stirrup**, one of the ossicles, exerts force on a thin membrane called the **oval window**, transmitting sound pressure information into the **inner ear**.

The inner ear can be thought of as two organs: the semicircular canals which serve as the body's balance organ and the cochlea which serves as the body's microphone, converting sound pressure impulses from the outer ear into electrical impulses which are passed on to the brain via the auditory nerve.

The basilar membrane of the inner ear plays a critical role in the perception of pitch according to the place theory.

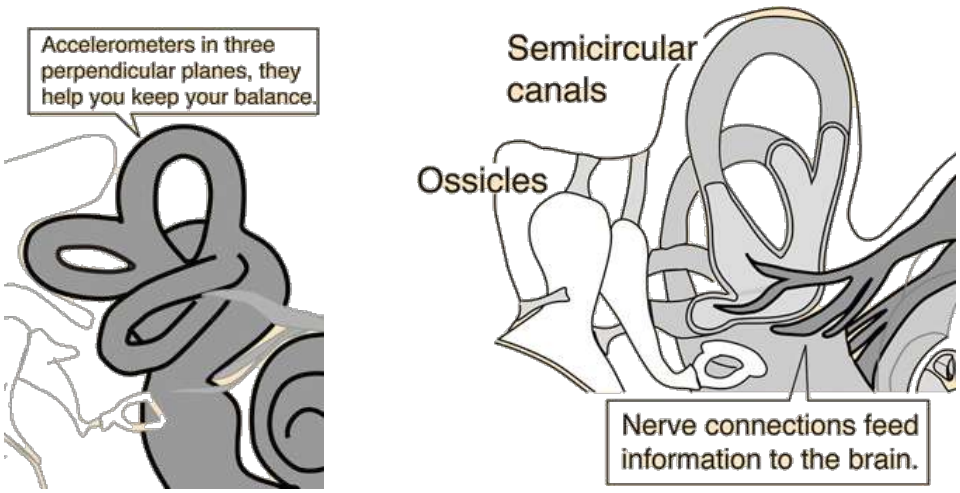

Fig. 7.6. The Semicircular Canals.

The semicircular canals, part of the inner ear, are the body's balance organs, detecting acceleration in the three perpendicular planes. These accelerometers make use of hair cells similar to those on the organ of Corti, but these hair cells detect movements of the fluid in the canals caused by angular acceleration

about an axis perpendicular to the plane of the canal. Tiny floating particles aid the process of stimulating the hair cells as they move with the fluid. The canals are connected to the auditory nerve.

The nerves stretching from these cells are collected into the auditory nerve bundle and from thence carries the impulses they give off when they receive sound at the right frequency in to the auditory cortex (not shown) where it becomes, eventually and through a process still not fully understood, our perception of sound. Our brains take this frequency resolved information – the biomechanical equivalent of a fourier transform of the sound signal, in a way – and synthesize it back into a detailed perception of sound and music within the general frequency range of 10 Hz to 20,000 Hz.

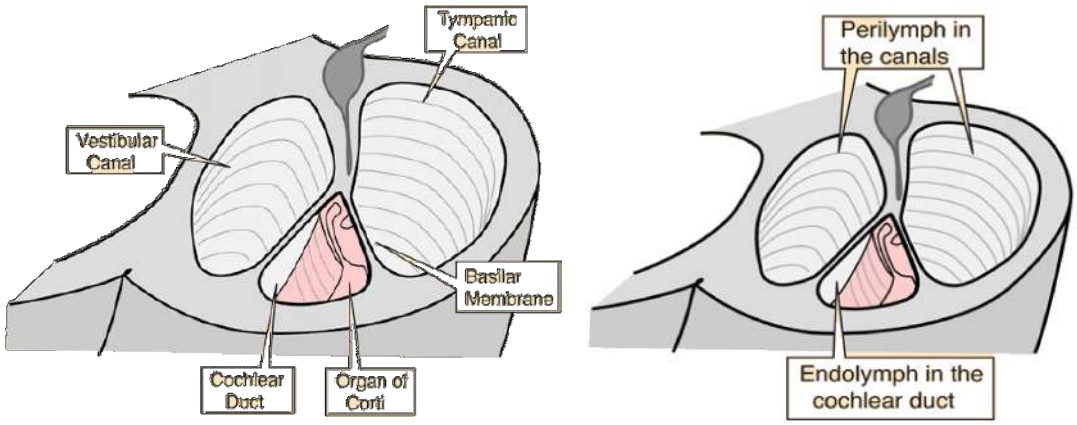

Fig. 7.7. The Semicircular Canals.

The inner ear structure called the cochlea is a snail-shell like structure divided into three fluid-filled parts. Two are canals for the transmission of pressure and in the third is the sensitive organ of Corti, which detects pressure impulses and responds with electrical impulses which travel along the auditory nerve to the brain.

The perilymph fluid in the canals differs from the endolymph fluid in the cochlear duct. The organ of Corti is the sensor of pressure variations

The pressure changes in the cochlea caused by sound entering the ear travel down the fluid filled tympanic and vestibular canals which are filled with a fluid called perilymph. This perilymph is almost identical to spinal fluid and differs significantly from the endolymph which fills the cochlear duct and surrounds the sensitive organ of Corti. The fluids differ in terms of their electrolytes and if the membranes are ruptured so that there is mixing of the fluids, the hearing is impaired.

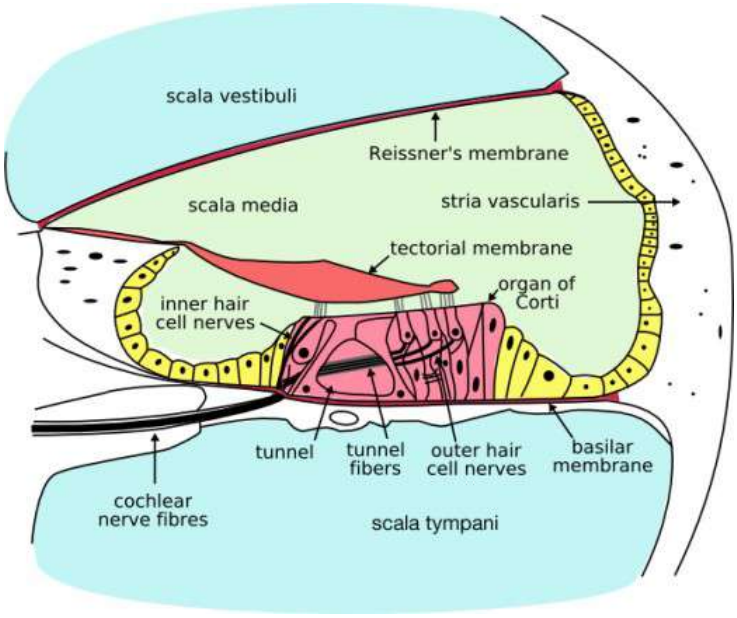

Fig. 7.8. A cross-section of the spiral structure of the cochlea.

Sound, amplifed by focal concentration in the outer ear, resonance in the auditory canal, and mechanical leverage in the ossicles, enters the **cochlea**, a shell-shaped spiral that is the primary organ of hearing that transduces sound energy into impulses in our nervous system through the oval window. The cochlea contains **hair cells** of smoothly varying length lining the narrowing spiral, each of which is **resonant** to a particular auditory frequency. The arrangement of the cells in a cross-section of the cochlea is shown in figure 7.8.

As you can see, there are many individual parts that can fail in the human auditory system. Individuals can lose or suffer damage to their outer ears through accident or disease. The ear canal can become clogged with cerumen, or earwax, a waxy fluid that normally cleans and lubricates the ear canal and eardrum but that can build up and dry out to both load the tympanic membrane so it becomes less responsive and physically occlude part of the canal so less sound energy can get through. The eardrum itself is vulnerable to sudden changes in sound pressure or physical contact that can puncture it. The middle ear, as a closed, warm, damp cavity connected to the throat, is an ideal breeding ground for certain bacteria that can cause infections and swelling that both interfere with or damage hearing and that can be quite painful. The ossicles are susceptible to physical trauma and infectious damage.

Finally, the hair cells of the cochlea itself, which are safely responsive over at least twelve to fourteen orders of magnitude of transient sound intensity (and safely responsive over eight or nine orders of magnitude of sustained sound inten-

sity) are highly vulnerable to both sudden transient sounds of still higher intensity (e.g. sound levels in the vicinity of 120 to 140 decibels and higher and to sustained excitation at sound levels from roughly 90 decibels and higher. Both disease and medical conditions such as diabetes (that produces a progressive neuropathy) can further contribute to gradual or acute hearing loss at the neurological level.

When hair cells die, they do not regenerate and hearing loss of this sort is thus cumulative over a lifetime. It is therefore a really good idea to wear ear protectors if, for example, you play an instrument in a marching band or a rock and roll band where your hearing is routinely exposed to 100 dB and up sounds. It is also a good reason not to play music too loudly when you are young, however pleasurable it might seem. One is, after all, very probably trading listening to very loud music at age seventeen against listening to music at all at age seventy. Hearing aids do not really fix the problem, although they can help restore enough function for somebody to get by.

7.2. Noise Health Effects

The damage done by noise depends on the loudness and the time of exposure. Noise that reaches the inner ear can provoke the temporary hearing loss. After certain period of time hearing may be restored. This is *temporary* hearing loss. Under long-term exposure to intense noise or short-term exposure to very intense noises the hearing loss will become *permanent*.

The elevated sound levels cause trauma to the cochlear structure in the inner ear. This hearing loss results from the destruction of the inner ear-cells which can never be replaced or repaired.

High intensity noise induces cardiovascular effects, rise in blood pressure, increased incidence of coronary artery disease.

Hearing loss can be evaluated by *noise-induced temporary threshold shift (NITTS)*, which is determined through the measurement of hearing sensitivity before and after noise exposure. This shift can be temporary or permanent depending on the intensity, frequency and duration of the noise.

Sonic impact (or **sound shock**) is a short-frequency sound with a continuous spectrum.

For example: explosion, thunder.

7.3. Physical characteristics of sound

1. Frequency (f) and frequency spectrum.

2. Velocity of sound describes how much distance such a wave travels in a given amount of time.

The speed of sound depends upon the compressibility, density and temperature of medium.

For instance,

in dry air the speed of sound is 333 m/s (t = 0°C) and 344 m/s (t = 20°C);

in water the speed of sound is 1402 m/s (t = 0°C) and 1500 m/s (t = 20°C).

In the human body tissues the average speed of sound is 1540 m/s.

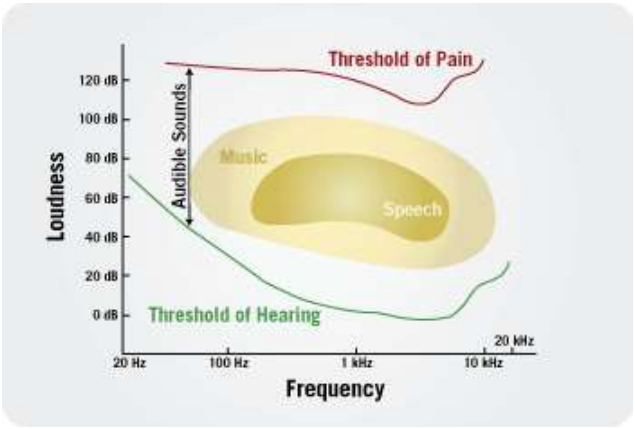

Fig. 7.9. Dependence of audibilityand pain thresholds on sound frequency.

3. Sound intensity (energy parameter) is the amount of energy that is transported through a given area of the medium per unit of time:

$$I = \frac{\Phi}{\Delta S} = \frac{\Delta E}{\Delta t \cdot \Delta S}, \qquad [I] = \frac{W}{m^2}.$$

The intensity range of hearing has two main characteristics:

I. The threshold of hearing (or **threshold of audibility**) is the minimum intensity of sound that is audible at a given frequency.

$$I_0 = 10^{-12}\frac{W}{m^2}$$

.

II. The threshold of pain (or **threshold of feeling**) is the intensity of sound wave, which gives rise to a painful annoying sensation in ears.

$$I_{max} = 10\frac{W}{m^2}$$

.

4. Sound overpressure (ΔP) is the excess pressure in the wave of atmospheric pressure.

$$\boxed{\Delta P = \sqrt{2\rho v I}}, \qquad I = \frac{\Delta P^2}{(2\rho v)},$$

where v is sound speed,

I is intensity of sound wave,

ρ is medium density,

$\rho v = X$ is the acoustic resistance of medium.

5. The level of intensity (L) is determined by formula

$$\boxed{L = lg\frac{I}{I_0}}, \qquad [L] = B \quad (bel),$$

where I is sound intensity; I_0 is sound intensity of hearing threshold.

$$L = lg\frac{10^{-11}}{10^{-12}} = 1B.$$

1 dB (decibel) = 0.1 B.

Norm is 40 – 60 dB.

Table 7.1 gives some typical values for the sound levels from various sources or causing specific symptoms in humans.

Table 7.1

Typical values of the sound levels from various sources or causing specific symptoms.

Source of sound	**Sound level, dB**	**Source of sound**	**Sound level, dB**
Damage of eardrum	160	Busy traffic	80
Nearby jet airplane	150	Vacuum cleaner	70
Threshold of pain	130	Normal conversation	50
Rock concert	120	Mosquito buzzling	40
Subway	100	Rustling leaves	10

7.4. Subjective characteristics of auditory sensation

1. Loudness. The sensation of loudness corresponds roughly to the intensity level of a sound. But the loudness of pure tones depends from their frequency. For example, the simple tone which frequency is about 4000 Hz seems louder than the tones with the same intensity level but other frequencies.

Weber - Fechner law may be formulated as follows: irritation sensation

E is in direct proportion to logarithm of intensity of the proper I. In other words, if irritation intensity I rise in geometric series that is by a factor of

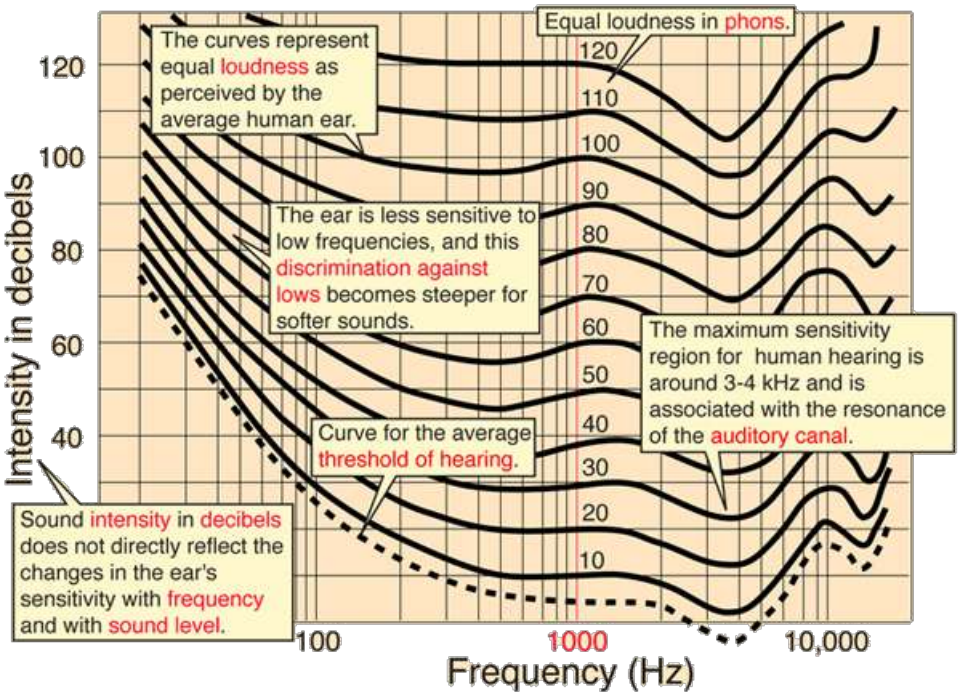

Fig. 7.10. Sound intensity.

100, 1000, ... (αI_0, $\alpha^2 I_0$, $\alpha^3 I_0$ ... $\alpha > 1$), then sensation of this irritation E increases in arithmetical progression, that is by a factor of 2, 3, ... (E_0, $2E_0$, $3E_0$ ...).

Physiologists Weber and Fechner proved that the relation between the sound intensity and level of its auditory sensation (loudness). The formula for loudness has the form (**Weber - Fechner law**):

$$\boxed{E = k(f) lg \frac{I}{I_0}}, \qquad [E] = phon,$$

$k(f)$ is proportionality coefficient which depends on frequency and intensity;

I is sound intensity; I_0 is sound intensity of hearing threshold.

The phone scale and decibel scale coincide for the frequency of 10^3 Hz ($k = 1$) but differ at other frequencies.

For $f = 1$ kHz, $k = 1$ *1 phon = 1dB = 0,1 B.*

2. Highness. Highness is the characteristic of sound that depends on its frequency. A low-pitched tone has the smaller frequency and a high-pitched tone has the greater frequency.

3. Timbre. Timbre is the characteristic of sound by which a man is able to differentiate between the sounds coming from different musical instruments even when the same note is played on all of them. This quality of sound depends upon the acoustic spectrum, the number of overtones and their intensity.

Test questions

1. Explain the device human ear.

2. What is the method of audiometry?

3. What sound methods are used in medicine?

4. What are the main factors that noise influence on human health?

5. How to reduce the harmful impact of noise on human?

6. What is the essence of the Weber-Fechner's law ? Write down the formula.

7. What is the threshold of acoustic sensation, pain threshold and boundary frequencies?

8. What is the nature of a sound? How does the speed of sound distribution depend on properties of the environment changes? Does the sound travel in vacuum?

9. Give the definition of characteristics of acoustic sensation: tone height, timbre, loudness. What is the connection between physical and acoustic characteristics?

Lecture No 8

Hydrodynamics of viscous liquid.

8.1. Basic physical quantities of hydrodynamics.

Volume velocity of current flow (Q) is called the volume of fluid, which flows through the section of the tube of current per unit of time:

$$\boxed{Q = \frac{V}{t}}, \qquad [Q] = \frac{m^3}{s}.$$

Linear velocity (v) is the path traversed by the particles of blood per unit time:

$$\boxed{v = \frac{l}{t}}, \qquad [v] = \frac{m}{s}.$$

The relationship between the linear and volume velocity

$$Q = S \cdot v,$$

where S is cross-sectional area of fluid flow.

Pressure is a physical quantity which characterizes the intensity of normal (perpendicular to the surface) force with which one body acts on the surface of another. If the force exhibits a uniform distribution along the surface, the pressure is determined as the ratio of force to area:

$$\boxed{p = \frac{F}{S}}, \qquad [p] = Pa.$$

where F is the magnitude of the normal force on the surface and S is the area of this surface.

The SI unit for pressure is the Pascal ($1\ N/m^2 = 1\ Pa$).

Blood pressure (p) is the force exerting by blood per unit area of blood vessel.

8.1.1. Stationary current of a liquid. Continuity equation of flow filament.

Hydrodynamics is the section of physics that study questions of motion of liquids (incompressible) and its interaction with environmental solid bodies. Liquid medium makes the greatest part of human organism, its motion provides metabolism and supply of cells by oxygen and therefore mechanical properties and current of liquids represent interest for physicians and biologists.

Consider the established or **stationary current** of liquid, i.e. such current at which speeds of particles of liquid in each point of a stream do not changes.

Stationary current is characterized by **lines of a current**, i. e. imagined lines, conterminous to trajectories of particles. Tangents to lines of a current shows direction of speed vector of particles of a liquid, density of these graphically represented lines is proportional to speed. A part of a stream of liquid limited from different directions by lines of a current forms **a tube of a current** or **jet** (fig.8.1)

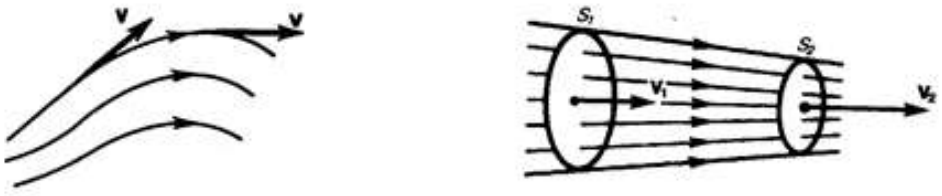

Fig. 8.1. Lines and tube of a current.

Suppose incompressible fluid which fills completely a channel such as tube and flows along it. Then if some amount of fluid enters one end of the channel, an equal amount must leave the other end.

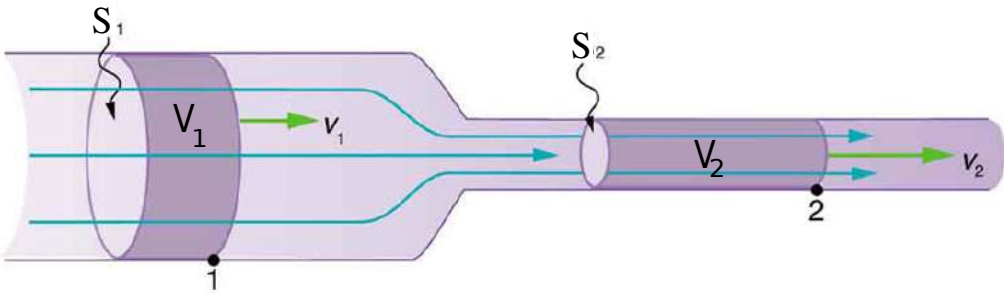

Fig. 8.2. The principle of continuity.

Flow rate Q is used as the measure of amount of fluid. It means a volume of fluid V, which moves through the cross section of a channel per second. If the

fluid enters one end of a tube at the flow rate Q_1 , it must leave the other end at the flow rate Q_2 , which is equal to Q_1.

$$Q_1 = Q_2$$

$$Q = \frac{V}{t} = \frac{S \cdot l}{t} = S \cdot v$$

where V is the volume, S is the cross sectional area of a pipe, v is the mean velocity of the liquid for a section.

Let's allocate a tube of the current so narrow, that speeds of particles v in its any section S perpendicular to axis of the tube is possible to count identical on all section. As speed of particles is directed along of line of a current, then particles of a liquid can not fall outside the limits of this tube. Then volume ΔV of incompressible liquid, proceeding through any section of a tube per unit of time, remains the constant:

$$\boxed{\Delta V = v \cdot S = const}.$$

The parity expresses **continuity equation of flow filament**, since only at continuous current through any section for same time pass identical volumes of incompressible liquid.

Then: $v_1 S_1 = v_2 S_2$ (see fig.8.1), whence $\frac{S_1}{S_2} = \frac{v_2}{v_1}$ i.e. average speeds of current in various sections of a pipe are inversely proportional to the areas of these sections.

From the continuity equation of flow follows, that also for a real liquid at the established current on a pipe of variable section, quantity of the liquid, proceeding for same time through any section of the pipe remains constant: $Q = const$. In particular, with big accuracy it is carried out for blood current in large blood vessels in time equal to several intimate cycles, directly following one after another.

8.2. Equation of Bernoulli and its consequences.

One of the most important equations used for the description of moving liquids, for the first time was received by Swiss mathematician and physicist Daniel Bernoulli (1700-1782). For deduction of the equation Bernoulli has assumed, that we deal with **ideal liquid**. It means that we neglect any forces of viscous resistance and friction.

Let's consider stationary current of an incompressible ideal liquid. We shall allocate the volume of liquid limited by walls of a narrow tube of current and perpendicular to lines of current by sections S_1 and S_2 (fig. 8.3). In time Δt this volume will be displaced along the tube of current and the border of volume S_1 will receive displacement Δl_1 and border S_2 displacement Δl_1. The work made at it by forces of pressure is equal to the increment of total energy $(E_K + E_P)$ in considered volume of liquid: $A = E_K + E_P$.

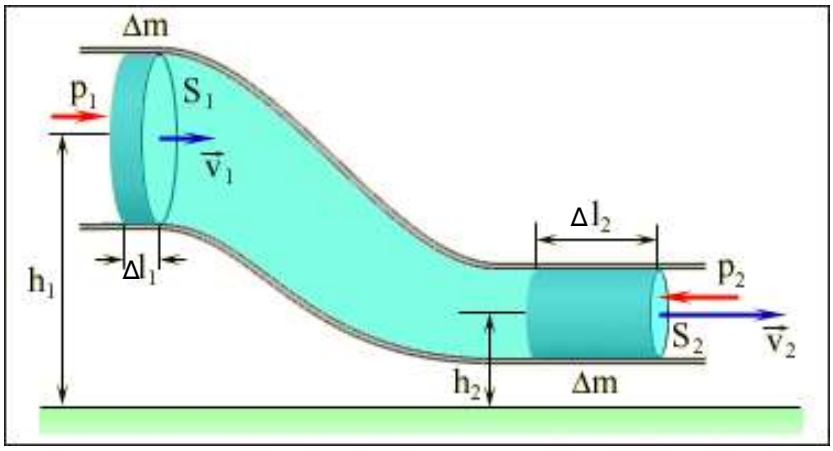

Fig. 8.3. Work-Mechanical Energy in Fluids: Bernoulli's Equation.

Forces of pressure upon walls of a tube of current are perpendicular in each point to the direction of moving of liquid owing to what works do not made. Work of forces of the pressure enclosed to sections S_1 and S_2 is equal to zero. This work is equal to:

$$A = p_1 S_1 \Delta l_1 - p_2 S_2 \Delta l_2 = (p_1 - p_2)\Delta V$$

Total energy of considered volume of liquid is composed from kinetic energy and potential energy in the field of forces of terrestrial gravitation. Owing to stationary current total energy of that part liquids, which is limited by sections $1'$ and $2'$ (the internal not shaded part of the tube of current on fig.) in time Δt does not change. Therefore the increment of total energy is equal to the difference of values of total energy of the shaded volumes ΔV_2 and ΔV_1, which mass is $\Delta m = \rho \Delta V$ (ρ is density of liquid).

Let's take section S of a tube of current and displacement Δt so small that all points of each of the shaded volumes could attribute the same speed v, pressure p and heights h. Then for the increment of total energy:

$$\Delta E = \left(\frac{\rho \Delta V v_2^2}{2} + \rho \Delta V g h_2 \right) - \left(\frac{\rho \Delta V v_1^2}{2} + \rho \Delta V g h_1 \right)$$

Change of total energy is equal to the work of external forces if we disregard friction within the system.

$$\boxed{\Delta E_{total} = A}$$

$$\boxed{E_{total} = \frac{mv^2}{2} + mgh}$$

$$\boxed{E_2 - E_1 = A_1 - A_2}$$

$$F = p \cdot S, \qquad A = p \cdot S \cdot \Delta l,$$

$$A = p_1 S_1 l_1 - p_2 S_2 l_2$$

$$\boxed{\frac{m_2 v_2^2}{2} + m_2 g h_2 - \frac{m_1 v_1^2}{2} + m_1 g h_1 = p_1 S_1 l_1 - p_2 S_2 l_2}$$

$$m = \rho V = \rho \cdot S \cdot \Delta l$$

$$\boxed{\frac{\rho S_2 l_2 v_2^2}{2} + \rho S_2 l_2 g h_2 - \frac{\rho S_1 l_1 v_1^2}{2} + \rho S_1 l_1 g h_1 = p_1 S_1 l_1 - p_2 S_2 l_2}$$

Divide by the expression $S_1 l_1 = S_2 l_2$.

$$\boxed{\frac{\rho v_2^2}{2} + \rho g h_2 + p_2 = \frac{\rho v_1^2}{2} + \rho g h_1 + p_1}$$

In according to Bernoulli's law the pressure in the flow is higher where speed is lower and vice versa.

For steady, irrotational flow, the velocity, pressure, and elevation of an incompressible, nonviscous fluid are related by an equation:

$$\boxed{p + \rho g h + \frac{\rho v^2}{2} = const}$$

p is static pressure,

$\rho g h$ is hydrostatic pressure,

$\frac{\rho v^2}{2}$ is dynamic pressure. This equation was discovered by Daniel Bernoulli.

Bernoulli's equation proves that the total pressure is constant along the tube of current during steady-state flow of perfect (or ideal) fluid.

8.2.1. Flow through a venturi tube

If an incompressible fluid moves through a venturi tube (i.e. a tube pur-

posefully built to be narrow in the middle), the continuity principle tells us the fluid velocity must increase through the narrow portion. This increase in velocity causes kinetic energy to increase at that point. If the tube is level, there will be negligible difference in elevation (z) between different points of the tube's centerline, which means elevation head remains constant. According to the Law of Energy Conservation, some other form of energy must decrease to account for the increase in kinetic energy. This other form is the pressure head, which decreases at the throat of the venturi:

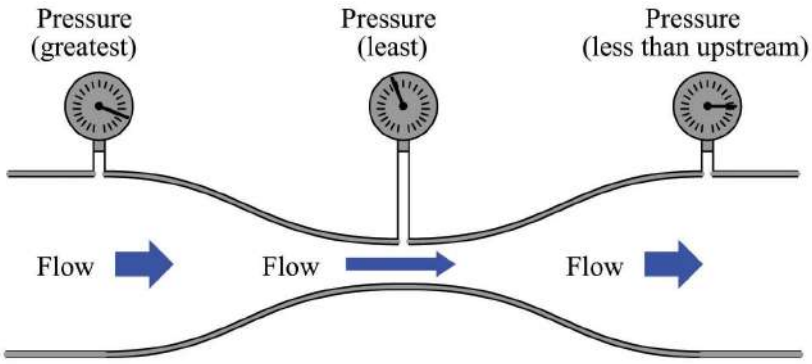

Fig. 8.4. The pressure head.

Ideally, the pressure downstream of the narrow throat should be the same as the pressure upstream, assuming equal pipe diameters upstream and down. However, in practice the downstream pressure gauge will show slightly less pressure than the upstream gauge due to some inevitable energy loss as the fluid passed through the venturi. Some of this loss is due to fluid friction against the walls of the tube, and some is due to viscous losses within the fluid driven by turbulent fluid motion at the high-velocity throat passage.

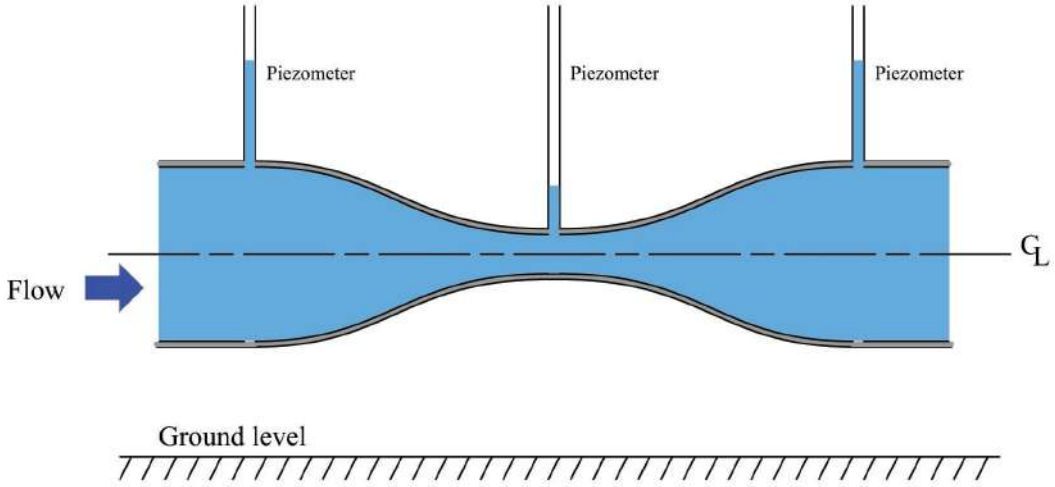

Fig. 8.5. Piezometer.

The difference between upstream and downstream pressure is called permanent pressure loss, while the difference in pressure between the narrow throat and downstream is called pressure recovery.

If we install vertical sight-tubes called piezometers along a horizontal venturi tube, the differences in pressure will be shown by the heights of liquid columns within the tubes. Here, we assume an ideal (inviscid) liquid with no permanent pressure loss:

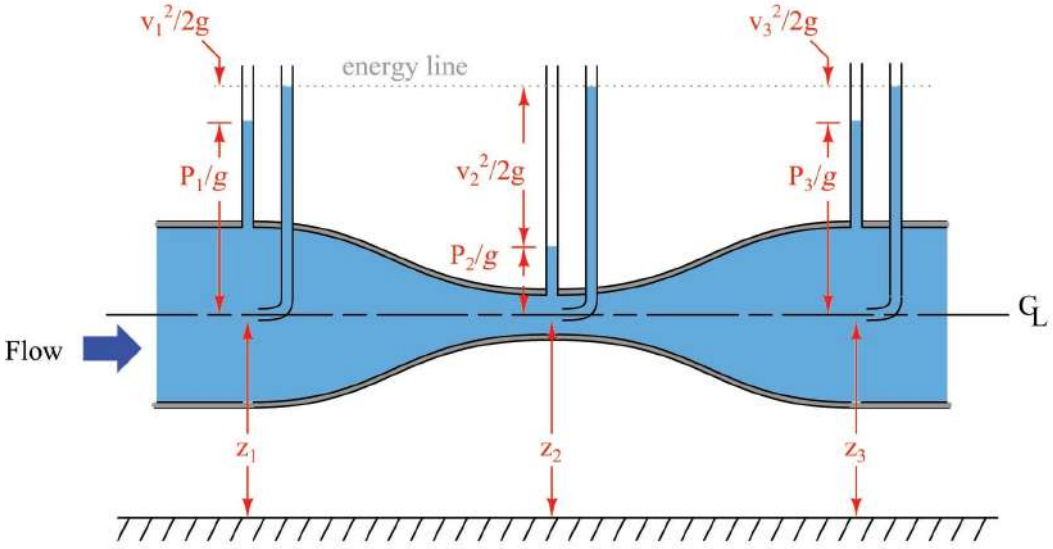

Fig. 8.6. The difference in liquid column height.

The height of liquid in each piezometer tube represents the amount of potential energy in the fluid at that point along the venturi tube.

We may gain more insight into the nature of energy in this moving fluid stream if we add three more piezometers, each one equipped with its own Pitot tube facing upstream to "catch" the velocity of the fluid. Rather than represent potential energy by liquid height as the straight-tube piezometers do, the Pitot tube piezometers represent the total energy (potential plus kinetic) of the fluid. As such, the liquid heights in these new piezometers are all equal to each other, showing that total energy is indeed conserved at every point in the system:

$$z + \frac{v^2}{2g} + \frac{P}{\gamma} = (\text{constant})$$

Here, each of the "heads" represented in Bernoulli's equation are shown in relation to the different piezometer heights. The difference in liquid column height between each Pitot tube piezometer (potential + kinetic energy) and its corresponding straight-tube piezometer (potential energy alone) reflects the amount of

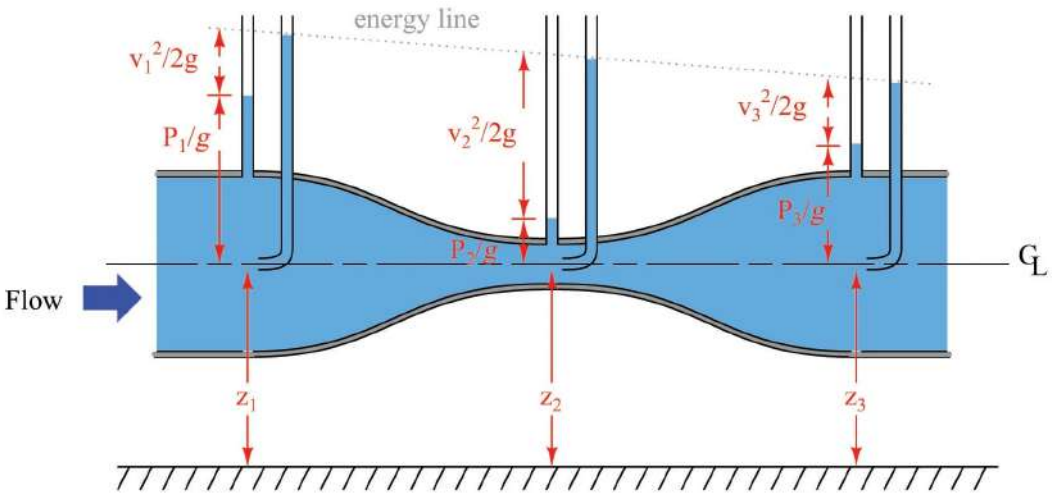

Fig. 8.7. The "energy line".

kinetic energy possessed by the fluid stream at that point in the venturi tube.

In a real venturi tube, there is some energy permanently lost in the moving fluid due to friction. Consequently the piezometer measurements in a real venturi tube would look something like this:

The "energy line" is seen to slope downhill from inlet to outlet on the venturi tube, showing a degradation in total energy content from beginning to end.

It is used in water-jet pumps, medical inhalers, sprays:

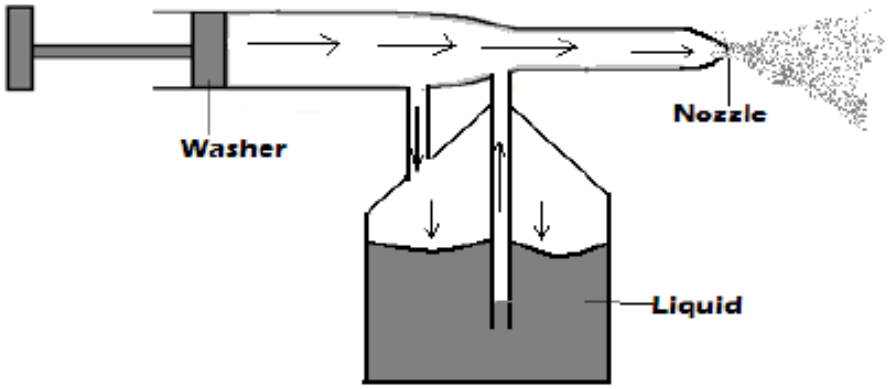

Fig. 8.8. The spray.

8.4. Fluid viscosity.

Viscosity is property of fluids owing to which they oppose any motion of their neighbouring portions relative to one another. Viscosity is created by **internal friction** between the molecules.

Various fluids differ greatly by the value of their viscosity. For instance, the viscosity of oil is greater than that of water. Viscosity is a major factor in determining the forces that must be overcome when fluids are transported in the tubes. Viscosity influences also the blood flow in the circulatory system.

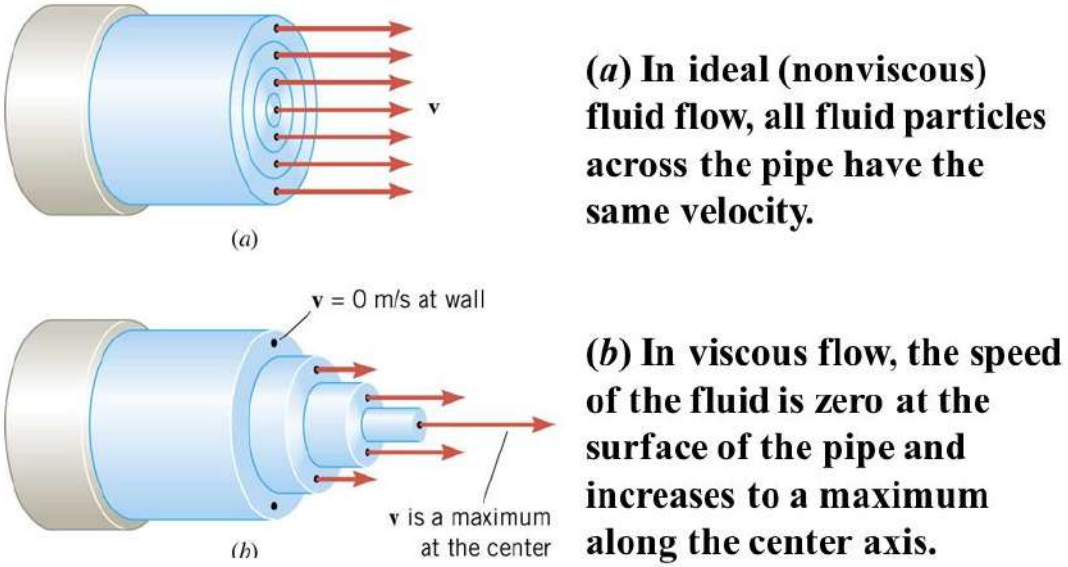

Fig. 8.9. Viscous flow.

In an ideal fluid there is no viscosity to hinder the fluid layers as they slide past one another. Within a pipe of uniform cross section, every layer of an ideal fluid moves with the same velocity, even the layer next to the wall, as fig. shows:

When viscosity is present, the fluid layers do not all have the same velocity, as part (b) of the drawing illustrates. The fluid closest to the wall does not move at all, while the fluid at the center of the pipe has the greatest velocity. The fluid layer next to the wall surface does not move, because it is held tightly by intermolecular forces.

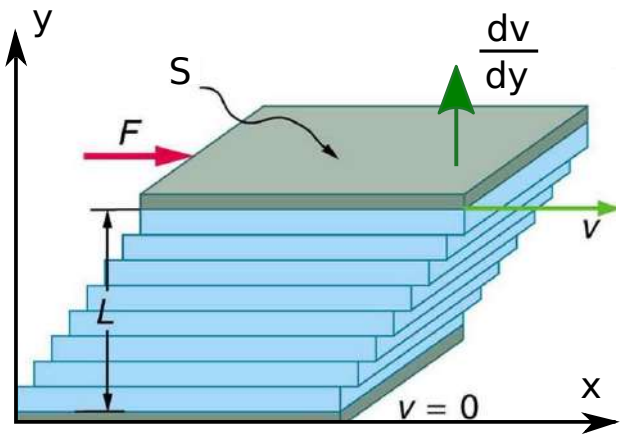

Fig. 8.10. Gradient of speed $\frac{dv}{dy}$.

In a real liquid between molecules operates the force of a mutual attraction causing **internal friction** or **viscosity**. Internal friction, for example, causes force of resistance at stirring of liquid, slows down speeds of falling body thrown in it.

Newton has established, that force F_{fr} of internal friction between two layers of liquid moving with var-

ious speeds is directly proportional to area S of adjoining layers and to gradient of speed $\frac{dv}{dy}$ between them (the gradient of speed between layers is the change of speed dv, divided to length dy in the direction, perpendicular of speed):

Fundamental law of viscous liquid was discovered by **Newton (1687)**:

$$\boxed{F_{fr} = -\eta \frac{dv}{dy} \cdot S} \qquad \textbf{(Newton's formula)}$$

where η is the coefficient of viscosity (or dynamic viscosity) and equal to the force of internal friction that acts on the unit area of the layers surface at the velocity gradient which is equal to one.

SI Unit of viscosity:

$[\eta] = P \cdot s \;\; (Pascal - seconds)$.

Common Unit of viscosity: poise (P).

1 poise (P) = 0.1 Pa·s

$[F] = N$ is the force of internal friction between the liquid layers appearing as result of their relative shift;

$\frac{dv}{dy}$ is a velocity gradient, which characterizes the degree of change in the velocity of the liquid flow in transfer from one layer to another; $[S] = m^2$ is the area of tangent layers.

The viscosity of blood: $\eta_{norm} = 0.004 \div 0.005 \;\; P \cdot s$.

$\eta_{pathology} = 1.7 \div 22.9 \;\; mP \cdot s$ **(anaemia, polycythemia).**

8.4.1. Newtonian fluids and non-Newtonian fluids.

The fluids whose viscosity may be defined by Newton's equation ($F \sim \frac{dv}{dy}$) are called *Newtonian fluids*. For *Newtonian fluids* the coefficient of viscosity (η) depends only on temperature and nature of fluid, and does not depend on conditions of fluidity. They are homogeneous fluids, such as water, spirits, solutions of electrolytes, mercury, glycerin etc.

There do exist more complicated heterogeneous fluids for which the Newtonian description is inadequate. The coefficient of viscosity of these fluids depends on conditions of fluidity. They are called *non-Newtonian fluids*. They includes some suspensions, emulsions, foams, protein solutions, polymer solutions, blood.

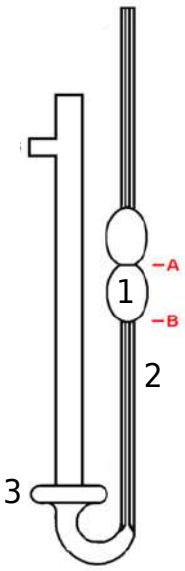

Fig. 8.11. Viscosimeters.

Relative viscosity is equal to the ratio of the coefficient of viscosity of the given fluid to the coefficient of viscosity of distilled water at one and the same temperature:

$$\boxed{\eta_{rel} = \frac{\eta}{\eta_w}}.$$

Viscosity of distilled water at room temperature is approximately equal to $10^{-3}\ Pa \cdot s$.

Kinematic viscosity is the ratio of the viscosity of a fluid to its density.

$$\boxed{\nu = \frac{\eta}{\rho}}, \quad (\rho \text{ is density of the liquid}), \quad [\nu] = \frac{m^2}{s}$$

Viscosity of a liquid is measured with the help of viscosimeters. With help of capillary viscosimeter of Ostwald (fig.) it is possible to measure viscosity of gases (from $10^{-5}\ Pa \cdot s$) and liquids (up to $10^4\ Pa \cdot s$). In the U-shaped tube one of which knees has the capillary 2 pour a researched liquid. With help of a rubber pear the liquid from the big tank 3 is sucked in the tank 1 above of mark A. Then, having removed the pear, is measured time of flow of liquid between labels A and . Comparing this time with time of flow of reference liquid (distilled water) can be determined viscosity of a researched liquid.

Method of incident blob (fig.) is used in the viscometers, which has been set up on the Stokes law. Measuring the velocity of the blob it is possible to find viscosity of a liquid.

8.4.2. Laminar and turbulent flow.

If each fluid layer slips over the other, different layers do not get mixed then the flow called **laminar**. The velocity of flow at any point of fluid remains stable. The streamlines do not intersect each other.

The other kind of flow is named **turbulent**. The turbulent flow is unstable. The flow of fluid is curled and all layers of liquid mixed. The turbulent flow needs more energy than the laminar one.

English physicist *Reynolds* investigated the conditions under which a flow becomes laminar or turbulent. The transition from laminar flow into turbulent

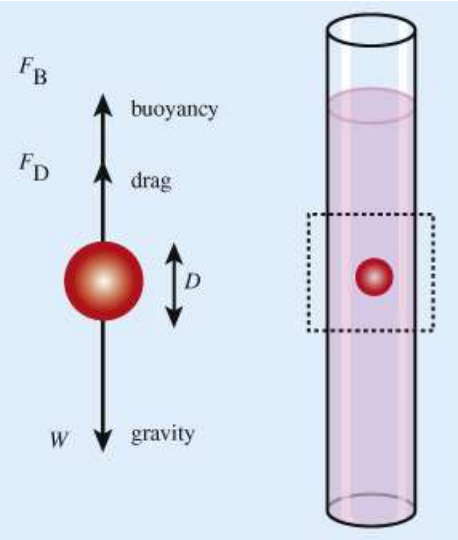

Fig. 8.12. The Stokes law.

one depends on the value of dimensionless quantity called the **Reynolds' number (Re)**. Reynolds'number for a liquid flowing in a cylindrical tube is defined by the formula:

$$\boxed{Re = \frac{\rho v D}{\eta}}$$

where v is an average velocity of flow; D is the diameter of a tube; ρ is density of fluid; η is the dynamic viscosity of fluid.

The transition from laminar to turbulent regime of flow is indicated by a critical Reynolds number (Re_{cr}), which depends on the exact flow configuration and is determined experimentally.

The critical value of Reynolds number for cylindrical tubes at which laminar flow turns into turbulent is 2000 – 2400.

8.4.3. Formula of change of fluid velocity along the tube radius $v = f(r)$.

Find linear and volume velocity of flow for steady-state stream of viscous fluid through a vessel with radius R, length l, pressure differential at its ends $p_1 - p_2$.

Mark elementary volume of a cylinder-like form with section S and length dx in fluid (fig.).

$$p = \frac{F}{S}, \qquad F = pS$$

$$F_{mouvement} = F_1 - F_2 = p_1 S - p_2 S = S(p_1 - p_2) = \pi \cdot r^2 (p_1 - p_2),$$

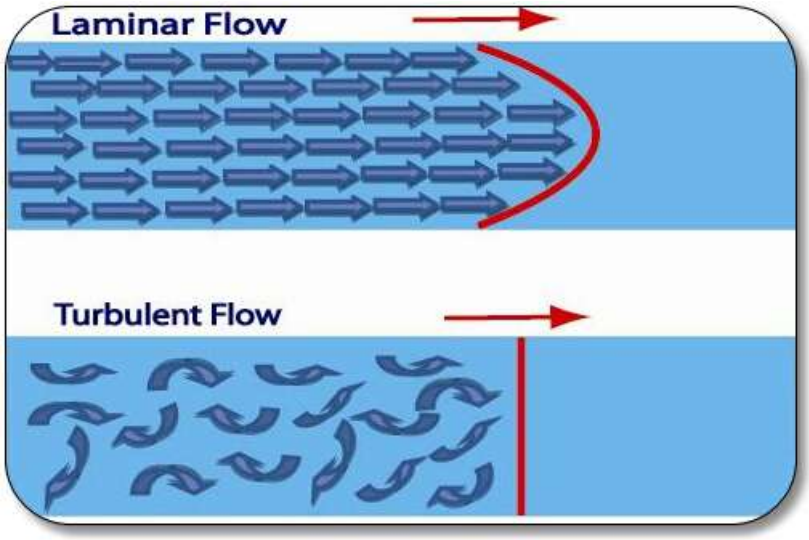

Fig. 8.13. Laminar and turbulent flow.

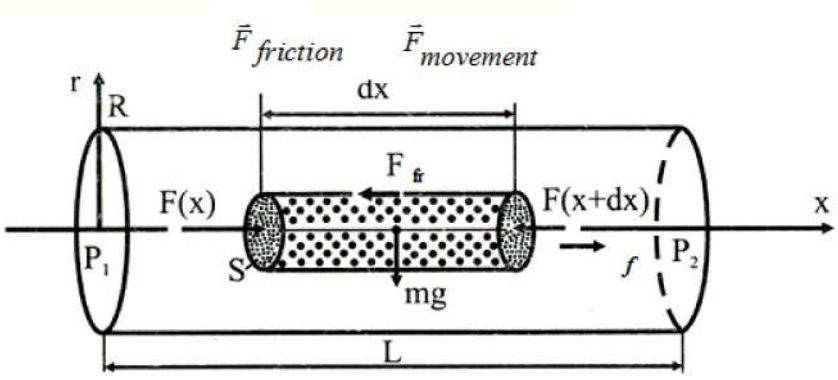

Fig. 8.14. Elementary volume of a cylinder-like form.

where S is area of sections.

$$F_{friction} = \eta \cdot \frac{dv}{dr} \cdot S = \eta \cdot \frac{dv}{dr} \cdot 2\pi r l,$$

where S is area of lateral surface of cylinder.

$$F_{movement} = -F_{friction}$$

$$\pi r^2 (p_1 - p_2) = -\eta \frac{dv}{dr} \cdot 2\pi r l$$

$$dv = -\frac{(p_1 - p_2)}{2\eta l} r \cdot dr$$

$$\int_0^v dv = -\frac{(p_1 - p_2)}{2\eta l} \int_R^r r \cdot dr$$

$$v = 0, \quad \text{when} \quad r = R$$

$$v = -\frac{(p_1 - p_2)}{2\eta l} \cdot \frac{r^2}{2}\Big|_R^r$$

$$v = -\frac{(p_1 - p_2)}{2\eta l} \cdot \left(\frac{r^2}{2} - \frac{R^2}{2}\right)$$

$$v = -\frac{(p_1 - p_2)}{2\eta l} \cdot \frac{1}{2} \cdot \left(-(R^2 - r^2)\right)$$

The dependence of the liquid layer velocity on the distance from the center of the pipe (r) will be follows:

$$\boxed{v = \frac{p_1 - p_2}{4\eta l} \cdot \left(R^2 - r^2\right)}$$

The particles moving along the tube (or pipe) axis ($r = 0$) have the maximum velocity:

$$\boxed{v_{max} = \frac{p_1 - p_2}{4\eta l} R^2}$$

8.8. Formula of Hagen-Poiseuille.

Mark cylinder layer with radius r and thickness dr (fig.8.15).

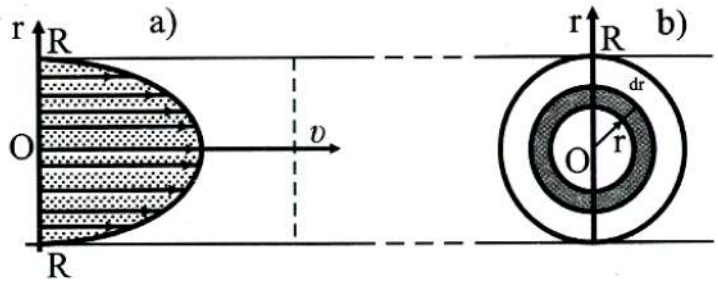

Fig. 8.15. The area of section this layer. Poiseuille's formula.

The area of section this layer:

$$dS = 2\pi r \cdot dr$$

$$dQ = v \cdot dS = v \cdot dS = v \cdot 2\pi r \cdot dr$$

$$Q = \int_0^R dQ = \int_0^R \frac{p_1 - p_2}{4\eta l} \cdot \left(R^2 - r^2\right) \cdot 2\pi r \cdot dr$$

$$Q = \frac{p_1 - p_2}{4\eta l} \cdot 2\pi \left(\int_0^R R^2 r \cdot dr - \int_0^R r^3 dr \right) = \frac{(p_1 - p_2)\pi}{2\eta l} \left(R^2 \cdot \frac{r^2}{2} \Big|_0^R - \frac{r^4}{2} \Big|_0^R \right)$$

$$Q = \frac{(p_1 - p_2)\pi}{2\eta l} \left(\frac{R^4}{2} - \frac{R^4}{4} \right) = \frac{(p_1 - p_2)\pi R^4}{8\eta l}$$

$$\boxed{Q = \frac{\pi r^4 (P_1 - P_2)}{8\eta l}} \qquad \textbf{(formula of Hagen-Poiseuille)}$$

η is dynamic coefficient of viscosity.

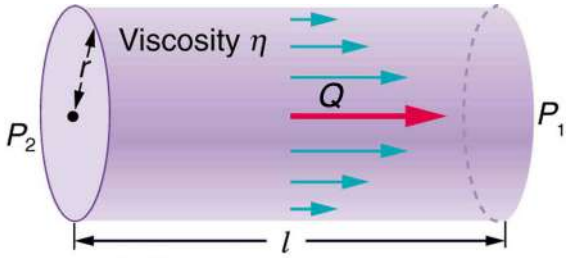

Fig. 8.16. The liquid flow in the pipe.

Quantity $X = \dfrac{8\eta l}{\pi r^4}$ is called **hydraulic resistance**.

$$\boxed{Q = \frac{\Delta p}{X}}, \qquad \textbf{(Poiseuille's formula)}$$

$$\Delta p = Q\dot{X}.$$

Sometimes an analogy can be drawn between the processes of the liquid flow in the pipe and electric current flow in a conductor. Then Q is the volume velocity, analogous to current I, and pressures difference $p_1 - p_2$ is analogous to potentials difference $\varphi_1 - \varphi_2$.

Taking into account that $R = \dfrac{\Delta\varphi}{I}$, where R is electrical resistance, quantity $X = \dfrac{\Delta p}{Q}$ is called hydraulic resistance.

Test questions

1. What do the terms "fluidity" define? "viscosity"?

2. Explain the mechanism of internal friction in the fluid and write down Newton's formula.

3. What is called a current tube?

4. What is the physical meaning of the continuity condition of the jet? How to get it out?

5. Derive the Bernoulli equation and explain its physical meaning.

6. What fluid flow is called laminar? turbulent?

7. What characterizes the Reynolds criterion?

8. How are Newtonian fluids different from non-Newtonian ones?

9. Derive the Poiseuille formula.

10. What physical phenomena can be quantified by using the Poiseuille formula?

11. Explain the essence of methods for determining the viscosity of a liquid.

Lecture No 9

Physical bases of hemodynamics.

Blood circulation is one of the most important processes occuring in human organisms. The section of biophysics studying laws of motion of blood on vascular system is named **hemodynamics** (Greek "haima" means blood). General laws of current of liquids investigated by hydrodynamics are established within the framework of classical physics and are the basis for description of complex hemodynamics processes in a human organism. However properties of blood in many respects are different from properties of liquids used in engineering and having elastic walls and repeatedly branching blood vessels considerably differ, for example, from system of water pipes. Therefore the biophysics considers only simplified model of blood circulation.

9.1. Elements of biomechanics of heart.

Let us consider the work and power of the heart. When contracting, the heart does work against the force of the blood pressure. Thus pushing the blood into the artery, and gives it kinetic energy. Work which done by the ventricle to push the blood per one contraction can be calculated using the formula $W = p \cdot \Delta V$, where p is the mean pressure in the ventricle at contraction, ΔV is a variation in the volume of the ventricle.

The kinetic energy of the blood pushed into the artery is $E_k = \dfrac{mv^2}{2}$. Calculate work of heart during time of systole.

$$W_{heart} = W_{left\ ventricle} + W_{right\ ventricle},$$

$W_{left\ ventricle}$ is the work of left ventricle;

$W_{right\ ventricle}$ is the work of right ventricle.

Work done by atrium being neglected, and assuming that work done by the

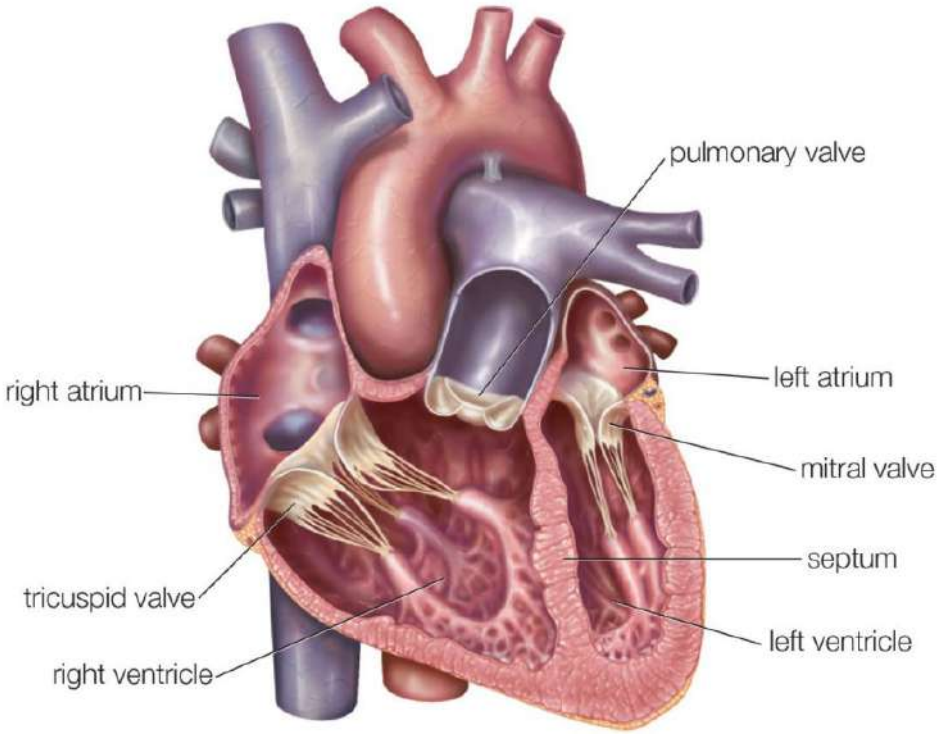

Fig. 9.1. The human heart.

right ventricle approximately equals 20% of that of the left one, we get

$$W_{right\ ventricle} = 0.2W_{left\ ventricle}$$

$$W_{left\ ventricle} = 1.2W_{left\ ventricle}$$

The work of heart is spent on moving of blood volume along aorta cross-section and given of blood of kinetic energy:

$$W_{left\ ventricle} = F\Delta l = p \cdot S \cdot \Delta l = pV + \frac{mv^2}{2},$$

$V = S \cdot \Delta l$ is heartbeat volume or stroke volume,

$m = \rho \cdot V$ is the mass of the pushed blood,

ρ is blood density, v is the mean velocity of the blood in the artery.

$$\boxed{W_{heart} = 1.2\left(PV + \frac{\rho V \cdot v^2}{2}\right) = 1.2V\left(P + \frac{\rho \cdot v^2}{2}\right)}.$$

The work done by the heart per one heart contraction approximately equals to 1 J.

$$W_{heart} = 1\ J$$

$$W_{heart} = 86400\ j_{(per24hours)}.$$

The time of one heart contraction being approximately 0.3 s, the power of heart approximately equals to 3.3 W.

$$N = \frac{W}{t_{systola}} = \frac{1}{0.3} \approx 3.3\ W.$$

9.2. The human cardiovascular system.

Haemodynamics studies the motion of blood in the circulatory system. One of features of physical model of cardiovascular system is elasticity of its walls. Elasticity is ability of material to test more or less significant elastic convertible deformations at rather small efforts. Walls of blood vessels differ on its structure. Aorta and large arteries have the walls consisting from muscular fibres, elastin and collagen. Elastin supposes deformations up to $200 - 300$ %, collagen up to 10 %. Arterioles consist completely only of muscular tissues which extensibility is much less. Walls of capillaries are not covered by elastic or muscular tissue.

Here is a list of True Facts about the human cardiovascular system:

• The heart, illustrated in the schematic in figure is the "pump" that drives blood through your blood vessels.

• The blood vessels are differentiated into three distinct types:

⊙ **Arteries**, which lead strictly **away from the heart** and which contain a muscular layer that elastically dilates and contracts the arteries in a synchronous way to help carry the surging waves of blood. This acts as a "shock absorber" and hence reduces the peak systolic blood pressure. Arteries split up the farther one is from the heart, eventually becoming **arterioles**, the very small arteries that actually split off into capillaries.

⊙ **Capillaries**, which are a dense network of very fine vessels (often only a single cell thick) that **deliver oxygenated blood throughout all living tissue** so that the oxygen can disassociate from the carrying hemoglobin molecules and diffuse into the surrounding cells in systemic circulation, or **permit the oxygenation of blood** in pulmonary circulation.

⊙ **Veins**, which lead strictly **back to the heart** from the capillaries. Veins also have a muscle layer that expand or contract to aid in thermoregulation and regulation of blood pressure as one lies down or stands up. Veins also provide "capacitance" to the circulatory system and store the body's "spare" blood; 60%

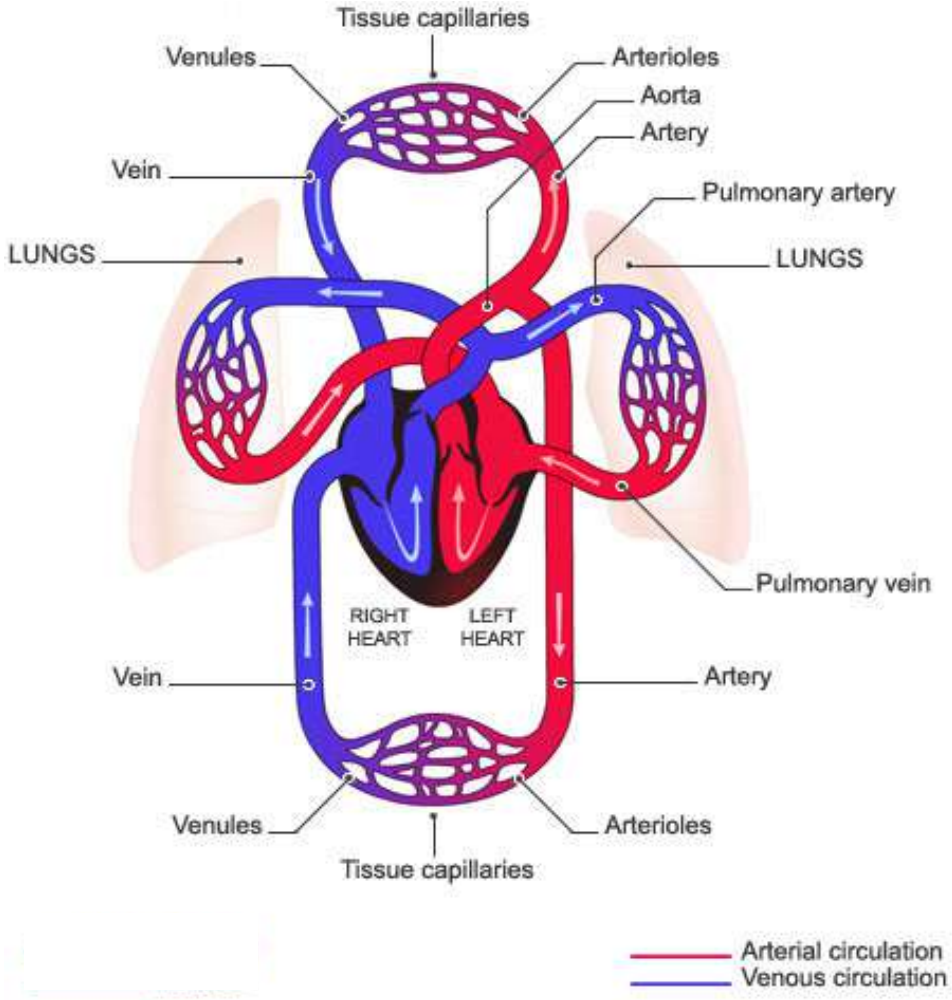

Fig. 9.2. The circulatory system.

of the body's total blood supply is usually in the veins at any one time. Most of the veins, especially long vertical veins, are equipped with one-way venous valves every $4 - 9$ cm that prevent backflow and pooling in the lower body during e.g. diastoli.

Blood from the capillaries is collected first in **venules** (the return-side equivalent of arterioles) and then into veins proper.

• There are two distinct circulatory systems in humans (and in the rest of the mammals and birds):

⊙ **Systemic circulation**, where oxygenated blood enters the heart via pulmonary veins from the lungs and is pumped at high pressure into systemic arteries that deliver it through the capillaries and (deoxygenated) back via systemic veins to the heart.

⊙ **Pulmonary circulation**, where deoxgenated blood that has returned from the system circulation is pumped into pulmonary arteries that deliver it to the lungs, where it is oxygenated and returned to the heart by means of pulmonary

veins. These two distinct circulations **do not mix** and together, **form a closed double circulation loop**.

• Blood pressure is generally measured and reported in terms of two numbers:

⊙ **Systolic** blood pressure. This is the **peak/maximum arterial pressure** in the wave pulse generated that drives **systemic circulation**. It is measured in the (brachial artery of the) arm, where it is supposed to be a reasonably accurate reflection of peak aortic pressure just outside of the heart, where, sadly, it cannot easily be directly measured without resorting to invasive methods that are, in fact, used e.g. during surgery.

⊙ **Diastolic** blood pressure. This is the **trough/minimum arterial pressure** in the wave pulse of systemic circulation.

"Normal" Systolic systemic blood pressure can fairly accurately be estimated on the basis of the distance between the heart and the feet; a distance on the order of 1.5 meters leads to a pressure difference of around 0.15 atm or 120 mmHg.

Blood is driven through the relatively high resistance of the capillaries by the **difference** in arterial pressure and venous pressure. The venous system is entirely a **low pressure return**; its peak pressure is typically order of 0.008 bar (6 mmHg). This can be understood and predicted by the mean distance between valves in the venous system – the pressure difference between one valve and another (say) 8 cm higher is approximately $\rho_b g \times 0.08 \approx= 0.008$ bar. However, this pressure is not really static – it varies with the delayed pressure wave that causes blood to surge its way up, down, or sideways through the veins on its way back to the atria of the heart.

9.3. Korotkov's method.

Blood flow in the circulatory system is laminar. In aorta it may become turbulent during a physical work which greatly increases velocity of blood. Blood flow may be turbulent also in arteries the cross-section area of which is diminished by some pathological process.

$$R_{cr} = 2000 \textbf{ (for blood)}, \qquad R_{cr} = 2300 \text{ (for water)}.$$

When Re is small ($Re < Re_{cr}$), the flow is laminar;
when Re is large ($Re > Re_{cr}$), the flow is turbulent.
The change of blood viscosity value (for example, in patients with anemia)

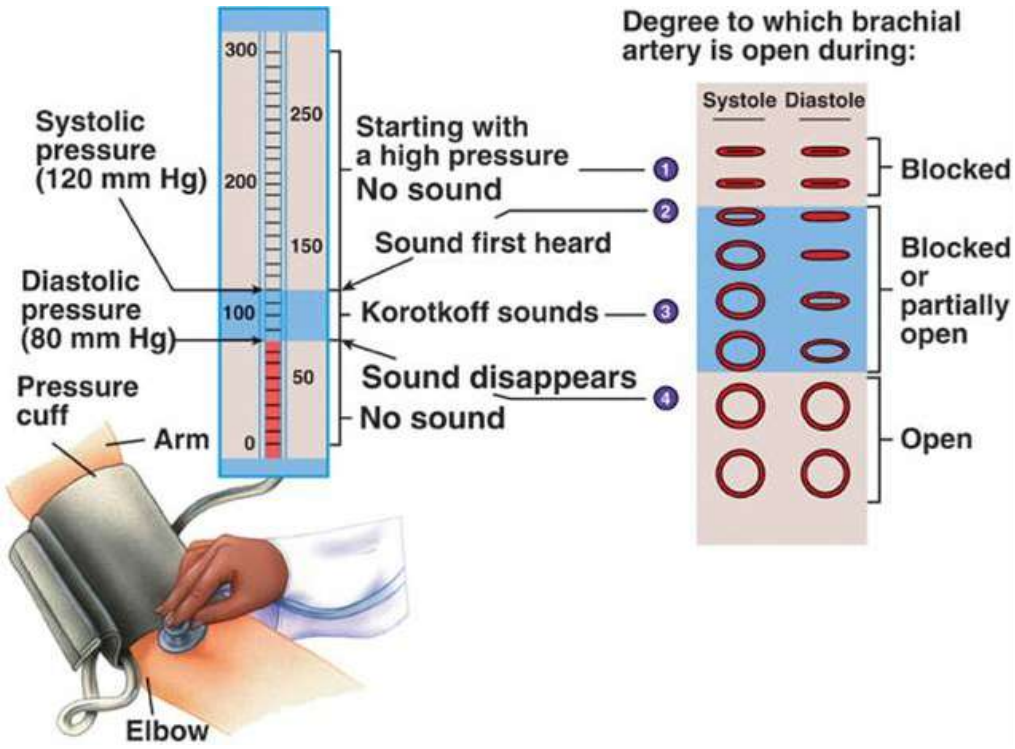

Fig. 9.3. Korotkov's method.

may be diagnosed owing to turbulent noises. It may be explained by the fact that at anemia viscosity coefficient decreases by a factor of $2-3$ and even more. Correspondingly, Reynolds number increases since $Re \sim 1/\eta$. As a result Reynolds number becomes greater than its critical value and the transition from laminar blood flow to turbulent one takes place.

Medical application of transition between laminar and turbulent flow of blood is connected with measuring of blood pressure by Korotkov's method.

In accordance with this method systolic (upper) pressure is measured at the moment when blood begins to squeeze through the hole in artery compressed by the cuff. Exactly at this moment noises appear resulting from turbulent flow of blood. Diastolic (lower) pressure is fixed at the moment when these noises disappear as a result of release of cuff and transition of flow from turbulent to laminar.

9.4. Atherosclerotic Plaque Partially Occludes a Blood Vessel.

Atherosclerosis – granular deposits of fatty material called **plaques** that attach to the walls of e.g. arteries and gradually thicken over time, generally associ-

ated with high blood cholesterol and lipidemia. The risk factors for atherosclerosis form a list as long as your arm and its fundamental causes are not well understood, although they are currently believed to form as an inflammatory response to surplus low density lipoproteins (one kind of cholesterol) in the blood.

In figure two arteries are illustrated.

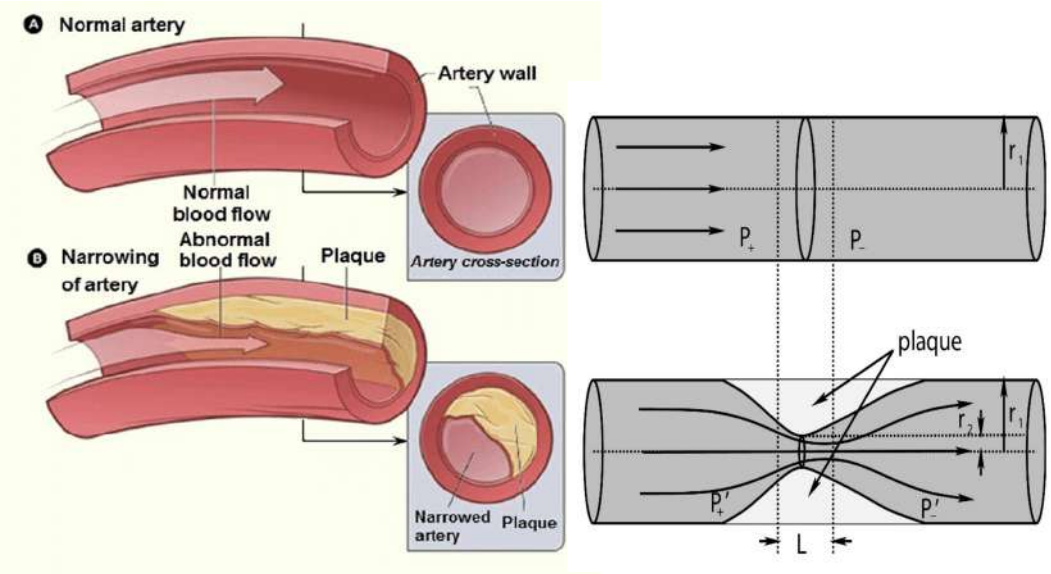

Fig. 9.4. Arteries.

Artery a) is "clean", has a radius of r_1 , and (from the Poiseuille Equation above) has a very low resistance to any given flow of blood. Because R_a over the length L is low, there is very little pressure drop between P_+ and P_- on the two sides of any given stretch of length L. The velocity profile of the fluid is also more or less uniform in the artery, slowing a bit near the walls but generally moving smoothly throughout the entire cross-section.

Artery b) has a significant deposit of atherosclerotic plaques that have coated the walls and reduced the effective radius of the vessel to $\sim r_2$ over an extended length L. The vessel is perhaps 90% occluded – only 10% of its normal cross-sectional area is available to carry fluid.

We can now easily understand several things about this situation. First, if the total flow in artery b) is still being maintained at close to the levels of the flow in artery a) (so that tissue being oxygenated by blood delivered by this artery is not being critically starved for oxygen yet) the **fluid velocity in the narrowed region is ten times higher than normal!** Since the Reynolds number for blood flowing in primary arteries is normally around 1000 to 2000, increasing v by a factor of 10 increases the Reynolds number by a factor of 10, causing the flow to become turbulent in the obstruction. This tendency is even more

pronounced than this figure suggests – I've drawn a nice symmetric occlusion, but the atheroma (lesion) is more likely to grow predominantly on one side and irregular lesions are more likely to disturb laminar flow even for smaller Reynolds numbers.

This turbulence provides the basis for one method of possible detection and diagnosis – you can *hear* the turbulence (with luck) through the stethoscope during a physical exam. Physicians get a lot of practice listening for turbulence since turbulence produced by *artificially* restricting blood flow in the brachial artery by means of a constricting cuff is basically what one listens for when taking a patient's blood pressure. It really shouldn't be there, especially during diastole, the rest of the time.

9.4.1. The total resistance of resistances in series and parallel.

The total resistance to flow posed by a number of individual resistances depends on the geometry of the system. Resistances in series are in sequence in the path of flow and all flow goes through all of the resistances one after another (Fig.9.5 A). The total resistance of resistances in series is calculated in according formula:

$$\boxed{R_{total} = R_1 + R_2 + \cdots + R_n}$$

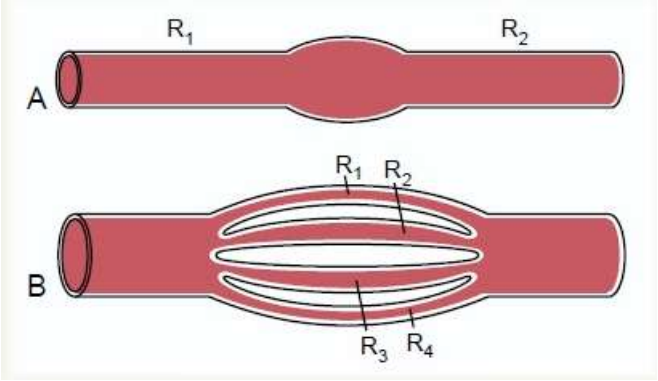

Fig. 9.5. Vascular resistances.

Parallel resistance exists when flow is divided between the resistances (Fig.9.5 B). Total resistance in this case is calculated in according formula:

$$\boxed{\frac{1}{R_{total}} = \frac{1}{R_1} + \frac{1}{R_2} + \cdots + \frac{1}{R_n}}$$

9.4.2. Distribution of mean pressure.

Mean arterial pressure is determined by the formula:

$$P_{mean} = P_d + \frac{P_s - P_d}{3},$$

where P_s is the systolic pressure, P_d is the diastolic pressure.

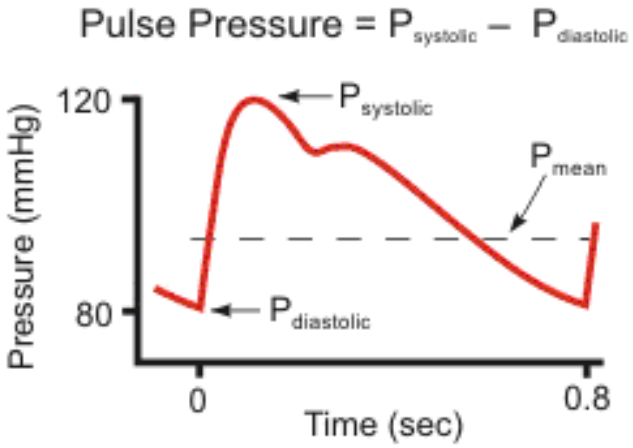

Fig. 9.6. Pulse pressure.

9.5. Pulse wave.

Pulse wave is the process of axial distribution of deformation of the cross-section sizes of a vessel in time.

The origin of pulse pressure is connected with reaction of elastic walls of a vessel to pulsing blood flow resulting from periodical work of the heat (Fig.).

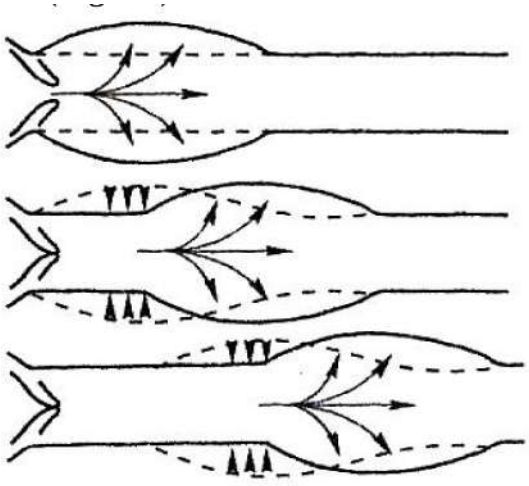

Fig. 9.7. Pulse wave.

Some characteristics of pulse wave

1. **Amplitude of pulse wave** $P_0(x)$ is named difference of maximum and minimum values of pressure in this point of vessel.

$P_0 = P_s - P_d$ is maximum amplitude of pulse wave in the beginning of aorta.

Amplitude of pulse wave is decreased in according formula:

$$\boxed{P_0(x) = P_0 \cdot e^{-x}}$$

where is damping coefficient.

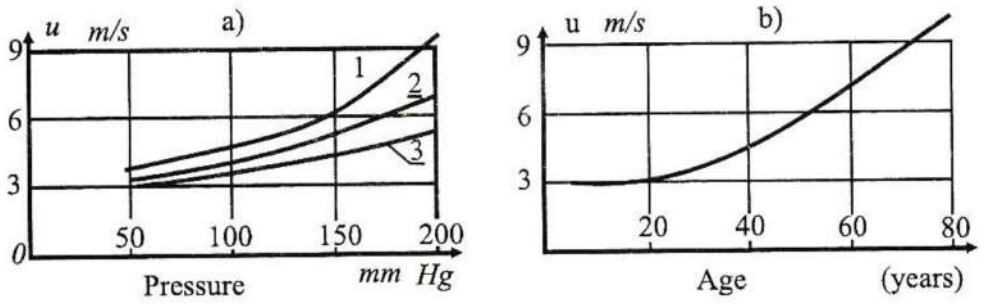

Fig. 9.8. Propagation speed change of pulse waves.

2. The value of propagate **velocity of pulse wave** is defined by the formula:

$$\boxed{v = \sqrt{\frac{E \cdot h}{\rho \cdot D}}},$$

where E is the Young's module of material of a wall of a vessel (or module elasticity);

h is the thickness of a wall of a vessel;

ρ is density of blood;

D is the diameter of vessel.

$v_{pulse\ wave} = 8 - 10\ m/s$, $v_{blood} = 0.3 - 0.5\ m/s$.

Blood speed is considerably lower than the speed of the pulse wave. Figure illustrates some examples of propagation speed change of pulse waves. The speed of pulse waves changes considerably at different vascular diseases. In this connection, its clinical determination allows getting additional information for estimation of functional state of vascular walls.

Test questions

1. What is the mechanism of fluid movement in the cardiovascular system?

2. What is a pulse wave?

3. What conditions of cardiovascular system characterize systolic and diastolic pressure?

4. What are the main mechanical properties that characterize vascular tissue?

5. Formulate the main features of blood flow in the human cardiovascular system.

6. What is the essence of the Korotkov method in determining blood pressure?

Lecture No 10

Biological thermodynamics

10.1. Thermodynamic System

Thermodynamics is branch of physics; it studies most common properties of macroscopic systems that are in thermodynamic equilibrium states, and processes of passage between them.

Thermodynamics studies mutual conversions of the different kinds of energy in macroscopic systems that are connected with transmission of heat or making work. Thermodynamics laws are universal for lifeless and living nature.

Thermodynamics is phenomenological science. Unlike atomically-molecular theories phenomenological ones learn general laws, not associated with concrete matter building, with the internal structure of the bodies or with the nature of the motion of the particles of which they are composed.

Thermodynamic system is a totality of macroscopic items (bodies or parts of bodies) that interchange energy in the form of heat and in the form of work with one another, as well as with the external medium (surroundings). For example, it can be: a cell, a heart, an organism, biosphere etc. There are following dominant classes of thermodynamic systems (fig.10.1):

a) isolated system when it can exchange neither energy nor matter with its surroundings.

b) closed system when it can exchange energy but cannot exchange matter with its surroundings.

c) open system when it can exchange either energy or matter with its surroundings.

Living organisms are open systems and only in separate parts of a cell there can exit conditions, typical for closed and even insulated system.

The state of a thermodynamic system is characterized by certain **thermodynamic parameters:**

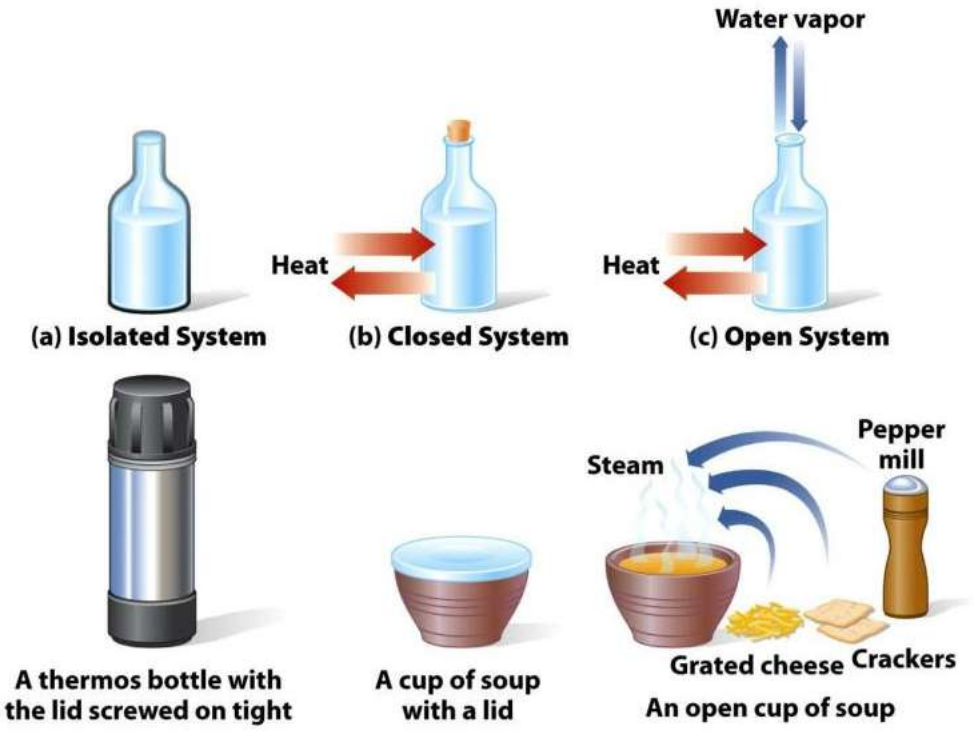

Fig. 10.1. Examples of thermodynamic systems

a) intensive parameters (independent on the mass or size of the system) are density, pressure, temperature, concentration;

b) extensive parameters (depend on the amount of material or size of the system) are volume (V), mass (m), internal energy (ΔU).

The state of a system is said to be **steady** if it does not change with time. Steady states are said to be equilibrium (quasi-static) ones if their invariability in time is no due to the occurrence of some process that is external in relation to the system.

Transfer of thermodynamic system from one state to another is called **thermodynamic process**.

Reversible and irreversible processes

A **reversible process** is defined as a succession of equilibrium states if the system passes from the initial state to the final state. The process is reversible if the system can return to its initial state without any residual exchanges in its surroundings.

An **irreversible process** passes through a series of nonequilibrium states and is accompanied with the irreversible changes; the system and its surroundings cannot be returned to their initial states. All real processes in nature are

irreversible.

10.1.1. 0th Law of Thermodynamics

Thermal Equilibrium

A system with many microscopic components (for example, a gas, a liquid, a solid with many molecules) that is isolated from all forms of energy exchange and left alone for a "long time" moves toward a state of *thermal equilibrium*. A system in thermal equilibrium is characterized by a set of macroscopic quantities that depend on the system in question and characterize its "state" (such as pressure, volume, density) that do not change in time.

Two systems are said to be in (mutual) thermal equilibrium if, when they are placed in "thermal contact" (basically, contact that permits the exchange of energy between them), their state variables do not change.

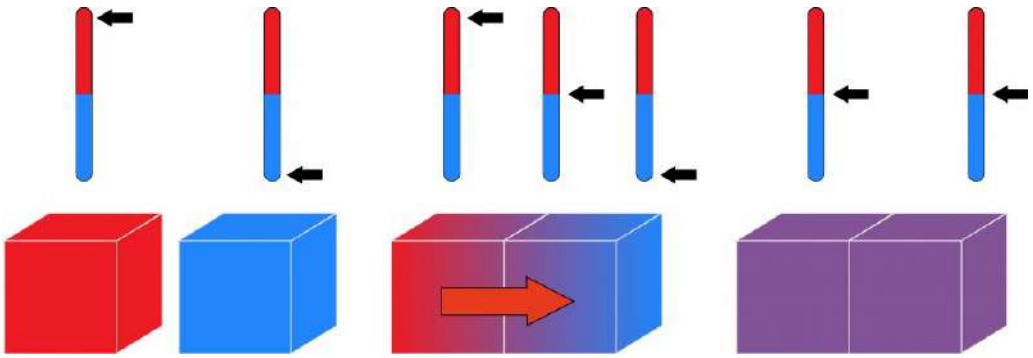

Fig. 10.2. Thermal equilibrium.

Zeroth Law of Thermodynamics

If system A is in thermal equilibrium with system C, and system B is in thermal equilibrium with system C, then system A is in thermal equilibrium with system B.

Temperature Scales

• **Fahrenheit:** This is one of the oldest scales, and is based on the coldest temperature that could be achieved with a mix of ice and alcohol. In it the freezing point of water is at 32^oF, the boiling point of water is at 212^oF.

• **Celsius or Centigrade:** This is a very sane system, where the freezing point of water is at 0^oC and the boiling point is at 100^oC. The degree size is thus $9/5$ as big as the Fahrenheit degree.

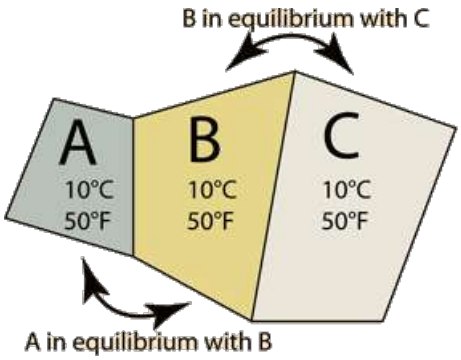

Fig. 10.3. Thermal equilibrium of three systems.

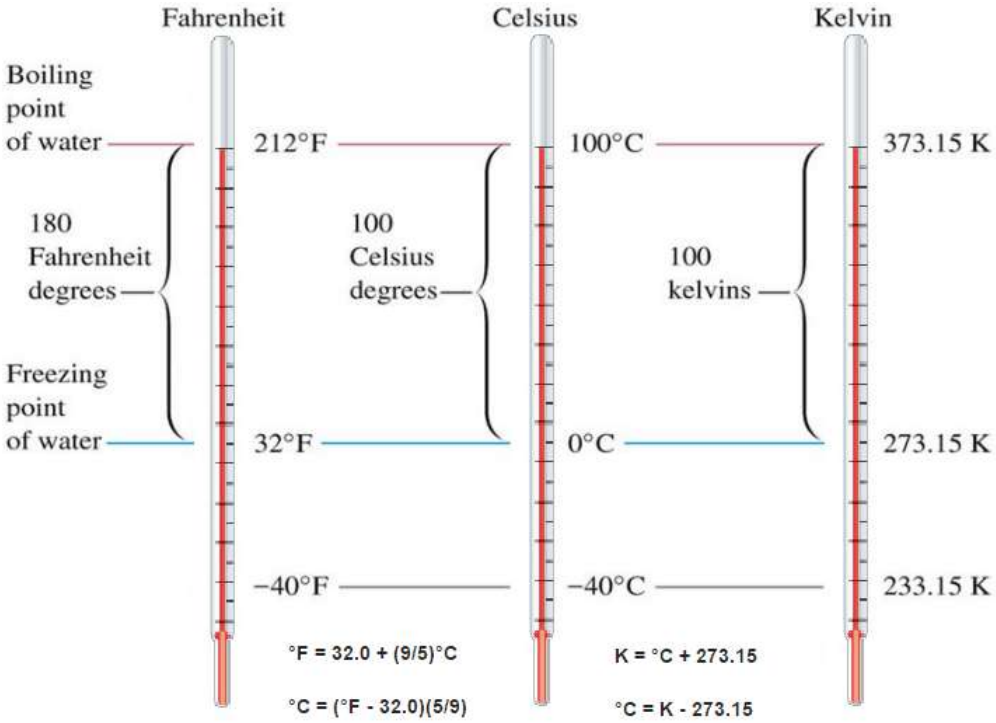

Fig. 10.4. Temperature Scales.

• **Kelvin or Absolute:** 0^oK is the lowest possible temperature, where the internal energy of a system is at its absolute minimum. The degree size is the same as that of the Centigrade or Celsius scale. This makes the freezing point of water at atmospheric pressure 273.16^oK, the boiling point at 373.16^oK.

Internal Energy

Internal energy is all the mechanical energy in all the components of a system. For example, in a monoatomic gas it might be the sum of the kinetic energies of all the gas atoms. In a solid it might be the sum of the kinetic and potential energies of all the particles that make up the solid.

Heat

Heat is a bit more complicated. It is internal energy as well, but it is internal energy that is transferred into or out of a given system. Furthermore, it is in some fundamental sense "disorganized" internal energy – energy with no particular organization, random energy. Heat flows into or out of a system in response to a temperature difference, always flowing from hotter temperature regions (cooling them) to cooler ones (warming them). Common units of heat include the ever-popular Joule and the calorie (the heat required to raise the temperature of 1 gram of water at 14.5^oC to 15.5^oC. Note that 1 cal = 4.186 J.

Heat Capacity

If one adds heat to an object, its temperature usually increases (exceptions include at a state boundary, for example when a liquid boils). In many cases the temperature change is linear in the amount of heat added. We define the heat capacity C of an object from the relation:.

$$\Delta Q = C\Delta T$$

where Q is the heat that flows into a system to increase its temperature by T.

Many substances have a known heat capacity per unit mass. This permits us to also write:

$$\Delta Q = mC\Delta T$$

where C is the specific heat of a substance. The specific heat of liquid water is approximately:

$$C_{water} = 1\frac{calorie}{kg \cdot^o C}$$

10.2. The First Law of Thermodynamics

The **First Law of Thermodynamics** presents a generalized law of conservation of energy; it considers the possible changes of the internal energy ΔU, heat Q and work W.

According to the law of conservation of energy the transition of the system that undergoes an infinitesimal change from the initial state to the final state the

small change in internal energy dU can be expressed as

$$dU = \delta Q - \delta W$$

where dU is change in internal energy;

δQ is a small amount of heat added to the system; δW is a small amount of work done by the system.

The change in internal energy of a system is equal to heat added to the system minus the work done by the system.

Here $dW > 0$, if work is done by system, and $dW < 0$, if work is done on the system.

The integral expression gives the exact area under the curve which is equal to the work:

$$W = \int_{V_1}^{V_2} pdV$$

Fig. 10.5. The work done during the expanding gas.

The elementary work dW done during the expanding gas is defined as follows

$$dW = Fdx = -pAdx = -pdV$$

where F is the force (for example, exerted by the gas on the piston of cross-sectional area A); p is pressure on the cylinder walls and piston; dV is the change in volume of the gas.

$$\Delta Q = \Delta U + W$$

The heat ΔQ transferred to the system is expended in changing its internal energy ΔU and in the work W done by the system to overcome external forces.

Cyclic Processes

Most of what we study in these final sections will lead us to an understanding

of simple heat engines based on gas expanding in a cylinder and doing work against a piston. In order to build a true engine, the engine has to go around in a repetitive *cycle*. This cycle typically is represented by a closed loop on a state e.g. $P(V)$ curve. A direct consequence of the 1-st law is that the **net work done by the system per cycle is the area inside the loop of the $P(V)$ diagram.** Since the internal energy is the same at the beginning and the end of the cycle, it also tells us that:

$$\Delta Q_{cycle} = W_{cycle}$$

the heat that flows into the system per cycle must exactly equal the work done by the system per cycle.

Adiabatic Processes are processes (PV curves) such that no heat enters or leaves an (insulated) system. The adiabatic condition:

$$PV^{\gamma} = const$$

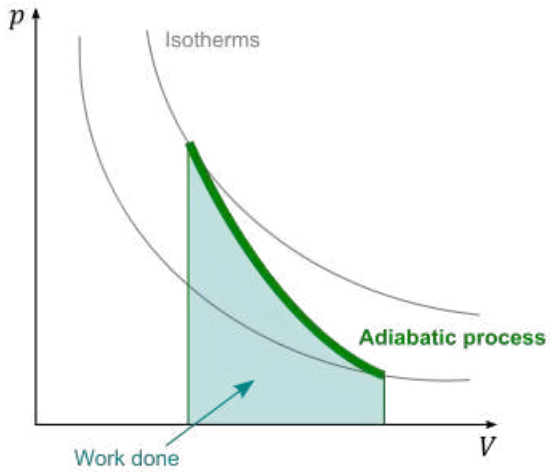

Fig. 10.6. Adiabatic Process.

Isothermal Processes are processes where the temperature T of the system remains constant.

$$PV = const$$

Isobaric Processes are processes that occur at constant pressure.

$$V \sim T$$

Isovolumetric Processses are processes that occur at constant volume

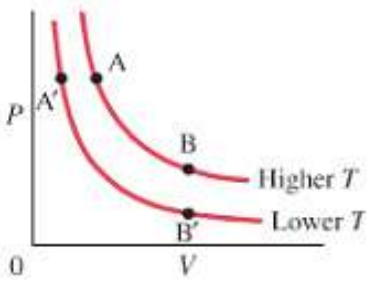

Fig. 10.7. Isothermal Process.

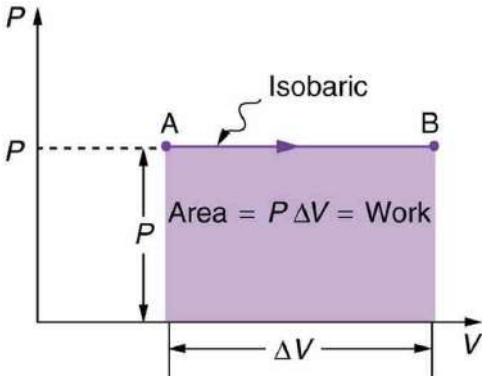

Fig. 10.8. Isobaric Process.

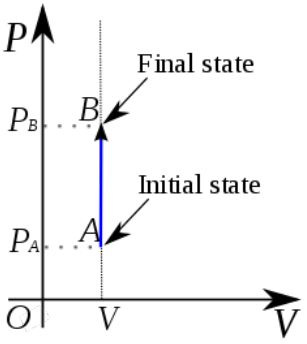

Fig. 10.9. Isobaric Process.

Work done by an Ideal Gas:

$$PV = NkT$$

where N is the number of gas atoms or molecules. Isothermal work at (fixed) temperature T_0 is thus:

$$W = \int_{V_1}^{V_2} \frac{NkT_0}{V} dV = NkT \ln \frac{V_2}{V_1}$$

Isobaric work is trivial. $P = P_0$ is a constant, so

$$W = \int_{V_1}^{V_2} P_0 dV = P_0(V_2 - P_1)$$

Adiabatic work is a bit tricky and depends on some of the internal properties of the gas (for example, whether it is mono- or diatomic).

A **heat engine** is a cyclic device that takes heat Q_H in from a hot *reservoir*, converts *some* of it to work W, and rejects the rest of it Q_C to a cold reservoir so that at the end of a cycle it is in the same state (and has the same internal energy) with which it began. The net work done per cycle is the area inside the PV curve.

The *efficiency* of a heat engine is defined to be:

$$\eta = \frac{W}{Q_H} = \frac{Q_H - Q_C}{Q_H} = 1 - \frac{Q_C}{Q_H}$$

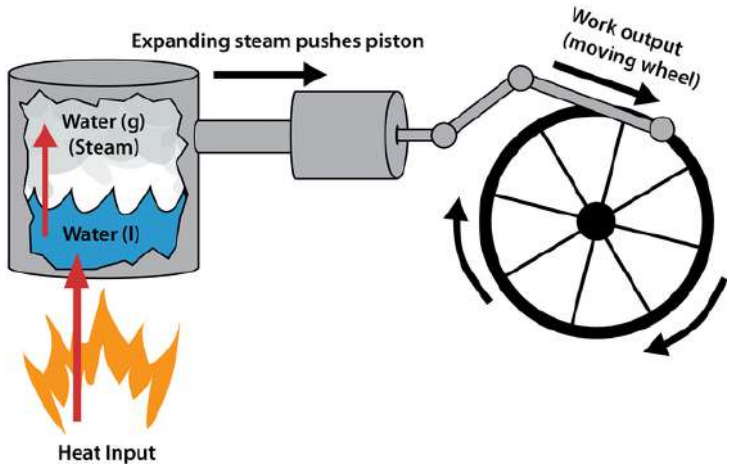

Fig. 10.10. Heat engine.

10.3. Entropy and its properties.

The directedness of thermodynamic processes is determined by **entropy**.

Rudolf Clausius postulated that the function S depends only on the initial and final states of a reversible process. For example, if S_A and S_B are the values of this function in the states A and B, a function S depends only on the initial

and final states A and B of a reversible process

$$S_B - S_A = \int_A^B \frac{dQ}{T}$$

where dQ is elementary heat that is absorbed by the system during small change of its state; T is temperature of heater.

The change in entropy, dS, between two equilibrium states is equal to the heat transferred, dQ, divided by the absolute temperature, T, of the system in this interval.

The entropy is the function S of the state of a system whose differential in an infinitesimal portion of a reversible process is equal to the ratio of an infinitesimal amount of heat Q, delivered to the system, to the absolute temperature T of the system. Thus:

$$dS = \frac{dQ}{T}$$

The unit of measurement of entropy is 1 J/K.

For an irreversible process the system expels more heat to the surroundings and the entropy increases

$$dS > \frac{dQ}{T}$$

Classical (equilibrium) thermodynamics describes systems which are in equilibrium or are undergoing *reversible* processes. Equilibrium systems are maintained without an exchange of energy or matter.

The main tendencies of the change of entropy can be formulated as:

1. *Entropy changes in a reversible process is zero; entropy remains constant in the isolated system during the reversible processes*

$$dS = 0; \qquad S = const.$$

2. *Entropy increases in the isolated system during the irreversible processes and reaches the maximal value*

$$dS > 0.$$

In such a way, any process in isolated system is accompanied with the increase of entropy

$$dS \geq \frac{dQ}{T}.$$

10.3.1. Non-equilibrium thermodynamics.

Living organisms are open thermodynamic systems that exchange either energy or matter with its surroundings. These organisms are able to absorb the energy, nutrients, to take part in gas exchange, to do work, to expel the products of metabolism. Thus, the viability of living organism presents the set of non-equilibrium processes. The result of such exchange of living organism with its surroundings is achievement of *stationary state*, which is characterized by the constant physical and chemical properties of thermodynamic system in spite of the absence of thermodynamic equilibrium.

10.4. The Second Law of Thermodynamics.

Although all natural processes must take place in accordance with the First Law, the principle of conservation of energy is, by itself, inadequate for an unambiguous description of the behavior of a system. Specifically, there is no mention of the familiar observation that every natural process has in some sense a preferred direction of action.

The Second Law of Thermodynamics states that the heat cannot of itself pass from a colder to a hotter body: i.e., it is impossible to transfer heat from a cold to a hot reservoir without at the same time converting a certain amount of work to heat. It is also impossible for any device that operates on a cycle to receive heat from a single reservoir and produce a net amount of work; it can only get useful work out of the heat if heat is at the same time transferred from a hot to a cold reservoir. This means that there is no possibility of a "perpetuum mobile" which is isolated. Also, from this it follows, that a reduction in the increase of entropy in a specified process, such as a chemical reaction, means that it is energetically more efficient.

The Second Law of thermodynamics defines that a system will always tend towards its chemical equilibrium or maximal entropy. If a spontaneous process leaves the overall system's energy the same ($dU = 0$), and the entropy of the end state is larger than the entropy of the initial state, the change in entropy dS must be larger than zero ($dS > 0$ for a spontaneous reaction).

If a process occurs in a closed system, the entropy, of the system increases for irreversible processes and remains constant for reversible processes. It never decreases. Although entropy may decreases in part of a closed system, there will always be an equal or larger entropy increase in another part of the system, so that the entropy of the system as a whole never decreases. This fact is one form

of the second law of thermodynamics and can be written as: $dS > 0$, where the greater than sign applies to irreversible processes, and the equals sign to reversible processes. This equation applies only to closed system.

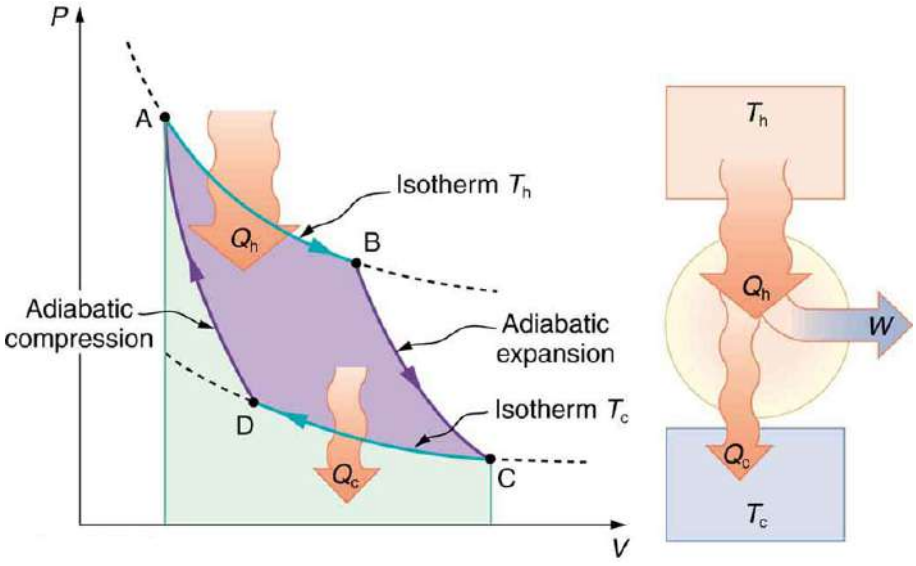

Fig. 10.11. The work done during the expanding gas.

In the real world almost all processes are irreversible to some extent because of friction, turbulence, and other factors, so entropy of real closed system undergoing real processes always increases. Processes in which the system's entropy remains constant are always idealizations.

Second law of thermodynamics – law of increase of entropy (law of degradation of energy):

in closed (isolated in the thermal and mechanical attitude) system the entropy either

• *remains to constant (if in system reversible, equilibrium processes proceed), or*

• *grows (at non-equilibrium processes) and*

• *in a state of equilibrium reaches a maximum.*

Other equivalent formulations:

1) transport of heat from a body of more cold to a body more heated is impossible without any changes in system or a surrounding medium;

2) it is impossible to create periodically acting machine which all activity would be reduced to a raising of some load (mechanical operation) and cooling of the thermal tank;

3) it is impossible to create a perpetuum mobile of 2-nd sort.

Kelvin-Planck statement of the Second Law of Thermodynamics

It is impossible to construct a cyclic heat engine that produces no other effect but the absorption of energy from a hot reservoir and the production of an equal amount of work.

Clausius Statement of the Second Law of Thermodynamics

It is impossible to construct a cyclic refrigerator whose sole effect is the transfer of energy from a cold reservoir to a warm reservoir without the input of energy by work.

Refrigerators (and Heat Pumps)

A **refrigerator** is basically a cyclic heat engine run backwards. In a cycle it takes heat Q_C in from a cold reservoir, does work W on it, and rejects a heat Q_H to a hot reservoir. Its net effect is thus to make the cold reservoir colder (refrigeration) by removing heat from inside it to the warmer warm reservoir (warming it still further, e.g. as a heat pump).

The **coefficient of performance** of a refrigerator is defined to be

$$COP = \frac{Q_C}{W}$$

Carnot Engine

The Carnot Cycle is the archetypical reversible cycle, and a Carnot Cycle-based heat engine is one that does not dissipate any energy internally and uses only reversible steps. Carnot's Theorem states that no real heat engine operating between a hot reservoir at temperature T_h and a cold reservoir at temperature T_c can be more efficient than a Carnot engine operating between those two reservoirs.

A Carnot Cycle consists of four steps:

a) Isothermal expansion (in contact with the heat reservoir)

b) Adiabatic expansion (after the heat reservoir is removed)

c) Isothermal compression (in contact with the cold reservoir)

d) Adiabatic compression (after the cold reservoir is removed)

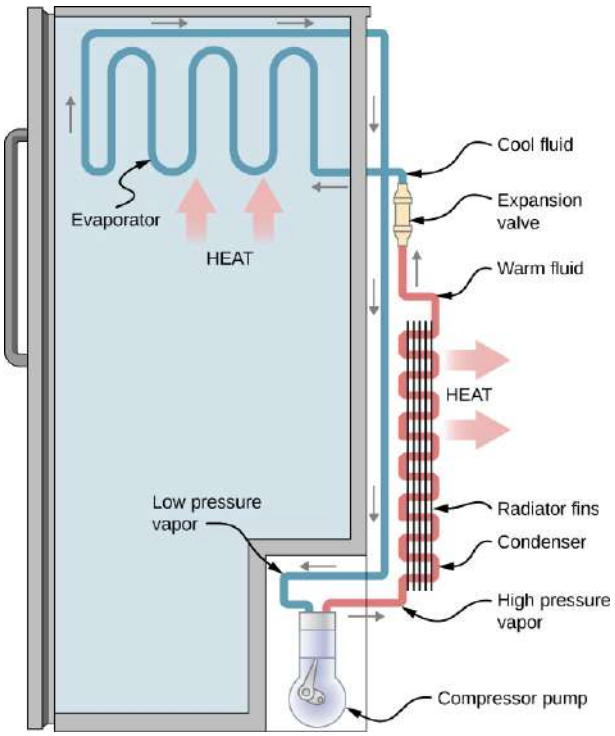

Fig. 10.12. Refrigerator.

The efficiency of a Carnot Engine is:

$$\eta_{Carnot} = 1 - \frac{T_c}{T_h}.$$

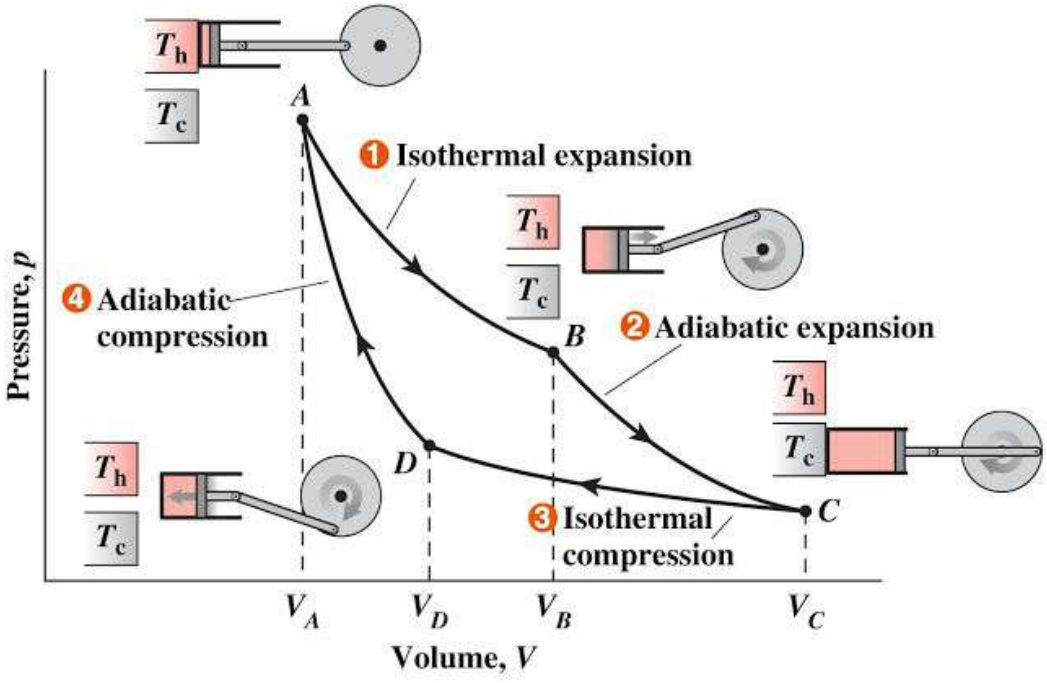

Fig. 10.13. The Carnot Cycle.

10.5. Entropy from a Statistical View.

Statistical mechanics explains entropy as the amount of uncertainty which remains about a system, after its observable macroscopic properties have been taken into account. For a given set of macroscopic quantities, like temperature and volume, the entropy is a function of the probability that the system is in various quantum states.

The entropy of a system can be defined in terms of the possible distribution of its molecules. For identical molecules, each possible distribution of molecules is called a microstate of the system. All equivalent microstates are grouped into a configuration of the system. The number of microstates in a configuration is the multiplicity W of the configuration. A basic assumption of statistical mechanics is that all the microstates are equally possible. Thus, configuration with a large multiplicity occur most often. The multiplicity W is defined as the thermodynamic probability.

Thermodynamics probability is a scalar physical quantity, which is numerically equal to number of microstates, for help of ones the given macrostate can be provided.

If the system consist of two parts and for first of them the thermodynamics probability is equal to W_1 , and for second of them – W_2, then, it is obviously, that for all system the thermodynamics probability W will be equal to ($W = W_1 \cdot W_2$).

The **entropy** is scalar physical quantity, which is numerically determined by formula $S = klnW$.

In 1877, Anstrain physicist *Ludwig Boltzmann* derived a relationship between the multiplicity W of a configuration of a system and the entropy S of the system in that configuration. This relationship is known as **Boltzmann's entropy equation**:

$$S = k \ln W + const,$$

where $k = 1.38 \cdot 10^{-23}\ J/K$ is the Boltzmann constant.

Every state of a system is characterized by definite values of internal energy and entropy. Therefore internal energy and entropy are called state functions. There are else three state functions used in thermodynamics that are called thermodynamic potentials (energetic quantities): enthalpy (H), Gibbs free energy (G), and Helmholtz free energy (F).

Potentials are derived from the energy balance equation on a thermodynamic

system and used to measure energy changes in systems as they evolve from an initial state to a final state. The potential used depends on the constraints of the system, such as constant temperature or pressure. Internal energy is the internal energy of the system, enthalpy is the internal energy of the system plus the energy related to pressure-volume work, and *Helmholtz* and *Gibbs* free energy are the energies available in a system to do useful work when the temperature and volume or the pressure and temperature are fixed, respectively. They are defined by the following equations:

$$H = U + PV;$$

$$G = H - TS;$$

$$F = U - TS,$$

where P is the pressure of the system and V is the volume of a system.

The last formula can also be written in the form:

$$U = F + TS.$$

In this equation the first term on the right side is called a free energy (a part of internal energy which could be used for performance of work), and the second term on the right side sometimes is called a bound energy (transition of a part of internal energy into a heat, that couldn't be transformed into a work).

Living systems are not at equilibrium, but metabolic processes are driven toward equilibrium. The total of all processes is spontaneous, while many individual steps are not spontaneous by themselves.

10.6. Prigogine's theorem.

This is the statement of the **Second Law of Thermodynamics** in the most general form **(Prigogine)**.

The entropy change d_iS due to irreversible processes in living organism can reach positive values only according to the Second Law.

The entropy change d_eS due to the exchange with the exterior can be positive, negative or equal zero.

If $d_eS \leq 0$ corresponds, for example, the disintegration of complex biological formations into simple structures which is accompanied with the cessation of vital processes and life as a whole.

If $d_eS < 0$ and $|d_eS| > |d_iS|$, total change of entropy becomes negative; this situation means complication of system organization, formation of more complex compounds, growth of tissues etc.

In such a way, contrary to equilibrium thermodynamics of isolated systems, which determines the entropy of equilibrium states only, nonequilibrium thermodynamic (of open systems) introduces the term of entropy flow.

One of the important achievements of thermodynamics of irreversible processes is that the temporal changes of entropy is discussed (For example, growth or aging of living organism).

Differentiation of equation leads to the expression:

$$\frac{dS}{dt} = \frac{d_eS}{dt} + \frac{d_iS}{dt}.$$

Here $\frac{dS}{dt}$ is called the *rate of entropy production.* The whole rate of entropy production in open system is equal a sum of entropy flow $\frac{d_eS}{dt}$ through the open system and the rate of entropy production $\frac{d_iS}{dt}$ due to irreversible processes inside the system.

Balance of entropy in stationary state of the system is determined as:

$$\frac{dS}{dt} = \frac{d_eS}{dt} + \frac{d_iS}{dt} = 0$$

Thus, a stationary state implies that the total entropy S of the system is constant.

Let's consider as an example of the relation of entropy change and irreversible processes an isolated system that consists of two parts of unequal temperature ($T_1 > T_2$).

Here dQ is amount of heat that flows from first part to second one during period of time dt. As soon as this is isolated system, $d_eS = 0$. The volumes of each part are constant also, that's why $dW = 0$.

The First Law of Thermodynamics can be written for each part as

$$dU_1 = dQ_1; \qquad dU_2 = dQ_2.$$

Since the heat lost (dQ_1) by first part of the system is equal to the heat gain

(dQ_2) by the second part, we can write

$$dQ_1 = dQ_2 = dQ.$$

The total change in entropy due to the irreversible processes, d_iS, can determined as

$$d_iS = -\frac{dQ}{T_1} + \frac{dQ}{T_2} = \left(\frac{1}{T_1} - \frac{1}{T_2}\right) dQ$$

The rate of entropy production can be obtained by dividing last equation by dt

$$\frac{d_iS}{dt} = \left(\frac{1}{T_1} - \frac{1}{T_2}\right) \frac{dQ}{dt}$$

Conclusion: the rate of entropy production due to irreversible processes $\frac{d_iS}{dt}$ is a product of the thermodynamic force Fk (the difference of inverse of temperature $\left(\frac{1}{T_1} - \frac{1}{T_2}\right)$) and thermodynamic flow $J_Q = \frac{dQ}{dt}$.

If the living organism is assumed as a system not far from the equilibrium state, than for this organism the Prigogine's theorem is valid: in steady state at fixed external parameters producing the entropy in open system is minimum, i.e.

$$d_iS \to min$$

If organisms are in steady state than the rate of entropy producing exceeds one in steady state. At return of organism is steady state there are processes of self-regulation that are accompanied by decreasing the rate of entropy producing. This assertion is called the **Prigogine's principle**.

10.7. Heat therapy.

Heat therapy, also called **thermotherapy**, is the use of heat in therapy, such as for pain relief and health. It can take the form of a hot cloth, hot water bottle, ultrasound, heating pad, hydrocollator packs, whirlpool baths, cordless FIR heat therapy wraps, and others. It can be beneficial to those with arthritis and stiff muscles and injuries to the deep tissue of the skin. Heat may be an effective self-care treatment for conditions like rheumatoid arthritis.

Heat therapy is most commonly used for rehabilitation purposes. The therapeutic effects of heat include increasing the extensibility of collagen tissues; decreasing joint stiffness; reducing pain; relieving muscle spasms; reducing inflammation, edema, and aids in the post acute phase of healing; and increasing blood flow. The increased blood flow to the affected area provides proteins, nutrients, and oxygen for better healing

Direct contact

Moist heat therapy has been believed to be more effective at warming tissues than dry heat, because water transfers heat more quickly than air. Clinical studies do not support the popular belief that moist heat is more effective than dry heat. Moist heat results in the perception that the tissue is heated more deeply. In fact, recent studies indicate that vasodilation, the expansion of the blood capillaries (vessels) to allow more blood flow, is improved with dry heat therapy. Expansion of the blood capillaries is the primary objective of heat therapy. Heat therapy increases the effect on muscles, joints, and soft tissue. Heat is typically applied by placing a warming device on the relevant body part.

Newer breeds of heat therapy devices combine a carbon fiber heater with a cordless rechargeable lithium battery and are built into the specific body wrap (i.e., shoulder wrap or back wrap) for targeted heat therapy. Such devices can be used as alternatives to chemical or plugged-in heating pads, but have not been shown to improve the clinical benefit. All devices primarily provide heat to promote vasodilation.

Infrared radiation

Infrared radiation is a convenient system to heat parts of our body. It has the advantage over direct contact in that radiation can heat directly the area where the blood capillaries and neuron terminals are. When heat comes from a direct contact source it has to heat the external layer of the skin, and heat is transferred to the deeper layer by conduction. Since heat conduction needs a temperature gradient to proceed, and there is a maximum temperature that can be safely used (around 42^oC), this means lower temperature where warming is needed.

Infrared (IR for short) is the part of the electromagnetic radiation spectrum comprised between 0.78 μm and 1 mm wavelength. It is usually divided into three segments:

IR-A, from 0.78 to 1.4 μm.

IR-B, from 1.4 to 3 μm.

IR-C, from 3 μm to 1 mm.

IR radiation is more useful than the visible radiation for heating our body, because we absorb most of it, compared to a strong reflection of visible light. Penetration depth of infrared radiation in our skin is dependent of wavelength. IR-A is the most penetrating, and reaches some millimeters, IR-B penetrates into the dermis (about 1 mm), and IR-C is mostly absorbed in the external layer of the epidermis (stratum corneum). For this reason the infrared lamps used for therapeutic purposes produce mainly IR-A radiation.

Mechanism of action, and indications

Heat creates higher tissue temperatures, which produces vasodilation that increases the supply of oxygen and nutrients and the elimination of carbon dioxide and metabolic waste.

Heat therapy is useful for muscle spasms, myalgia, fibromyalgia, contracture, bursitis.

Moist heat can be used on abscesses to help drain the abscess faster. A study from 2005 showed heat therapy to be effective in treating leishmaniasis, a tropical parasitic skin infection.

Heat therapy is also sometimes used in cancer treatment to augment the effect of chemotherapy or radiotherapy, but it is not enough to kill cancer cells on its own.

Heat therapy is contraindicated in case of acute injury and bleeding disorders (because of vasodilation), tissues with a severe lack of sensitivity, scars and in tissues with inadequate vascular supply (because of increased metabolic rate and demand which a tissue with poor blood supply may fail to meet resulting in ischemia).

The use of Heat therapy for deep-seated tissue can be treated with shortwave, microwave, and ultrasonic waves. This produces a high temperature that penetrates deeper. Shortwave produces a 27 MHz current, microwaves use 915 and 2456 MHz, and ultrasound is an acoustic vibration of 1 MHz. The way ultrasonic waves work is they selectively superimpose the incoming wave and increase the energy for absorption, and the significant part of the longitudinal compression gets converted into shear waves. When they are rapidly absorbed, the interface

between soft tissue and bone is selectively heated.

Therapeutic Benefits

Thermotherapy increase the extensibility of collagen tissues. Using heat, it can relieve the stiffness in joints in different cases. Shortwave and Microwave heat application may reduce muscle spasms, and selective heating with microwaves can accelerate absorption of hematomas. This will, in turn, allow the stiff muscle to stretch. Ultrasounds are not absorbed significantly in homogenous muscle. Heat therapy using hyperthermia has been used to treat cancer in combination with ionizing radiation.

10.7.1. Cryotherapy.

Cryotherapy, sometimes known as **cold therapy**, is the local or general use of low temperatures in medical therapy. Cryotherapy may be used to treat a variety of tissue lesions. The most prominent use of the term refers to the surgical treatment, specifically known as cryosurgery or cryoablation. Cryosurgery is the application of extremely low temperatures to destroy abnormal or diseased tissue and is used most commonly to treat skin conditions.

Cryotherapy is used in an effort to relieve muscle pain, sprains and swelling after soft tissue damage or surgery. It can be a range of treatments from the application of ice packs or immersion in ice baths (generally known as cold therapy), to the use of cold chambers.

While cryotherapy is widely used, there is little evidence as to its efficacy that has been replicated or shown in large controlled studies. Its long term side effects have also not been studied

Cryosurgery

Cryosurgery is the application of extreme cold to destroy abnormal or diseased tissue. The application of ultra-cold liquid causes damage to the treated tissue due to intracellular ice formation. The degree of damage depends upon the minimum temperature achieved and the rate of cooling. Cryosurgery is used to treat a number of diseases and disorders, most especially skin conditions like warts, moles, skin tags and solar keratoses. Liquid nitrogen is usually used to freeze the tissues at the cellular level. The procedure is used often as it is relatively easy and quick, can be done in the doctors surgery, and is deemed quite low risk. If a cancerous

lesion is suspected then excision rather than cryosurgery may be deemed more appropriate

Ice pack therapy

Ice pack therapy is a treatment of cold temperatures to an injured area of the body. Though the therapy is extensively used, and it is agreed that it alleviates symptoms, testing has produced conflicting results about its efficacy.

An ice pack is placed over an injured area and is intended to absorb heat of a closed traumatic or edematous injury by using conduction to transfer thermal energy. The physiologic effects of cold application include immediate vasoconstriction with reflexive vasodilation, decreased local metabolism and enzymatic activity, and decreased oxygen demand. Cold decreases muscle spindle fiber activity and slows nerve conduction velocity; therefore, it is often used to decrease spasticity and muscle guarding. It is commonly used to alleviate the pain of minor injuries, as well as decrease muscle soreness. The use of ice packs in treatment decreases the blood flow most rapidly at the beginning of the cooling period, this occurs as a result of vasoconstriction, the initial reflex sympathetic activity.

Ice is not commonly used prior to rehabilitation or performance because of its known adverse effects to performance such as decreased myotatic reflex and force production, as well as a decrease in balance immediately following ice pack therapy for 20 minutes. However, if ice pack therapy is applied for less than 10 minutes, performance can occur without detrimental effects. If the ice pack is removed at this time, athletes are sent back to training or competition directly with no decrease in performance.

Cold spray anesthetics

In addition to their use in cryosurgery, several types of cold aerosol sprays are used for short-term pain relief. Ordinary spray cans containing tetrafluoroethane, dimethyl ether, or similar substances, are used to numb the skin prior to or possibly in place of local anesthetic injections, and prior to other needles, small incisions, sutures, and so on. Other products containing chloroethane are used to ease sports injuries, similar to ice pack therapy.

Test questions

1. Explain the basic concepts of thermodynamics: thermodynamic system (open, closed), the amount of heat, internal energy.

2. Formulate and write down the first law of thermodynamics. What does this law express?

3. What is the difference between reversible and irreversible processes? Why are all real processes irreversible?

4. Is it possible the process by which heat is taken from the heater, fully converted into work?

5. In what direction can the entropy of a closed system change? open system?

6. What is the meaning of the concept of "entropy"?

7. Formulate and write down the second law of thermodynamics.

8. What are you known methods of heat therapy?

9. What are the applications of low temperatures in medicine?

Lecture No 11

Membrane transport.

11.1. Structure of biological membranes

A bioplast is an open system, which continuously exchanges different substances with the environment. This exchange is effected by transport of substances across the surface (plasma) membrane of the cell. The processes of substance transport ensure cell breathing, feeding, electric activity, and a host of other important functions.

Studying the electric properties of biological tissues and the electric phenomena related to the cell's vital functions is impossible without studying the processes of substance transport across the membrane, which, in turn, requires a preliminary consideration of the cell membrane structure.

Cell membranes are superfine structures with the thickness of 6 to 10 nm.

The membrane separates the cell content from the environment. The cell also has internal membranes, which surround such intracellular structures as the mitochondria, the endoplasm reticulum, the Golgi apparatus, lysosomes, and others. The plasma membrane ensures relative, self-sufficiency and stability of the cell's chemical composition, and, at the same time, it effects and regulates substance transport from the environment into the cell and therefrom to the environment. The plasma membrane also provides formation of contacts between cells.

The specialized plasma membrane of nerve cells plays a key role in transmission of nerve impulses. The plasma membranes of the epithelial cells of the gastrointestinal tract and kidneys take part in the processes of absorption and secretion.

The main components of membranes are proteins and lipids. Membranes also contain a small amount of carbohydrates.

Phospholipids comprise the main part of membrane lipids. They consist of two parts, viz. the "head", which is polar, and therefore has hydrophilic properties.

The other part is "tails" comprising the residuals of fatty acids. The tails are non-polar, and so they are hydrophobic. The structure and functions of the membrane proteins are very different. It is they who determine the functional diversity and specialization of biological membranes.

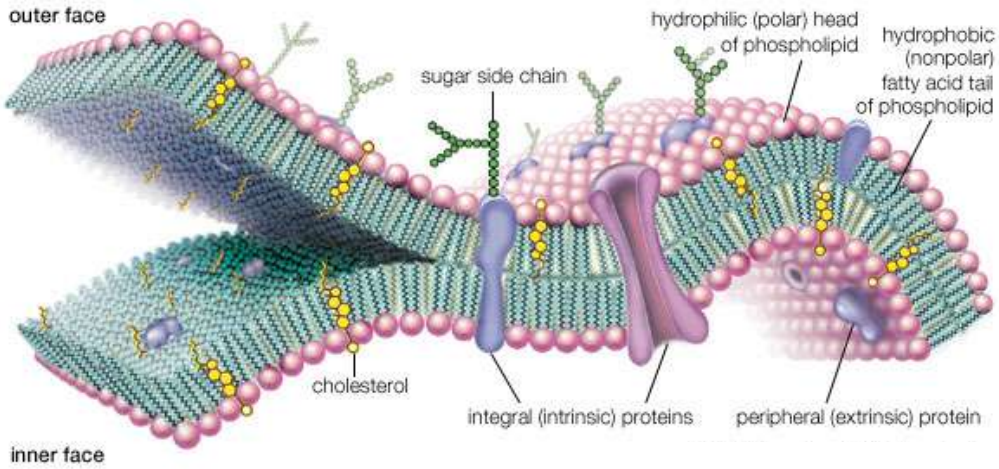

Fig. 11.1. The main components of membranes

Currently, the majority of scientists accept the liquid-mosaic model of the structure of biological membranes (Fig.11.1). According to this model, the key continuous part of the membrane is the double layer (bilayer) of phospholipids, in which the polar heads are directed out from the layer and come into contact with the water medium (the extracellular fluid outside the cell and the cytoplasm inside the cell), and the tails, which are directed into the bilayer, and linked by hydrophobic interactions.

Table 11.1

Parameter of a membrane

Properties of a membrane	**units**	**value**
Thickness	nm	$6.0 - 10.0$
Relative viscosity	–	$30 - 100$
Resistance	Ohm/cm	$10^2 - 10^5$
Strength of electric field	V/m	$10^6 - 10^9$
Parameter of refraction	–	1.6

Over a certain temperature range, the phospholipid structures combine the properties of liquids and crystals. As liquids, they possess fluidity, but as crystals, they maintain order in the arrangement and orientation of molecules. Therefore, phospholipid systems can possess liquid-crystalline properties. At a temperature

drop below a certain critical value, a transition of phospholipids from the liquid-crystalline state to gel occurs. At the temperature, which is common for the cell, the phospholipid bilayer is in the liquid-crystalline state According to the liquid-mosaic model, proteins do not cover the surface of the lipid bilayer completely, but as if float therein.

The character of arrangement with and relation to lipids classifies membrane proteins into peripheral and integral proteins. Peripheral proteins are arranged on the outer or inner, surface of the lipid layer and interact with the heads of phospholipids. Molecules of integral proteins have large hydrophobic sections. These sections are embedded within the lipid layer. Some integral proteins permeate the entire membrane. Integral proteins are connected with the membrane by hydrophobic interactions.

11.2. Functions of Membranes

1) Structural function. The outer membrane of a cell separates its contents from environment, determines the form of a cell. Sheaths (covers) of organoids divide contents of various regions of a cell, providing their different composition, and, accordingly, different biochemical reactions in them.

2) Protective function. If the cytoplasmic membrane protects contents of a cell from undesirable influences of environment that membranes of organoids carry out similar function in relation to own contents. For example, membranes of lysosomes protect a cell from its lytic enzymes.

3) Transport function. It is carried out due to presence of pores, canals and proteins - transmitting agents. Transport of substances on these structures is carried out selectively and depends on a structure of substances. Pores and canals have small diameter (less than 1 nm), are covered by proteins which polar groupings cooperate with substances penetrating into a pore, promoting or, on the contrary, hindering from their transmission. Hydrophobic groups cooperate with lipids of a membrane. The permeability of membranes can vary in dependence on a cell state.

4) Enzymatic function. The most part of enzymes of a cell is concentrated on membranes. Thus membranes organize biochemical reactions, including multistage (multiphasic, sequential). They provide strictly certain directivity of reactions due to the certain sequence in a locating of enzymes. A vivid example is respiratory chains of mitochondrions.

5) Electrogenic function. of membranes consists that the outer membrane of cells due to presence ATPase-carriers of ions creates a difference of concentrations $K+$, $Na+$, $Ca2+$ and $Cl-$ between cytoplasm and intercellular medium. The different permeability of a membrane for these and other ions results in appearance of a superfluous negative charge inside a cell (because of a diffusion of ions $K+$ in environment) and, hence, to a potential difference between the intrinsic and outside sides of a membrane.

ATPase (adenosine triphosphatase) - the common name of the proteins decomposing ATP on ADP and phosphate. Energy releasing at it is used for realization of endothermic (thermonegative) reactions.

6) Adhesion function. This property causes existence of multicellular organisms. The adhesion happens, despite of presence of the same charge of cells. The basic mechanisms of an adhesion:

a) Mechanical interaction - engagement (gearing) - of acting parts of membranes;

b) "Coagglutination" of membranes at participation of organic salts of the calcium cooperating with carboxylic groups of proteins and phosphatic groups of lipids;

c) Interaction of the proteins coating (covering) membranes – formation of peptide bonds.

7) Receptor function. On membranes there are receptor proteins or complexes which accept influences from environment (for example, cooperate with hormones or mediators) and accordingly vary metabolism of a cell.

8) Antigenic function consists that on a cellular membrane there are peptide structures, characteristic for cells of the given tissue of the given organism. These structures are intended for mutual "recognition" of cells. Due to their presence exercise of a host (immune) defence of an organism is possible.

11.3. Mechanisms of substance transport across the membrane.

The mechanisms (kinds) of substance transport can be classified into two groups.

1. Passive transport of substances. It is named so because it does not demand cell energy input for transport of substances across the membrane. The

mechanisms of passive transport include different kinds of diffusion: ordinary diffusion, facilitated diffusion, electrodiffusion, osmosis, filtration.

2. Active transport of substances. This kind of transport involves cell energy input.

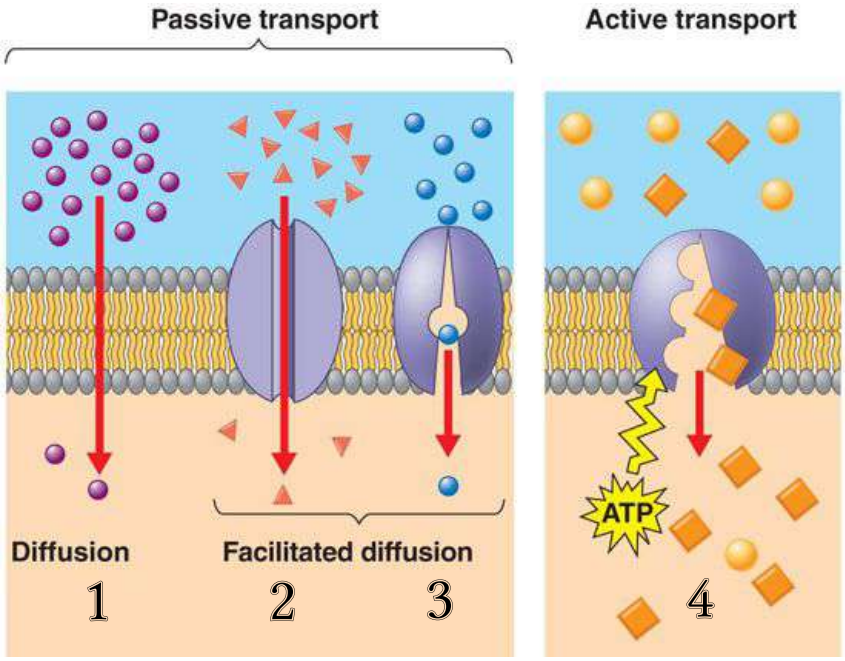

Fig. 11.2. Kinds of passive and active transport through membrane: 1 – simple diffusion through membrane; 2 – diffusion through channels; 3 – diffusion using carriers; 4 – active transport

11.3.1. Passive Transport

Passive transport through the membranes is related to some kind of diffusion. In the simplest understanding, diffusion is transfer of substance from a space with higher concentration of the substance to one with its lesser concentration.

Ordinary (simple) diffusion of uncharged particles is described by Fick equations. At ordinary diffusion, the direction of substance transfer is opposite to that of the substance concentration gradient. If the substance concentration in space depends only on one coordinate (for example, on coordinate x), then the first **Fick equation** is

$$\boxed{\frac{dm}{dt} = DS\frac{dC}{dx}}$$

where dm/dt is rate of a diffusion (mass of the substances born in unit of time through the area S); dC/dx – gradient of concentrations; x – distance between points in which concentrations C are observed; D – diffusion coefficient which

depends by substance nature, densities and viscosity of medium, temperature and other factors and is measured in cm^2/s. If one introduces the notations:

$$\Phi = \frac{dm}{dt}, \qquad J = \frac{\Phi}{S},$$

where Φ is the substance flow (more exactly, the substance mass flow), and J is the substance flow density, then **the first Fick's equation** can be written in the form

$$\boxed{J = -D\frac{dC}{dx}}$$

Direct diffusion across the phospholipid bilayer. The number of substances, which are capable of diffusing directly across the phospholipid bilayer, is very limited. These are some gases (oxygen and carbon dioxide) and non-electrolytes, which are soluble in lipids. Most substances that are transported across the membrane cannot diffuse across the phospholipid layer.

Diffusion through pores. Diffusion of most substances across the membrane is effected through specialized membrane formations known as pores or channels. They are formed from integral proteins. Diffusion of water, some simple organic compounds and ions occurs through the pores. It is typical for pore diffusion that the membrane permeability for a given molecule or ion depends on the channel and the molecule (ion) dimensions. For ions, transport also depends on the ion charge and the potential difference between the membrane sides.

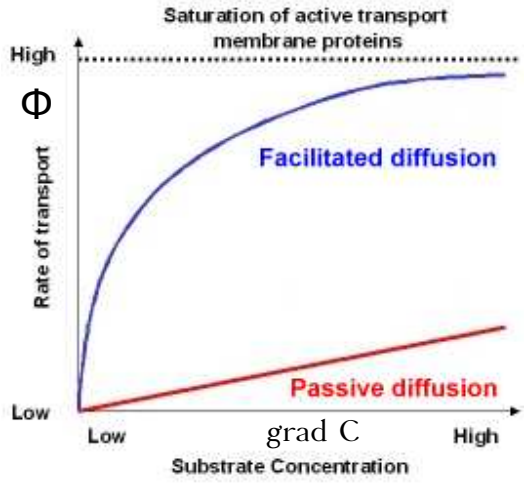

Fig. 11.3. Versions of passive transport

Facilitated (lightly) diffusion. Facilitated diffusion is effected by means of so-called proteins-carriers. At facilitated diffusion, the molecules of the transported substance form a complex compound with the molecules-carriers. Facilitated diffusion is characterized by the following: 1) permeability of membrane (a stream of substance Φ) is significantly (in 100 – 1000 and more time) higher than diffusions of Fick determined by the law; 2) saturation – a stream of substance Φ increases only up to achievement of some limiting value of gradient of concentration; 3) high specificity – the certain substance carried only (with similar structure); 4) the expressed sensitivity to low concentration of inhibitors

and activators of a carrier; 5) integrity with streams of ions through a membrane.

Electrodiffusion is diffusion of charged particles, namely, ions. The phenomenon of electrodiffusion involves not only availability of an ions concentration gradient in the medium, but also the presence of an electric field potential gradient. The electrodiffusion phenomenon is described by the **Nernst-Planck's equation**, which has the form

$$\boxed{J = -D\frac{dC}{dx} - UZF\frac{d\varphi}{dx}}$$

where U is ion mobility,

Z is ion charge expressed by elementary charges (i.e. valency);

F is Faraday constant; φ is electric field potential; $\dfrac{d\varphi}{dx}$ is potential gradient modulus.

Taking into account that ion mobility is equal to

$$U = \frac{D}{RT}$$

where R is universal gas constant, T is absolute temperature, the **Nernst-Planck's equation** can be rewritten as

$$\boxed{J = -D\Big(\frac{dC}{dx} + \frac{ZF}{RT}\cdot\frac{d\varphi}{dx}\Big)}$$

Osmosis is locomotion of a dissolvent (in biological objects – water) from range with smaller concentration of solute in range with the greater concentration. As a matter of fact, it is a diffusion of a dissolvent.

This appearance is observed in the presence of semipermeable membrane between specified areas.

• *Electroosmosis* is locomotion of a dissolvent in an electrical field if moleculas of a dissolvent are charged, or owing to the osmosis accompanying locomotion of dissolved charged particles.

• *Abnormal osmosis* happens at presence both a concentration gradient, and conditions for electroosmosis. If the electrical gradient interferes with osmosis, and locomotion of a dissolvent goes on an electrical gradient against concentration gradient, negative abnormal osmosis is observed. If electrical gradient only reduces rate of transmission on a concentration gradient or even enlarges, it speak about

a positive abnormal osmosis.

• *Filtration* is movement of water and solute molecules across the cell membrane due to hydrostatic pressure generated by the cardiovascular system. Depending on the size of the membrane pores, only solutes of a certain size may pass through it.

11.3.2. Active Transport

Active transport of substances of substances across the membrane is related to the action of ion channels. The channels are formed by integral membrane proteins whose hydrophilic polar groups are directed inward the channel. Most channels have intrinsic selectivity, which is the capacity to let only certain ions pass. Sodium, potassium, calcium, and chlorine channels are known.

Ion channel action is regulated by the cell. Due to conformational alterations of the proteins, which form the walls of the channel, it can switch from the open state to the closed one, and vice versa. One of the causes of such change, for example, is the membrane electric field change.

Active transport of substances, unlike passive transport, is related to the transfer of substances across membranes in the concentration gradient direction, i.e. from the field where the substance concentration is less to the field where it is higher. Such transfer is impossible without cell energy input.

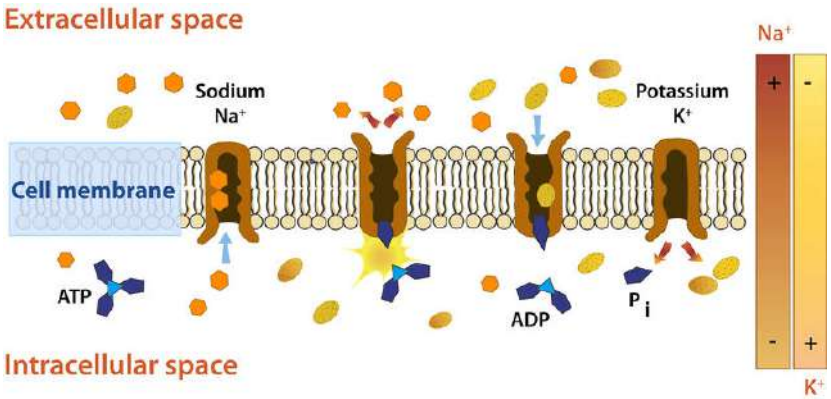

Fig. 11.4. Sodium-potassium pump

Active transport of substances is often related to the action of so-called ion pumps: the sodium-potassium pump of plasma membranes, the calcium pump of sarcoplasmic reticulum membranes, and the proton pump of mitochondria and chloroplasts. Active transport includes transfer of different substances, for example, sugars, across the membranes of intestine epithelium cells. We will consider

the sodium-potassium pump ($Na+-K+$- pump).

Three sodium ions and two potassium ions are usually transported during one pump action cycle, though under artificial conditions this relation can be another one.

Sodium and potassium ions are transported across the membrane by the complex integral protein-carrier that hydrolyzes ATP, whose energy activates the $Na+-K+$-pump. This protein (enzyme) is known as the transporting ATPase or $Na+-K+$-ATPase. Ion transport is caused by conformational changes in the protein-carrier.

Ion transport during sodium-potassium pump action results in an essential difference in sodium and potassium concentrations in and beyond the cell. One distinguishes intracellular and extracellular concentrations of some other ions, particularly chlorine ions. Such a difference in intracellular and extracellular concentrations of ions with different charges yields a difference of potentials between the outer and inner surfaces of the plasma membrane.

11.4. Biopotentials.

In a state of rest, the cell's inner medium is charged negatively relative to its environment. The difference of potentials between the inner and outer sides of a membrane is known as the membrane potential ($\varphi = \varphi_i - \varphi_0$).

If a cell is in a state of rest, the membrane potential is known as the resting potential (φ_r). The resting potential of different cells of various organisms can have different negative values in the range of about – 50 to – 100 mV.

At present, the best model that describes the formation of the resting potential is the model, which assumes that, in the stationary state, the sum of flows of basic ions ($K+$, $Na+$, $Cl-$) is equal to zero, i.e.

$$\Phi_K + \Phi_{Na} + \Phi_{Cl} = 0$$

Proceeding from this equation, and using the Nernst-Planck equation, one obtains the Goldman-Hodgkin-Katz formula for the membrane resting potential:

$$\varphi_r = \frac{RT}{F} \ln \frac{P_K[K^+]_o + P_{Na}[Na^+]_o + P_{Cl}[Cl^-]_i}{P_K[K^+]_i + P_{Na}[Na^+]_i + P_{Cl}[Cl^-]_o}$$

where φ_r is the membrane resting potential; P_K, P_{Na}, and P_{Cl} are the membrane

permeability coefficients for potassium, sodium and chlorine ions, respectively; $[K^+]$, $[Na^+]$ and $[Cl^-]$ are active concentrations of potassium, sodium and chlorine ions, respectively.

In this formula, the membrane permeability for an ion is considered as the proportionality factor between the density of flow of these ions across the membrane and the difference of concentrations of these ions on both sides of the membrane. In the simplest model of ion diffusion across the membrane, the permeability of the membrane is equal to

$$p = \frac{D \cdot k}{\lambda}$$

where k is the distribution factor of the particles of substance among the membrane and the environment (usually water); λ is membrane thickness.

The permeability of a cell membrane for different kinds of ions is essentially different. If one takes membrane permeability for potassium as unit, then the following relation: $P_K : P_{Na} : P_{C1} = 1 : 0.04 : 0.45$ can be written. This relation shows that potassium ions have a key role in forming of the resting potential.

The process of action potential formation is related to a significant and non-simultaneous change of membrane permeability for sodium and potassium ions. First, after reaching the threshold value, the sodium channels open, and the bulk of sodium ions pour into the cell. This causes depolarization and subsequent membrane potential reversion. Then the potassium channels open and the sodium channels close. The bulk of potassium ions exit the cell. Then the $Na+$- $K+$- pump activates to recover the sodium and potassium concentrations, which are close to the initial

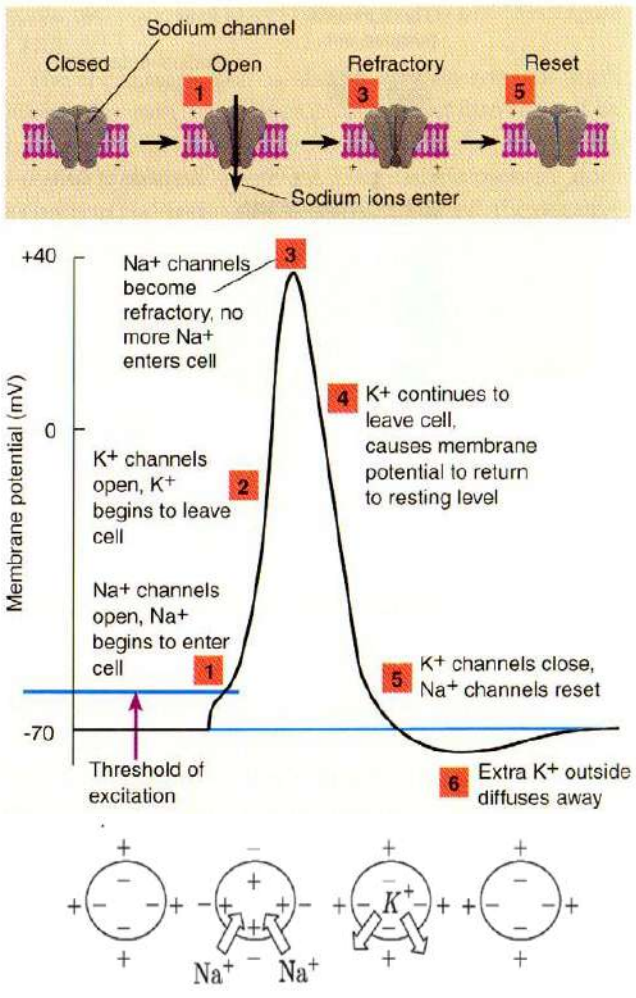

Fig. 11.5. Action potential generation.

ones. As a result, the membrane potential again becomes negative, i.e. membrane repolarization occurs.

I phase – a phase of depolarization: change of polarity of membrane potential differences;

II phase – a phase of repolarization: reduction of amplitude potential of action, and restoration of polarity of membrane potential differences.

III phase – a phase of hyperpolarization: value of module of membrane potential differences more than stationary value of potential of rest.

Duration, the form and value of potential of action essentially differ for membranes of various cells.

Action potential generation in a membrane section causes membrane potential changes in adjacent sections of the membrane. If these changes exceed the threshold value, this, in turn, causes action potential generation in these adjacent sections. As a result, the action potential wave propagates across the membrane. In a muscle cell, this process causes its contraction; in the nerve cell, action potential wave propagation across the membrane represents nerve impulse transmission.

Test questions

1. What processes are called transport phenomena?

2. What is the diffusion mechanism? Give examples of experiments diffusion in various environments.

3. Show the derivation of the Fick equation. Explain his physical meaning.

4. Show the derivation of the Nernst-Planck equation. Explain its physical meaning.

5. What are the different types of transport substances through the membrane?

6. What is active transport?

Lecture No 12

Electric Dipole. Physical bases of electrocardiography.

12.1. Basic characteristics of electric field

Electric field is a field generated by electric charges and performing their interaction.

Electrostatic field is a field generated by fixed electric charges, the quantity of which does not change with the interval of time.

Major characteristics of electric field:

1. The force characteristic of electric field is called **strength tension** or **intensity of electric field** (E). In some point of the electric field the intensity is equal to the ratio of the force (F) effecting from the electric field on a charge (q) placed in this point to the value of this charge, i. e.

$$\vec{E} = \frac{\vec{F}}{q}$$

q is a test charge (a unit of positive charge);

F is a force acting from the direction of field on the test charge located at the given point.

E is a vector physical quantity being the force characteristic of the electric field.

The dimension of strength: $[E] = \frac{N}{C} = \frac{V}{m}$

In metric system (SI) the unit of intensity is volt per meter (V/m), the unit of charge is coulomb (C).

Schematically, electric field is represented using strength lines (force lines).

Force lines are lines, tangens to which at each point of electric field coincides with the field vector at this point.

Forse lines of electrostatic field are not closed: they start on the positive charges and finish on the negative charges or continue into infinity (fig.12.1).

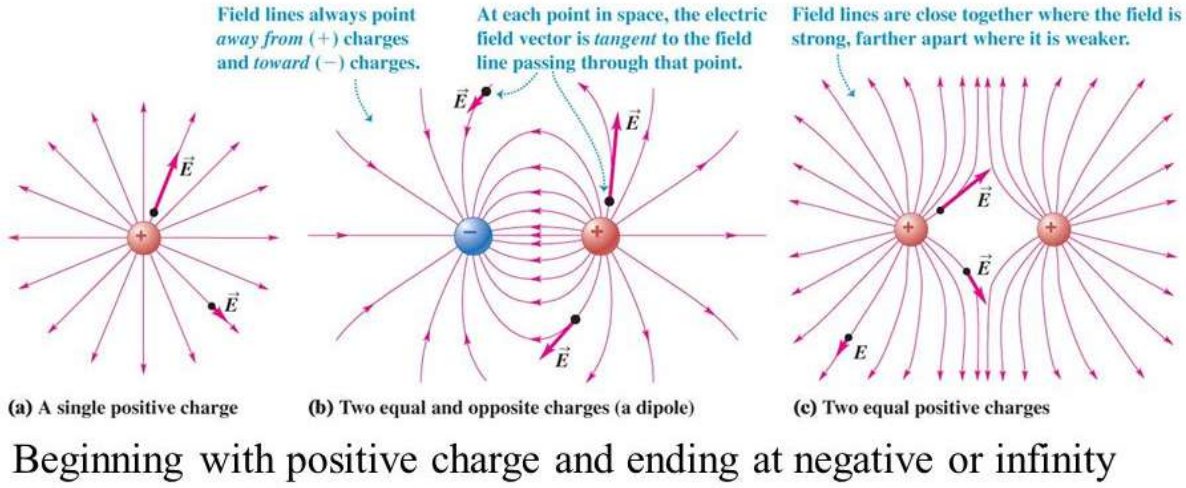

Fig. 12.1. Electric field lines - examples

2. The power characteristics of electric field are called **potential** (φ), **potential difference** ($\Delta\varphi$).

The potential of electric field (φ) is a scalar quantity being the power characteristic of the electric field. In some point the potential of the electric field is equal to the ratio of the potential energy (W_p) of a charge placed in this point to the value of this charge, i. e.

$$\varphi = \frac{W_p}{q}, \qquad [\varphi] = \frac{J}{C} = V,$$

where W_p is potential energy of test charge q located at the given point of the field.

The electrostatic field is potential, that is why work, executed the field, equals diminishing of potential energy:

$$A = W_1 - W_2 = q(\varphi_1 - \varphi_2) = qU,$$

where $U = \varphi_1 - \varphi_2$ is a difference of potentials, it is named also **tension** or **voltage**:

$$\Delta\varphi = \varphi_1 - \varphi_2 = \frac{A_{1-2}}{q}$$

where A_{1-2} is the work done when a unit electric charge is moved between two points having different electric potentials.

The SI derived unit of electric potential difference is the volt $[\varphi] = V$.

Geometric position of points having equal potential is called **equipotential surface** (fig.12.2 show them in dotted lines).

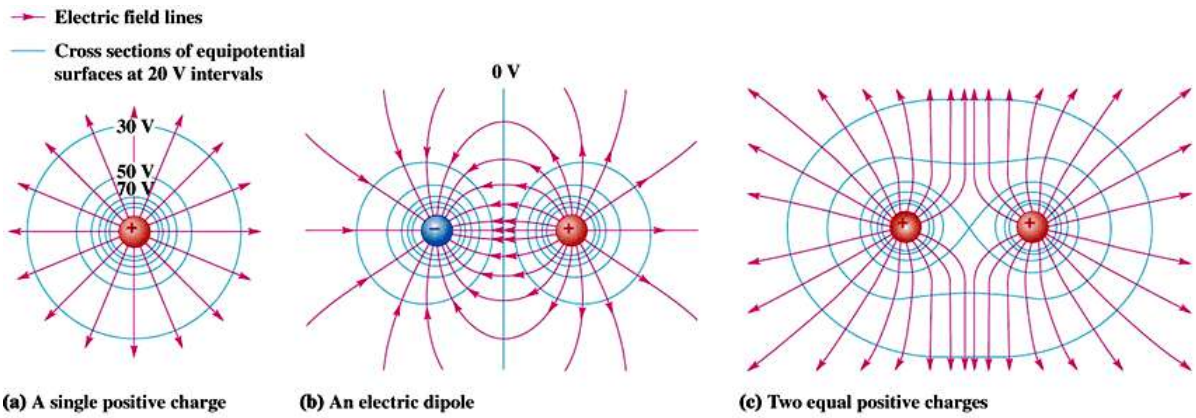

Fig. 12.2. Equipotential surface

Correlation between characteristics of uniform field ($E = const$)

The intensity is connected with the potential by the ratio

$$E = -grad\ \varphi$$

where grad φ is the differential operator named a gradient and expressed through the partial derivatives on coordinates. It is rather

$$grad\ \varphi = \frac{\partial\varphi}{\partial x}\vec{i} + \frac{\partial\varphi}{\partial y}\vec{j} + \frac{\partial\varphi}{\partial z}\vec{k}$$

where i, j and k are the vectors of unit length (orts) directed along coordinate axes OX, OY, and OZ, accordingly. Thus components of the vector E along the specified coordinate axes are equal to $E_x = -\frac{\partial\varphi}{\partial x}$, $E_y = -\frac{\partial\varphi}{\partial y}$, $E_z = -\frac{\partial\varphi}{\partial z}$ accordingly.

$$\boxed{E = -\frac{\Delta\varphi}{d}}, \qquad [E] = \frac{V}{m},$$

where d is distance along direction E between points with potentials φ_1 and φ_2. Sign "-" in formula indicated that vector is directed towards decrease of potential.

$$E = \frac{\varphi_1 - \varphi_2}{d} = \frac{U}{d}$$

The direction of the vector of the potential gradient coincides with the direction of the fastest increase of the potential.

Electric charges and the variable magnetic field generate the electric field. The intensity of the electric field created in some point by an isolated point charge q is calculated by the formula

$$E = \frac{|q|}{4\pi\varepsilon\varepsilon_0 r^2}$$

where ε is the *dielectric permittivity of medium*; ε_0 is the electric constant ($\varepsilon_0 = 8.85 \cdot 10^{-12}\ \ F/m$), and r is the distance from the point in which the intensity of the field is determined up to the point in which there is a charge.

At the same time the potential of the field in the given point is calculated by the formula

$$\varphi = \frac{q}{4\pi\varepsilon\varepsilon_0 r}.$$

If the field is created by several charges its characteristics are calculated from **the principle of superposition of fields**. According to this principle the intensity of the electric field, created in some point by several charges, is equal to the vector sum of intensities of the fields created in this point by each of charges, and the potential is equal to the scalar sum of potentials created by each of charges, i. e.

$$E = E_1 + E_2 + \cdots + E_n$$

$$\varphi = \varphi_1 + \varphi_2 + \cdots + \varphi_n$$

where E and φ are the resulting intensity and potential of the field created by charges; E_1, E_2, ... ,E_n are the intensities of fields created by each of the charges, φ_1, φ_2, ..., φ_n are the potentials of fields created by each of the charges. Based on the principle of superposition of fields it is possible to calculate characteristics of the field, created by various systems of charges.

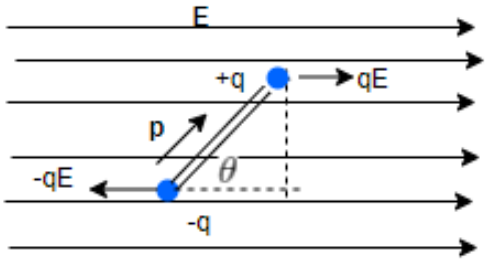

Fig. 12.3. Equipotential surface

Electric dipole is aggregate of two-point charges ($+q$ and $-q$) of the opposite sign equal by quantity and placed at a small distance d from one another, which is called dipole arm.

The characteristic of dipole: *electric (dipole) moment* () is a vector directed from the negative to the positive end, with the amount (P) that is equal to the product of charge q and the distance between them d:

$$\boxed{P = q \cdot d}, \quad [P] = C \cdot m.$$

The dipole generates electric field which differs from field of a dot charge a direction of the force lines. They begin on the positive charge and finish on the negative charge.

12.2 Einthoven's theory of leads.

Current dipole (dipole electrical generator)

In real condition electric dipole cannot live long time: in electric field charges begin move and dipole or will neutralized or will screened.

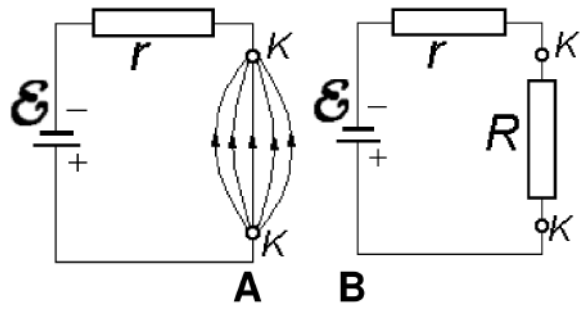

Fig. 12.4. Charge and current dipoles.

As current dipole can be a system, in which electric potential difference is remained, for example, part of electric circuit, in which direct current present. On Fig.12.4(a) E – voltage source, r – equivalent of internal resistance of voltage source. Between points KK current dipole is present. On Fig.12.4(b) R is equivalent of resistance of conducting medium. If $R << r$, current in the circuit will be constant and will not depend on medium properties. Such double-pole system consists of spring and gutter of current placed on small distance d is named as current dipole.

Potential drop $\Delta\varphi$ between two field points equidistant from current dipole is

equal to:

$$\Delta\varphi = \frac{p\cos\alpha}{4\pi\varepsilon\varepsilon_0 r^2}.$$

In other words, potential drop between two field points equidistant from current dipole in electroconductive homogenous medium is proportional to projection of dipole moment onto the straight line which connects these points.

Electrography.

All live tissues are sources of electrical potential named as biopotentials.

Electrography is general name of all biopotential registering methods that used with research and diagnostic purposes.

Electroencephalography (EEG) is examinations of electrical activity of brain.

Electromyography is examinations of electrical activity of muscles.

Electrogastrography (EGG) – examinations of electrical activity of digestive tract, first of all, stomach.

Electrooculography is examinations of electrical activity of eyes, where oculomotor muscles are main search of signals.

Most used method in everyday doctor's work is **electrocardiography**.

The ECG records (indirectly) the electrical activity of the heart. This activity reflects the action of the cardiac muscle as it depolarizes and repolarizes during the cardiac cycle.

The ECG represents the temporal and spatial summation of the action potentials of the myocardial fibers typically measured with body-surface electrodes.

ECG's are used to diagnose arrhythmias, abnormal electrolyte (potassium, calcium) levels, and conduction abnormalities. They are also used for screening and therapy guidance for heart disease as well as cardiac gating for imaging.

In order to get an electrical signal from the body, suitable electrodes, amplification and appropriate display are required. Some cardiac cells generate action potentials (pacemakers). Once generated, and under physiological conditions, the action potential propagates through the cardiac muscle. The temporal and spatial summation of the monophasic action potentials of the myocardial fibers produces an electrical signal known as the ECG.

Electrical activity of the heart.

The human body can be considered, for purposes of an ECG, a large-volume conductor. It is basically filled with tissues surrounded by a conductive ionic fluid. You can imagine that the heart is suspended inside of that conductive medium.

During the cardiac cycle, the heart contracts in response to action potentials moving along the chambers of the heart. As it moves, there will be one part of the cardiac tissue that is depolarized and another part that is at rest or polarized. This results in a charge separation, or dipole, which is illustrated in Fig. 1.

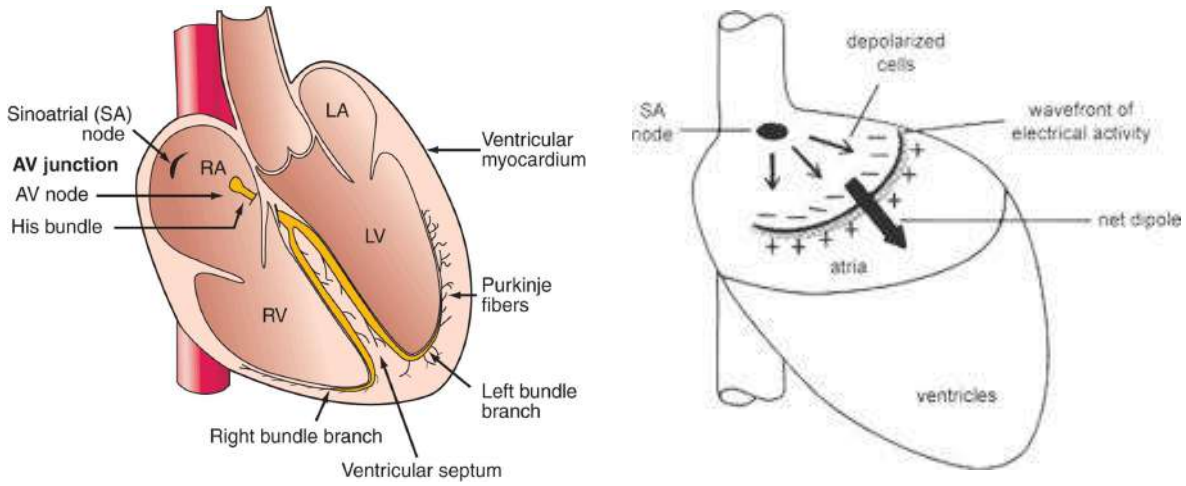

Fig. 12.5. After conduction begins at the sinoatrial (SA) node, cells in the atria begin to depolarize. This creates an electrical wavefront that moves down toward the ventricles, with polarized cells at the front, followed by depolarized cells behind. The separation of charge results in a dipole across the heart (the large black arrow shows its direction).

The dipole causes current flow in the surrounding body fluids between the ends of the heart, resulting in a fluctuating electric field throughout the body. This is much like the electric field that would result, for examples, if a common battery were suspended in a saltwater solution (an electrically conductive medium). The opposite poles of the battery would cause current flow in the surrounding fluid, creating an electric field that could be detected by electrodes placed in the solution. A similar electrical field around the heart can be detected using electrodes attached to the skin. The intensity of the voltage detected depends on the orientation of the electrodes with respect to that of the dipole ends. The amplitude of the signal is proportional to the mass of tissue involved in creating that dipole at any given time. Using electrodes on the surface of the skin to detect the voltage of this electrical field is what provides the electrocardiogram.

Cells within the sinoatrial (SA) node are the primary pacemaker site within the heart. These cells are characterized as having no true resting potential, but instead generate regular, spontaneous action potentials. Unlike non-pacemaker action potentials in the heart, and most other cells that elicit action potentials (e.g., muscle cells, nerve cells), the depolarizing current is carried primarily by

relatively slow, inward Ca^{++} currents instead of by fast Na^{+} currents (fig.12.6). There are, in fact, no fast Na^{+} channels operating in SA nodal cells. Na^{+} enters node cells through opened Ca^{++} channels. This results in slower action potentials in terms of how rapid they depolarize.

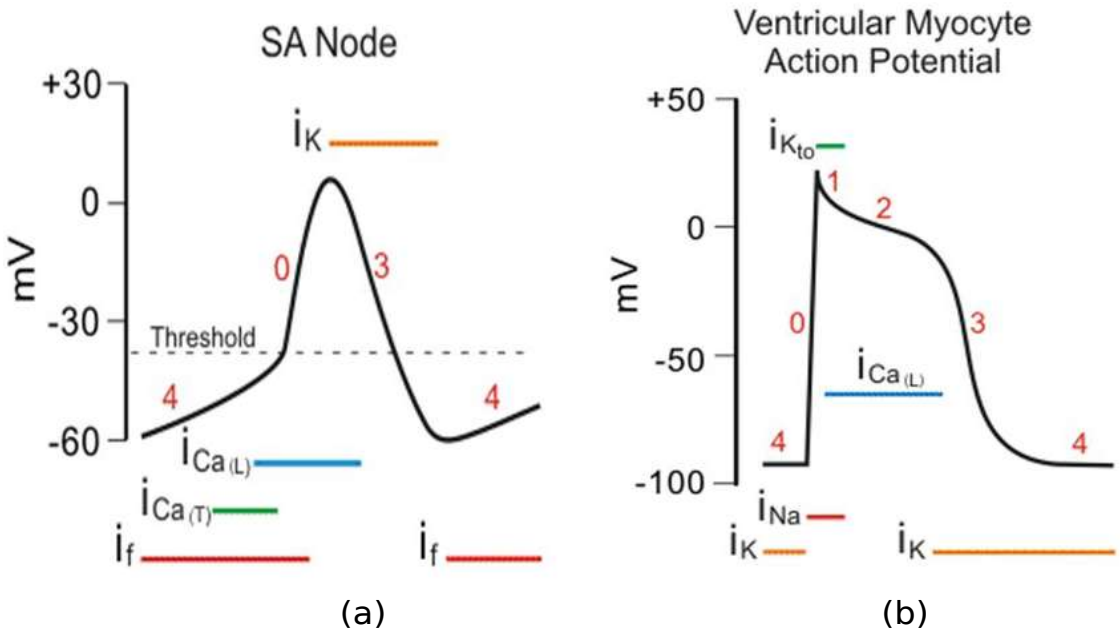

Fig. 12.6. The working myocardiocyte (a) and ion currents in various periods of action potential of pacemaker (b). Phases: 1 – depolarization, 2 – initial fast repolarization, 3 – slow repolarization, 4 – ended repolarization, 5 – slow diastolic depolarization (b) or rest (a). I_f – Na^{+} "funny" current, I_a – Ca^{++} currents, I_K – K^{+} current, I_{Na} – fast Na^{+} current.

Therefore, these pacemaker action potentials are sometimes referred to as "slow response" action potentials. SA nodal action potentials are divided into three phases. Phase 5 is the spontaneous depolarization (pacemaker potential) that triggers the action potential once the membrane potential reaches threshold between ($-40\ mV$ and $-30\ mV$). Phase 1 is the depolarization phase of the action potential. This is followed by phase 4 repolarization. Once the cell is completely repolarized at about $-60\ mV$, the cycle is spontaneously repeated.

The changes in membrane potential during the different phases are brought about by changes in the movement of ions (principally Ca^{++} and K^{+}, and to a lesser extent Na^{+}) across the membrane through ion channels that open and close at different times during the action potential. When a channel is opened, there is increased electrical conductance (g) of specific ions through that ion channel. Closure of ion channels causes ion conductance to decrease. As ions flow through

open channels, they generate electrical currents that change the membrane potential.

Fig. 12.7. An early ECG device

Willem Einthoven was a Dutch physiologist who pioneered the ECG and won the **Nobel prize in Medicine** in 1924 for this work. It was Dr. Willem Einthoven from the Netherlands who innovated the EKG for clinical use by refining the electrometer and delineating PQRST waves. He used three limb leads, which became known as **Einthoven's triangle**.

It is experimentally established that heart – dipole during excitation generates action potentials, which on a surface of human skin give the lines of equal potentials (equivalent lines). On Fig.12.8 position of the dipole moment of current dipole of heart and equipotential lines is shown.

If through a heart – dipole to lead the straight line "ab" and through the center of arm of dipole perpendicular "00", then on the part of positive pole of dipole – heart are settle down positive and on the part of negative – negative lines of identical potential.

The line "00" have zero potential. From the drawing is visible that if to apply electrodes on various points of surface of body of the person it is possible to determine the potential difference ΔU of these points.

For the first time theoretically proved points on a surface of human body from which it is possible to take potentials of heart have been offered by Einthoven. Dipole representation about heart underlies the theory of leads of Einthoven.

The main postulates of Einthoven's theory are:

1) the electric field of the heart is represented as electric field of current dipole $\vec{P}$ which is called the integral electric vector of heart;

2) $\vec{P}$ is in homogeneous isotropic conducting medium that is tissues of organism;

3) $\vec{P}$ permanently changes its direction and value.

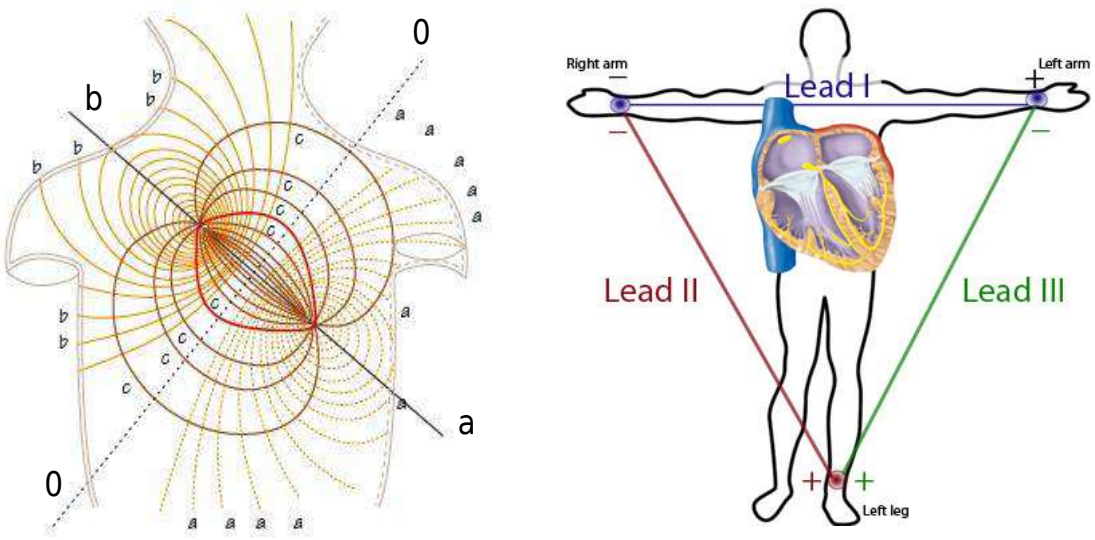

Fig. 12.8. The integral electric vector of heart and Einthoven's triangle

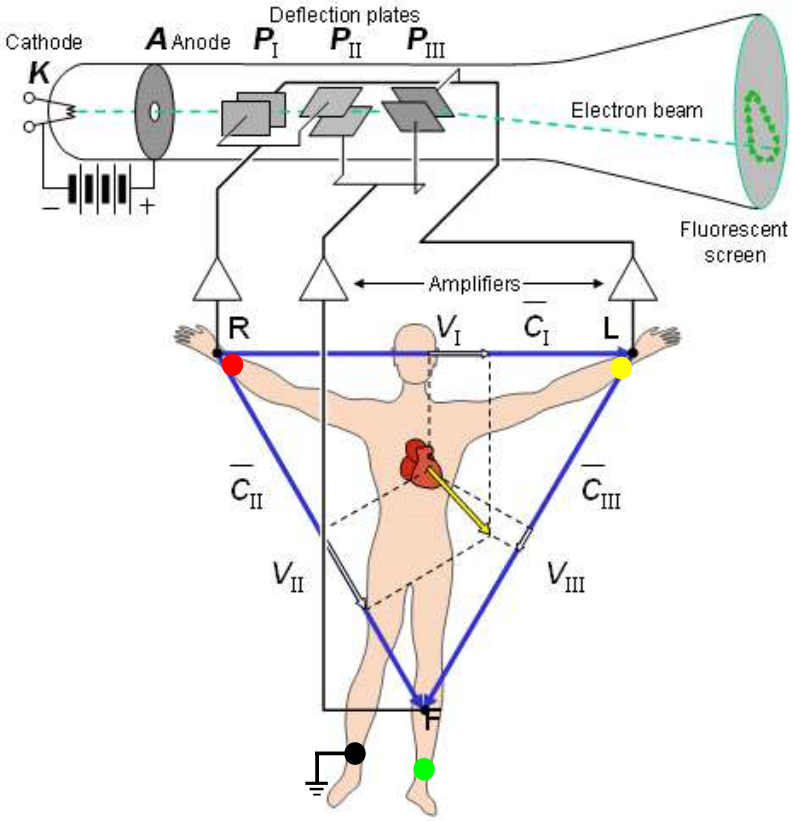

Fig. 12.9. Registration of ECG

The electrodes established in each chosen points on a body surface, are connected to a galvanometer of electrocardiograph. One of electrodes attach to a positive pole of a galvanometer (it is *positive*, or *active*, an electrode of lead), the second electrode — to its negative pole (a *negative* electrode of lead).

Now in clinical practice 12 leads are used most widely of an electrocardiogram which record is obligatory at everyone electrocardiographic inspection of the pa-

tient: **3 standard** leads, **3 amplified** (strengthened) unipolar leads from limbs and **6 chest** leads.

For record of these leads electrodes impose on the right hand, the left hand and on the left leg (see Fig.12.9). These electrodes are in pairs connected to electrocardiograph for registration of each of three standard leads. The fourth electrode is placed on the right leg for connection grounded wires.

Marks (+) and (−) here designate corresponding connection of electrodes to positive or negative poles of a galvanometer, i.e. are specified a positive and negative pole of each lead.

Three standard leads are connections with the electrodes established in form an equipotential triangle (Einthoven's triangle).

In the center of equipotential Einthoven's triangle the electric center of heart, or the dot of dotted sole heart dipole equidistant from all three standard leads is located.

Amplified leads and chest leads have a single positive electrode and uses combination of other electrodes as a composite negative electrode; they have conventional name **unipolar leads**.

Amplified (strengthened) leads from limbs.

Amplified leads register a potential difference between one of limbs on which the active positive electrode of the given lead (the right hand is established, the left hand or the left leg), and average potential of two other limbs. Thus, as a negative electrode in these leads use so-called joint (incorporated) Goldberger's electrode which is formed at connection through additional resistance of two limbs. Three amplified unipolar leads from limbs are designated as follows: aVR — amplified lead from the right hand; aVL – amplified lead from the left hand; aVF – amplified lead from the left leg.

Electrical axis of heart is vector that coincides with heart electric dipole in time moment, when dipole moment is maximal. Usually it manifests in maximal R wave in corresponding axis. In other cases it position calculated as algebraic sum of two largest neighboring R waves (need remember that aVR vector is used with opposite direction in this action).

Chest leads

The remaining 6 of the 12 lead recordings are the 6 chest leads. These leads are also unipolar; however, they measure electrical activity in the traverse plane instead of the frontal plane. Similar to the unipolar limb leads, a neutral reference lead is "created," this time using all 3 limb leads connected to the negative

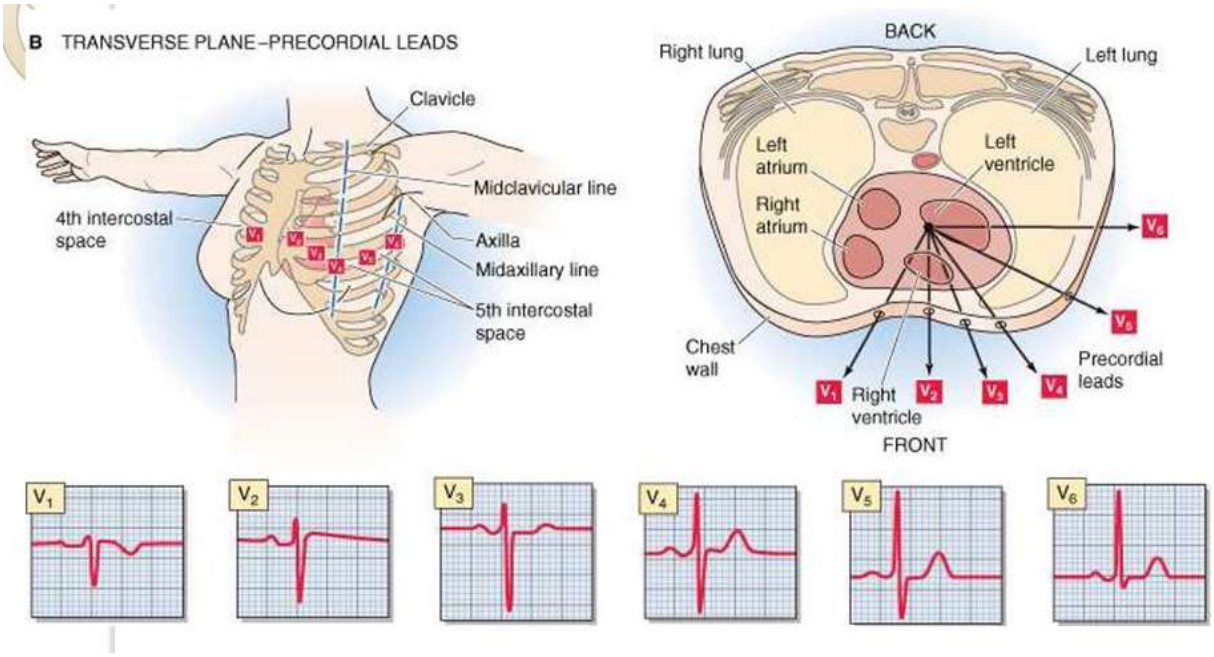

Fig. 12.10. A cross-section of the chest shows the relative position of the six precordial leads in the traverse plane, along with a typical waveform detected for ventricular depolarization. An anterior view of the chest shows common placement of each precordial lead, V_1 through V_6.

ECG lead, which basically puts it in the center of the chest. The 6 positive, or "exploring," electrodes are placed as shown in Fig.12.10 (around the chest) and are labeled V_1 through V_6 (the V meaning voltage). These chest leads are also known as theprecordial leads. Fig.12.10 shows a simple cross-ection (looking superior to inferior) of the chest, depicting the relative position of each electrode in the traverse plane. Fig.12.10 also shows a typical waveform obtained from each of these leads. The 3 bipolar limb leads, 3 unipolar limb leads, and 6 precordial leads make up the 12-lead ECG.

12.3 The ECG waveform

When an ECG is recorded, a reading of voltage vs time is produced, which is normally displayed as millivolts (mV) and seconds. A typical lead II ECG waveform is shown in Fig.12.11 For this recording, the negative electrode was placed on the right wrist, and the positive electrode was placed on the left ankle, giving the standard lead II ECG.

It shows a series of peaks and waves that corresponds to ventricular or atrial depolarization and repolarization, with each segment of the signal representing a different event associated with the cardiac cycle. The cardiac cycle begins with the

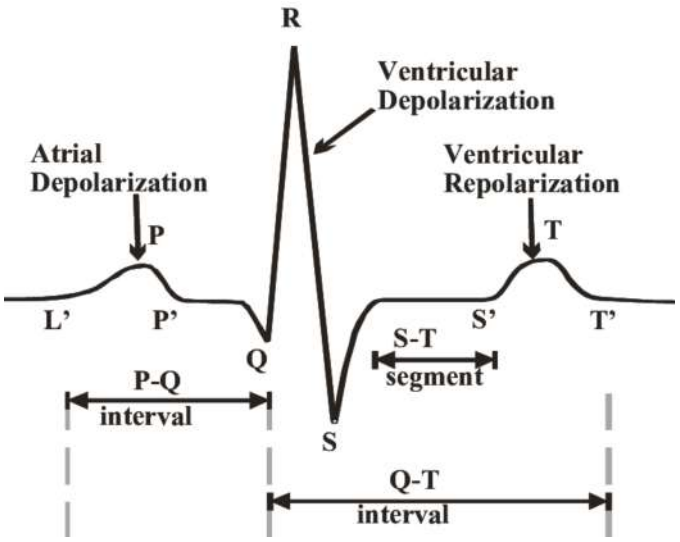

Fig. 12.11. A typical ECG waveform for one cardiac cycle measured from the lead II position. The P wave denotes atrial depolarization, the QRS indicates ventricular depolarization, and the T wave denotes ventricular repolarization. The events on the waveform occur on a scale of hundreds of milliseconds.

firing of the sinoatrial node in the right atrium. This firing is not detected by the surface ECG because the sinoatrial node is not composed of an adequately large quantity of cells to create an electrical potential with a high enough amplitude to be recorded with distal electrodes (signal amplitude is lost as it dissipates through the conductive medium). The atria then depolarize, giving rise to the P wave. This represents the coordinated depolarization of the right and left atria and the onset of atrial contraction. The P wave is normally around $80 - 100\ ms$ in duration. As the P wave ends, the atria are completely depolarized and are beginning contraction.

The signal then returns to baseline, and action potentials (not large enough to be detected) spread to the atrioventricular node and bundle of His. Then, roughly 160 ms after the beginning of the P wave, the right and left ventricles begin to depolarize, resulting in what is called the QRS complex, representing the beginning of ventricular contraction, which is around $80\ (60 - 100)\ ms$ in duration. Typically, the first negative deflection is the Q wave, the large positive deflection is the R wave, and if there is a negative deflection after the R wave, it is called the S wave. The exact shape of the QRS complex, depends on the placement of electrodes from which the signals are recorded.

Simultaneous with the QRS complex, atrial contraction has ended, and the atria are repolarizing. However, the effect of this global atrial repolarization is sufficiently masked by the much larger amount of tissue involved in ventricular

depolarization and is thus not normally detected in the ECG. During ventricular contraction, the ECG signal returns to baseline. The ventricles then repolarize after contraction, giving rise to the T wave. Note that the T wave is normally the last-detected potential in the cardiac cycle; thus, it is followed by the P wave of the next cycle, repeating the process.

Of clinical importance in the ECG waveform are several notable parameters (regions), which include the P – R interval, the S – T segment, and the Q – T interval. The P – R interval is measured from the beginning of the P wave to the beginning of the QRS complex and is normally $120 - 200\ ms$ long. This is basically a measure of the time it takes for an impulse to travel from atrial excitation and through the atria, atrioventricular node, and remaining fibers of the conduction system. The S–T segment is the period of time when the ventricles are completely depolarized and contracting and is measured from the S wave to the beginning of the T wave. The Q - T interval is measured from the beginning of the QRS complex to the end of the T wave; this is the time segment from when the ventricles begin their depolarization to the time when they have repolarized to their resting potentials and is normally about $400\ ms$ in duration.

An obvious observation made concerning the QRS complex is that it has a much higher peak and shorter duration than either the P or T waves. This is because ventricular depolarization occurs over a greater mass of cardiac tissue (i.e., a greater number of myocytes are depolarizing at the same time); furthermore, the ventricular depolarization is much more synchronized than either atrial depolarization or ventricular repolarization. For additional details relative to the types of action potential that occur in various regions of the heart.

It is also very important to note that deflections in the ECG waveform represent the change in electrical activity caused by atrial or ventricular depolarization and repolarization and not necessarily generalized cardiac contractions or relaxations, which take place on a slightly longer time-scale (see Fig.12.12). Shown in Fig.12.12 are the certain points on the ECG waveform and how they relate to other events in the heart during the cardiac cycle.

One last thing that should be noted relative to the ECG waveform is the sometimes-detected potential referred to as the U wave. Its presence is not fully understood, but is considered by some to be caused by late repolarization of the Purkinje system. If detected, the U wave will be toward the end of the T wave and have the same polarity (positive deflection). However, it has a much shorter amplitude and usually ascends more rapidly than it descends (which is

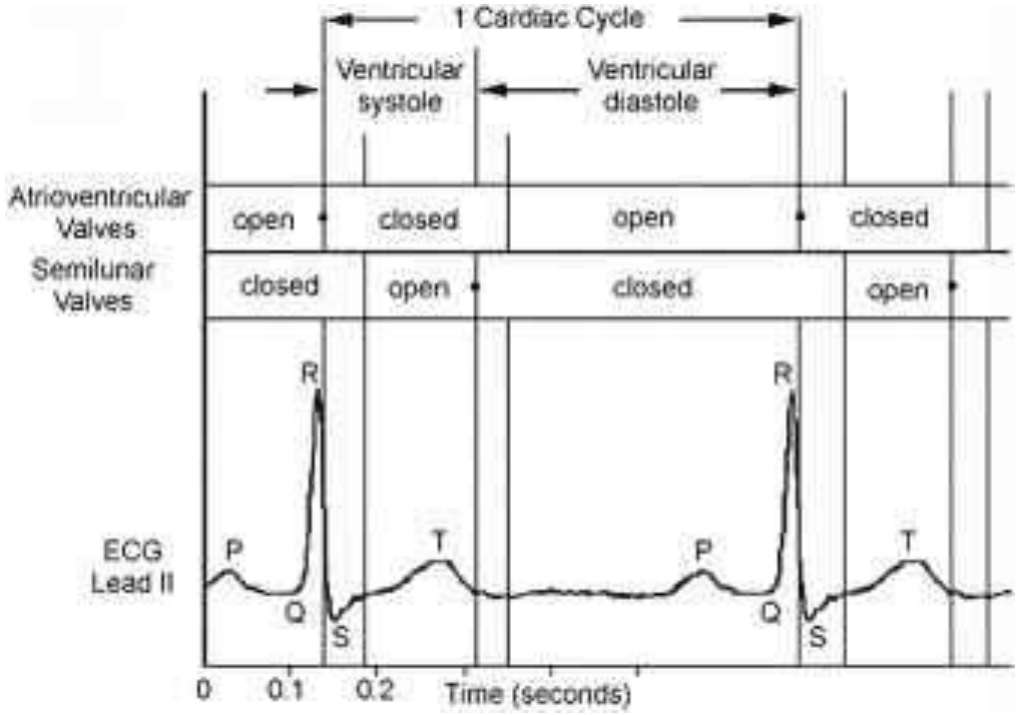

Fig. 12.12. A typical lead II electrocardiogram (ECG) waveform is compared to the timing of atrioventricular and semilunar valve activity, along with which segments of the cardiac cycle the ventricles are in systole/diastole.

the opposite of the T wave).

The important information can be received at other pathologies of heart. At the present time other diagnostic techniques based on registration of changes of biopotentials at functioning those or other organs are also applied. So the potentials generated by the cerebrum are registered at the **electroencephalography** (EEG), the potentials generated by various muscles of a body are registered at the **electromyography** (EMG), the potentials generated by the eye retina are registered at the **electroretinography** (ERG). Biopotentials registered at the electroencephalography are connected basically with the functioning of pyramidal neurons of the cortex of cerebrum. Electroencephalograms have the complex form. The maximum value of the biopotentials registered at EEG is about several microvolts.

Test questions

1. What is an Electric Dipole? What physical quantities is it characterized by?

2. What is the interaction of an electric dipole with an external uniform and inhomogeneous electric field?

3. What is called a current dipole?

4. How to determine the dipole moment and potential created by a current dipole?

5. What are the basics of electrocardiography?

6. What are the basics of Einthoven's theory?

7. What does an electrocardiogram show?

Lecture No 13

Electric properties of tissues and bodies

13.1. Characteristics of electric current

The electric phenomena in a human organism are connected with passing of currents in tissues of a human organism being in most cases the mediums well conducting the electric current (except for skin and bone tissues). The basic characteristics of currents are the **current strength** (I) and the **current density** (j).

Electric current is an ordered motion of electric charges. Electric current has properties of thermal, chemical and magnetic action.

Current strength I is determined by the ratio of charge quantity dq, which is carried through the cross-section of conductor, to the time interval dt, required to carry this charge:

$$\boxed{I = \frac{dq}{dt} = q^{'}(t)}$$

If for any equal time intervals the equal quantities of electric charge are carried, then the current will be called continuous (or direct). Then

$$\boxed{I = \frac{q}{t}}, \qquad [I]_{SI} = A, \quad [q]_{SI} = C$$

The current density is the ratio of the current strength passing through the cross section of a conductor, to the area of this section (S), i. e.

$$\boxed{j = \frac{I}{S}}$$

The unit of the current density is ampere per square meter (A/m^2).

Ohm's Law. In most cases at constant temperature the ratio of voltage at

Table 13.1

Effect of Current strength

Current strength	Effect
0.1 – 1 mA	No
1 – 1.5 mA	Irritating action. 1mA – perceptible current threshold
1.5 – 2 mA	Loss of perceptibility
2 – 16 mA	Pain, contraction of muscles
16 – 20 mA	16 mA – "motionless" current, the higher quantity of which make impossible for a human being to get free from electrodes.
20 – 100 mA	Respiratory paralysis
0.1 A – 3 A	Lethal ventricular fibrillation (urgent reanimation is required)
More then 3 A	Cardiac arrest, severe flash-burn

the ends of conductor U to quantity of current I in it is a constant, i.e.

$$R = \frac{U}{I}$$

Quantity R is called conductor resistance. The unit of resistance is ohm (Ohm or Ω). Resistance of uniform conductor with unchangeable section is in direct proportion to its lengths l and is inversely proportional to the area of cross-section S, i.e.

$$\boxed{R = \rho\frac{l}{S}}$$

where ρ is **specific resistance**.

Quantity σ, reciprocal to specific resistance is called **specific conductivity**: $\sigma = \rho^{-1}$.

The EMF can be written in terms of the internal resistance of the battery (r) where: $\epsilon = I(r + R)$

Which from Ohm's law, we can then rearrange this in terms of the terminal resistance: $\epsilon = U + Ir$

The EMF of the cell can be determined by measuring the voltage across the cell using a voltmeter and the current in the circuit using an ammeter for various resistances.

13.2. Conductivity of biological tissues and fluids

Conductors are substances which have the free charges, capable to move under the influence of electric field.

Examples: blood plasma, lymph, intercellular liquid, spinal liquid, cytoplasm.

Dielectrics (insulators) are substances which have no free charges, therefore do not spend an electric current.

Examples: a dry skin, sinews, bone tissues, cellular membrane.

Biological tissue is of rather heterogeneous conductivity. It has areas with high conductivity (biological fluids) and low conductivity (skin, bone and adipose tissue, membranes of cells and organelles), that alternate in a complicated manner. Table 13.2 illustrates quantity resistance of some tissues and fluids to direct current.

Electric conductivity measurement (conductometry) is widely used in study of processes occurred in live cells and tissues at the time of physiological change of state as a result of action of particular chemical substances and pathologic processes. Considering dynamics of change of electric resistance of skin, we can judge of the so-called skin-galvanic reactions, which reflect emotions, tiredness and other states of an organism. Electric resistance for determination of "active points" is used in reflexotherapy. Year in year, arsenal of researches of electric properties of biological tissue has been growing steadily.

Table 13.2

Effect of Current strength

Tissue	$\rho\ [Ohm \times m]$
Cerebrospinal fluid	0.55
Blood	1.66
Muscles	≈ 2
Brain tissue	≈ 14
Adipose tissue	≈ 33
Dry skin	10^5
Bone without periosteum	10^7

In an organism there are no such systems which would be similar to inductance coils, therefore tissues of the person do not possess inductance resistance.

Reception of alternating current. Its basic characteristics

Alternating current (AC) is a current periodically changing on value and direction. The most widespread is the sine wave alternating current which instant value varies in time under the law of sine (or cos).

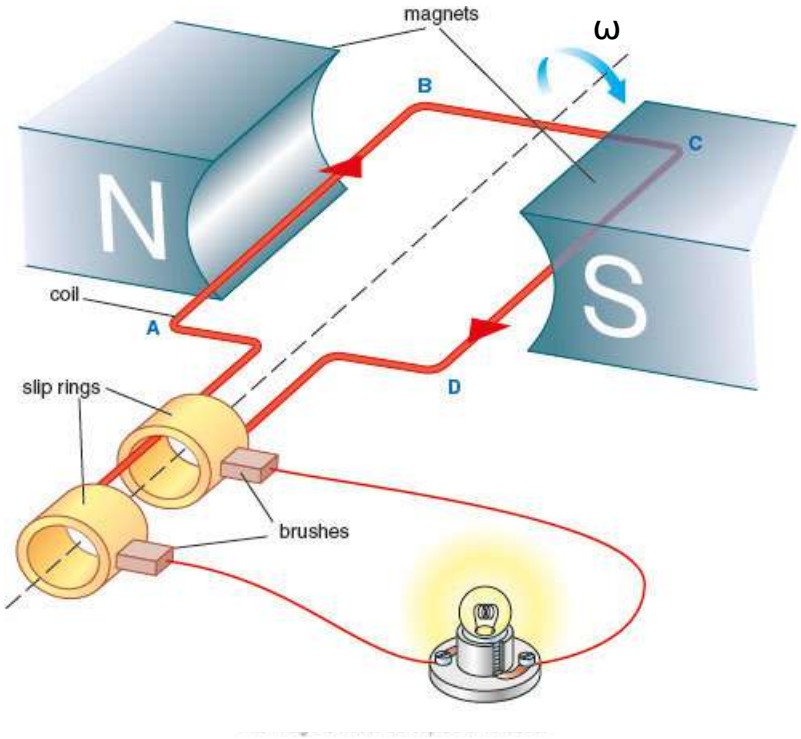

Fig. 13.1. Simple AC-Generator

Let's consider physical bases of reception of an alternating current and the principle of device of a generator. Let in the field of a constant magnet ($B = const$) rotates the conducting framework with constant angular speed ω (fig.13.1). Then instant value of the stream connected with the contour will be equal to:

$$\Phi = BS\cos\varphi = BS\cos\omega t$$

, where S is the area limited by the contour; is an induction of the magnetic field; $\varphi = \omega t$ is the angle of rotation of the contour, counted from its initial position, when $S \perp B$. According to the **law of Faraday** in the framework arises EMF of induction:

$$E = -\frac{d\Phi}{dt} = BS\omega\sin\omega t,$$

where $BS\omega = E_m$ is maximal (peak) value of EMF of induction, i.e. $E = E_m\sin\omega t$. Hence, if in the uniform magnetic field rotates the conducting contour, in it there is variable EMF changing under the law of sine in regular intervals.

This EMF creates a sine wave alternating current in the contour:

$$I = \frac{E}{R} = \frac{E_m}{R} \sin \omega t = I_m \sin \omega t,$$

where R is resistance of the contour and of the electric circuit in which the electric current (by means of brushes N sliding on rings K) is allocated; I_m is peak value of an alternating current; ω is circular frequency; $\varphi = \omega t$ is a phase of a current.

The alternating current also is characterized by period T and frequency ν and $\omega = \dfrac{2\pi}{T} = 2\pi\nu$. Graphically value EMF and an alternating current will be represented by two sinusoids (values change in identical phases). The considered way of reception of alternating current underlies at the heart of device of the industrial generator of alternating current. In industrial generators the magnetic field is created by a powerful electromagnet. The rotating contour consists from n coils (connected in series)of the wire which has been reeled up on the ferromagnetic core (rotor of the generator). Therefore EMF, excited in such generator will be equal: $E = BSn\omega \sin \omega t$.

For the characteristic of an alternating current the concept of working (effective) or root-mean-square value of current is entered. The effective (virtual) value of AC current is defined as that steady (constant) current which would develop the same quantity of heat ia the same time in the same resistance as is done by the DC:

$$I_{ef} = I_{w.} = \frac{I_m}{\sqrt{2}}, \qquad U_{ef} = U_{w.} = \frac{U_m}{\sqrt{2}}.$$

The devices included in a circuit of alternating current (ammeter, voltmeter) show effective values of a current and a voltage. When it is stated that the AC potential difference between the supply mains is 220 Volts, this means the effective potential difference is the 220 Volts and the peak value $U_0 = \sqrt{2} \times 220 = 311$ Volts.

13.3. Various kinds of electric resistance in a circuit of alternating current

a) Active resistance in a circuit of alternating current.

Resistance R in a circuit of alternating current (Fig.13.2) is called the **active** since at passage of a current in it there is an irreversible loss of energy. At presence

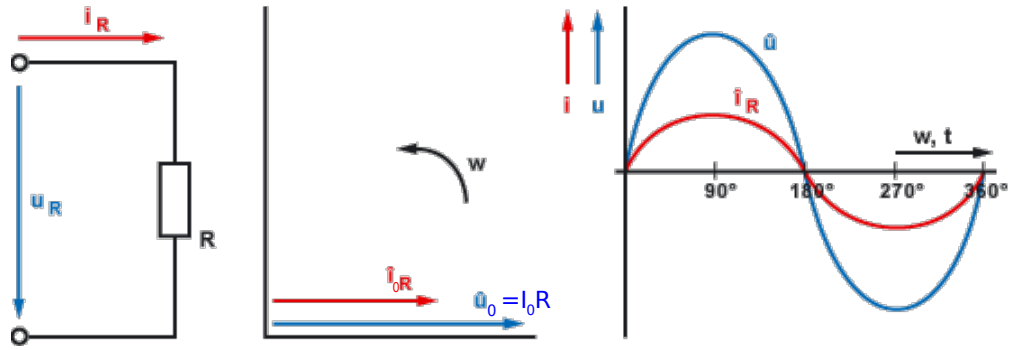

Fig. 13.2. Pure resistance

in a circuit only R a voltage is $U = U_0 \sin \omega t$ and $I = \dfrac{U}{R} = \dfrac{U_0}{R} \sin \omega t = I_0 \sin \omega t$, i.e. the current and the voltage coincide on a phase. The graph of a current and a voltage, and also the vector diagram of amplitudes of a current and a voltage are shown on Fig.13.2.

b) Inductive resistance (inductance) in a circuit of alternating current.

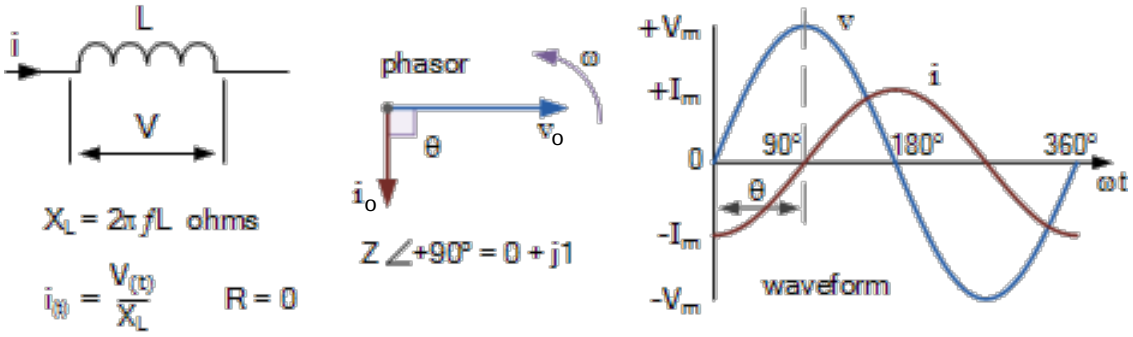

Fig. 13.3. Pure inductance

Let's consider a case, when the circuit contains only the coil of inductance with the small resistance R ($R = 0$) and significant inductance L (Fig.13.3). Let in the circuit there is alternating current: $I = I_0 \sin \omega t$. It causes in the coil EMF of selfinduction E_L , which at any moment is opposite to enclosed voltage U_L and counterbalances it: $U_L = -E_L$, but $E_L = -L\dfrac{dI}{dt}$,then:

$$U_L = I_0 \omega L \cos \omega t = U_{0L} \sin (\omega t + 90^o),$$

where $U_{0L} = I_0 \omega L$ is peak value of the voltage. From this formula follows,

that

$$I_0 = \frac{U_{0L}}{\omega L} = \frac{U_{0L}}{R_L},$$

where $R_L = \omega L$ is inductive resistance of the coil. At only induced resistance in a circuit calorification is not presence, as $R = 0$. The role of inductance is reduced to accumulation of magnetic field energy and returning of this energy back to a cell. There is a periodic energy transfer from a current source to a circuit and from the circuit to the current source, in ideal case loss-free of energy.

R_L increases with growth of frequency of alternating current. Dimension of inductive resistance is Ohm.

For a circuit with inductance in which *the voltage outstrips a current on* 90^o, wave and vector diagrams are submitted on Fig.13.3.

c) A capacitance in a circuit of alternating current.

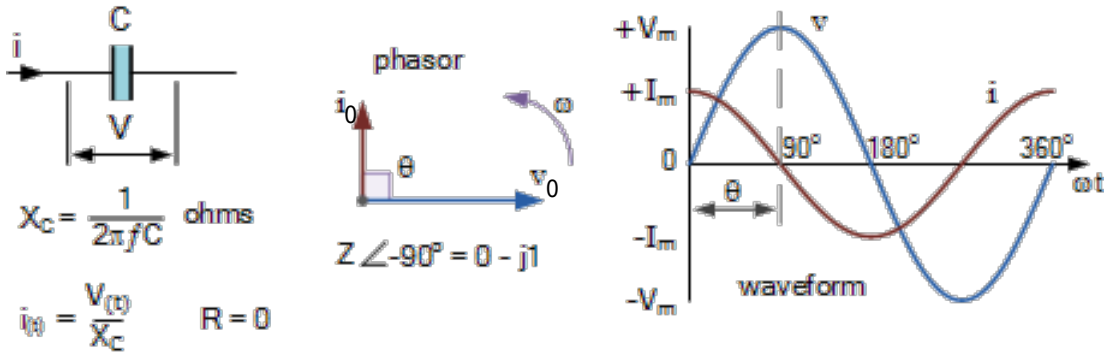

Fig. 13.4. Pure capacitance

Let's consider a case, when the condenser C is included in a circuit (Fig.13.4) (by resistance and inductance of bringing wires is possible to neglect). Let the current in the circuit changes under the law: $I = I_0 \sin \omega t$. The voltage on plates of the condenser can be presented by the following formula: $U_C = \frac{q}{C}$. Current in the circuit: $I = \frac{dq}{dt}$, $dq = Idt$, $q = \int I_0 \sin \omega t dt = -\frac{I_0}{\omega} \cos \omega t$, then

$$U_C = \frac{q}{C} = -\frac{I_0}{\omega C} \cos \omega t = \frac{I_0}{\omega C} \sin (\omega t - 90^o) = U_{0C} \sin (\omega t - 90^o),$$

where $U_{0C} = \frac{I_0}{\omega C}$ is the peak value of the voltage enclosed to the condenser. Peak value of the current is $I_0 = \frac{U_{0C}}{1/(\omega C)} = \frac{U0C}{R_C}$, where $R_C = \frac{1}{\omega C}$ capacitance.

It decreases with growth of frequency. R_C has dimension of Ohm.

In a circuit with only capacitance the voltage enclosed to plates of the condenser lags behind on phase of current on 90^o. It is reflected on the wave and vector diagrams on Fig.13.4. In the circuit with the condenser calorification is not presence, as ohmic resistance of conductors equal to null (warming of a dielectric in the alternating electric field is not taken into account, it will be surveyed later). The role of capacity is reduced to accumulation of electrical field energy of the condenser and returning of this energy back to a current source. There is a periodic energy transfer from a current source to the circuit and from the circuit to the current source, in ideal case loss-free of energy.

d) Total resistance (impedance) in a circuit of alternating current Resonance of a voltage.

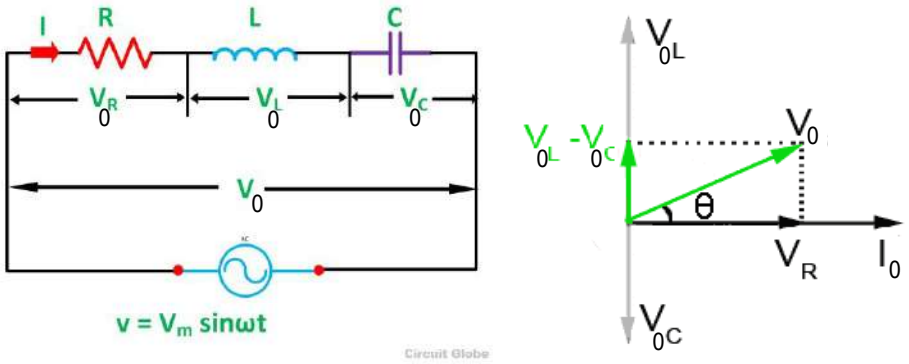

Fig. 13.5. Series Resonance Circuit

Let's consider a circuit consisting from R, L and C joined in series (fig.13.5). The current is equal $I = I_0 \sin \omega t$. We shall define how the voltage will change. The sum of falling of voltages on R, L and C is equal to the enclosed voltage:

$$U_0 = U_{0R} + U_{0L} + U_{0C} = I_0 R + I_0 R_L + I_0 R_C = I_0 R + I_0 \omega L + I_0 \frac{1}{\omega C}.$$

Owing to presence of a phase difference between U_L, U_C and the current $I(U_R)$ is in identical phase with the current) these voltages can be put vectorially and under theorem of Piphagor enclosed voltage U_0 (fig.13.5) is equal:

$$U_0 = I_0 \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} = I_0 Z,$$

where $Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}$ is known as total resistance, or *impedance of a circuit*. The law of Ohm for the given circuit will be written down so:

$$I_0 = \frac{U_0}{\sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}}.$$

The difference of phases between the current I and the voltage U is defined by the angle θ between vectors U_0 and U_R. Then: $U = U_0 \sin(\omega t + \theta)$. From diagram on Fig.13.5 follows that $\tan\theta = \frac{I_0 \omega L - I_0 \frac{1}{\omega C}}{I_0 R} = \frac{\omega L - \frac{1}{\omega C}}{R}$.

From the formula for Z follows that than closer on value ωL and $\frac{1}{\omega C}$, the less impedance Z and the is more current in this circuit. At $R_L = R_C$ or $\omega L = \frac{1}{\omega C}$ total resistance is $Z = R$ and current achieves of the greatest value caused only by active resistance of a circuit: $I_{0,RES} = \frac{U_0}{R_0}$. This phenomenon is known as *electric resonance*, which is provided by selection corresponding L and C. Resonance in the series circuit is called *resonance of voltage*, as thus occurs mutual indemnification of voltage U_L and U_C (they are directed opposite), each of U_L and U_C can significantly exceed voltage U enclosed to the circuit.

13.4. Total resistance (impedance) of tissues of human organism. Using of the method of electroconductivity in medicine.

At work with biological objects found that on high frequencies (10^7 Hz) electroconductivity is much higher than for low frequencies. At increase of frequency, electrical conductivity increases up to some maximal value. On fig.13.6 is resulted a curve of dependence of resistance of a muscle from frequency (curve of dispersion).

The zone of dispersion of electroconductivity usually varies in the interval $10^2 - 10^8\ Hz$. The dispersion of electroconductivity of living tissues on low frequencies is connected with polarization, but with increase of frequency the polarizing phenomena will decrease.

For a damaged tissue the steepness of dispersion decreases and for a dead tissue the graph is represented by a line parallel to axis of frequency $\ln \nu$ (Fig.13.6).

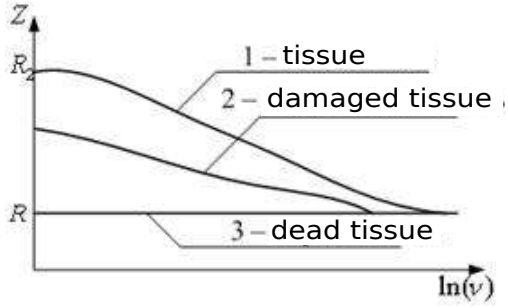

Fig. 13.6. The dependence of total impedance of a tissue on frequency.

Steepness of **dispersion** express by the ratio of value of the resistance measured on low frequency to the value of the resistance measured on high frequency. If two resistances are measured on different frequency under the same conditions, the ratio between them appears to constants for a normal condition of the given tissue. Usually choose for measurement of frequencies 10^4 Hz and 10^6 Hz, as at frequency of 10^6 Hz in many cases is observed maximal electroconductivity and for frequency of 10^4 Hz is observed change of a curve dispersion: $K = \dfrac{Z_{10^4}}{Z_{10^6}}$. For a tissue the specified coefficient aspires to 1.

Explaining passage of an alternating electric current through biological objects recognize that **resistance of living cells consists from ohmic and capacitor resistance**. Inductive elements in biological objects are absent.

Tissues of human organism will consist of the cells washed by a tissue liquid. Such element represents two mediums with good conductivity of current (a tissue liquid and cytoplasm of a cell) divided by badly conducting layer of a cellular membrane. Such system has electric capacity. In tissues there are the macroscopical formations consisting of various connecting mediums and partitions (badly conduct a current), on which both parties there are tissues well conducting electric current. It gives to tissues capacitor properties also.

Ohmic resistance does not depend from frequency, and capacitor resistance considerably decreases with increase of frequency and it results to increase of conductivity of all capasitance-resistance system. For connection in series of R and C total resistance is defined under the formula: $Z = \sqrt{R^2 + \left(\dfrac{1}{\omega C}\right)^2}$.

Presence at biological systems of capacities proves by the presence of *shift of phases* between a current and a voltage. The angle of shift of phases is defined by a ratio between capacitor and ohmic resistance, and for their series connection is equal: $\tan \dfrac{1/(\omega C)}{R}$. For biological systems the big value of this angle is characteristic. It shows that the share of a capacitance in tissues is great.

Let's result examples of value of angles of shift of the phases received on

frequency of 10^3 Hz: Human skin: 55^o; muscle of rabbit: 65^o; nerve of frog: 64^o.

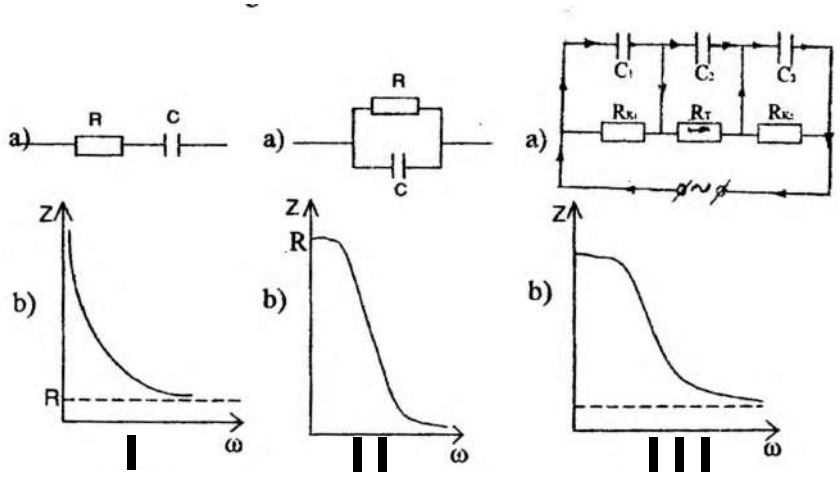

Fig. 13.7. The equivalent circuits

Taking into account that the total value of resistance (impedance) of living objects is submitted only by the geometrical sum of ohmic and capacitor resistance, for characteristic of conductivity of a current by living tissue use the *equivalent circuits*, i.e. to such combinations of ohmic resistance and capacity, which in some approximation can model electric parameters of tissues. Elementary of them are circuits with connection of R and C in series (Fig.13.7 Ia) and with parallel connection of these elements (Fig.13.7 IIa).

But these elementary circuits cannot be completely applicable for living cells. As follows from the graph of dependence of Z from ω for the first circuit (Fig.13.7 Ib): if $\omega \to 0$, then resistance $Z \to \infty$, that contradicts to experience.

From the graph of dependence of Z from ω for the second circuit (Fig.13.7 IIb) is visible, that at $\omega \to \infty$ $Z \to 0$, that on experience does not prove to be true.

The circuit combining first two circuits is most successful. One of them is represented on Fig.13.7 IIIa. On this circuit R_{K_1} and R_{K_2} are resistance of a skin; R_T is resistance of a tissue; C_1, C_2 and C_3 are capacity shunting these resistance. Arrows show the direction of alternating current in one of half-cycles. Resistances R_{K_1} and R_{K_2} are very great and alternating current through them does not pass. On Fig.13.7 IIIb the graph of dependence of Z from ω for this circuit is given that corresponds to the skilled data. There are other complex equivalent circuits, however any of them in accuracy cannot reproduce the laws inherent to complex biological systems.

On Fig.13.8 is presented the graph of the frequency dependence of impedance of a muscular tissue. For compactness the curve builds in logarithmic coordinates.

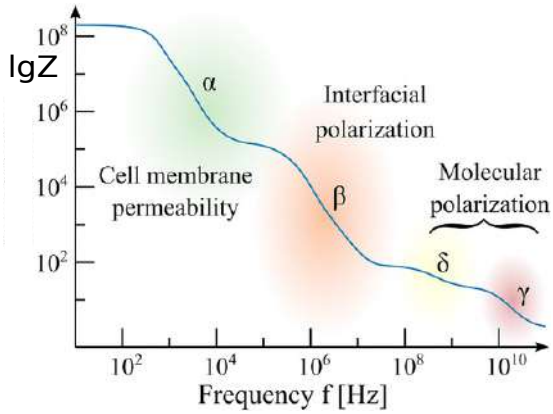

Fig. 13.8. The frequency dependence of impedance of a muscular tissue.

From the graph two singularities of this association are visible:

1) the smoothly varying decrease of impedance with increase of frequency; 2) presence of three areas of frequencies in which the deviation from a common course of dependence occurs: Z is not varies.

They have been called accordingly as fields of $\alpha-$, $\beta-$ and $\gamma-$ dispersions of impedance. Presence of these fields of impedance speaks, that with increase of frequency of alternating electric field in appearance of polarization participate different structures of biological tissues: at low frequencies all structures react to change of a field (α - dispersion), with increase of frequency large molecules - dipoles of organic junctions and molecules of water (β - dispersion) react, and at the most major frequencies molecules of water (γ - dispersion) react only. With increase of frequency of an electric field of ever less structures will react to a change of this field, the capacity of tissues decreases that leads to increase of impedance Z. Hence, at a common trend to decreasing of Z there are fields with smaller decreasing of Z.

The method of electroconductivity on an alternating current in living tissues and cells is used in biological researches and medicine for estimation of pathological processes. For example, at measurements in the field of low frequencies is observed increase of resistance of a tissue at an inflammation at the first stages. Cunrent of low frequency goes mainly through intercellular spaces. At inflammation as result of swelling of cells the section of intercellular intervals decreases, that attracts increase of ohmic resistance, while the capacity of cells at early stages of inflammation remains constant.

In diagnostics it is used as the method of measurement of the angle of shift of phases. At some diseases (thyroid gland) or at physiological changes (ageing of an organism) appreciable change of the angle of shift of phases is found out.

For characteristic of physiological condition of tissue is used value of a steepness of dispersion (K). This criterion applies for example at estimation of viability of the tissues intended for transplantation.

13.5. Physical bases rheography.

An alternating current of frequency 30 Hz is used in diagnosis for detection of filling tissues with the blood. Currents, which strength is less then threshold current, are used in this case, i.e. currents that do not irritate excitable tissues. Diagnostic method based on registration of changing of tissue impedance that takes place due to change in tissue filling with the blood (that was caused by heart functioning) is called **rheography** (or impedance-plethysmography, or rheoplethysmography).

Fig. 13.9. The rheography.

When cerebrum is examined with the help of rheography we have a rheoencephalogram, when heart is examined with the help of rheography, we have a rheocardiogram. Arterial vessels of lungs, liver and extremities can be examined with the help rheography. In stomatology vessels of paradont, mouth mucous membrane, salivary glands etc. can be examined with the help of rheography.

Rheodentography, a method that is similar to rheography, is used in stomatology. Tooth pulp is examined with the help of rheodentography. An alternating current of $0.5 - 1$ MHz frequencies is used for this purpose.

Rheography is a method of an estimation of a condition (parameters) of the blood channel by measurement of resistance (impedance) of site of tissue or a specific region of body (hand, leg, head, etc.) to an alternating current.

High frequency is used (30 kHz).

$$Z = R = \rho\frac{l}{S},$$

where $\rho = 1.5\ Ohm \cdot m$ is specific resistance of blood.

Let's deduce dependence $\Delta V = f(\Delta R)$: $R = \rho\dfrac{l}{S}$

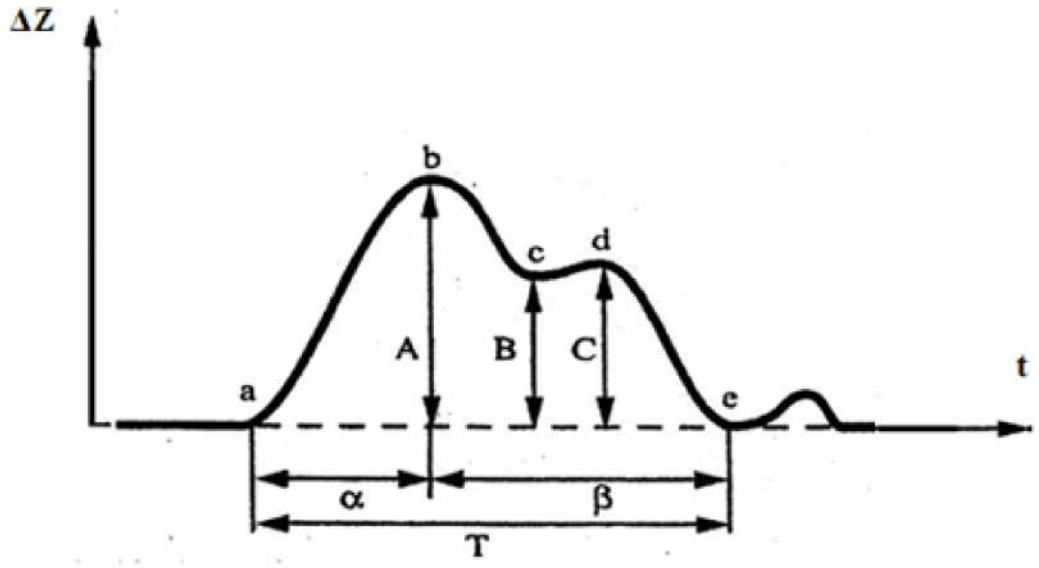

Fig. 13.10. The Rheogram: ab – anacrotic; bcd – incesture; bcde – catacrota; α – duration of anacrota (characterizes the tone and elasticity of the arteries); β – duration of catacrota (characterizes the contractility of blood vessels and their elasticity) ; A – anacrotic amplitude; B – incision amplitude; C – catacrota amplitude; T – duration of one heartbeat cycle.

We multiply numerator and denominator on l, where l is a length of a vessel.

$$R = \rho\frac{l^2}{S \cdot l}, \quad (V = S \cdot l)$$

$$R = \rho\frac{l^2}{V}, \quad V = \rho\frac{l^2}{R} \quad (1)$$

To find changes of volume ΔV differentiate the left and right part of the equation (1).

$$\frac{d}{dt}\left(V\right) = \frac{d}{dt}\left(\frac{\rho l^2}{R}\right)$$

$$\boxed{d(V) = \rho \cdot l^2 (R^{-1})' dR = -\frac{\rho \cdot l^2}{R^2} dR} \quad \Rightarrow \quad \boxed{\Delta V = -\frac{\rho \cdot l^2}{R^2} \Delta R}$$

This is fundamental rheography formula. Kedrov's Formula

ΔV is change of volume of the blood channel to which there corresponds change of resistance of blood channel (ΔR); l is distance between electrodes; R is base resistance.

$$Z \approx R, \quad \Delta R \approx \Delta Z, \quad \boxed{\Delta V = -\frac{\rho \cdot l^2}{Z^2}\Delta Z}, \quad \Rightarrow \quad \boxed{\frac{\Delta V}{V} = -\frac{\Delta Z}{Z}}$$

Rheogram is a plot of changing of impedance ΔZ depends on time t Fig.13.10.

13.6. Electrophysiotherapy and electrosurgery.

Electrical current effect on tissues depends on current type. The following currents can influence on tissues: direct, pulse and alternating ones. A current strength (I) and a current density (j) are direct current characteristics. Current strength (or simply current) in tissues is defined by an applied voltage (U) and tissue specific resistance (ρ) or specific conductance (σ).

When direct current is flowing through tissues, ions in tissues flow uninterruptedly in one and the same direction. Positive ions shift to one side and accumulate in particular parts of tissues; negative ions shift to opposite side and accumulate in other parts of tissues. So the main mechanism of direct current effect on biological tissues is change of ion concentrations in different parts of tissues in comparison with common concentration. It is necessary to note that direct current flowing in tissues can significantly decrease in time even if voltage is constant. It is related to that fact that ions shift and ion concentration in different parts of human body changes. Due to ion shifting the ions form an electric field in tissues. This opposite field partly compensates an external field i.e. tissue polarization EMF (electromotive force) is induced and it reduces current.

Medical methods based on use of direct current:

Galvanization

Voltage $U = 60 - 80\ V$ is used for **galvanization**. In this method current density must not exceed $j < 1\ A/m^2$. An alternative current rectifier is used as a device for galvanization. It enables to regulate output voltage and control (measure) induced current. Electrodes made of sheet lead or foil are commonly used to get current to a patient. Hydrophilic layers wet with water or physiological solution are placed between a patient's skin and the electrodes. The application of layers reduces skin resistance due to skin moisturizing; it also improves the contact between the electrodes and a human body. The application of layers protect

patient's skin from being burned by the products of sodium chloride ($NaCl$) electrolysis during which alkalis are deposited on cathode and acids are deposited on anode.

Human body reacts on the change of concentration of different ions in tissues when direct current flows through them. It, in particular, stimulates circulation of the blood and the lymph as well as metabolic processes in cells. Besides, nerve and muscular cells excitability increases under cathode and decreases under anode. Nerve cell excitability diminution under anode causes, in particularly, local analgesia (anesthetization) effect.

Galvanization can be used in treatment of some nervous diseases, bronchial asthma etc.

Medical electrophoresis is a method of treatment when direct current is used for introducing medical substances through skin or mucous membranes.

Medical electrophoresis is performed similar to galvanization but one of hydrophilic layers is wet not with water but with a certain medicine solution. Introducing of medical substances by means of electrophoresis is possible if a medicine dissolving in water forms ions. In this case anions are introduced to patient when a layer under cathode is wet with medicine solution, cations are introduced to patient when a layer under anode is wet with medicine solution.

Electrophoresis method of introducing medicines has a lot of advantages in comparison with other methods. Electrophoresis does not hurt skin and ensures local effect in a required place. The medicine is introduced in ion form, and it is ions that make therapeutic effect. The medicine introduced by means of electrophoresis accumulates in hypodermic (subcutaneous) cellular tissue and is washed out from it slowly that ensures prolonged uninterrupted medicine action on a pathological center.

Electrophoresis can be used for diagnostic purposes, for example, for plasma protein fraction separation and detection that is highly informative for diagnostic of a number of diseases.

Pulse current effect on biological tissues

Currents that can change in time are divided into pulse and alternating. An alternating current is a current that changes in time according to a harmonic low, i.e. current strength changes in time according to a sine or a cosine law as $i(t) = i_m \cos(2\pi f t + \varphi_0)$. A pulse current is a current that depends upon

time periodically but not harmonically. However, single pulses of electric current (electrical pulses) i.e. currents that change in time but do not flow for a long time are also used in medicine.

Electrical pulses can have different shape (different dependence of voltage and current upon time). So, different kinds of pulses are used – rectangular, triangular, trapezoid and other pulses.

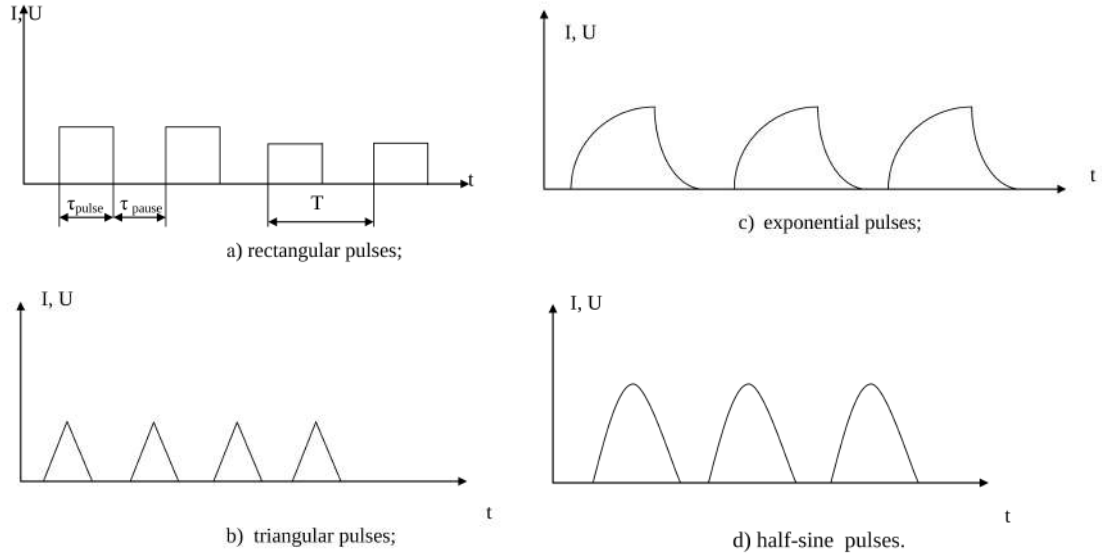

Fig. 13.11. The impulse currents graphs at different impulse shapes.

The most common pulse current characteristics are pulse amplitude, pulse duration, and pulse front steepness.

A front is a pulse section that corresponds to voltage or current growth as well as their diminution (back front or cut).

Amplitude is modulus of voltage or current maximum value (U_{max} or I_{max} correspondingly).

Pulse duration is a time interval during which voltage (or current) exceeds 0.1 U_{max} (or 0.1 I_{max} for current).

Front steepness (s) is determined as follows:

$$s = \frac{0.8 \cdot U_{max}}{\tau_f}$$

where τ_f is a duration of growth (or diminution) of pulse front when voltage ranges between 0.1 U_{max} and 0.9 U_{max}.

Pulse current is characterized by a period of pulse repetition (T) and pulse repetition frequency (f), at that $f = \frac{1}{T}$.

Pulse current cause a quick change of potential in intercellular fluid that can reduce cell membrane potential and causes action potential generation starts on cell membrane. Thus, the basic mechanism of pulse currents effect on biological objects is the irritation of excitable tissues. The excitable tissues are muscular, nerve, and glandular ones.

Irritation action of pulse current depends on its amplitude, frequency, steepness and pulse duration. The tissue irritation (action potential generation in tissue cells) is possible if pulse amplitude exceeds certain minimum value. Minimum current when response reaction (irritation) starts is called **threshold current** (i_{thr}).

Dependence of pulse current irritation action on frequency becomes apparent at significantly great frequencies. In this case the more pulse repetition frequency the less current irritation action (that is threshold current grows).

The more pulse front steepness is, the more current irritation action. **Du Bois-Reymond's law** describes this phenomenon: an electric current irritation action is directly proportional to the rate of current rise (or diminution) that is proportional to a derivative of current with respect to time. As the speed of pulse current growth depends on a front steepness, we can be assured that irritation action of pulsating current is directly proportional to pulse steepness.

As pulse duration increases its irritation action on excitable tissues increases, i.e. threshold current decreases. This conclusion makes a foundation of electrodiagnosis – a method of examining of tissue excitability properties by means of determination of dependence of a threshold current upon pulse duration when single rectangular pulses irritate tissue.

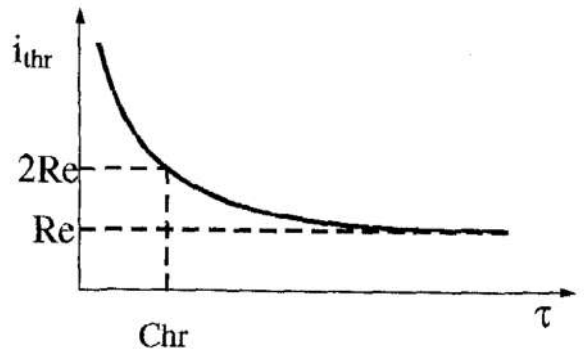

Fig. 13.12. Diagram of Dependence of threshold current i_{thr} on pulse duration τ.

This dependence (its diagram is shown in fig.13.12) is described by **Weiss-**

Lapicque's equation as follows:

$$\boxed{i_{thr} = \frac{q}{r} + Re},$$

From the diagram that describes that dependence we can see that the more pulse duration, the less the threshold current depends on τ and tends to a certain limiting value that is called a **rheobase** (Re) in electrophysiology.

Thus, rheobase is equal to a limit to which threshold current tends when pulse duration tends to infinity, i.e.

$$Re = \lim_{\tau \to \infty} i_{thr}.$$

Other words, rheobase is determined as the value of threshold current when it does not depend upon pulse duration.

Chronaxie (Chr) is one more feature that describes tissue exciting properties as well as rheobase. Chronaxie is such a pulse duration when threshold current is equal to doubled rheobase.

Rheobase and chronaxie certain values are characteristic for the exciting tissues state. At different pathological states these parameters change. For instance, rheobase decrease when chronaxie is constant is typical for beginning of inflammation. Necrotic changes in tissues are accompanied by rheobase increase and chronaxie decrease.

Medical methods based on use of pulse current

Pacing. In normal state pulses causing systoles are produced by a sine node (nodus sinuatrialis) called a rhythm driver and come to cardiac muscle through conducting system.

If sine node does not execute its function or conducting system is bad, it is necessary to have an external rhythm driver. Carrying or implanting cardiological electrical stimulator (heart pacemaker) is used as an external rhythm driver. Heart pacemaker is a generator of electrical pulses with repetition frequency of $1-1.2\ Hz$ and pulse duration of $0.8 - 3\ ms$.

Defibrillation. Defibrillators are used during cardiac arrest or heart ventricle fibrillation, i.e. when separate muscle fibers are contracted irregularly as a result of their irritation by each other. Defibrillator produces single high voltage electrical

pulses (discharges) that cause great contractions of heart muscle that often result in restoration of regular heart rhythm. Voltage equal to 8 kV is commonly used. In case of unsuccessful attempt to start the heart a greater voltage is used for the following attempts.

Electrogymnastics. Electrical exercises for muscles support muscle tonus, improve circulation of the blood and metabolism in bad muscles or in muscles with bad innervation, support their ability for contraction. Pulse current with pulses of a triangle shape with pulse duration of $1 - 1.5$ ms and repetition frequency of 100 Hz as well as pulses of an exponential shape with pulse duration of $3 - 60$ ms and repetition frequency of $8 - 80$ Hz are used for electrical exercises.

Electrosleeping. Electrical sleep is a method of inhibition of central nervous system by a pulse current of a rectangular shape with pulse duration of $0.1 - 1$ ms and repetition frequency of $5 - 150$ Hz.

Alternating current effect on biological tissues

Alternating current is characterized by voltage, current strength, amplitudes of these quantities, frequency (and quantities, referred to frequency – cycle frequency and period), and phase.

Dependency on frequency, different effects of alternating current on tissues can be apparent. If frequencies are low alternating current similar to pulse current causes irritation of excitable tissues. If frequencies are high when charged particles shift is not great in tissues, thermal effect takes place, i.e. tissue heating as a result of current flowing.

Alternating current like pulse current causes irritation in excitable tissues only then when current is greater or equal to threshold current. Threshold current grows as alternating current frequency grows (i.e. alternating current irritant action decreases as frequency grows).

Medical methods based on use of alternating current

Diathermy is based on phenomenon of heating when alternating current flows through tissues. There are therapeutic diathermy and surgical diathermy (electrical surgery), which, in its turn, is divided into diathermotomy and diathermocoagulation. When diathermy is used a specific heat power (an amount of heat evolved per 1 second in 1 m^3 of tissue) of current flowing through the tissue is defined by the formula

$$q = j^2 \cdot \rho$$

where q is specific heat power; j is current density; ρ is resistivity of tissue through which current flows.

An alternating current of 1 MHz frequency and $1 - 1.5$ A is used for therapeutic diathermy. The current is applied to a patient with the help of metal electrodes that have a sufficiently great area of contact with patient's body. The electrodes must be tightly attached to patient's body. It is evident from the specific, heat power formula, the more specific resistance is, the greater amount of heat is evolved. That results in useless evaluation of a great amount of heat in the skin. Besides, therapeutic diathermy is not a safe method, if the contact between an electrode and the skin surface is lost and the contact area decreases, it results in current density increase and local heating. It can cause skin burn instead of skin heating. Because of these problems, therapeutic diathermy is rarely used in our time.

The current of 10 MHz frequency is used in electrical surgery. Two electrodes are applied for it: one electrode have a great area of contact with patient's body and the other electrode is sharp. In the very place where a sharp electrode touches patient's body, current reaches great density and a great heat power occurs.

During diathermocoagulation ($j = 6 - 10\frac{mA}{mm^2}$ it gives the possibility to 'seal' blood vessels due to the effect of coagulation.

During diathermotomy ($j \sim 40\frac{mA}{mm^2}$) a sharp electrode acts as an electrical knife, which cuts tissue by burning it. Diathermotomy is good for a surgeon, as this method is practically bloodless because the vessels are cut and sealed simultaneously. At the same time such cuts are healed up slowly then the usual cuts made by a common scalpel.

Local D'Arsonvalisation. When certain methods used currents are applied, the current affecting the patient is an alternating and a pulse at the same time. So, at local D'Arsonvalisation the alternating current of $100 - 400$ kHz frequency acts at a patient, but the current is applied to him like pulses with repetition frenquency of 50 Hz. In this case the voltage is 10 kV.

The current is applied to a patient with the help of an electrode having a very great resistance (commonly the electrode is hollow or it is glass electrode filled with graphite). When such electrode is approaching to a patient skin, a spark occurs, i.e. the current starts to flow.

However, when current flows, at a glass electrode a great voltage reduces that results in a sharp decrease of voltage between the skin and the electrode. When a discharge (a spark) goes down, the current flowing stops and a great voltage

occurs again between the skin and the electrode. That leads to the next discharge again etc.

A local D'Arsonvalisation gets a local irritation to nerve- endings in the skin that stimulates the local circulation of the blood in skin and causes a lot of other positive results.

Alternating electromagnetic field effect on biological tissues

Electrical currents can be induced in tissues without electrodes. If tissues (some parts of human body) are placed into an alternating electromagnetic field, alternating currents are induced in them. Heating of tissues with the help of currents induced by an alternating field is the base of the following methods, such as inductothermy, **UHF-therapy** (UHF - ultra-high frequencies), and **microwave therapy**.

An important point is that effect of high-frequency electromagnetic oscillations on the human body is not only thermal. The impact of radiation on human body is present even then when thermal effect is insignificant. The action of electromagnetic radiation on human body is not studied enough.

The result of unfavorable effect of electromagnetic radiation of radio frequency range can be both direct pathological phenomena (internal diseases, internal dysfunction) and weakening of human body protection and adoption. It is accepted as correct a negative effect of high intensity electromagnetic radiation on cardiovascular, central nervous, endocrine, haematogenic and other systems. Alternating electromagnetic fields effect can cause dizziness, high fatigability, high irritability, memory weakening, insomnia, general weakness and other negative results. If even small electromagnetic fields act upon human body for a long time, they cause strong dysfunction in cerebral cortex.

Strict hygienic standards of permissible levels of electromagnetic fields acted upon human bodies are developed. Within the frequency range from 30 kHz to 300 MHz electric field intensity (E) (characteristic of electrical component of electromagnetic field) is standardized. Within the frequency range from 300 MHz to 30 GHz electromagnetic radiation energy flux (i.e. energy of electromagnetic radiation acted upon unit of surface area during unit of time) is standardized. Tolerance limits of electromagnetic radiation are given in the table:

Medical methods based on use of alternating electromagnetic field

UHF-therapy. In the course of UHF-therapy an alternating electric field of

Table 13.3

The frequency range from 300 MHz **to** 30 GHz **electromagnetic radiation**

Frequency (except the lowest limit, include the highest limit)	$30-300$ kHz	$0.3-3$ MHz	$3-30$ MHz	$30-300$ MHz	$0.3-3$ GHz
Wavelength	$10-1$ km	$1-0.1$ km	$100-10$ m	$10-1$ m	$100-1$ cm

UHF-range (frequencies of $30-300$ MHz) effect on a patient's tissues.

A standard instrument for UHF-therapy induces electromagnetic oscillations of 40.58 MHz frequency. In this case the wavelength is about 7.5 m.

The UHF-field induces electric currents in a patient's tissues (more exactly, charged particles oscillations) of the same frequency as the frequency of change of the UHF-field is. Emerged currents heat patient's tissues (organs), moreover when these frequencies of electromagnetic radiation are used, the hottest are those tissues which have less conductivity, i.e. tissues-dielectrics.

A specific heat power for tissues-dielectrics under UHF-therapy is determined by the formula

$$q = E^2 \cdot \omega \cdot \varepsilon \cdot \varepsilon_0 \cdot \tan\delta$$

where E is a root-mean-square value of electric-field intensity ($E = \frac{E_0}{\sqrt{2}}$, where E_0 is the amplitude of electric-field intensity); ε is the electrical constant; δ is the dielectric loss angle.

If an alternating current flows through media having capacitance (as stated, biological tissues posses capacitive properties), current phase differs from voltage phase producing that current. In this case, current vector is a vector sum of active and reactive components current. Active component current phase coincides with voltage phase; reactive component current phase differs from voltage phase by $\pi/2$.

Superhigh frequency therapy (SHF-therapy). If SHF-therapy is used for heating patient's tissues, it means that patient's tissues are affected by electromagnetic waves with frequency within the range of 300 MHz – $30-30$ GHz.

The commonly used devices are those ones that produce electromagnetic waves from the following standard values of wavelength: for **decimetric waves** (DMW) therapy – 65.2 cm ($\nu = 460$ MHz), for **microwave** (MW) therapy – 12.6 cm

($\nu = 2375\ MHz$).

Decimetric waves and microwave therapies differ from each other in the depth of radiation penetration in tissues. When decimetric therapy is used, the depth of penetration is 9 cm; when microwave therapy is used, the depth of penetration is $3 - 5$ cm.

At SHF-therapy the muscle tissues and the blood are heated well.

Direct electric and magnetic fields effect on biological tissues

Not only alternating electromagnetic fields but also direct electric and direct magnetic fields can be used for medical purposes.

Direct electric field influence on tissues causes dielectric polarization due to molecule reorientation; in this case they behave as dipoles. It results in ions shift and change of their concentration in different sections of tissues. Ions shift lasts until electric field that they have developed will not compensate external electric field effect on ions. Such methods as *electrostatic shower* (or *franclinization*) and aeroionic therapy are used. At that a patient is placed in a strong electrostatic field (voltage up to 50 kV is used) where partial air ionization occurs. As this takes place, aeroions are produced as well as air ionization products - ozone and ozone oxides that irritate skin receptors and mucous membrane receptors of respiratory tract.

This results in change of functional stat of central nervous system that is shown in predominance of inhibitory processes, ascension of neurologic level, sleep improvement etc.

Direct and low frequency magnetic fields biological effect on human body is not studied enough. The change of tissue agent diffusion rate, change of rate and direction of biological reactions, change of water structure and some other effects happened upon subcellular level are supposed to take place in magnetic fields.

There is information on some essential process stimulation as well as depression in human body under action of magnetic fields.

Direct and low frequency alternating magnetic field medical effect on human body is called *magnetic therapy.*

Test questions

1. What is alternating current? What is the principle of obtaining it?

2. Explain the law of variation of alternating current in an active circuit.

3. What are the phase relationships of current and voltage in a circuit with an inductor and a capacitor?

4. What is meant by inductive and capacitive resistance?

5. How do I get a law for a complete AC circuit?

6. What is impedance? How is it calculated?

7. Explain the conditions of resonance in the AC circuit and write the expression that defines the resonant frequency.

8. What is the conduction mechanism of biological tissues?

9. What is the effect of direct and alternating current on body tissues?

10. What methods of exposure to direct and alternating current are used for diagnostics and therapeutic purposes?

Lecture No 14

Magnetic field

14.1. Magnetic field in vacuum and its characteristics

The source of macroscopic magnetic field is magnetised bodies, conductors with current and movable electric charges. The nature of these sources is unified: magnetic field appears as a result of motion of charged microparticles (electron, protons, ions). Magnetic field may be determined by the action on movable electric charges of magnets permanent.

Magnetic field is characterized by vector of magnetic field density (magnetic induction) B. *Magnetic induction* at some field point is equal to the ratio of maximum torque acting on the test frame (of infinitesimal size) with current at the given point to frame magnetic moment:

$$B = \frac{M_{max}}{p_m} = \frac{M_{max}}{I \cdot S}$$

Magnetic field intensity H is another quantitative characteristic of magnetic field. Magnetic field intensity (unlike $\mathbf{B}$) does not depend on magnetic properties. In vacuum vectors and $\mathbf{B}$ are unidirectional and connected with each other by the ratio:

$$B = \mu_0 H,$$

where $\mu_0 = 4\pi \times 10^{-7}$ H/m is an absolute magnetic permeability of vacuum (magnetic constant). Unit of magnetic field intensity in the SI system: $[H] = A/m$.

Vector flux of magnetic induction (magnetic flux) $d\Phi$ through elementary area element dS of open surface S is the quantity:

$$d\Phi = B \cdot dS \cdot cos\alpha,$$

$$\Phi = \int_S B_n dS,$$

where α – angle between normal line to the area element and vector B. Magnetic flux unit in SI system is 1 Weber (Wb) – magnetic flux generated by magnetic field with induction 1 Tl, that goes through flat surface 1 m^2, placed perpendicularly to vector B.

14.2. Biot-Savart-Laplace's law

Magnetic field intensity H generated by direct current may be defined using the law experimentally opened by French physicists J.B. Biot and F. Savart in 1820 and formulated in general form by P.S. Laplace. Let's mark in conductor with current a rather small part dl, which may be considered as vector directed in the same way as current.

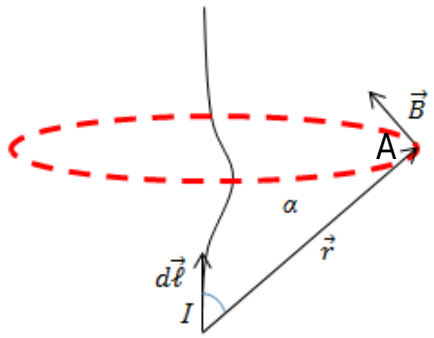

Fig. 14.1. Biot-Savart-Laplace's law.

The product Idl is current element. Draw from the current element Idl radius-vector r at the investigated point A (Fig. 14.1). Then the quantity ofmagnetic field intensity at point A, generated by the given current element is equal to:

$$dH = k\frac{Idl \cdot sin\alpha}{r^2}$$

where α – angle between vectors dl and r, coefficient k depends on the chosen unit system, $k = \frac{\mu_0}{4\pi}$. In SI system: $k = 1/4\pi$.

Force action of magnetic field on conductor with current may be defined in accordance with the empirical Ampere's law (1820):

$$F_A = I \cdot B \cdot dI \cdot sin\alpha,$$

where α – angle created by vectors dl and B.

14.3. Electromagnetic method for measurement of blood velocity

We have seen effects of a magnetic field on free-moving charges. The magnetic field also affects charges moving in a conductor. One result is the Hall effect, which has important implications and applications. Figure 14.2 shows what happens to

charges moving through a conductor in a magnetic field. The field is perpendicular to the electron drift velocity and to the width of the conductor.

The creation of a voltage across a current-carrying conductor by a magnetic field is known as the Hall effect, after Edwin Hall, the American physicist who discovered it in 1879.

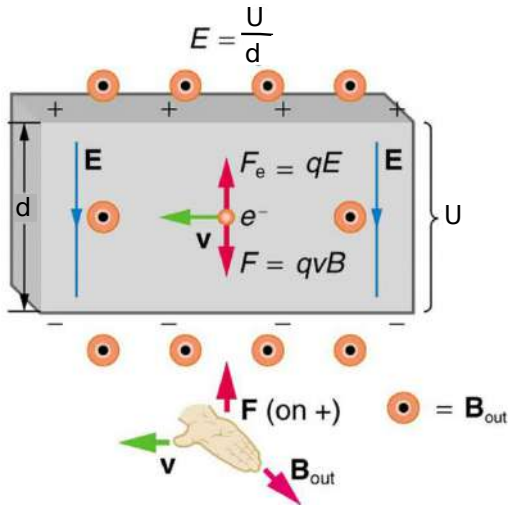

Fig. 14.2. The Hall voltage produces an electric force that balances the magnetic force on the moving charges. The magnetic force produces charge separation, which builds up until it is balanced by the electric force, an equilibrium that is quickly reached.

There are a lot of methods for measurement of blood velocity in vessels of blood circulation system. One of them is based on the action of magnetic field on movable charges. Blood contains considerable quantity of electric charges: ion concentration $Na^+ \sim$ 145 mmol/l, $Cl^- \sim$ 125 mmol/l.

If we place artery of d diameter between magnet poles then Lorentz force $F = evB$ will influence univalent ions. Under its action ions of different signs move towards opposite artery walls and create potential difference U along the vertical line, i.e. electric field with strength $E = U/d$ (**Hall effect**).

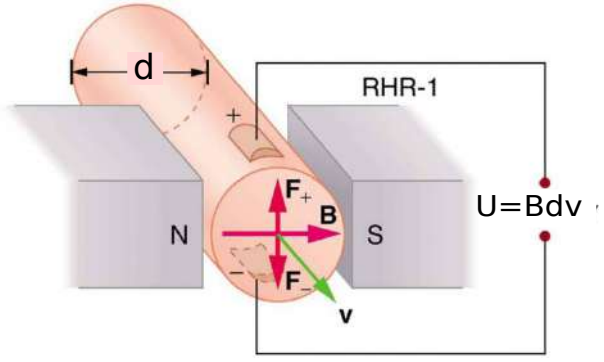

Fig. 14.3. The Hall effect can be used to measure fluid flow in any fluid having free charges, such as blood. The Hall emf is measured across the tube perpendicular to the applied magnetic field and is proportional to the average velocity v.

Charge concentration on the opposite artery walls will increase until force $F_{el} = eE$, on the field generated by them, compensates Lorentz force. From the equality $F = F_{el}$ we may determine ion flow velocity, and accordingly blood flow speed:

$$v = \frac{U}{d \cdot B}$$

Hence, blood flow velocity is in direct proportion to voltage (potential difference at Hall effect), which occurs across the artery if it is brought into magnetic field.

Biocurrents appeared in organism is the source of weak magnetic fields, which sometimes may be registered. Thus, for example, there is a diagnostic approach

called **magnetocardiography**, i.e. registration of magnetic field change of heart during cardiocycle. Advantage of such approach is the absence of electrical contacts that creates necessary prerequisites for remote diagnostics.

Primary physical process are in the basis of action of magnetic field on biological objects. First and foremost they include: a) change of concentration of molecules in heterogeneous medium; b) action of Lorentz force on ions moving together with biological fluid; c) Hall effect that appears in magnetic field at the time of spread of electric impulse; d) different kinetic processes (for example, cross and longitudinal effects of Ettinsgausen related to origination of temperature gradient in the presence of electric current and magnetic field).

It should be stressed that magnetobiology is still in the making and physical nature of action of magnetic field on biological objects has not been sufficiently studied up to now.

14.4. Medical method based on use of alternating magnetic field

Inductothermy. In the course of inductothermy an alternating magnetic field which oscillation frequency is within the range of $10 - 15$ MHz effect patient's tissues. A standard instrument for inductothermy produces magnetic field changing with a frequency of 13.56 MHz. An alternating magnetic field induces eddy currents in tissues, when those currents flow tissues are heated and heat is evolved.

In the course of inductothermy a specific heat power is defined by the formula

$$q = k \cdot \frac{\omega^2}{\rho} \cdot B_0^2,$$

where k is proportionality factor; ω is cyclic frequency of alternating magnetic field; B_0 is amplitude of magnetic induction.

It follows from the formula that tissues having less specific resistance (i.e. those tissues that are good conductors) are heated better.

The tissues are heated effectively up to the depth of $6 \sim 8$ cm. An increase of temperature in tissues intensify the circulation of the blood in them, causes different ferments activation. In the course of inductothermy human body immune system is being stimulated.

In the course of inductothermy a magnetic field is produced with the help of coil (an inductor) through which an alternating electrical current of a corresponding frequency flows. For example, to heat the extremity tissues means to make a coil with the extremity within, so the extremity should be winded with a wire in such a way that a coil will form.

Test questions

1. What are the main characteristics of the magnetic field?
2. Explain the essence of the bio-savar-laplace law.
3. What is the Hall effect?
4. What is the electromagnetic method for measuring blood velocity?
5. Explain the Medical method based on use of alternating magnetic field.

Lecture No 15

Interference and diffraction of light.

15.1. Interference of light waves. Coherence.

Interference of light is used in medicine in interferometers, interference microscopes, at determination of sensitivity of retina and in many other cases.

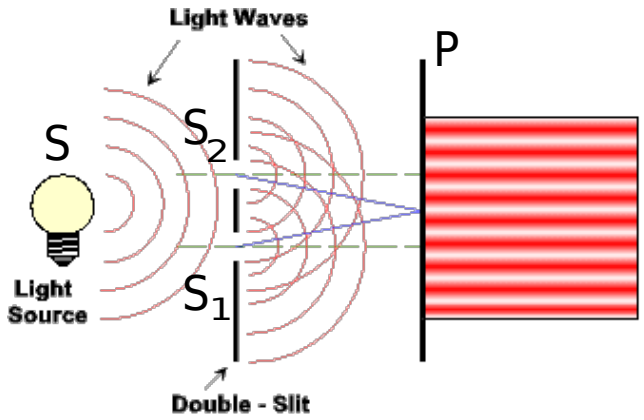

Fig. 15.1. Interference of light waves.

Superposition of two or several waves resulting to steady in time amplification of oscillations in some points of space and attenuation m others, is known as **interference**. In usual conditions often there is superposition of light waves from various sources, but interference of light is not observed. Each such source (a lamp, a flame, the Sun, etc.) represents set of big number of radiating atoms. Difference of phases of oscillations which radiate such sources *will not be constant* and quickly and randomly changes in time. The interference of light can arise only from the coordinated **coherent sources**, which provide constant in time difference of phases $\Delta\varphi$ of composed waves in various points. The waves adequate to this condition are known as *coherent*, i.e. **at identical frequency (length of wave) have the constant phase difference**. The interference could be carried out from two sine wave waves of identical frequency. It is possible to receive coherent waves having divided a wave from one source on two parts (by

reflection or refraction) and then to reduce these two waves together (Fig.15.1). Pinholes are equidistant from S and are close to each other. Spherical waves spread out from S. Spherical waves also spread out from pinholes.

Two waves received by such way will be coherent and at superposition can interfere. In practice division of one light wave on two waves can be carried out by means of the opaque screen with two small apertures. According to principle of Huygens-Fresnel, source S (Fig.15.1) creates in apertures of the screen sources of secondary waves S_1 and S_2, which will coherent. The first supervision of interference was carried out by T.Young in 1802, having passed solar beams through very small aperture in the opaque screen.

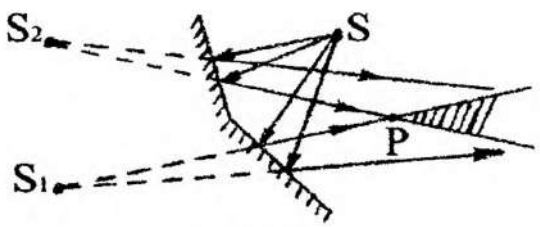

Fig. 15.2. Mirrors of Fresnel.

The second way of reception of coherent sources is based on reflection of light from two flat minors established under the angle α, close to 180^o. This optical system is called **mirrors of Fresnel**. Imaginary images S_1 and S_2 serve as coherent sources of basic source S (Fig.15.2). The picture of interference arises in the point P.

Next way consists in reception of imaginary image $S^{'}$ of the source S by the special single-layered mirror (**Lloyd's mirror**). Sources S and $S^{'}$ (fig.15.3) can be considered as coherent. They create the interference picture in the point A of screen Э.

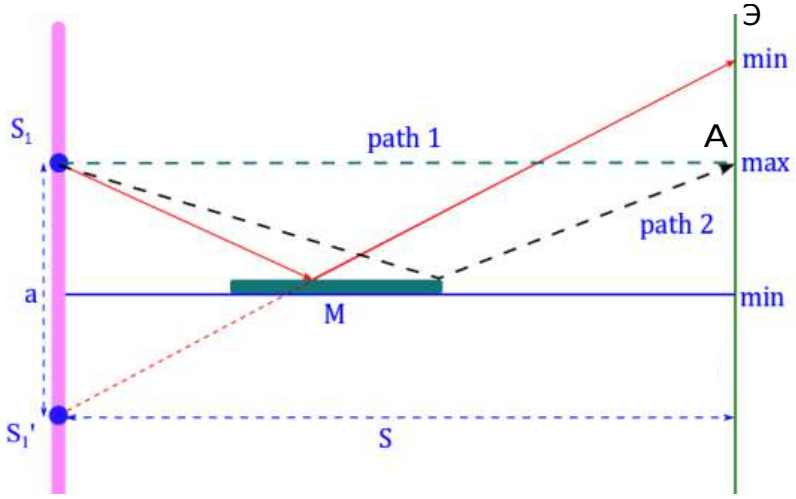

Fig. 15.3. Lloyd's mirror.

For monochromatic light the interference picture represents a number of alternating dark and light strips (maxima and minima). In case of white (not

monochromatic) light maxima for different λ settle down in the different places, because of what interference strips extend and get iridescent colouring.

Formation of coherent waves and their interference occurs also at hit of light on thin transparent plate or film. Due to reflection of light from both surfaces of the film there is a splitting of falling light beam, and there are conditions for interference of light. It speaks occurrence of iridescent colouring of soap bubbles, coloring of film of oily substances on surface of water or of wings of butterflies and other insects, coloring of internal surface of bowls, feathers of some birds (humming-bird, peacocks).

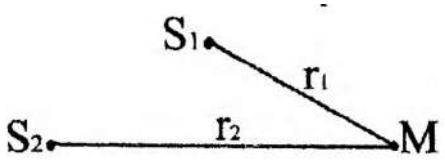

Fig. 15.4. Geometrical difference of ways.

Let's execute calculation of interference picture, when two coherent waves from sources S_1 and S_2 pass different ways r_1 and r_2 and interfere in the point M (fig. 15.4), i.e. between them arises a **path difference** $\Delta = r_2 - r_1$ (geometrical difference of ways). If waves are propagated in the medium with refractive index of n, we can speak about **optical path difference** $\delta = n \cdot \Delta$. Oscillations of vectors of electric intensity E in the point M removed from sources on distance r_1 and r_2 accordingly from each source, occurs under the harmonious law (amplitudes of both oscillations we shall accept identical and we shall designate by E_m). Then $E_1 = E_m \cdot \cos 2\pi(\nu t - \frac{r_1}{\lambda})$, $E_2 = E_m \cdot \cos 2\pi(\nu t - \frac{r_2}{\lambda})$.

Addition of harmonious oscillations of one direction and identical frequency with the phase difference $\Delta\varphi = \varphi_1 - \varphi_2$ as it has been shown earlier, gives resulting oscillations of the same frequency. The amplitude of resulting oscillation of light vector is expressed as:

$$E^2 = E_1^2 + E_2^2 + 2E_1E_2\cos\Delta\varphi, \tag{15.1}$$

and for the case considered by us:

$$E = \sqrt{E_m^2 + E_m^2 + 2E_mE_m\cos\Delta\varphi} = E_m\sqrt{2(1+\cos\Delta\varphi)} = \\ = 2E_m\cos\frac{\Delta\varphi}{2}. \tag{15.2}$$

From the formula $\left[\cos^2\varphi = \frac{1+\cos 2\varphi}{2}\right]$ follows $\frac{1+\cos 2\varphi}{2} = \cos^2\frac{\Delta\varphi}{2}$.

Let's determine:

$$\Delta\varphi = \varphi_1 - \varphi_2 = 2\pi(\nu t - \frac{r_1}{\lambda}) - 2\pi(\nu t - \frac{r_2}{\lambda}) = 2\pi(r_2 - r_1) = \frac{2\pi}{\lambda}\Delta.$$

If light is propagated in the medium with the refractive index n:

$$\Delta\varphi = \frac{2\pi}{\lambda}\delta$$

Substituting in the formula (2) value of Дер, we shall receive:

$$E = 2E_m \cos\left(2\pi\frac{r_2 - r_1}{2\lambda}\right) = 2E_m \cos\frac{\pi\cdot\Delta}{\lambda},$$

i.e. E depends on value Δ.

In the points where Δ *is equal to* **odd number of lengths of half waves, i.e.***:* $\boxed{\Delta = (2k+1)\frac{\lambda}{2}}$ $k = 0, 1, 2, 3, ...,$ *value is* $\cos\frac{\Delta\varphi}{2} = 0$ *and the amplitude of resulting oscillation is equal to zero. In these points are formed interference minima. If the difference of pathes is equal to* **even number lengths of half waves** *(or to the integer of lengths of waves)* $\boxed{\Delta = 2k\frac{\lambda}{2}} = k\lambda$ *that* $\cos\frac{\Delta\varphi}{2} = 1$ *and* $E = 2E_m$ **(interference maximum)**.

15.2. Diffraction of light. Diffraction of light on a slit in parallel beams

At propagation of waves in the medium containing heterogeneity the phenomenon of *diffraction* is observed. **Diffraction** *is bending by waves of the obstacles meeting on their paths, or in more comprehensive sense diffraction is any deviation of advance of waves near to hindrances from laws of geometrical optics.*

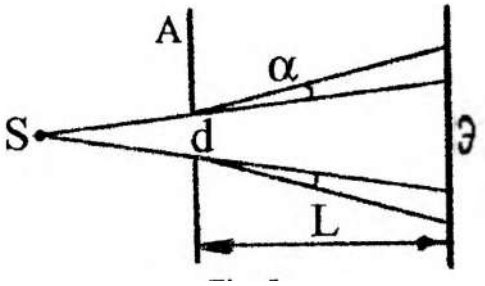

Fig. 15.5. The passage of light through an aperture of small size.

The opportunity of supervision of diffraction depends on *a parity of wavelength of light and the sizes of obstacle.* The phenomenon of diffraction is shown more strongly, if the **sizes of an obstacle (slit) are comparable to length of a wave** λ. The phenomenon of diffraction of light naturally is not

observed almost, because the sizes of the most part of bodies environmental us are *incommensurable to the wavelength of light*. Owing to diffraction the shadow image of object ceases to be similar to the subject.

In experience diffraction can be observed at formation of shadow from an obstacle as thin wire or hair and also at passage of light through an aperture of small size (share of millimeter) (fig. 15.5).

If between the screen Э and a source of monochromatic light S to place other opaque screen A with the small aperture d, the border of the geometrical shadow will not be sharp. It is especially appreciable, when the size d of aperture is very small in comparison with distance L from the screen up to the aperture ($d << L$). Then the stain on the screen will be submitted as system of the alternating light and dark rings gradually passing each other, grasping the area of a geometrical shadow and also leaving for its limits. It speaks about not rectilinear propagation of light from the source S, about bending of light waves at edges of aperture in the screen A. At use of white light diffraction picture gets iridescent colouring. Diffraction of light speaks occurrence of iridescent rings around of light source, when air is sated with a fog or dust, colouring of pearls (diffraction of white light on the alien smallest contained in it particles).

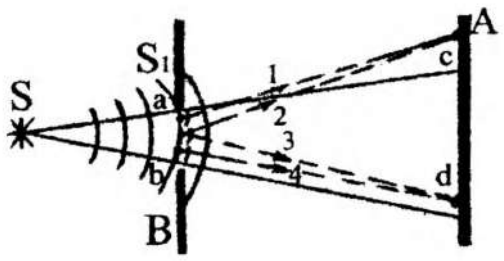

Fig. 15.6. The principle of Huygens-Fresnel.

Diffraction is defined by wave properties of light and to explain this phenomenon is possible by *principle of Huygens-Fresnel according to which:* **points of medium of which were reached with front of a wave are sources of elementary secondary waves, which are coherent**. Let light from the source S falls on the screen through the round aperture "ab" in the screen (fig.15.6). Each point of site "ab" of front of light wave S_1 is the secondary light source. Secondary sources are coherent, therefore beams (waves) proceeding from them $1, 2$; $3, 4$, etc. will be interfere among themselves.

Depending on path difference of beams on the screen A in the points c; d, etc. will arise interference maxima and minima, i.e. ring-shaped diffraction picture.

To define result of diffraction in some point of space, it is necessary to calculate according to principle of Huygens-Fresnel the interference of the secondary waves, which have got in this point from a wave surface. For a wave surface of the any form such calculation is difficult. But on occassion (the spherical or flat wave

surface, a symmetric arrangement of a point concerning wave surface S_1 and opaque barrier A) calculations are simple. The wave surface thus is breaking on separate sites (*zones of Fresnel*) that simplifies mathematical calculation.

Let's consider result of diffraction of flat monochromatic waves on the slit which have been cut out in the opaque barrier (fig. 15.7) and having constant width $BC = a$. Wave falls normally to slit as plane-parallel bunch of light. All points of the wave surface open by the slit are the centers of the secondary waves extending behind the slit on all directions. To represent all these secondary waves is impossible, therefore on fig. 15.7 are shown only secondary waves extending under the angle φ.

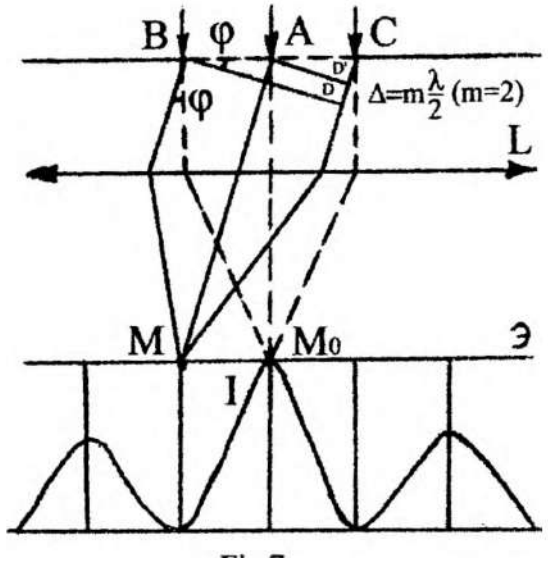

Fig. 15.7. The diffraction of flat monochromatic waves on the slit which have been cut out in the opaque barrier.

Beams, diffract under the angle φ will be collected in the point M of the screen and interfere. At $\varphi = 0$ all waves will come in the point M_0 in the identical phase and will strengthen each other, on the screen will appear the most light strip (the central maximum). To define result of interference at $\varphi \neq 0$, from the point we shall lead perpendicular BD to the direction of bunch of secondary waves. Optical ways of beams from BD up to point M of the screen are identical (lens L of additional path difference does not bring), therefore the path difference DC of extreme beams is equal: $\Delta = a \sin \varphi$. We break DC into the pieces equal to $\lambda/2$ (in figure 2 pieces are shown). Generally this path difference contains m half waves: $a \sin \varphi = m\dfrac{\lambda}{2}$. Let's lead from the point D' straight line $D'A$ parallel of BD and we shall divide BC into two equal *zones of Fresnel* $BA = AC$. To any secondary wave going from any point of one zone of Fresnel it is possible to find in the next zones such secondary waves, that the path difference between them will be $\lambda/2$. For example, the secondary wave going from the point C in the chosen direction passes up to the point M distance on $\lambda/2$ more, than the wave going from the point A, etc. Hence, the secondary waves going from two next zones will extinguish each other, since differ on phase on π. The number of zones stacked in a slit, depends on length of wave λ, and

angle φ. If slit BC is broken at construction on odd number of Fresnel zones $m = (2k + I)$, and DC on odd number of the pieces equal to $\lambda/2$, in the point M is observed diffraction maximum, i.e.: $\boxed{a \sin \varphi = (2k+1)\frac{\lambda}{2}}$ $k = 0, 1, 2....$ is order of the maximum. Condition of diffraction minima:

$$\boxed{a \sin \varphi = 2k\frac{\lambda}{2} = k\lambda} \quad (m = 2k).$$

On fig. 15.7 is shown the case, when $m = 2$, that corresponds in the point M to the diffraction minimum. Thus, on the screen $\Im$ the system of light (maxima) and dark (minima) strips symmetrically located to the left and to the right from central ($\varphi = 0$) the brightest strip will turn out. Intensity I of other maxima decreases on measure of removal from the central maximum (see fig. 15.7). If the slit is shined with white light, on the screen $\Im$ is formed the system of color strips; only central maximum will keep color of frilling light, as at $\varphi = 0$ amplify all length of waves of light.

15.3. Diffraction grating. Diffraction spectrum

Diffraction grating is the optical device representing a glass plate on which by the diamond edge renders a plenty of parallel lines with intervals between them.

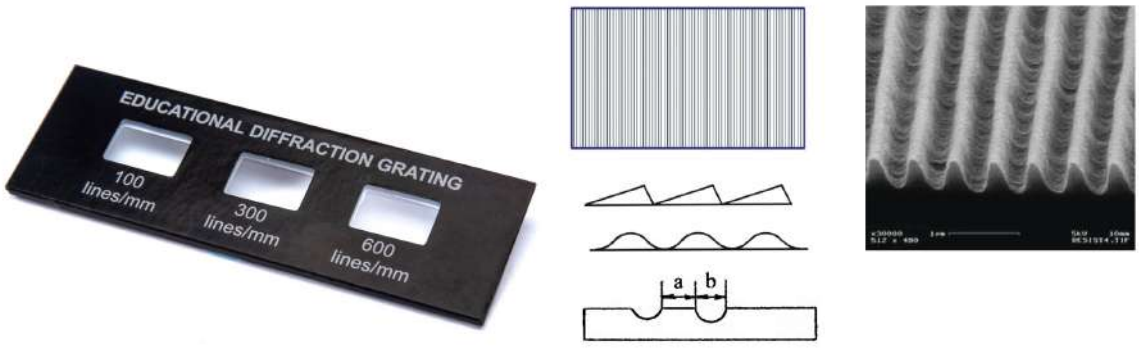

Fig. 15.8. Diffraction grating.

The intact glass between lines serves as slits of the grating. Total width of the slit "a" (fig. 15.8) and interval "b" between slits is named *the constant or the period of diffraction grating:* $d = a + b$. The best diffraction gratings have up to $1200 - 1500$ slits per millimeter (n), n is general number of lines of a diffraction grating.

Let's consider diffraction (fig. 15.9) of the flat monochromatic wave falling normally on the diffraction grating. If on the grating falls the beam of monochromatic beams, the secondary waves going from slits on all directions are coherent and will interfere, forming a diffraction picture. If between the screen and the diffraction grating to place a collecting lens L (the screen is located in the focal plane of the lens), there is diffraction picture which is growing out of two processes: diffraction of light from each separate slit and interference of light from slits. The basic features of this process are defined by the second phenomenon.

Let's consider the beams falling on the left edges of slits ($N = 3$). Due to diffraction, light from slits will propagated in every possible direction. The path difference of beams from extreme points of two next slits and diffracting under the angle φ, is defined: $\Delta = d \sin \varphi$.

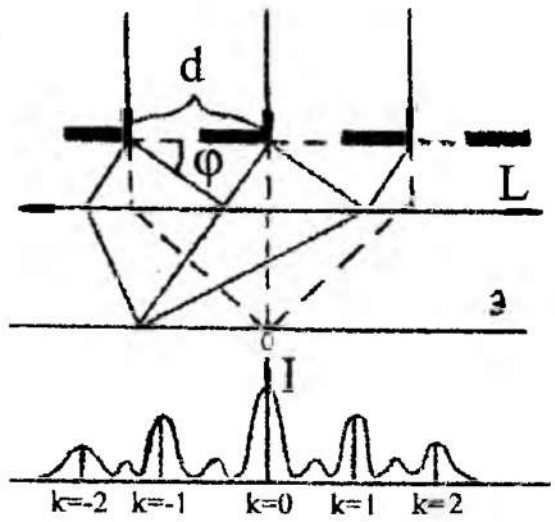

Fig. 15.9. The diffraction of the flat monochromatic wave falling normally on the diffraction grating.

If this path difference will be equal to zero or an integer number of wavelengths, at interference there are main maxima for which the condition satisfies: $d \sin \varphi = k\lambda$, where k is the order of the main maxima ($k = 0; 1; 2; 3; ...$). At performance of the condition $d \sin \varphi = (2k+1)\dfrac{\lambda}{2}$ will arise interference minima.

Maxima will symmetrize concerning central ($k = 0;\ \varphi = 0$). **Expression $d \sin \varphi = k\lambda$ named the basic formula o f a diffraction grating.** Possible number a maximum is limited, it cannot be more than d/λ. Between the main maxima are formed minima (additional), which number depends on number of slits of a grating (N). Between the main next maxima is settled down $(N-1)$ additional minima. On fig. 15.9 distribution of intensity of maxima on the screen for the grating with $N = 3$ is shown.

At plenty of slits separate additional minima practically do not differ and all space between the main maxima looks dark. As the amplitude of light oscillations in maxima is proportional to number of slits, intensity of maxima is proportional to the square of number of slits (N^2), i.e. than is more number of slits of a diffraction grating, than the main maxima are especially sharp: $E_m \sim N$; $I \sim E_m^2$;

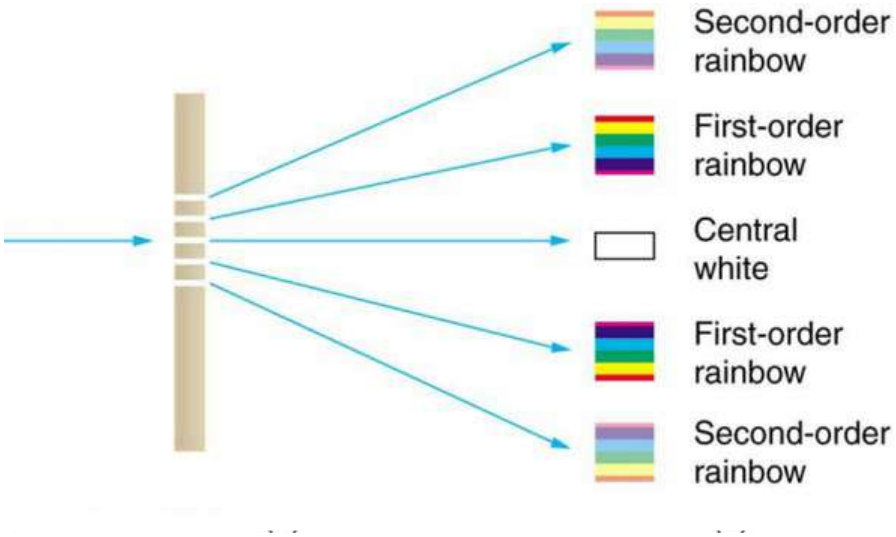

Fig. 15.10. The central maximum will look like a narrow white strip and each of lateral maxima represents multi-coloured strip of certain width.

$I \sim N^2$.

If on a grating falls white light for all values of wavelengths position of maxima of zero order coincide ($k = 0$ falls; $\varphi = 0$), position of maxima of higher order will be various (more λ than is more φ) for the given value k. Therefore the central maximum will look like a narrow white strip (fig. 15.10) and each of lateral maxima represents multi-coloured strip of certain width.

All spectra are symmetric concerning zero maximum and are inverted to it by short-wave (violet) edge. The most intensive is the spectrum of 1-st order ($k = 1$).

Thus, diffraction grating decomposes complex light into a spectrum and consequently with success is applied in spectral devices, for example, in diffraction spectroscope; it is a device, employee to measurement of length of light waves, i.e. for carrying out of the spectral analysis.

15.4. Diffraction of electromagnetic waves on spatial structures. Bases of X-ray crystal analysis

Diffraction of waves can occur on small heterogeneities and particles. Most simple case is when heterogeneity forms periodic structure. Diffraction gratings are examples of the periodic structure. 3 – measured structure crystals are natural, in which as the scattering centers serve units (atoms, ions) of a crystal grating.

In a crystal it is possible to allocate directions along which diffract waves reinforce each other. The beam of monochromatic radiation, passing through such structure forms on the screen plane two –dimensional diffraction picture, i. e. system of the light spots (maxima) located in the certain order. On the

location of these maxima, their relative intensity and the wavelength it is possible on the basis of corresponding calculations to define spatial 3 – measured structure of the object, which has caused diffraction. As such objects there can be large molecules.

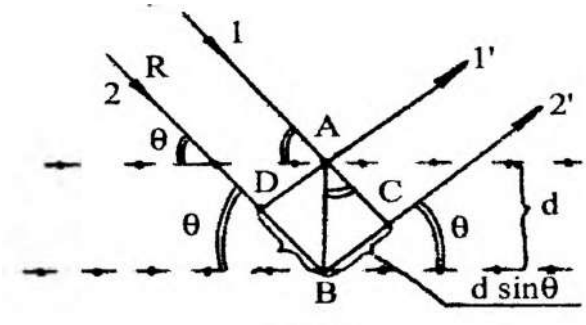

Fig. 15.11. The diffraction of X-rays can be observed at reflection from a crystal.

However precise diffraction picture can be received only, if the period of structure d will be a little bit more of the wavelength λ ($d > \lambda$). This restriction does not allow carring out diffraction of light on crystals, as the period of a crystal grating (distance between planes of crystal) in thousand times ($\sim 10^{-10}\ m$) is less than length of light wave. However for X-rays the condition $d > \lambda$ is carried out. In 1912 M. Laue having passed the narrow beam of X-rays allocated by lead diaphragms through the monocrystal (the crystal plays a role of a spatial grating) has received on photographic plate the diffraction picture (lauegram) as dark spots (diffraction maxima). *The order of arrangement ofparticles in a grating of the crystal defines the order and symmetry in arrangement of diffraction maxima.*

Diffraction of X-rays can be observed at reflection from a crystal (fig. 15.11). Beams 1 and 2 interfere having reflected from two next layers (d is distance between the next nuclear layers). Path difference $\Delta = DB + BC$ of beams is equal to $2d\sin\theta$, where θ is the angle of sliding. These waves reinforce each other, i.e. form maxima in directions for which the path difference is multiple of λ. Therefore is possible to write down:

$$2d\sin\theta = k\lambda, \tag{15.3}$$

where $k = 1; 2; 3; ...$ is the order of maxima. **It is Woolf - Bragg's formula.** At falling of monochromatic -ray radiation on a crystal under various angles, the maximum will take place for the angles adequate to the formula of Woolf - Bragg. The angle θ is measured in photo of diffraction pictures (on position of diffraction maxima).

X-ray crystal analysis is the method which on diffraction picture received on unknown crystal structure by means of X-rays of known length allows to find

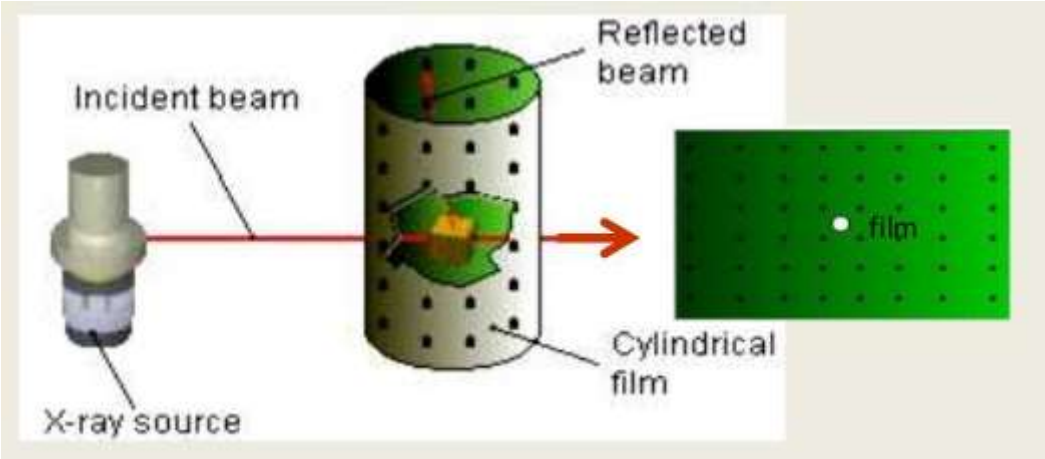

Fig. 15.12. The method of "rotating crystal".

the arrangement of particles making this structure (to determine d). It has the big practical value for biology, as it is the most effective method of definition of spatial structure of crystal connections. For monocrystal the method of "rotating crystal" is usually used (fig.15.12). At rotation of the crystal the different systems of planes get under the falling beam of X-rays in position at which is carried out the condition (15.3). It means that at the certain angles θ on the photographic plate will appear diffraction maxima. Knowing λ and having determined from experience the θ and order of spectrum k, under the formula (15.3) calculate distance d between corresponding layers of structural particles.

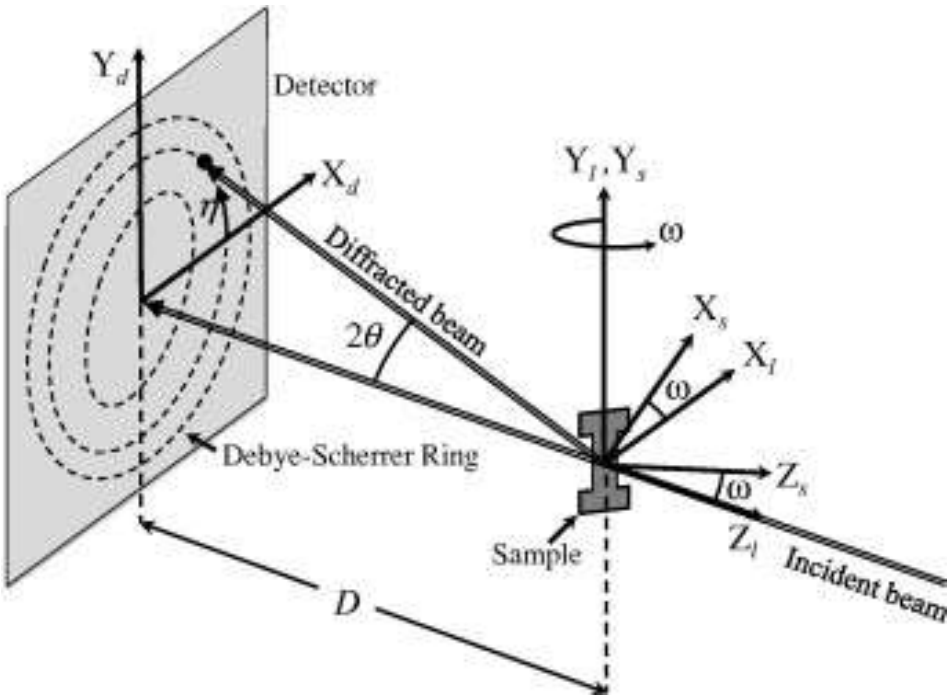

Fig. 15.13. The method of Debuy and Sherer.

For polycrystalline bodies use the method of powders (method of Debuy and

Sherer) (fig. 15.13). Among the big number of fine crystals always will be such crystals, for which the condition (15.3) satisfies. As the condition (15.3) is identical for many crystals on miscellaneous oriented, diffract beams form the cone, which top lays in researched object in space and the angle is equal to 4θ. **The X-rays crystal analysis is widely applied at research of structure of biological molecules (DNA, proteins).** By described method deciphers spatial structures of hemoglobin, ribonuclease, etc. Being based on the analysis of roentgenograms, F. Kric and J. Watson have reproduced spatial structure of DNA and have been awarded with the Nobel Prize. With help of X-rays crystal analysis it was possible to understand functioning of molecules of enzymes, to find out structural basis of many hereditary diseases, structure of viruses, etc.

Test questions

1. What is the phenomenon of light diffraction?

2. Formulate the Huygens – Fresnel principle. What relation has this principle to explain the diffraction of light?

3. What are called Fresnel zones? For what purpose is the wave surface (wave front) divided into these zones?

4. What diffraction pattern arises when a plane-parallel light beam hits a narrow long slit located in a flat opaque barrier? Indicate the appropriate formulas.

5. Explain the diffraction of light by a diffraction grating. Provide formulas for major highs and incremental lows.

6. What is the role of a minimum from one slit in the formation of a diffraction pattern from a diffraction grating?

7. What is called a diffraction spectrum?

8. Why are characteristics such as angular dispersion and resolution introduced?

9. Explain an example with the diffraction of a spherical wave by a round holes.

10. Why does a diffraction grating decompose white light into a spectrum?

11. How to determine the highest order of the spectrum of a diffraction grating?

12. Why is not visible light diffraction observed on crystals and X-ray diffraction is observed?

13. What practical application does the Wolfe-Bragg formula have?

Lecture No 16

Polarization and dispersion of light

16.1. Polarization of light. Light natural and polarized. Malus law

The electromagnetic wave emitted by separate atom can be presented as oscillation of 2 mutual perpendicular vectors of intensities of electrical (E) and magnetic (H) fields. Both vectors change in the plane, perpendicular to velocity vector. Such electromagnetic wave is called *plane-polarized*. By reviewing phenomenon of polarization we further shall conduct all reasonings concerning of vector E, as experience and the theory shows, that chemical, physiological, etc. actions of light on material are stipulated mainly by vector E. Light from the Sun, from filament of bulb, etc. is *unpolarized*, natural. In such light vectors of E from the different elementary radiators have various orientations of oscillations. Projections of vectors E of natural light on the plane, perpendicular to velocity, will look like figured on Fig.16.1a All orientations are equiprobable and amplitudes of values E are equal in all directions.

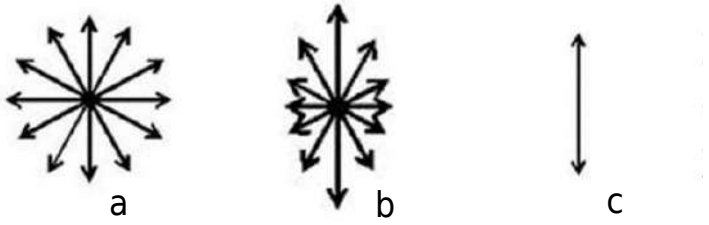

Fig. 16.1. The natural (a), partially polarized (b) and linearly polarized (c) light.

If there is a preferable direction of oscillations, such light is called partially polarized (Fig.16.1b). Natural light is possible to turn into polarized, i.e. to polarize with the help of the devices called as polarizers. The diagram of linearly polarized light is presented on Fig.16.1c. Polarizers are capable to pass only component of vector E laying in the certain plane PP', called as the *principal*

plane of a polarizer (Fig.16.2). Thus through the polarizer passes polarized light, which intensity is equal to half of intensity of incident light: $I = I_{NAT}/2$.

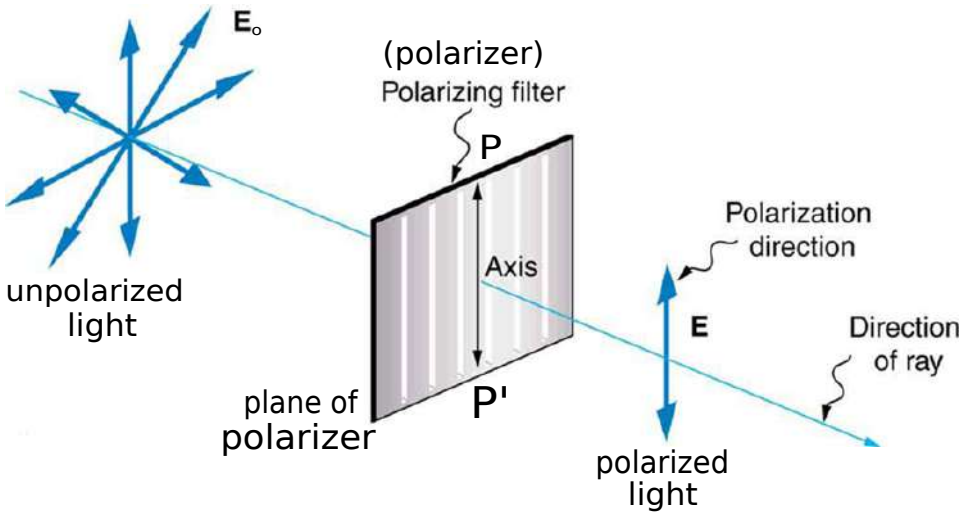

Fig. 16.2. The polarizer.

At rotation of polarizer concerning of the ray of natural light, the vibration plane of plane-polarized light rotates also, but intensity of it does not change.

The polarizer can be used for analysis of polarized light, then it is called as the *analyzer*.

If the plane-polarized light falls on the analyzer of intensity E_o, it passes only component $E = E_o \cos\varphi$, where φ is angle between principal planes of the analyzer and polarizer. As light intensity is proportional to square of amplitude of oscillations, then

$$\boxed{I = I_o \cos^2 \varphi}. \tag{16.1}$$

The equation (16.1) is **Mains' law**, where I_o is intensity of the plane-polarized light, which has left polarizer, I is the light intensity, which has left the analyzer.

Properly from the equation (16.1), that at rotation of the analyzer concerning the beam, intensity of light which has left the analyzer is changed from 0 up to I_o. If at rotation of analyzer concerning the ray intensity of light, which has left it does not change, the incident light is natural; if light changes under the law of Malus, then light is plane-polarized.

Eye of a person is not capable to distinguish polarized light from natural, but approximately $25 - 30\%$ of people have this ability, though almost never suspect about it.

16.1.1. Polarization of light at reflection and refraction on border of

two dielectrics

If the angle of incidence of light on the border of 2 dielectrics is not equal to zero, the reflected and refracted rays appear in part polarized. In the reflected ray oscillations, perpendicular to the plane of incidence prevail (on Fig.16.3a they are marked out by points); in the refracted ray prevail oscillations parallel to the plane of incidence (on Fig.16.3a they are marked out by two-sided arrows). The natural ray is conditionally designated by alternating arrows and points.

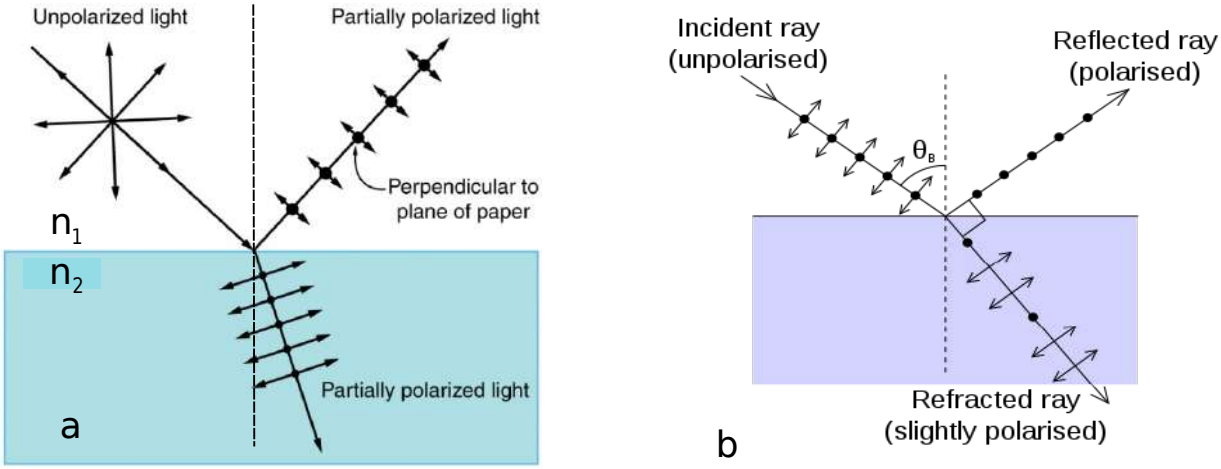

Fig. 16.3. The polarization of light at reflection and refraction.

Degree of polarization depends on the angle of incidence of rays and on the refractive index of reflecting medium. At angle of incidence, satisfying to the requirement

$$\boxed{\tan\theta_B = n} \tag{16.2}$$

where n is refractive index of the second medium concerning the first, the reflected ray is completely polarized, refracted is in part polarized, but degree of its polarization the greatest. The **equation (16.2) is Brewster's law**, angle θ_B is called *Brewster's angle*. It is easy to test, that at falling of light under the Brewster's angle, the *reflected and refracted rays are mutually perpendicular* (Fig.16.3b).

Thus, the border of two dielectrics or a dielectric and vacuum is a polarizer. The polarity effect of a reflected light is used, for example, at detection from air of a film of petroleum on water plane.

The Brewster's law is inapplicable in case of reflectance of light from a surface of conductors (metals).

It is possible to achieve, that the refracted ray will be completely polarized. For this purpose as polarizer is used the pile of the glass plates located one after another. At realization of the Brewster's law the degree of polarization of refracted ray will increase in accordance with transit of plates.

16.1.2. Polarization at birefringence (double refraction)

At transit of light through some crystals the light ray is parted on two rays. This appearance has received the name of **birefringence** (Fig.16.4).

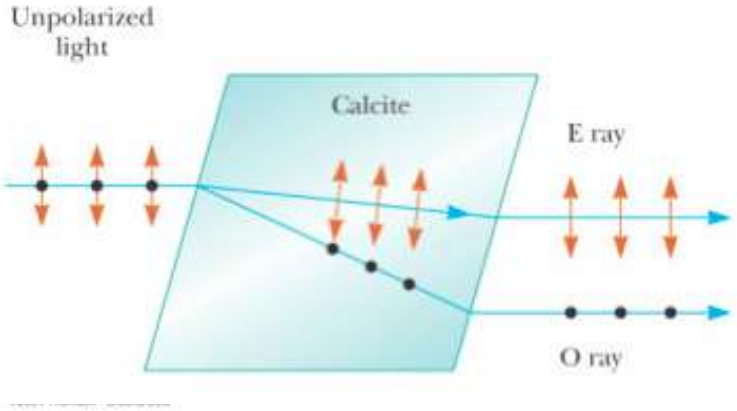

Fig. 16.4. The polarization at birefringence.

One of rays on Fig.16.4 (O) fulfils to the Snell refraction law, it is called **ordinary**, for the second ray which has been marked out by E (which is known as **extraordinary**) the ratio $\frac{\sin\theta_{INC}}{\sin\theta_{REFR}}$ remain the constant and depends on the direction of ray incidence.

The ordinary ray at normal falling of light on a surface of crystal passes not refracting as it follows from the Snell refraction law, extraordinary ray (fig.16.4) refracts.

Crystals have directions for which ordinary and extraordinary rays are propagated not being parted and with equal velocity. These directions are termed as *optic axises* of a crystal. If such direction is one, crystals are termed *uniaxial*, if two are termed *diaxonic*. To uniaxial crystals concerns the Iceland spar (type of calcium carbonate $CaCO_3$), quartz, tourmaline (the composite aluminosilicate). On Fig.16.4 this direction is shown by the shaped line. Plane which is taking place through optic axis and light ray is called the **main plane**. Both rays, which have left a crystal are completely polarized in the mutually perpendicular planes. Oscillations of an ordinary ray are perpendicular to the main plane and oscillations

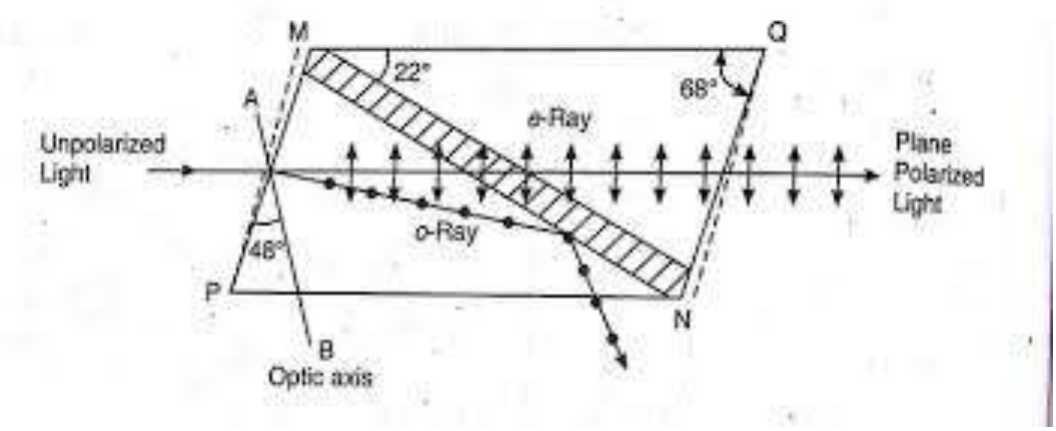

Fig. 16.5. The Nicol prism.

of extraordinary are in the main plane, Birefringent crystals immediately are not used as polarizers, however from them produce special polarization prisms.

In some crystals one of rays is absorbed more strongly another. This appearance is known as **dichroism**. The strong dichroism has crystal of the *tourmaline*. In it the ordinary ray is practically completely absorbed on length of 1 mm and light which has left is plane-polarized. The same property has the **polaroid**, it is celluloid film on whom is lined the quantity of equally oriented crystals of herapathite (iodine sulphate – quinine). The layer of polaroid depth of 0.1 mm completely absorbs the O-ray. Hence, the polaroid can be used as polarizer.

High-quality polarizer is the polarization prism of the Nicol (or primely the **Nicol**). Action of the Nicol sets up on birefringence of the Iceland spar $CaCO_3$. The Nicol is the prism intagliated by the special mode from iceland crystal, cut almost on diagonal and pasted together by Canada balsam (pitch of spruce fir), which *refractive index* ($n = 1.550$) *lays between values of refractive indexes of iceland crystal for ordinary ray* ($n_0 = 1.658$) *and extraordinary ray* ($n_e = 1.486$). It allows having puck up in appropriate way angles of the prism to provide *total reflection* of O-ray on the border with Canada balsam. This ray is absorbed by black lower plane of prism. The extraordinary ray goes out from prism parallel to lower plane (fig.16.5).

Defect of the tourmaline and polaroids in comparison with the Nicol is their bad wavelength characteristics. The white light after their transit becomes painted, while the Nicol is transparent in the visual part of spectrum.

Double refraction is explained by **anisotropy** of crystals (it is dependence of optical properties from direction). The majority of diaphanous crystals are optically anisotropic. In them speed of light and, hence, refractive index are

various on different directions.

16.2. Optical rotation. Polarimetry

At transit of the plane-polarized light through some materials is observed rotation of the plane of oscillation of E. Such materials are termed *optically active*. To them belong crystals (quartz, film-pitch), pure liquids (turpentine, nicotine) and solutions (water solutions of Saccharum, tartaric add, etc.). Light after escaping of material is plane-polarized, but the vibration plane of its vector E appears rotated on the angle φ. Optical rotation for the first time was revealed on crystals of *quartz*.

For crystals the rotation angle of polarization plane is proportional to the path l, traversed by the ray in the crystal:

$$\boxed{\varphi = \alpha l}$$

where α is specific rotation, it is equal to rotation angle of polarization plane of material by the layer of unit length. It is accepted to express it in $degree/mm$.

Specific rotation depends on wave length λ of light in which observation is conducted. This dependence is called the *dispersion of rotatory power*, it is individual for each material.

In solutions the rotation angle of polarization plane is proportional to the path of ray in solution l and concentration of active material C:

$$\boxed{\varphi = \alpha \cdot C \cdot l},$$

where α is specific rotation. l is accepted to express in $dm.$, C in g/cm^3, φ in degree.

If concentration presented in grammes in 100 cm^3 of solution, then

$$\boxed{\varphi = \frac{\alpha \cdot C \cdot l}{100}}.$$

In dependence on direction of rotation of polarization plane, optically active materials are sectioned into the right and laevorotatory. There is left and dextrorotary quartz (if to look towards to a ray), saccharum, etc. materials. Numerical values of rotational constant for both types are equal.

Optical activity of materials is stipulated by *asymmetry* of their molecules, which are not having neither planes, nor the center of symmetry. Optical activity of many biopolmers is stipulated, in particular, by casual frame of their moleculas. All proteins built only from laevorotatory aminoacidic oddments. Apparently because of it organic isomers can strongly discriminate on physiological action. For example, the dextrorotary nicotine is more toxicant, than the laevorotatory nicotine, laevorotatory epinephrine renders more the strong hormonal action, than dextrorotary epinephrine.

Except of natural optical activity, the material can have synthetic optical activity, which originates in it under effect of external actions, for example, at addition of material into a magnetic field (the **phenomenon of Faraday**).

If between the crossed polarizers to locate optically active material the visual field will brighten up. Again to receive the dark field is necessary to rotate the second polarizer on the angle φ. Knowing specific rotation α of the given material, length l, having measured φ, it is possible to find C. Such method of determination of concentration of material is known as **polarimetry** or **sacharimetry**. The devices used for this purpose are known as **polarimeters**.

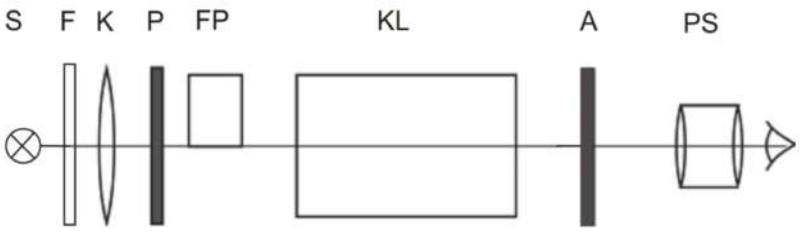

Fig. 16.6. The Schema of polarimeter.

In the physics lab half shadow polarimeters are used. Their optical scheme is presented in Fig.16.6. Light from the light source S first passes the filter F, which separates the sodium D-line from the spectrum. Next, before reaching the eye, the light passes the condenser K, polarizer (polaroid) P, the phase plate FP which divides the field of view into two parts, the test cell KL filled with liquid, analyzer (polaroid) A and a spyglass PS. The phase plate is an optical device which separates the light into two partial waves with different speeds, causes the phase difference (gearchange), and then interfers them. In polarimeters the half-wave plate (with gearchange $(2k+1)\frac{\lambda}{2}$), which is oriented so that the polarization plane of the polarized light is rotated by a relatively small angle.

The telescope's field of view of the polarimeter which is divided into two halves by a phase plate, at different positions of the main plane of the analyzer AA is

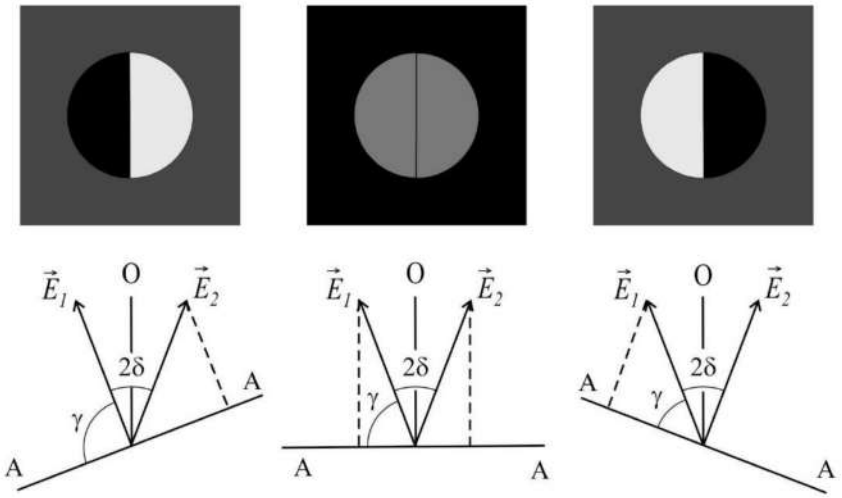

Fig. 16.7. The oscillation plane of the E-vector in the left and right half of the field of view of polarimeter.

shown in the figure Fig.16.7.

Since the phase plate FP covers the left side of linearly polarized light coming out from the polarizer, only right half of the light falls directly on the analyzer (if there is no cuvette with a liquid solution). The left side of the beam passes the half-wave plate before reaching the analyzer. The half-wave plate rotates the oscillation plane of the E-vector anticlockwise by the angle $2\delta = 7^o$ (Fig.16.7). In the uncovered half of the field of view, the oscillation plane of the E-vector remains unaffected.

Thus, the oscillation plane of the E-vector is different in the left and right half of the field of view. (In Fig.16.7 $\vec{E_1}$ is the amplitude of the E-vector in the left half, and $\vec{E_2}$ on the right.)

According to the Malus'law, the light intensity of the left side of the field of view (with index 1) and the right side (with index 2) manifest as follows:

$$I_t = I_0 \cos^2 \gamma, \tag{16.3}$$

$$I_2 = I_0 \cos^2(180^o - \gamma - 2\delta) = I_0 \cos^2(\gamma + 2\delta). \tag{16.4}$$

Here I_0 is light intensity falling on the analyzer in each half of the field of view. On the left diagram of Fig.16.7 $\gamma = 90^o$, $\gamma = 90^o - \delta$ on the central and $\gamma = 90^o - 2\delta$ on the right diagram.

According to formulas (16.3) and (16.4) the condition of $I_1 = I_2$ is fulfilled only in two cases: first, when $\gamma = -\delta$ and, second when $\gamma = 90^o - \delta$.

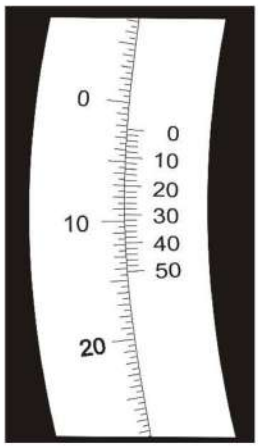

Fig. 16.8. The circular verniers of the polarimeter

In the first case, both sides of the field of vision are of maximum brightness, in the second case the minimum. Both states can be considered as zero positions of the polarimeter. However, considering characters of the eye together with the changing light intensity, the measurements (rotating the analyzer's main plane AA) are made nearby $\gamma = \frac{\pi}{2}$. Rotating the analyzer by a small angle near $\gamma = \frac{\pi}{2}$, there will be a considerable change of intensity in the halves of the field of vision (once the left, once the right side darker, as it is shown in Fig.16.7).

Based on the foregoing rotation angle of the analyzer is always measured in the polarimeter zero position (Fig.16.7 Average share), where both sides of the field of view are equally weakly lighted ($I_1 = I_2$). Then the analyzers main plane AA is perpendicular to the bisector O of the angle 2δ. In the future we will name this position – the zero position.

Angle readings are taken from two symmetrically placed circular verniers of the polarimeter. One of them is shown with a main scale in the Fig.16.8. From the main scale grades and half-grades are taken, smaller parts from the vernier. It is important that readings must be taken with accuracy of the vernier. In the Fig.16.8 we can see the reading 2.24^o.

In medicine the polarimeter is used for determination of saccharum in urine, for biophysical researches. By this method is possible to distinguish dextrorotary paravariations of materials from laevorotatory. For example, laevorotatory Chloromycetin is the fissile antibiotic, while dextrorotary Chloromycetin has no medical properties. Use of dispersion of rotation gives good outcomes at research of biopolimers. It is very sensitive to any changes in molecular composition.

16.2.1. Polarizing microscope

Viewing transparent biological objects in a microscope, it is difficult to reveal different structures, therefore sometimes use polarization microscopy.

The polarizing microscope (Fig.16.9) is analogous to a routine biological microscope, but has analyzer A between objective and eyepiece and polarizer P before condenser. Thus, the object is illuminated by polarized light and is looked through the analyzer. If to cross polarizer and analyzer the visual field be dark.

Series of tissues (muscle, osteal, nervous, optical mediums of eye) have ***optical anisotropy*** *(difference of optical properties on different directions).*

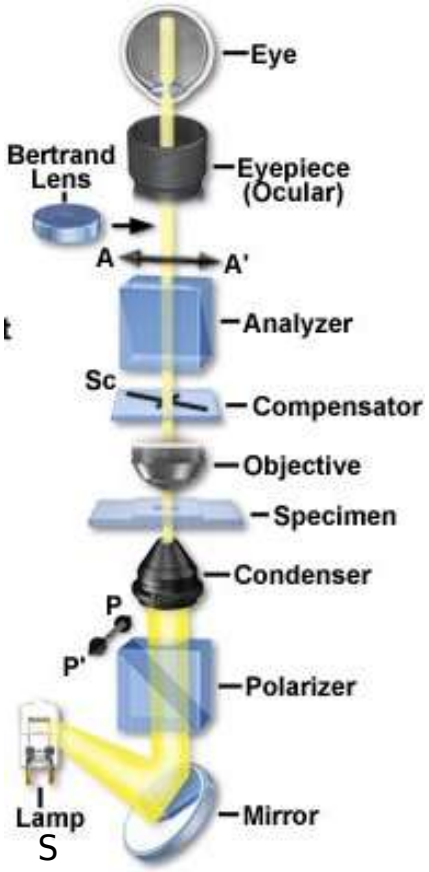

Fig. 16.9. Polarized Light Microscope Optical Configuration Schematic Diagram

If between polarizer and analyzer to locate any tissue with the anisotropic structures then light, past polarizerf) be have double refraction in tissue also. In this connection polarized light is not extinguished completely by the analyzer, and the relevant structures appear light on a blanket dark background. For example, for estimation of the mechanical strain incipient in osteal tissues, from the transparent isotropic material, for example, from the Perspex frame flat model of bone. Loading, we can produce anisotropy of the Perspex that becomes noticeable on the reference pattern of strips and spots. On this pattern it is possible to make conclusions about mechanical strains incipient in models, so, and in a nature. The polarizing microscope makes available observation of objects, which are difficult for observing by other methods (chromosomes, process of division, etc.).

16.3. Dispersion of light

Dispersion of light is called the phenomenon caused by dependence of refractive index of material n from frequency (or lengths) of light wave λ: $n = f(\lambda)$.

In most cases on the border of two different transparent mediums the short wave radiation refracts more strongly, than the long-wave radiation.

Distribution of any radiation on lengths of waves (or on frequencies) is known *as spectrum* of this radiation.

By corollary of light dispersion is decomposition of white light to spectrum at passage through a prism, for the first time explored by **Newton** (Fig.16.10).

To explain dispersion is possible as follows. Under action of transiting electromagnetic wave electrons of medium start to make the harmonic forced oscillations with frequency equal to frequency of the transmitted wave. Oscillating electrons

Fig. 16.10. Dispersion of light.

radiate secondary waves of the same frequency. Between the primary wave and secondary waves the phase shift is formed, caused by retardation of oscillations of electrons. The resultant wave (from initial and secondary waves) is dephased also in comparison with the initial and consequently has other speed of propagation.

This phase shift depends on the oscillation frequency of electromagnetic field, i.e. *light of various lengths of waves will have different speeds of propagation v and so, different refractive indexes n, as* $n = \frac{c}{v}$, where c is speed of light in vacuum. The least refractive index will have red colour and the greatest is violet.

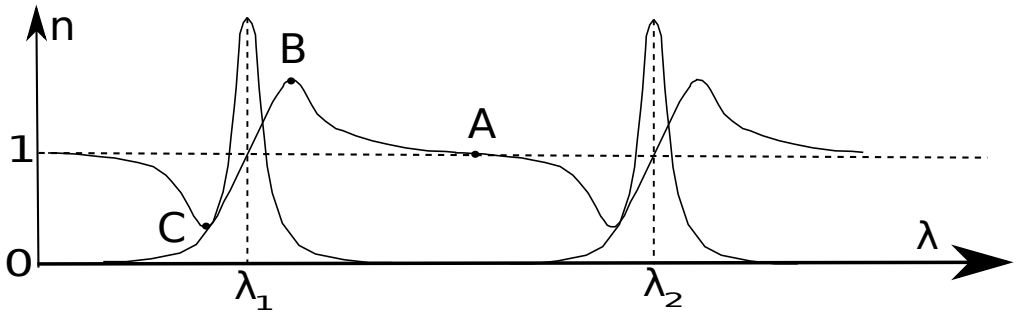

Fig. 16.11. The wavelength dependence of the refractive index.

For all transparent colourless materials function $n = f(\lambda)$ is shown on Fig.16.11.

With decrease of wave length the refractive index in the beginning is incremented. The quantity $\frac{dn}{d\lambda}$ is known as **dispersion of material**, it is $\frac{dn}{d\lambda} < 0$ on part AB. Such dispersion is called the *normal*. At very small lengths of waves the part, where $\frac{dn}{d\lambda} > 0$ can appear. On Fig.16.11 it is part BC. Such dependence n from λ is termed as **abnormal dispersion**.

The abnormal dispersion is observed on those parts of lengths of waves, where

there is light absorption (the absorption coefficient on Fig.16.11 is shown by the dot line) that corresponds to resonant requirement $\omega = \omega_0$, where ω is frequency of transmitted wave. ω_0 is natural frequency of oscillations of electrons of medium. The abnormal dispersion enables to find frequencies of eigentones of electrons in atoms and molecules and on this basis to judge their structure.

The phenomenon of dispersion in various optical systems plays positive and negative role. In lenses of cameras, microscopes dispersion of light produces chromatic aberration that worsens the image, as at chromatic aberration the luminous point emitting the white light looks like the iridescent stain.

16.4. Spectroscopic devices

Dispersion of light finds practical application in spectral prismatic devices.

The elementary scheme of the spectral device with a prism is figured on Fig.16.12.

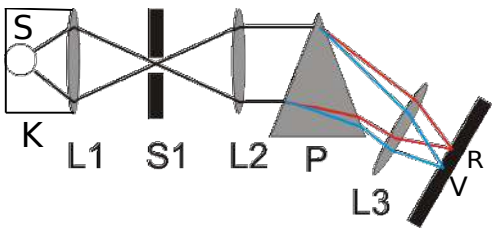

Fig. 16.12. The Early spectroscopic devices for observing solar.

Light from slit S illuminated by a light source falls on the lens L_1 and, transiting through it forms parallel bundle, since the slit is posed in the focal plane of the lens. The slit and lens L_1 are posed in a tube termed as *collimator* K. After refraction in prism P, cones of light of different lengths of waves are focalized by lens L_3 on the screen, where a series of monochromatic images of the slit is gained.

Depending on the way of recording of a spectrum, spectroscopic instrumentations are sectioned on the following types:

a) Spectrographs in which obtain *photos* of a spectrum;

b) Spectroscopes in which the spectrum is considered by human eye through eyepiece;

c) Monochromators intended for separation of radiation in a particular narrow part of spectrum; it has the second slit in the screen, on which the smoothly rotating the prism output various fields of spectrum. Variable width of exit slit

allows to separate the convenient interval of wavelengths $\Delta\lambda$ for examination. Monochromators usually are amounting part of the more composite devices.

d) Spectrophotometers intended for deriving and simultaneous photometric measurement of spectral lines (measuring of their relative intensity). With this purpose explored radiation with help of a monochromator sequentially is output on a photoelectric cell or the photomultiplier, conversing light signal into electrical.

One of the basic characteristics of the spectral device is its **resolving ability**. Resolving ability R of the prism as well as a diffraction grating, characterizes property of the device to divide the radiations distinguished on lengths of waves on quantity $\Delta\lambda$. The less this interval is more resolving ability of the device. It expresses through a dimensionless quantity, equal to $R = \lambda/\Delta\lambda$.

There are three types o f spectrums: continuous, striped (band) and line.

Incandescent solid and fluid bodies and gases at major pressure give *continuous spectrum* in which one colour gradually transfers to another. An example of continuous spectrum is the spectrum of white light. In it is conditionally accepted to discriminate seven primary colours.

Line spectrums will consist of the separate narrow lines of various colour parted by dark gaps. Such spectrums receive from atoms of luminous gases or the steams, which are taking place in the unloaded state. They appear as a result of electronic transitions inside atoms and ions of some elements. For observation of ruled spectrums sometimes use, that gases shine, when through them transits electric current. To receive a line spectrum of materials which in routine requirements are in solidity, it is possible to inject their grains into a torch flame. The study of line spectrums has shown that *each chemical element gives the line spectrum, which is not conterminous to spectrums o f other elements.*

Band spectrums are look like the separate light lines parted by dark gaps. Many of strips, by viewing through a spectroscope with major resolving ability, disintegrate on a series of separate lines. Band spectrums are characteristic *for molecules* of heated gases and steams and grow out changes of electronic, oscillatory and rotary energies of molecules.

All these three views of spectrums are **emission spectra**. Besides them there are absorption spectrums, which gain as follows. White light from a radiant pass through investigated material (gas, pairs, solution) and guide on a spectroscope. In this case on backgrownd of continuous spectrum will be visible the dark lines

posed in the particular order. Their number and the order allow to judge about composition of investigated material. Experiments confirm that *lines o f absorption always precisely correspond to lines of emission in the spectrum of gas or the steam, immersing light.* This dependence expresses **Kirchhoffs law**: *any materia absorb those beams, which itself can emit.* This law originating the dark lines apparent in a spectrum of sunlight speaks. They always borrow the same place and in basic represent lines of absorption of steams and the gases environmental the Sun, which temperature is much lower, than on its surface.

16.5. Spectrum analysis

On a line spectrum of steams of any material it is possible to establish, what chemical elements enter its composition. Such method of definition of chemical composition of material is termed **as qualitative spectrum analysis**. Precise location of lines in the radiation spectrum of each element is given in express tables of spectral lines. The spectrum analysis of gases can be carried out and on absorption spectrums.

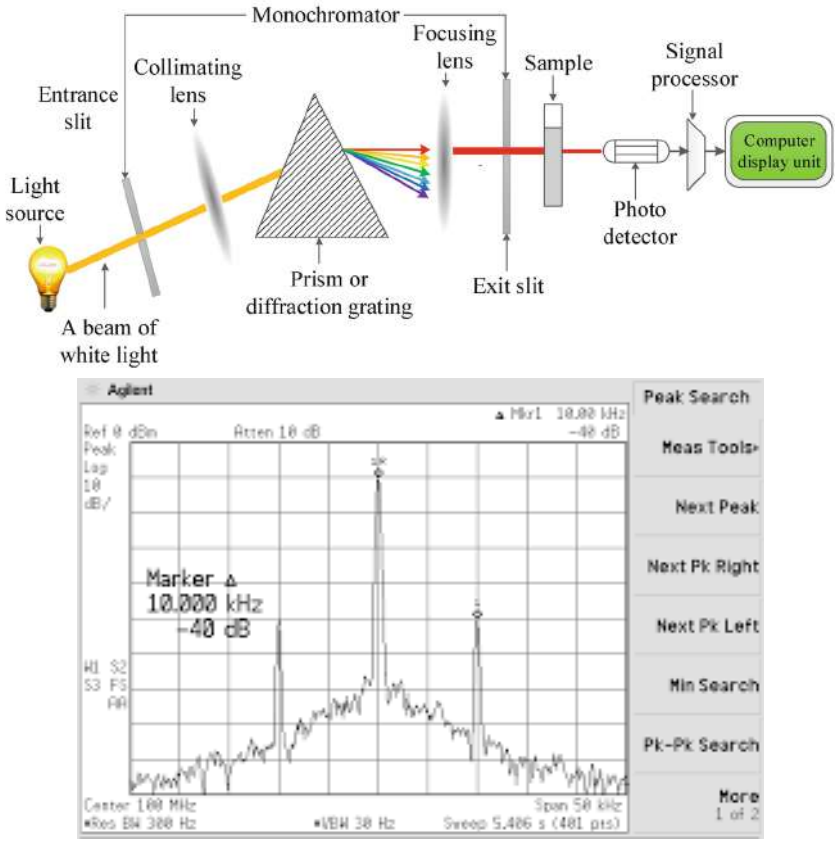

Fig. 16.13. The Spectrum Analysis.

The methods *of the quantitative spectrum analysis* permitting on *intensity* of

light emission of lines of chemical element to determine its percentage in assay designed. It allows to find presence of very much mass of element (**up to** $10^{-6} - 10^{-9}g$).

For definition of elements with low excitation energy ($2 - 4\ eV$) is used the photometry of flame. The abbreviated plan is on Fig.16.13.

Further light guide on a photoelectric cell, connected with a sensing galvanometer. Having a gang of replaceable light filters, it is possible to study various fields of spectrum and to establish, whether there is at assay any element. In a flame with major precision are defined all alkaline and alkali-earth elements, manganese, chromium, etc. Photometry of flame is applied to the analysis of connatural waters, biological fluids, chemical fertilizings and medical specimens.

Test questions

1. What is a plane-polarized wave (plane-polarized light)? What is a plane of polarization?

2. How is natural light different from plane polarized light?

3. What is called a polarizer and analyzer?

4. Formulate Malus' law.

5. Write down the expression for Brewster's Law and explain its meaning.

6. Name and explain the methods of obtaining polarized light.

7. Explain the phenomenon of birefringence. What ray should you call ordinary and what extraordinary?

8. Give a definition of anisotropy.

9. Explain how the Nicolas prism works and how it works.

10. Give examples of polaroids. What is their mechanism of action?

11. What is the phenomenon of rotation of the plane of polarization?

12. Explain which studies of biological tissues should be carried out in polarized light.

13. What is polarimetry and spectropolarimetry?

Lecture No 17

Optical system of eye.

17.1. Optical system of eye. Accommodation

Human eye is the original optical device borrowing in section of geometrical optics special place. It speaks that many optical tools are designed for visual perception of their indications. On the other hand human eye (and animals) as the biological system advanced during evolution, gives some ideas on designing and improvement of optical systems.

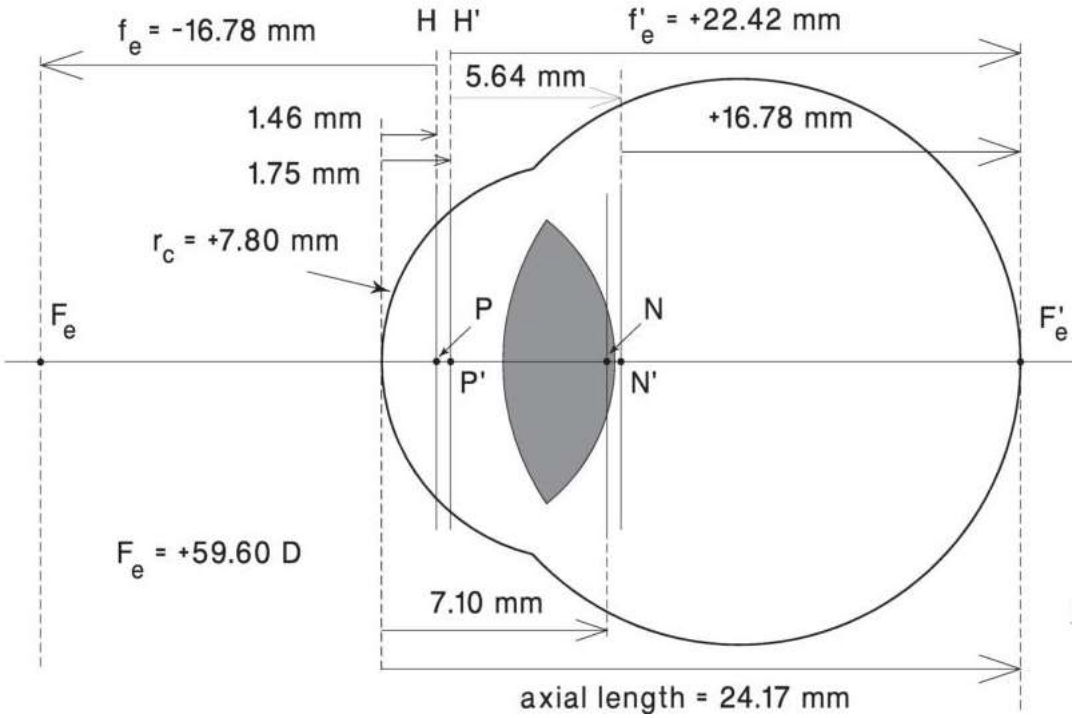

Fig. 17.1. The simplified schematic eye.

Let's consider the structure of an eye. The eyeball has almost spherical form of diameter in axial direction $24 - 25$ *mm* (fig. 17.1). Eye contains *photoconductive* and *photoperceplible* device. Walls of eye consist of three concentric located mediums: external, average and internal. External medium (**sclera**) in the forward part of eye (**cornea**) passes to the transparent convex cornea (fig. 17.2).

Separated from sclera, cornea has the form of a spherical cup of diameter about 12 mm; thickness of cornea is about 1 mm. Radius of it curvature is $7-8$ mm, the refractive index is 1.38. To sclera adjoins **vascular medium**, which internal surface is covered by a layer of the pigmentary cells, interfering internal dispersion of light in eye. To vascular medium in the back part, named bottom of the eye, adjoins **retina**, containing photoperceptible device of an eye. It consists of the smallest cells – rods and cones, providing twilight and color vision. In the forward part the vascular medium passes to iridescent, painted at various people differently and having in the center small round aperture that is known as *pupil*.

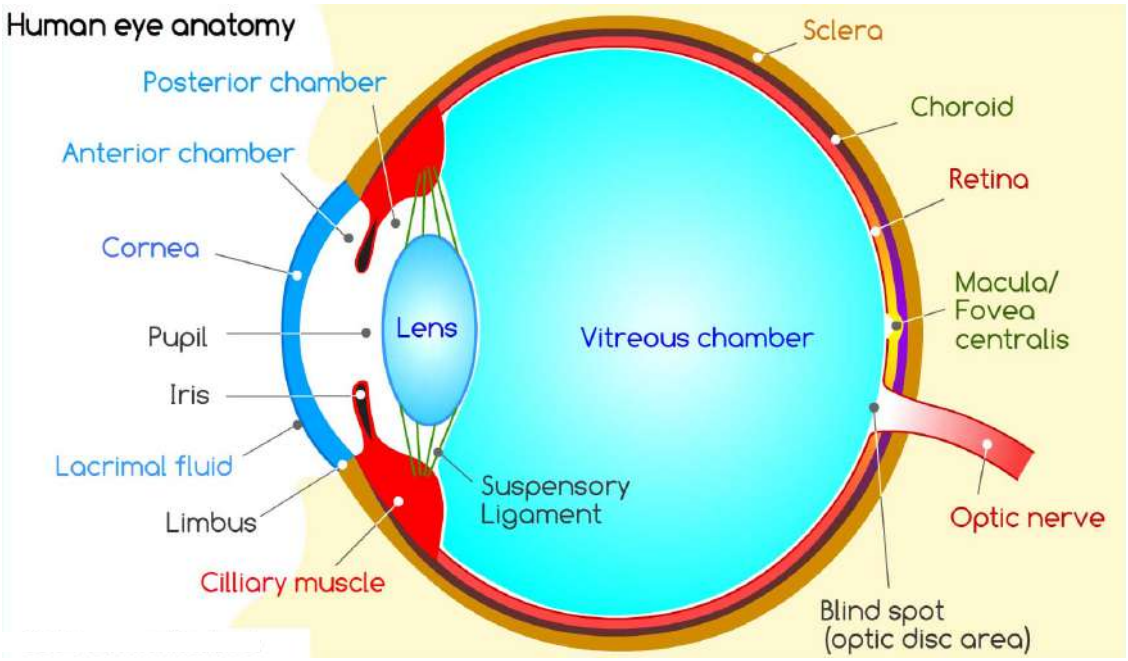

Fig. 17.2. Human eye anatomy

The *iris* of an eye is the original diaphragm regulating diameter of the pupil (from $2-3$ mm at bright up to $6-8$ mm at weak illumination) and by that the light stream getting to the eye. The space between the iris of eye and the cornea (**the forward chamber**) is filled by the transparent liquid close on optical properties to water. Directly behind the pupil is located the **crystalline lens**, it is the elastic transparent body having the form of the biconvex lens ($n = 1.4$). Diameter of crystalline lens is $8-10$ mm, radius of curvature of the forward surface is 10 mm and back is 6 mm. Curvature of crystalline lens can change with help of the *circular ciliary muscle*. All internal cavity of eye is filled by transparent jellylike liquid – the **vitreous body** ($n = 1.33$).

Retina serves as photoperceptible screen on which turns out the valid and reduced image of a subject considered by eye. *Refracting system of an eye: the*

cornea, the moisture of the forward chamber, the crystalline lens, the vitreous body represent the centre optical system with the optical axis, which is taking place through the geometrical centers of crystalline lens, pupil and cornea. For optical system of an eye, as well as for any optical system is possible to specify six cardinal points with which help determine direction of rays of light: it are two main points (H_1 and H_2), two central points (N_1 and N_2) and two focuses (F_1 and F_2) (fig. 17.3).

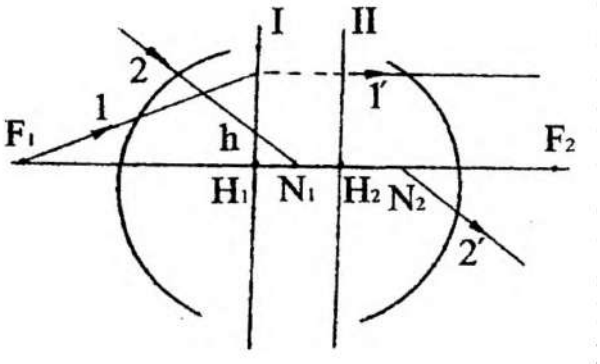

Fig. 17.3. The optical system of an eye.

Through two main points perpendicularly to optical axis pass two main planes (I and II). On identical distance from them there are focuses. At construction of images use the following rules. The beam going from a subject through focus and crossing the first main plane on height h from the optical axis leaves the plane II on the same distance h from axis and in parallel to it. The beam going from a subject to one central point N_1 leaves other central point N_2 in parallel the initial direction (see fig.17.3). Average position of cardinal points in the human eye is defined by results of research of set of people with normal sight. Both main and both central points in the average eye are located close from each other.

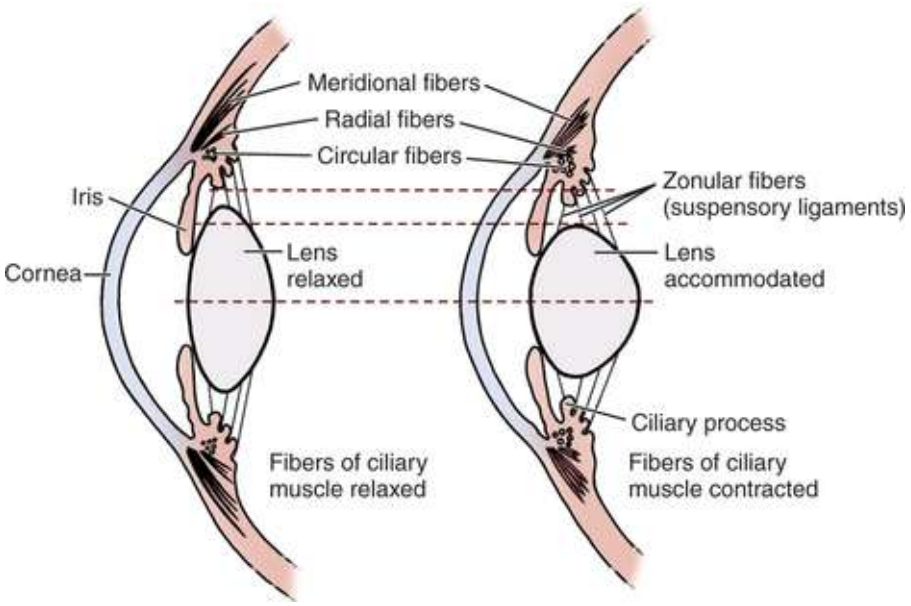

Fig. 17.4. The accommodation.

In an eye still distinguish the **visual axis**, which is taking place through the center of crystalline lens and the *yellow spot* (site on retina of the greatest light sensitivity) and determining direction on which the eye has the best sensitivity. The angle between and visual axes is 5^o. This angle has practical value. If it is great, there is impression of squint. It should be taken into account at determination of distance of centre to centre of spectacles lenses.

The healthy eye adapts to vision of the subjects located from it on distance of $10-15$ *cm* and up to infinity. This ability of an eye is known as *accommodation.*

Accommodation *is the ability of eye to change optical force due to change of curvature (shape) of crystalline lens, that allows to receive precise images of subjects on retina.* When the eye is accommodated for distant vision, both the circular and meridional fibres of the ciliary muscle are relaxed, thus stretching the zonula which squeezes the elastic lens to a flattened shape (fig. 17.4). In order to increase the refractive power of the lens when viewing a close object, both sets of fibres of the siliary muscle contract. Each has the effect of releasing the tension in the zonula, which then allows the lens to bulge. Thus there is increase of optical force (D): $D = (n-1)\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$ where R_1 and R_2 are radiuses of curvature of crystalline lens. Proceeding from the formula of lens $D = \frac{1}{d} + \frac{1}{f}$ it allows at increase of D to reduce the distance up to object (d), not changing distance up to the image (f).

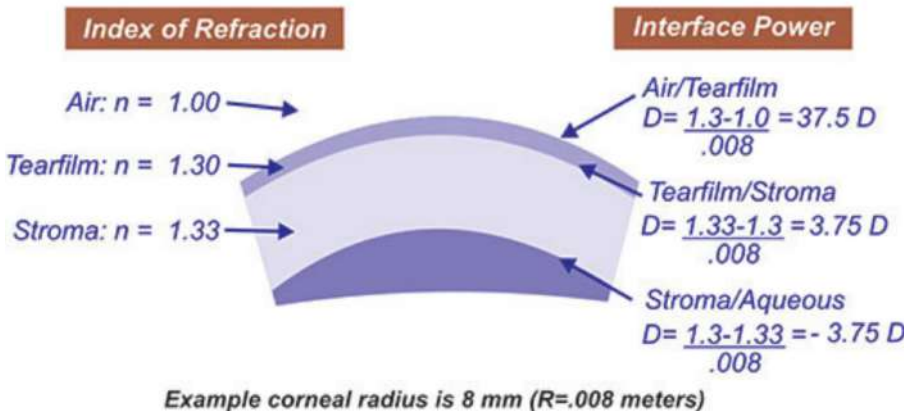

Fig. 17.5. The refractive interfaces of the Cornea.

As whole the optical system of an eye operates as the *converging lens with variable focal length.* All system of eye in not intense condition (rest of accommodation) has *optical force (refractive power) about 65 diopters. The basic refraction of light occurs on external surface of cornea on border with air (fig. 17.5). Cornea has optical force about 43 diopters. Optical force o f crystalline lens is*

$\approx$ 20 *diopters* (at vision of the removed subjects). At vision of close subjects curvature of crystalline lens increases and optical force of eye can reach of $70-75$ diopters (*limit of accommodation*).

At the adult person (healthy) at approach of subject to eye up to distance of 25 *cm* accommodation is made without a pressure and due to a habit to descry the subjects which are taking place in hands, the eye more often accommodates on this distance named as **distance of the best vision** (or **least distance of distinct vision**). The minimal distance up to a subject, corresponding to the maximal accommodation, defines position of a so-called **near point of clear vision** (near point of the eye). Position of near point depends on age of a person. With the years this distance increases and accommodation decreases.

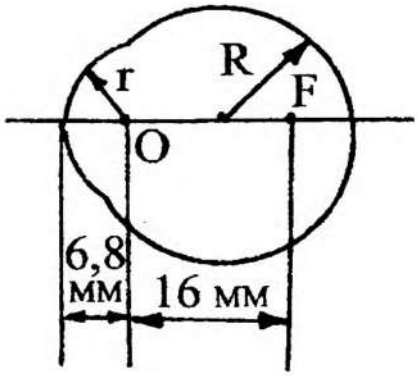

Fig. 17.6. The eye on Verbitsky.

For construction of the image of subjects on retina of an eye and the analysis of the phenomena connected to it is used the *reduced or the resulted eye*, which is considered as *homogeneous spherical lens*. Some circuits of the resulted eye rather close among themselves are offered. On fig.17.6 (the eye on Verbitsky) such circuit is resulted: radius r of the forward refracting surface is 6.8 *mm*, radius R of sphere is 10.2 *mm*, length on the axis is 23.4 *mm*; $n = 1.4$. The optical center of lens is on the distance of 6.8 *mm* from top of the forward refracting surface, and main focus F on distance of 16 *mm* from the optical center, i.e. inside the lens.

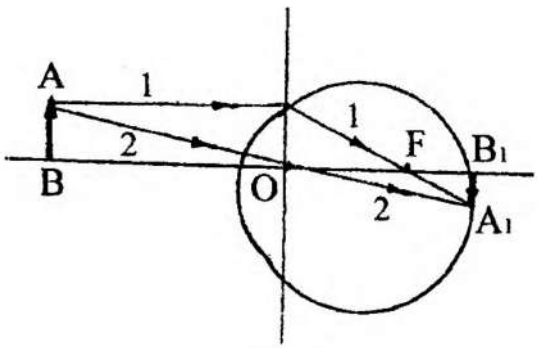

Fig. 17.7. The construction of the image of subjects in the eye.

In the resulted eye on Verbitsky the main and central points for simplification of constructions, as well as for a thin lens are combined and are in the point (the center of the lens) (fig. 17.6).

Construction of the image of subjects in the resulted eye is done by rules for a single thin lens. The subject usually settles down behind the double focal length, and the image A_1B_1 results on the back surface of the reduced eye by *valid, inverted and reduced* (fig. 17.7).

17.2. Resolution of eye. Defects of optical system of eye

The size of the image on the retina depends not only from the size of a subject, but also from distance to the eye, i.e. from the angle under which the subject is seen (fig. 17.8). For the characteristic of size of the image on the retina enter concept of the *visual angle.*

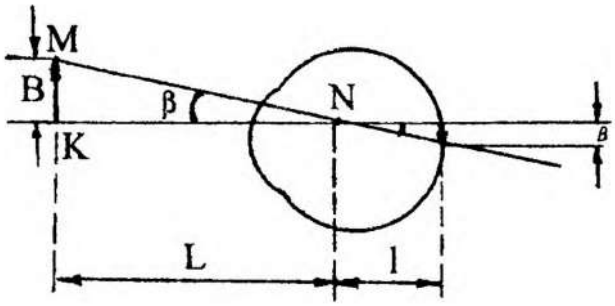

Fig. 17.8. The size of the image on the retina.

It is the angle β between the beams going from extreme points of the subject through conterminous central points (fig. 17.8). The size of the image on retina is $в = l \cdot \beta$, where l is distance between the uniform central point and the retina ($l = 17\ mm$). The formula is fair, if the visual angle is small. From the resulted figure is easy to establish connection of size of the subject with the distance L of subject from the central point of the eye and visual angle of sight: $B = L\beta$, then $в = \dfrac{lB}{L}$.

Resolution of eye can be estimated by the minimal visual angle β_{min} under which two next points of a subject are visible separately. This angle finally defines the size of the image on the retina. Within the limits of the yellow spot at good illumination, eye of the person starts to perceive two points if $\beta \geq 1'$. *The value $\beta_{min} = 1'$ characterizes resolved ability of an eye (the maximal visual acuity) and is defined by structure of the retina.* Two next points are visible separately, if their images get on *different* receptors. This condition also defines the limit of resolution of eye. At normal vision the person can see separately from distance of 25 *cm* two points, taking place from each other on distance of 70 microns. The size of the image on the retina thus is equal to 5 microns that corresponds to average distance between cones. Therefore, if the image of two points on the retina will borrow smaller distance of 5 microns, these points are not resolved, i.e. eye of them does not distinguish separately. In medicine resolution of an eye estimate **visual acuity $V = 1/\beta_{min}$. For norm of visual acuity is accepted l and in this case the least visual angle is equal to** $1'$. At infringements visual acuity in so much time is less than norm, in how many times the least angle of sight at deviation from norm more than one minute. If for the patient the minimal angle of sight is equal $4'$, visual acuity is equal $1 : 4 = 0.25$. Visual acuity is the basic function of eye by which is guided at selection of glasses.

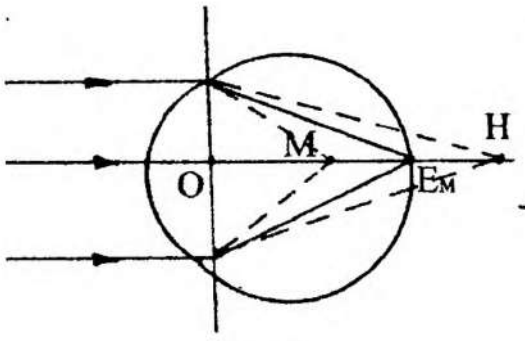

Fig. 17.9. The clinical refraction of eye.

Main optical characteristic of eye is represented with position of back focus concerning of retina. It is known as clinical refraction of eye (fig.17.9). If the point of focus lays on the retina, the refraction is called **emmetropic** (Em), if behind the retina the eye is **hypermetropic** (H), if before the retina the eye is **myopic** (M). Only the first refraction provides (at rest of accommodation) the precise image of far subjects on the retina and, hence, normal vision.

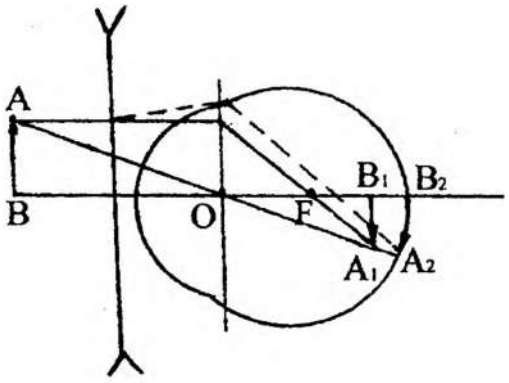

Fig. 17.10. The short-sightedness.

The most widespread defect of vision is **short-sightedness (myopia)**, connected with 1) *the lengthened form of the eyeball* and less often with 2) *excessive optical force of refracting mediums* of the eye at his normal form. Thus sharp image A_1B_1 of far subjects is formed in the plane laying a little bit ahead of retina (fig. 17.7). Such eye does not see distinctly far subjects. Accommodation thus is useless, as it increase excessive for the given form of eyeball optical force of an eye even more.

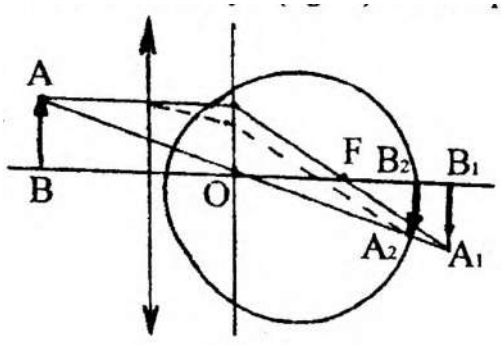

Fig. 17.11. The far-sightedness.

For correction of short-sightedness is necessary to reduce optical force of the eye by application of glasses with **disseminating** (negative) lenses. Thus image A_2B_2 of the distant subjects turns out on the retina (fig. 17.10). Continuous lines the path of beams without the lens is shown, by a combination of shaped and continuous lines are shown with the lens.

The **far-sightedness (hypermetropia)** is connected with insufficient refracting ability of the eye or with the short form of the eyeball. Image A_1B_1 of the far subjects turns out behind of retina (fig. 17.11). With help of accommodation of eye in part eliminates this defect, however limits of accommodation are

limited and such eye does not see distinctly close located subjects. Glasses are applied to elimination of far-sightedness with **collecting** (positive) lenses, which strengthen refracting ability of the eye and provide sharp image A_2B_2 on the retina (fig. 17.11).

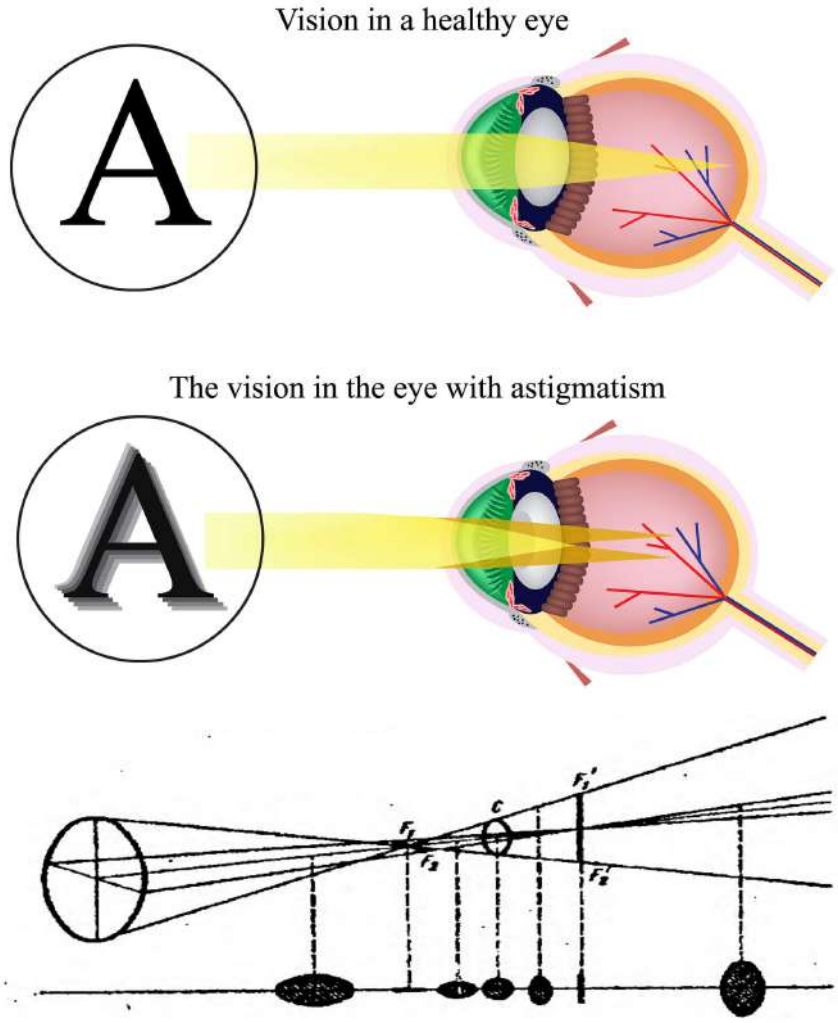

Fig. 17.12. Astigmatism..

With the years there is change of elastic properties of a crystalline lens and there is the *age far-sightedness*.

More seldom there is defect of vision **astigmatism**, which is connected with non-uniform refraction in various meridian planes of the eye (fig. 17.12). This phenomenon is caused by defect of the correct spherical form of external surface of the cornea and the crystalline lens. If on such eye falls the bandle of parallel rays, instead of one focal point rays are agglomerated in two segments.

Segments lays in the focal planes of principal cuts. Measure of astigmatism (so-called **astigmatic difference**) is difference of refractive force of two principal cuts (in diopters). The more astigmatic difference is more distance between horizontal

- Plano cylinders
- Cross cylinder
- Sphero – cylinders
- Toric lenses

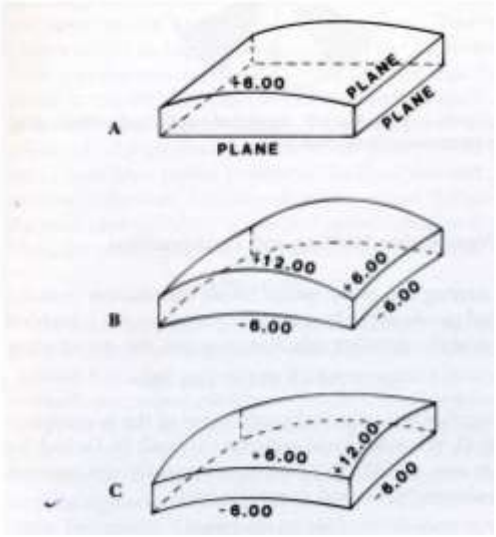

Fig. 17.13. Cylindrical and toric lens forms. A. plus plano-cylinder front surface and flat back surface; B. plus toric front surface and concsvespherical back surface with net power in the vertical meriddian of zero; C. plus toric surface and spherical back surface with net power in the horizontal meridian of zero.

and vertical segments. Significant astigmatism causes distortion of the form of a subject on the retina, stretching them at length or width.

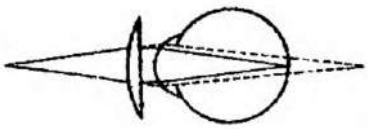

Fig. 17.14. The age weakening of accommodation.

Astigmatism demands correction of vision also, though it is dependent kind of clinical refraction, and accompanies to emmetropia, hypermetropia or myopia more often. If to look on an astigmatic eye in front and mentally to dissect his by planes, which are taking place through the pole of cornea, appears that the refractive index of this eye smoothly changes from the biggest value in one of sections up to the smallest in the other section, which will be perpendicular to the first.

If in each section the refraction remains the constant it is the *correct astigmatism*. If for various beams within the limits of the certain section the refraction appears various, this kind is named wrong astigmatism. Glasses can correct only correct astigmatism. Correction of astigmatism is carried out by cylindrical and spherocylindrical lenses (fig. 17.13).

To defects of sight also concerns presbyopy or age weakening of accommodation (because of loss of elasticity by crystalline lens). For its correction are used positive lenses (fig. 17.14).

Special case is the **aphakia**; it is condition after extraction of the grown turbid crystalline lens (*cataract*). Thus there is the far-sightedness of very high degree

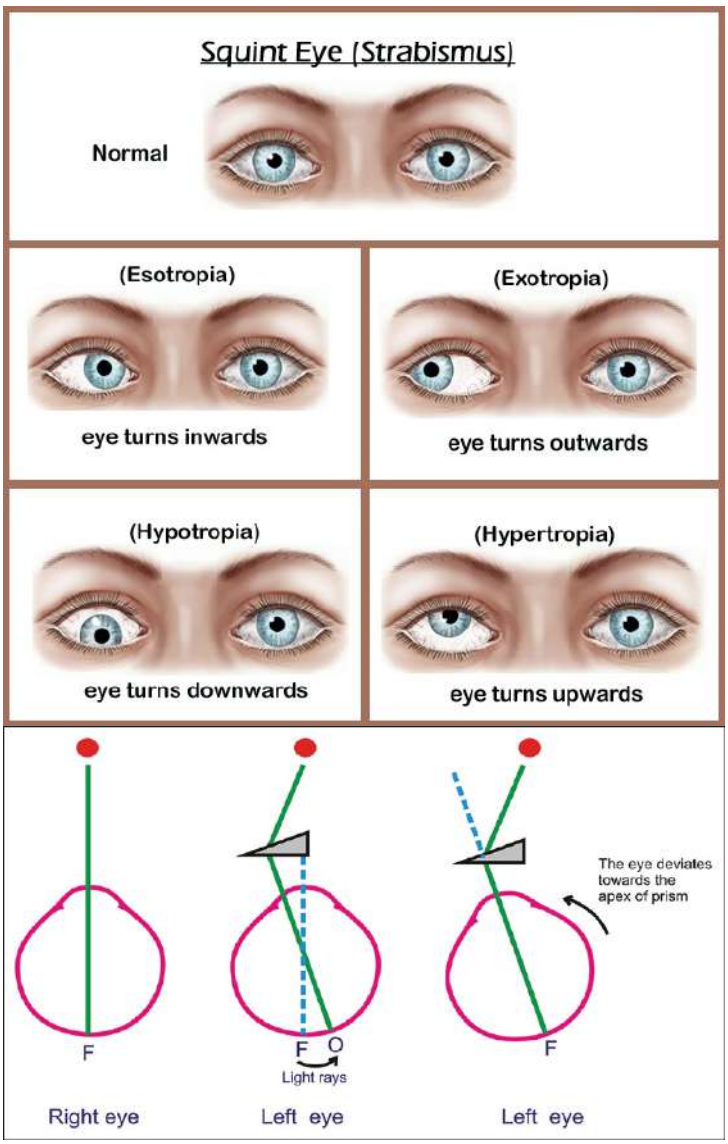

Fig. 17.15. Squint eye.

demanding correction by strong positive lenses (of $8 - 13$ diopters).

Both eyes work for the person as the coordinated system forming a uniform image of the seen subject. Ability to form such image from images of two eyes is known as **binocular sight**. Simultaneously, binocular sight allows to estimate remoteness of mediumal subjects. This ability of an eye is called three-dimensional sight. The place in space where visual lines of two eyes are crossed is known as a point of fixing.

Infringement of binocular sight is shown as the **squint** more often. The squint is the deviation of the visual line of one of eyes from the joint point of fixing. On the direction of deviation of visual line of the eye distinguish the squint converging, missing and vertical. For correction of the strabismus use prismatic lenses (fig. 17.15).

17.3 Bases of photometry

Discussion of questions of biophysical processes of visual perception demands of knowledge of the basic concepts concerning measurement of light values. This circle of questions studies *photometry* (photo = light, metry = measurement).

One of photometric quantities is the **light flux (illuminating power)**(Φ), which represents energy of (dW) the light wave, taking place per unit of time through the given surface: $\Phi = \dfrac{dW}{dt}$. The light flux is measured in **lumens (lm)**.

Light exposure (intensity of illumination) is a ratio of the light flux to the area of this surface: $E = \dfrac{d\Phi}{dS}$. Light exposure is measured in **lux**.

Physiological action of light on the person substantially depends on light exposure. At small light exposure the eye hardly distinguishes fine subjects and quickly gets tired. At big light exposure light renders harmful action on the retina and excite of nervous system. Therefore hygienic norms of light exposure of inhabited and industrial rooms are established. For example, in an auditorium (at the level of surface of a table) light exposure should be **150 lux**, in rooms in a hostel is **50 lux**, etc.

Light intensity of a source (I) is the ratio of the light flux to value of the solid angle inside which this flux is propagated: $I = \dfrac{d\Phi}{d\Omega}$, where Ω is the solid angle, measured in **steradians**. Light intensity is measured by **candela (cd)**. It is basic unit in system of SI.

Brightness (L) is the value numerically equal to the ratio of the light flux inside the solid angle to value of this angle and to the area of the radiating platform: $L = \dfrac{d\Phi}{d\Omega dS\cos\alpha}$, where α is the angle between the normal to the surface and axis of the solid angle. As $\dfrac{d\Phi}{d\Omega} = I$, then $L = \dfrac{I}{dS\cos\alpha}$.

Unit of measurements of brightness is 1 $Nit = 1\ cd/m^2$. Brightness of a sheet of the white paper at reading should be not less than 10 cd/m^2.

17.4 Sensitivity of eye for light and color. Adaptation.

Light getting to the eye is focused with the help of crystalline lens on the layer of photosensitive cells of retina (fig. 17.16). Photosensitive elements of the retina **cones** are located in the yellow spot. **Rods** settle down on edges of yellow spot and on other surface of retina. The number of cones is equal approximately to 7

million and number rods are approximately 130 million.

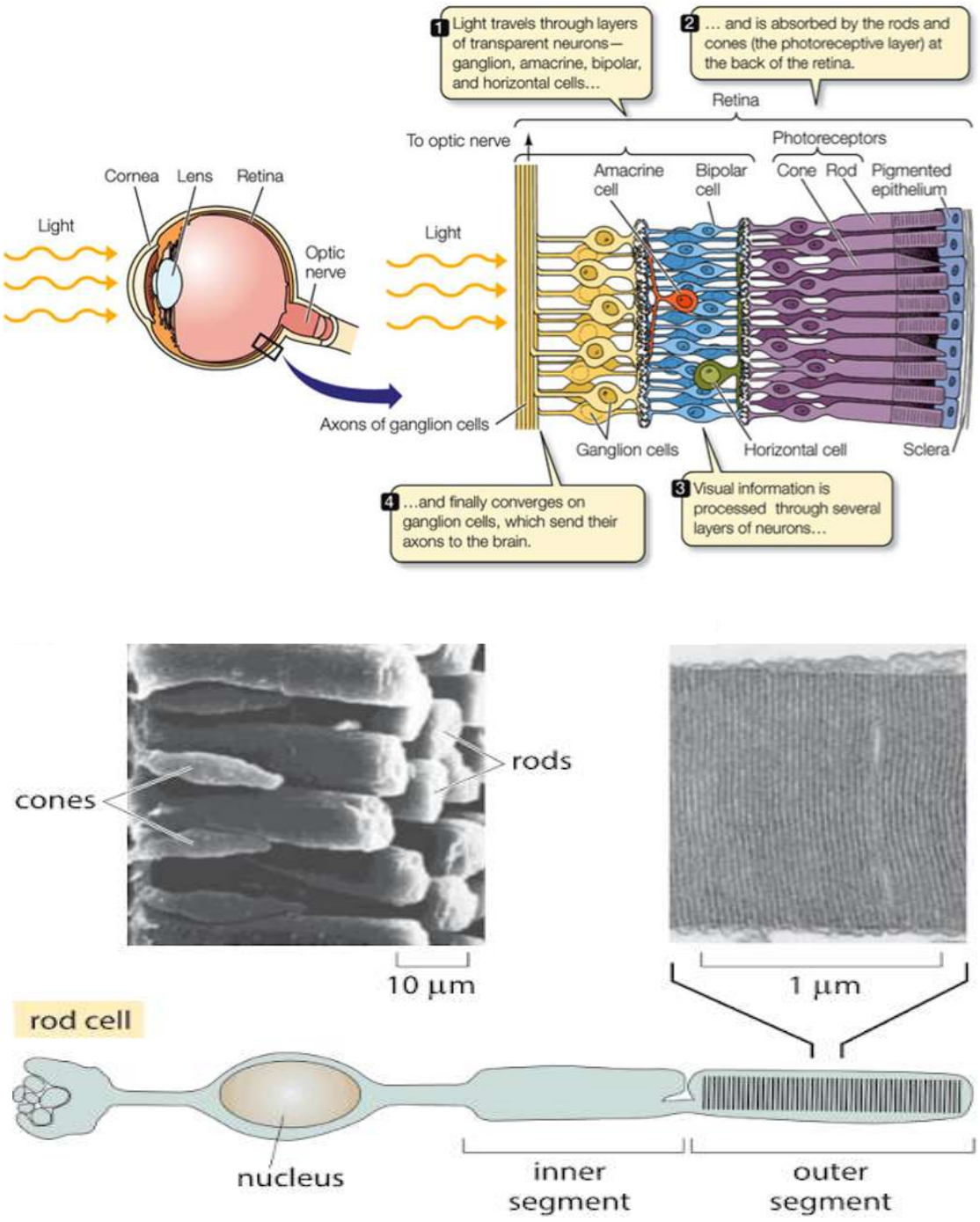

Fig. 17.16. Photosensitive cells of retina.

Rods do not distinguish color and responsible for black-white vision or twilight vision. Only the light irritation of cones causes sensation of color and due to presence of cones is carried out day time sight. Cones are concentrated in the center of the retina because color vision is carried out in conditions of bright illumination at the narrowed pupil, passing light basically on the central part of the retina.

Sensitivity of rods is much higher, than cones that function only at light exposure more than 10^{-2} lux, whereas rods react to light even at light exposure up to 10^{-6} lux. It is easy to be convinced in it on twilight, when it seems to us, that all subjects lose the colouring.

Perception of light as well as the perception of sound, submits to law of **Weber-Fechner** according to which the change of force of light sensation is proportional to the logarithm of the ratio of intensity of two compared light streams.

The photosensitivity of the eye changes over wide range due to visual adaptation. It is known, that at input in poorly covered room, in the beginning the person does not distinguish subjects and for their distinction needs certain time, i.e. transition from day time vision to twilight vision demands certain time. This process is known as adaptation. **Adaptation** is the ability of eye to adapt to various brightnesses. Before full adaptation time of 30-40 minutes sometimes is necessary.

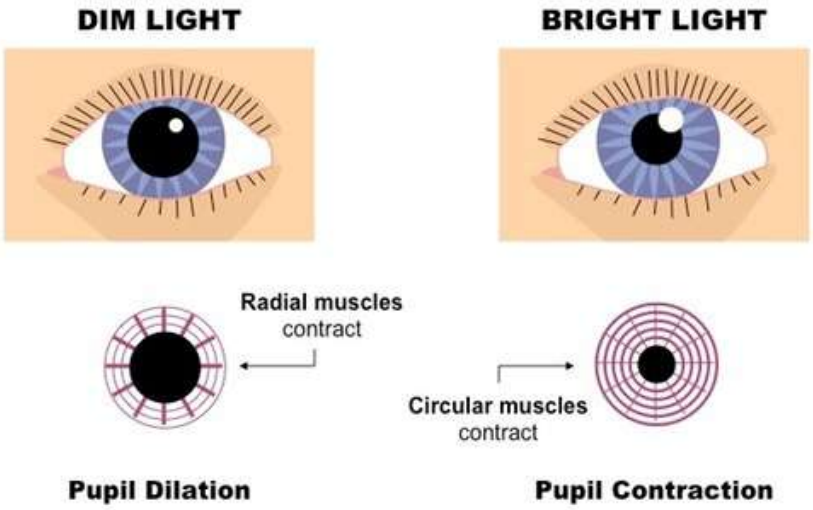

Fig. 17.17. Adaptation.

Adaptation allows to the eye to function normally in the range of brightness $10^{-7} - 10^{5}$ cd/m^2.

Researches have shown that minimum quantity of light, which should fall on the surface of an eye for creation of light sensation, makes from 60 up to 150 photons of yellow - green light To retina reaches even less photons. About 50% is absorbed by crystalline lens, about 4% is reflected from the cornea. Thus, on share of photoreceptors there are some photons from number of the photons, falling on the cornea. Recent researches have shown that threshold of sensitivity of the eye adapted to darkness, for yellow - green light makes only **2-3 photons**.

Photoreceptors will transform light energy into electric with coefficient of amplification of $10^5 - 10^6$. Such big amplification allows even to individual photons to create the nervous impulse and accordingly light sensation.

Thus, the eye is one of the most sensitive devices.

Human eye reacts to electromagnetic waves with length of wave approximately from 400 up to 780 nanometers. And even in the specified interval sensitivity of eye for different lengths of waves is difference. Sensitivity of the eye for longer and shorter waves is sharply reduced. The greatest sensitivity human eye has to wavelength of $\lambda_{max} = 555$ nanometer, i.e. to green color. If to take some sources of the different colors of identical capacity, they will be submitted to the eye not equally brightly. For example, that red light seem so bright, as well as green, is necessary, that its capacity exceeded capacity green in some times. Therefore, for the characteristic of spectral sensitivity of an eye enter value, which is equal to the ratio of capacity of radiation with length of wave $\lambda_{max} = 555$ nanometer to capacity of radiation with wavelength of λ, causing brightness of the same sensation, as well as radiation of wavelength λ_{max}. This value is called *relative spectral light efficiency* V_λ (sometimes is used the old name: relative luminosity): $V_\lambda = \dfrac{P_{\lambda=555\ nm}}{P_{\lambda\ nm}}$. Value of function V_λ for various wavelengths have been determined by averaging results of numerous measurements. For green color of $\lambda_{max} = 555\ nm$ value $V_\lambda = 1$.

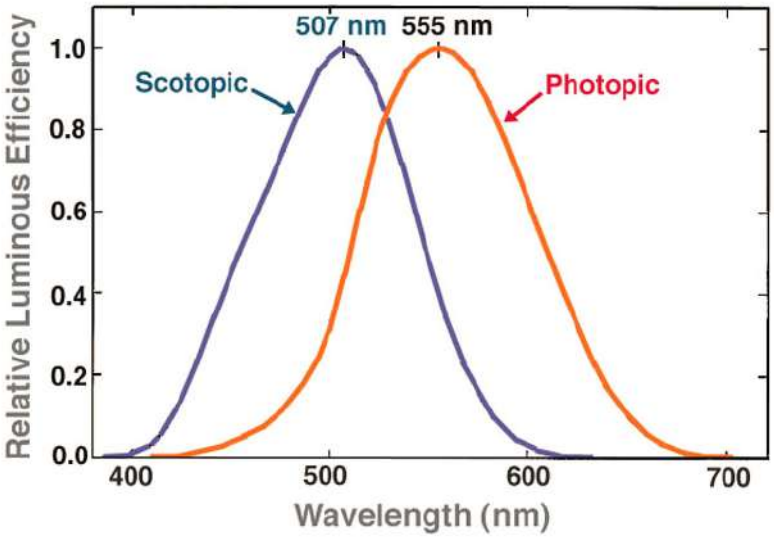

Fig. 17.18. Relative spectral light efficiency.

The graph of dependence of V_λ from wavelength λ is called *spectral sensitivity* of the eye.

Curves (fig. 17.18) of spectral sensitivity in conditions of day time (photopic) vision are received and compared with curve of sensitivity in conditions of twilight vision (scotopic). At day time vision the maximum of sensitivity corresponds to wavelength of $\lambda_{max} = 555\ nm$ at twilight to wavelength of 510 nanometers, i.e. the curve (photopic) is displaced concerning the curve (photopic) of day time vision to the direction of short wavelengths and the maximum is in the field of dark blue color. Therefore, at consideration of a subject at first at strong illumination, and then at weak illumination, notice displacement of colouring of the subject in the dark blue part of spectrum.

It is not necessary to think, however, that the eye is not sensitive to the radiations laying outside of range of $400 - 780$ nanometers. The person can perceive radiation in ultra-violet area up to 300 nanometers and in infra-red area up to 950 nanometers, but sensitivity of the eye to these waves in billion times is less, than for $\lambda_{max} = 555$ nanometers.

Crystalline lens and vitreous body almost completely absorb ultraviolet. Therefore at ablation of a crystalline lens (concerning of cataract), sensitivity of the eye to the ultraviolet considerably grows.

The maximum of curve luminosity of day time vision corresponds to the maximum of the sunlight, which past through atmosphere and has got on the surface of Earth. In it the expediency of the organization of eye as photosensitive device is shown.

17.5 Biophysical bases of visual reception.

Light, getting to the eye is fixed by optical system of the eye on the retina, which represents multilayered cellular system (fig. 17.16). Photoreceptor cells are in the back layer of the retina, basing by the photosensitive segments on the layer absorbing photons of epithelial cells painted by dark pigment. To get in photoreceptors for light is necessary to pass preliminary through the layer of nervous cells that however does not reduce sensitivity of the eye, as these cells are transparent for seen light.

Such position ofphotoreceptors protects these cells from external influences better and prevents hit on them of the photons, reflected and absent-minded by other sites of the eye; it improves visual acuity. Each rod and cone will consist of the external and internal segment containing the nucleus and mitochondrias, providing poweT processes (fig. 17.19) cross section of eye; 2) cone; 3) rod; 4)

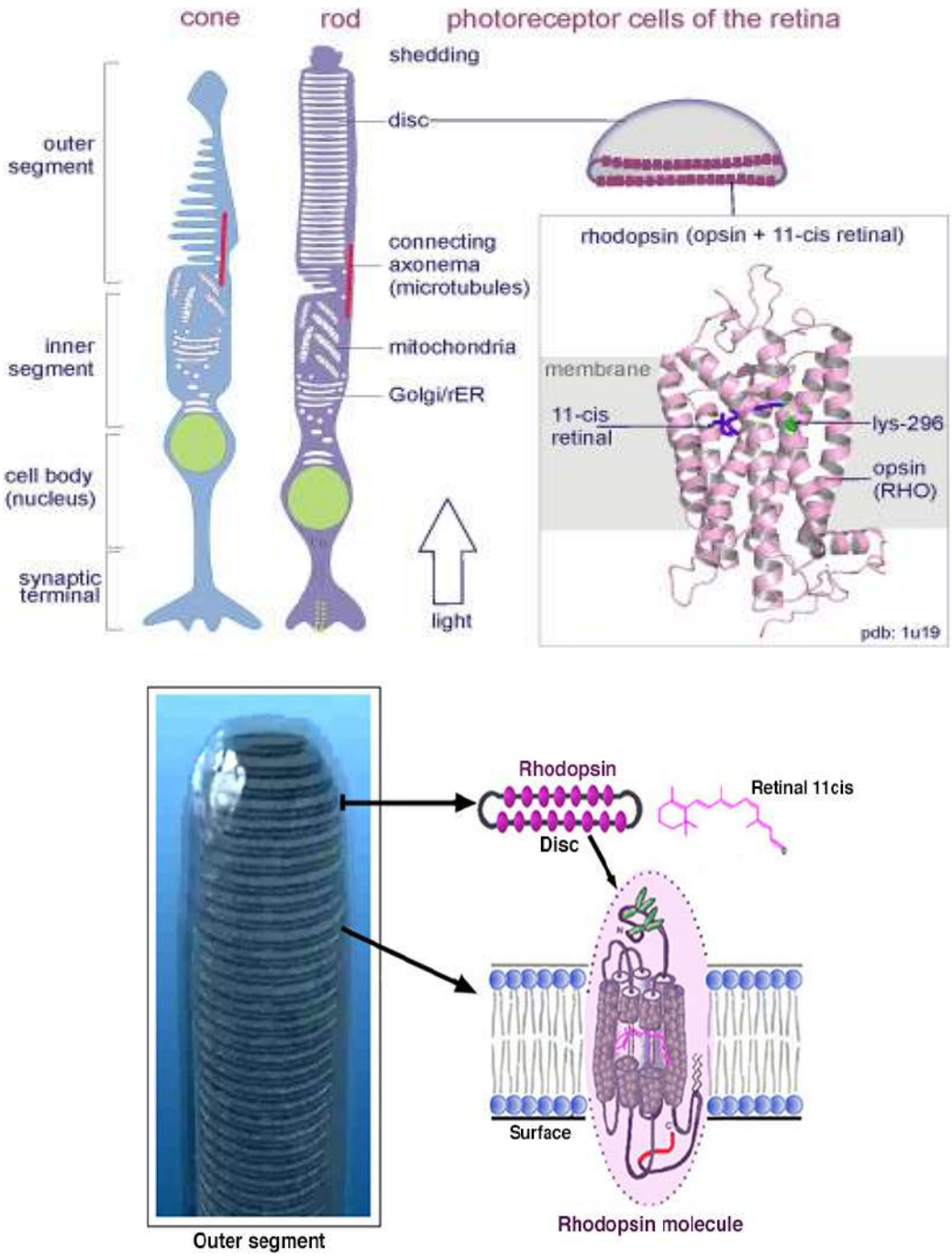

Fig. 17.19. Rod and cone.

disk of the outer segment of the rod; 5) a fragment of the membrane of the disk with the molecula of rhodopsin built-in it; 6) retinal in two states; the M is accumulation of mitochondrions. At the end of the internal segment inverted to light, there is the synoptic contact with the nervous fibre.

Let's consider the structure and functions of rods. The external segment of rods consists from the pile of photosensitive disks, in which is built the visual pigment **rhodopsin** (protein) of red color (fig. 17.19).

Each disk by thickness about of 20 nanometers, the flattened out balloon

reminding by self, will consist from bilayers lipids membranes with molecules of proteins penetrating it. A plenty of disks in the pile increases the general photosensitive surface sensory of the cell that raises probability of absorption by cell of a photon.

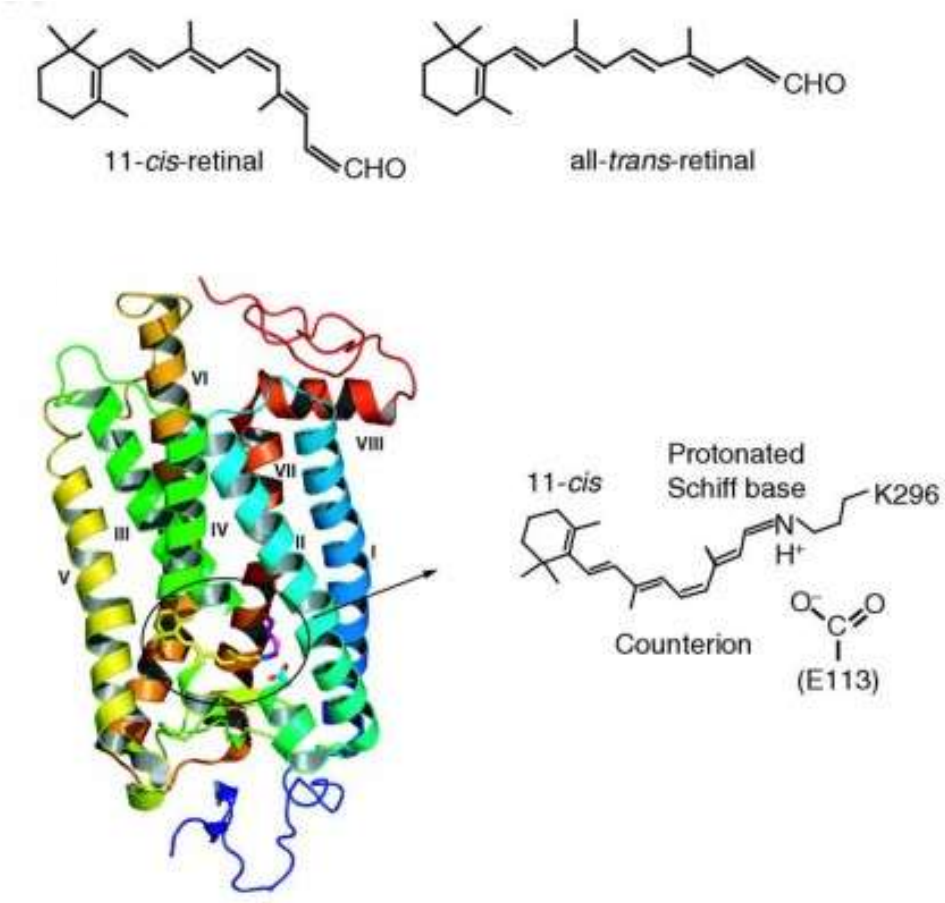

Fig. 17.20. Retinal can be in two isomer configurations: cis- or trance- configuration.

Visual disks are formed during all human life. They gradually move along the segment and on the end are separated, then are absorbed by cells of pigmentary epithelium and then disks are blasted.

The visual pigment **rhodopsin** is the complex protein. It will consist of protein opsin and chromophor group retinal (aldegid group of vitamin A). If with food in an organism acts insufficient quantity of vitamin A, process of synthesis of visual pigments is broken, that results to deterioration of the twilight vision named *"night blindness"*.

Retinal can be in two isomer configurations: cis- or trance- configuration (fig. 17.20). The molecule of trance - retinal has the straightened form. Cis-retinal has the bent form, turn of group of atoms begins from the eleventh atom of carbon therefore isomer is called 11-cis-retinal. The bent molecule of 11-cis-retinal in

darkness forms the complex with opsin and densely enters into the corresponding deepening in the molecule of opsin. At illumination cis-retinal passes to the steadier transform and the straightened molecule of the trance - retinal is not located in the deepening, leaves it and chip off from opsin. **Disintegration of rhodopsin on retinal and opsin results to excitation of receptor cell and to occurrence of generating potential**. Disconnection between opsin and retinal results to decolouration of rhodopsin. Return process of transformation the trance - retinal into cis-retinal occurs under action of enzyme retinalisomerase, then cis-retinal joins to opsin. In the retina at constant illumination takes place stable equilibrium at which rate of decay of rhodopsin is equal to speed of its restoration takes. In darkness speed of regeneration of rhodopsin reaches of maximum and the eye gets the maximal sensitivity.

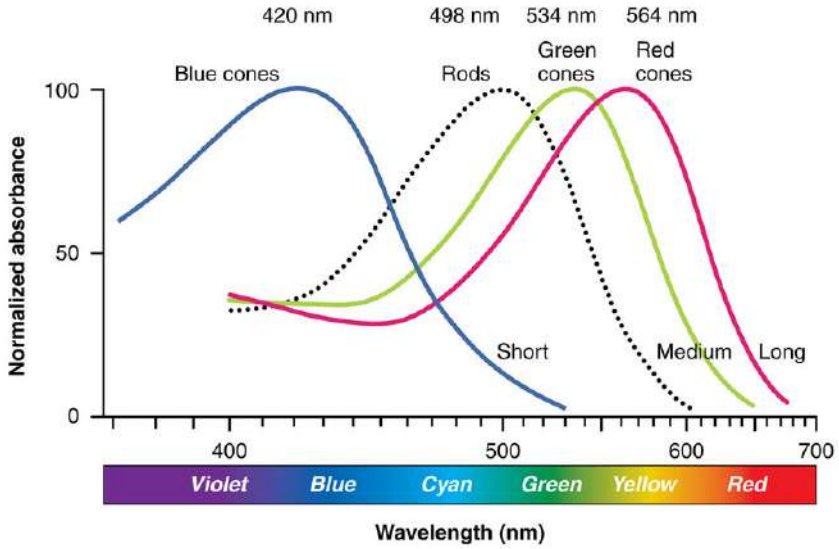

Fig. 17.21. Three types of cones.

Such reorganizations of rhodopsin for the first time have been investigated by Wolt, received for it the Nobel Prize in 1966.

Registration of electroretinogramme has allowed to establish that right after illumination of a rod by short flash of light is observed early receptor potential, then approximately through 1ms develops late receptor potential. The nature of these potentials is completely various and is still insufficiently investigated.

These pulses act to the axon of the optic nerve and are transferred to the central nervous system, where the sensation of light is formed.

Color sight is caused by cones. According to theory of Young-Gelmgolts, *there are three types of cones with various curves of spectral sensitivity with maxima of* 440, 540 *and* 590 *nanometers* (fig. 17.21). *Each kind o f cones creates sensation*

only one color: red, green or dark blue. At simultaneous excitation of receptors, to the brain enter signals of different intensity that create sensation of intermediate colors. All variety of color sensations is defined by the parity between number of the pulses sent by excited cones.

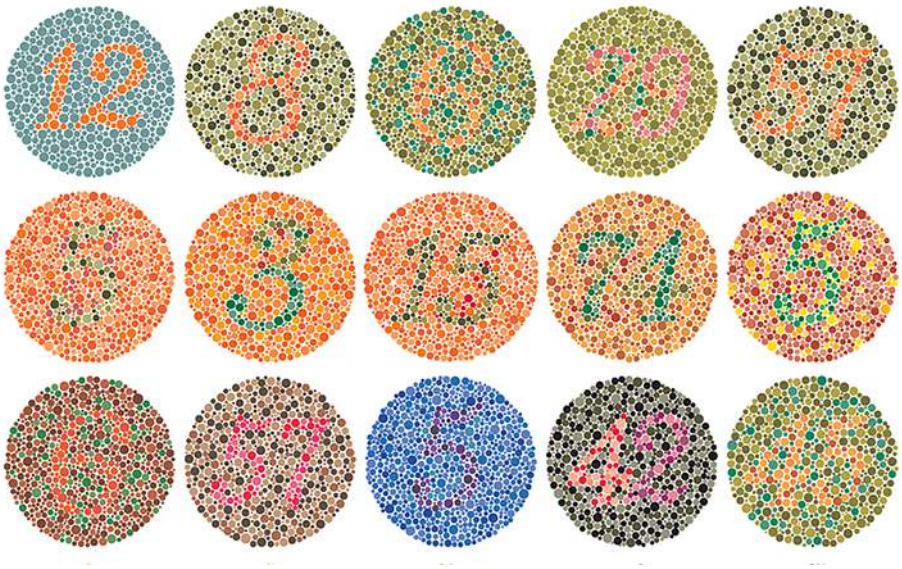

Fig. 17.22. The check table ability to distinguish colors.

The pigment of cones contains also 11-cis-retinal, as well as rhodopsin, but the albuminous part of pigment differs, therefore pigments of cones is called **iodineopsin**, they have violet colouring. Everyone cone contains *only one kind of iodineopsin.* Absorption of light by iodineopsin, as well as in the case with rhodopsin, results to occurrence of potentials in cones.

Three componental theory of color vision explains the majority of the facts of physiology and pathology of color vision. At some genetic diseases synthesis of proteins - iodineopsins is broken therefore some pigment of color vision is not formed. The person loses ability to distinguish colors (fig. 17.22). This illness is known as daltonism.

Test questions

1. What are the main elements of the eye and their function?
2. What is the mechanism of vision?
3. What is accommodation? How is it carried out?
4. What are the phenomena of myopia and hyperopia?
5. What is special about binocular vision?
6. What is relative spectral luminous efficiency?

Lecture No 18

Optical and electronic microscopy. Fiber optics.

18.1. Lenses. Optical microscope. Path of rays. Magnification.

A lens is a transmissive optical device that focuses or disperses a light beam by means of refraction. A simple lens consists of a single piece of transparent material, while a compound lens consists of several simple lenses (elements), usually arranged along a common axis. Lenses are made from materials such as glass or plastic, and are ground and polished or molded to a desired shape. A lens can focus light to form an image, unlike a prism, which refracts light without focusing. Devices that similarly focus or disperse waves and radiation other than visible light are also called lenses, such as microwave lenses, electron lenses, acoustic lenses, or explosive lenses.

Types of simple lenses. Lenses are classified by the curvature of the two optical surfaces. A lens is biconvex (or double convex, or just convex) if both surfaces are convex. If both surfaces have the same radius of curvature, the lens is equiconvex. A lens with two concave surfaces is biconcave (or just concave). If one of the surfaces is flat, the lens is plano-convex or plano-concave depending on the curvature of the other surface. A lens with one convex and one concave side is convex-concave or meniscus. It is this type of lens that is most commonly used in corrective lenses.

If the lens is biconvex or plano-convex, a collimated beam of light passing through the lens converges to a spot (a focus) behind the lens. In this case, the lens is called a positive or converging lens. For a thin lens in air, the distance from the lens to the spot is the focal length of the lens, which is commonly represented by f in diagrams and equations. An extended hemispherical lens is a special type of plano-convex lens, in which the lens's curved surface is a full hemisphere and

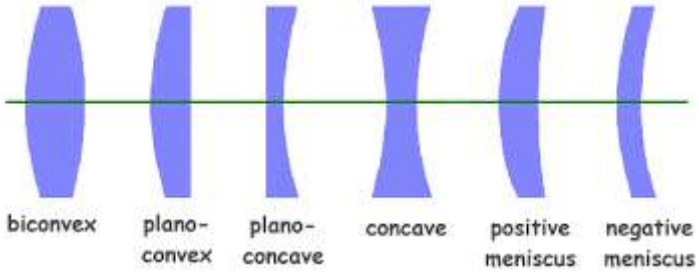

Fig. 18.1. Types of simple lenses.

the lens is much thicker than the radius of curvature.

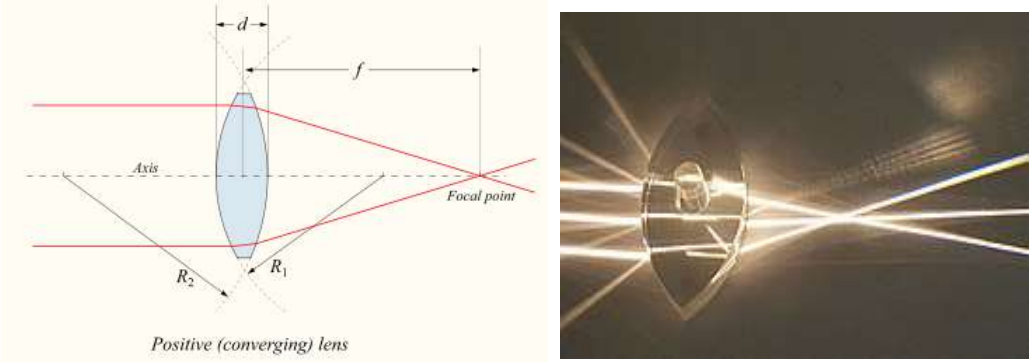

Fig. 18.2. The positive or converging lens.

If the lens is biconcave or plano-concave, a collimated beam of light passing through the lens is diverged (spread); the lens is thus called a negative or diverging lens. The beam, after passing through the lens, appears to emanate from a particular point on the axis in front of the lens. For a thin lens in air, the distance from this point to the lens is the focal length, though it is negative with respect to the focal length of a converging lens.

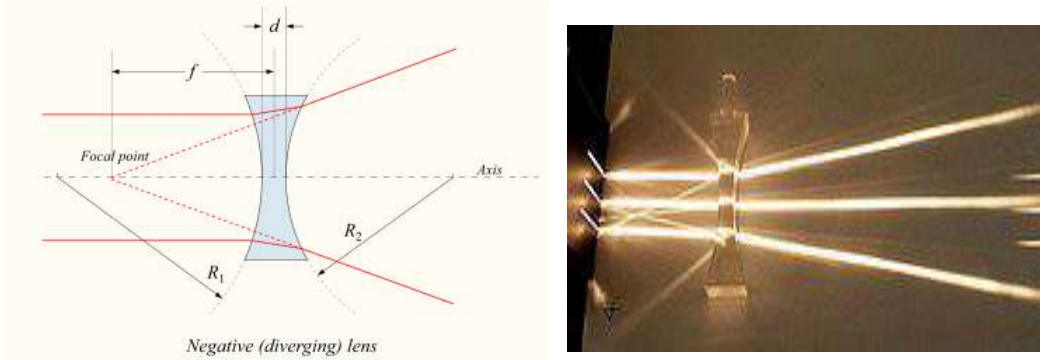

Fig. 18.3. The negative or diverging lens.

Convex-concave (meniscus) lenses can be either positive or negative, depending on the relative curvatures of the two surfaces. A negative meniscus lens has a

steeper concave surface and is thinner at the centre than at the periphery. Conversely, a positive meniscus lens has a steeper convex surface and is thicker at the centre than at the periphery. An ideal thin lens with two surfaces of equal curvature would have zero optical power, meaning that it would neither converge nor diverge light. All real lenses have nonzero thickness, however, which makes a real lens with identical curved surfaces slightly positive. To obtain exactly zero optical power, a meniscus lens must have slightly unequal curvatures to account for the effect of the lens' thickness.

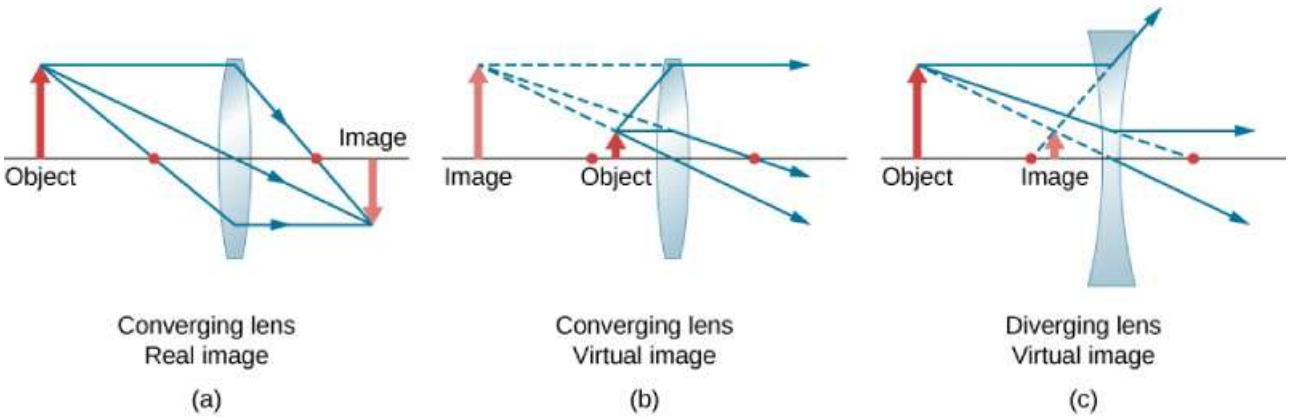

Fig. 18.4. The red dots show the focal points of the lenses. (a) A real, inverted image formed from an object that is farther than the focal length from a converging lens. (b) A virtual, upright image formed from an object that is closer than a focal length from the lens. (c) A virtual, upright image formed from an object that is farther than a focal length from a diverging lens.

A lens is formed by two curved surfaces or one curve and the second flat surface and a transparent medium like the glass are filled or exists between the two surfaces. The lens can have any type of curved surface such as spherical, parabolic or cylindrical. But we usually talk about or use a spherical lens. That is, when someone tells us about a lens, we generally think of it as a spherical lens. And in the middle there is a transparent medium, so we can also say clearly that there is the refraction of light from both the surfaces. The side from which light enters or enters the lens is called the first surface of the lens and after refraction from both surfaces, the page from which the light exits the lens is called the second surface.

Derivation of thin lens formula.

From the thin lens ray-tracing methods we can derive algebraic expressions relating quantities such as object distance, focal length, image distance, and magnification. We will call the distance that the object is to the left of the lens S_1 (or u). We'll call the distance that the image is to the right of the lens S_2 (or v). And the distance of the focal point to either side of the lens is f.

Let's also call the height of the object h_1 and the height of the image h_2.

Now we need to use geometry to get a relation between these distances. That is, we'll derive the equation.

There are two triangles here (fig.18.5,a) that we can make use of, one on the left and one on the right. These two triangles are similar. They have the same angles in all their corners. We know that the sides all scale in proportion. So we can conclude one useful piece of information and that is the $h_2/h_1 = S_2/S_1$.

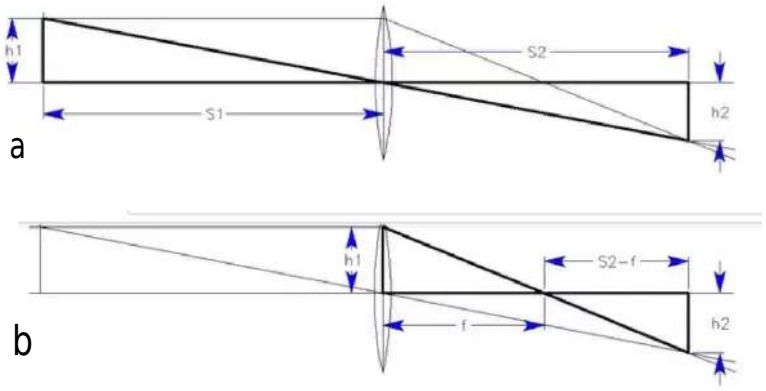

Fig. 18.5. The geometry of problem.

Next we'll construct two other triangles and make use of them (fig.18.5,b).

Again we have two similar triangles, so we can take ratios of sides:

$$h_1/f = h_2/(S_2 - f).$$

We are going after an equation that relates the S_1, S_2, and f, so we want to eliminate h_1 and h_2. That's quite easy to do. The first equation is the ratio of the h's, and we can rearrange the second equation to make it the same ratio of h's:

$$h_2/h_1 = (S_2 - f)/f$$

Great, now just equate this expression with the first expression:

$$h_2/h_1 = S_2/S_1 = (S_2 - f)/f.$$

Now it's a matter of fiddling these around to get the form of the equation you asked about. Simplify the right side dividing out the f:

$$S_2/S_1 = S_2/f - 1$$

Now divide both sides by S_2:

$$1/S_1 = 1/f - 1/S_2$$

Just move the $1/S_2$ to the left and we have it:

$$\boxed{\frac{1}{S_1} + \frac{1}{S_2} = \frac{1}{f}}, \quad \text{or} \quad \boxed{\frac{1}{u} + \frac{1}{v} = \frac{1}{f}}.$$

Magnification. For reception of a big magnification as *magnifier* are used shortphocus lenses. However such lenses have the small sizes and significant aberrations that impose restrictions on their magnification. Substantial magnification can be carried out examining the valid image of a subject, created by additional lens or by system of lenses. Such optical system is the *microscope*, in the elementary case consisting of 2 lenses. The magnifier in this is named the *eyepiece* and the additional lens or system of lenses is called the *objective*. They settle down from each other on distance of $15 - 20$ cm.

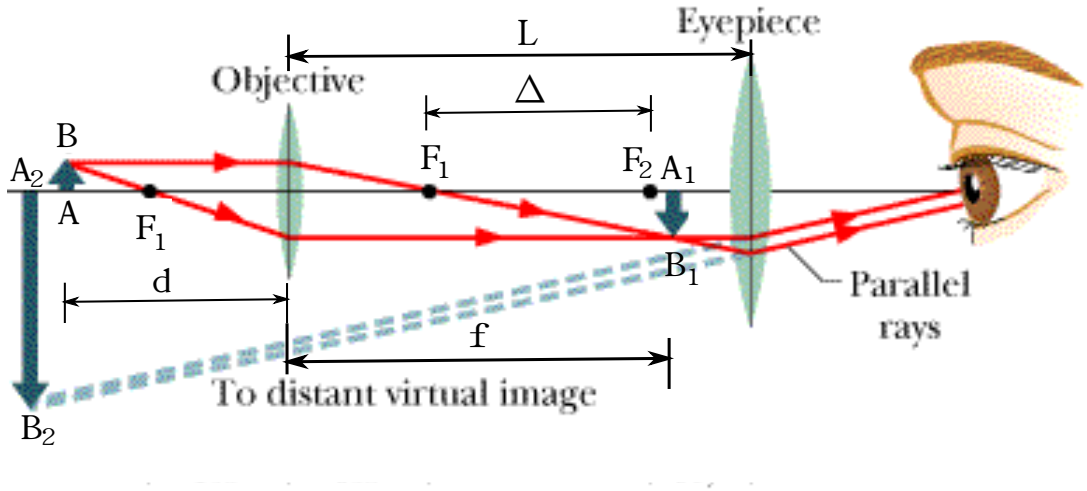

Fig. 18.6. The ray diagram to show the working of a compound microscope.

Usually, the objective of a biological microscope will consist of system of shortphocus lenses which reduces spherical and chromatic aberrations. The eyepiece of a microscope consist from several lenses, it is usual two lenses. Considered object AB (fig.18.6) is located on distance d near to focus of the objective. Its

valid and inverted image A_1B_1 is on distance f from the objective. The eyepiece settles down so, that image A_1B_1 was between forward focus F_2 and the eyepiece. The image AiBi descry through the eyepiece as through the magnifier. The final image is imaginary, increased and inverse concerning the considered object.

Position of the objective concerning of the object is selected so, that final image A_2B_2 settled down from the eye on the distance of the best vision $a_0 = 25\ cm$. The distance between internal focuses of the objective and the eyepiece is known **as optical length of a tube** ($\mathbf{\Delta}$). The optical length of a tube usually is shorter than **geometrical length** L for the sum of focal lengths F_1 and F_2. **Magnification of a microscope eyepiece is equal to product of magnifications of an objective and an eyepiece** $\Gamma = \Gamma_{obj} \cdot \Gamma_{eyep}$. The eyepiece of the microscope is used as magnifier and its magnification defined under the formula $\boxed{\Gamma_{eyep} = \frac{a_0}{F_2}}$. The magnification of the objective can be found taking into account, that the linear magnification of a lens ($\Gamma_{eyep} = \frac{f}{d}$) is equal to the ratio of distance from its optical center up to the image (f) and up to the subject d. Applying this formula to an objective of microscope, it is possible to count that $d = F_1$, $f = F_2 + \Delta$, or neglecting by focal length of objective F_1 in comparison with optical length of the tube (last approximately ten times is more), it is possible to count that $f \approx \Delta$. Then magnification of an objective: $\boxed{\Gamma_{obj} = \frac{F_1 + \Delta}{F_1} \approx \frac{\Delta}{F_1}}$. Hence, magnification of a microscope is equal: $\boxed{\Gamma = \frac{\Delta \cdot a_0}{F_1 \cdot F_2}}$, i.e. magnification of a microscope is equal to the ratio of products of optical length of the tube on distance of the best vision to product of focal lengths of the objective and the eyepiece.

The magnification of objective and eyepiece is called their *own magnifications* and are specified on the frame of lenses.

18.2. Resolution and useful magnification of a microscope.

It is possible to provide rather big magnification of a microscope with corresponding selection of lenses. However in practice seldom use the magnification exceeding of $1500 - 2000$ times. It speaks, that the opportunity to distinguish a fine details of the object is broken by diffraction phenomena that limits useful magnification of a microscope. At passage of light through the smallest details of subject

their image owing to diffraction can lose sharpness, there can be infringement of geometrical similarity of the subject and at last probably full disappearance of the image.

Resolution **R** (**resolving power**) is the ability of a microscope to give separate images of fine details of researched object.

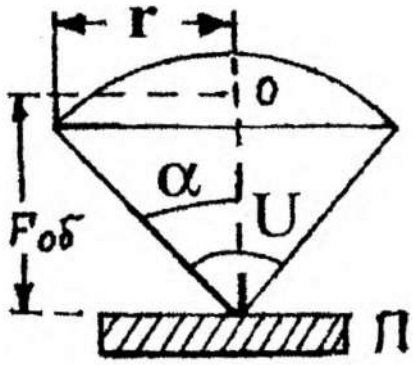

Fig. 18.7. The aperture angle U.

Limit of resolution (**or distinction limit**) (Z) is such least distance between two points of the object, when these points are visible in the microscope separately. Resolution is inversely proprtional to the limit of resolution: ($R \sim \frac{1}{Z}$). Resolution of a microscope is caused by the wave properties of light, therefore expression for limit of the resolution is possible to receive taking into account the diffraction phenomena.

The diffraction theory of resolution of a microscope is developed by E. Abbey, L. Mandelshtam and D. Rozhdenstvenski. Resolution of microscope as whole is defined by resolution of the objective in which rays of light directly enter, diffracting on an object.

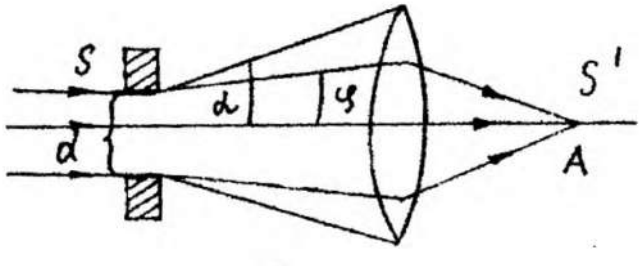

Fig. 18.8. The image of luminous aperture.

The basic element causing resolution of an objective is its **aperture angle U** (fig. 18.7), equal to the angle, formed by the beams going from the preparation to edges of the objective (angle U). The aperture angle defines the angular size of the objective and plays the big role at achievement of the big resolution of a microscope. We shall consider formation with help of the objective of the image of luminous aperture S of small diameter d, on which falls the bunch of parallel monochromatic beams (fig. 18.8). Passing through the aperture, light has diffraction. The objective collects beams and in the connected plane forms in the point A the image of the aperture S'. Two cases are possible:

1) The half of aperture angle α of the objective is more than angle of diffraction φ or is equal to it ($\alpha \geq \varphi$), then *all diffract beams take part in formation of the image and it will be similar to the subject.*

2) The angle $\alpha < \varphi$, then not all starting from aperture beams will take part in

formation of image of the subject. The image will not be completely geometrically similar to the subject.

The degree of infringement of the image will depend on what part of diffract beams does not get into objective and does not take part in formation of the image. The angle φ of diffraction is more, than more length of the wave λ and than is less diameter d of the aperture. Then $\varphi \sim \dfrac{\lambda}{d}$, in the limiting case when $\alpha = \varphi$ it is possible to establish the similar parity $\alpha \sim \dfrac{\lambda}{d}$, whence $\boxed{d \sim \dfrac{\lambda}{\alpha}}$.

Thus, *diameter of the aperture at which similarity of image to the subject is kept, can be than less, than more shortly wavelength and than more aperture angle.*

Having transferred reasonings on conditions of work with microscope, it is possible to count, that diameter d of aperture corresponds to the least size of structural details of a preparation, i.e. to equate it to the limit of resolution of objective of microscope $Z = d$. Then the similar statement can be resulted concerning the resolution of an objective.

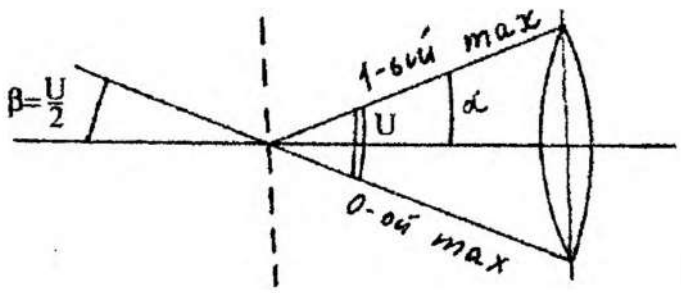

Fig. 18.9. The light incidens on the diffraction grating under the angle of β

In theory of Abbey the diffraction grating undertakes as considered object. In optical devices including a microscope, bunches of light are always limited, therefore it is important to know, how it will affect on distortion of image of a subject, what minimum quantity of beams is capable to transfer the full information about subject. Abbe in the experiments shielded in plane F of converging lens a part of the beams giving the image of diffraction grating. *He has established, that for the resolution of slits in the image of the diffraction grating received with help of a lens on the screen it is necessary,* **that for formation of its image participated beams from maxima of zero and first orders, even on the one hand**. In the limiting case, it agrees to Abbe, extreme beams of the limited conic light bunch will be the beams corresponding to the central (zero) and to the 1-st main maximum. Limit of resolution in this case can be equal to the period of the diffraction grating (d). Then, using the formula of diffraction grating $k\lambda = d\sin\varphi$ ($k = 1; = \alpha; d = Z$) for perpendicular incidence

beams in air, it is possible to write down: $\boxed{Z = \frac{\lambda}{\sin\alpha}}$; i.e. *distinction limit at direct incidence beams is numerically equal to the ratio of wavelengt of light to the sine of half of the aperture angle of the objective.*

Let light incident on the diffraction grating under the angle of $\beta = \frac{U}{2} = \alpha$ (fig. 18.9). In this case the formula of diffraction grating will look like: $d(\sin\beta - \sin\alpha) = \pm k\lambda$. As $\sin(-\alpha) = -\sin\alpha$, from this formula at $B = \frac{U}{2}$ also $\alpha = -\frac{U}{2}$, we shall receive $2d\sin\left(\frac{U}{2}\right) = X$, or $Z = \frac{0.5\lambda}{\sin\alpha}$. If light is distributed not in air and in the medium with the refractive index n, then $\lambda_n = \frac{\lambda}{n}$ (λ is length of wave of light in air).

For this case the **distinction limit** is equal: $\boxed{Z \geq \frac{0.5\lambda}{\sin\alpha} = \frac{0.5\lambda}{n \cdot \sin\alpha}}$.

At the other approach to deduce of the formula for definition of the distinction limit at *inclined incident beams* on the objective this formula looks like: $Z = \frac{0.61\lambda}{n \cdot \sin\alpha}$, where value $\boxed{A = n \cdot \sin\alpha}$ is named the **numerical aperture**.

From this formula for Z follows, that **one of the ways of reduction of the distinction limit is reduction of wavelength of light**. In this connection the *ultra-violet microscope* in which microobjects are investigated in *UV-light* is applied. In it the optics transparent for **UV**-beams (quartz optics), and for fixing the image photographic plate, luminescent screens or electron-optical converters is used.

Other way of reduction of limit of resolution is increase of the numerical aperture. It can be increased having increased the aperture angle. It can be made by 1) approaching the subject to the objective. However the distance from subject to lens of the objective cannot change any way, it is constant for each objective.

The numerical aperture can be increased with help of the special liquid medium (**immersion** (I)) in space between the objective (O) and integumentary glass (C) of microscope (fig. 18.10), (K is condenser). In immersion systems in comparison with "dry" systems receive the big numerical aperture. As immersion mediums is used water ($n = 1.333$), cedar oil ($n = 1.515$), monobrominenaphthaline ($n = 1.66$). At immersion light from the subject up to the objective passes on homogeneous optical medium and does not give losses on reflection. It considerably *raises brightness* of the image that has rather essential value especial

for the microscope with big magnification. In modern microscopes the aperture angle can have the greatest value of 70^o. In this case the distinction limit of the optical microscope is **0.2 – 0.3 microns**.

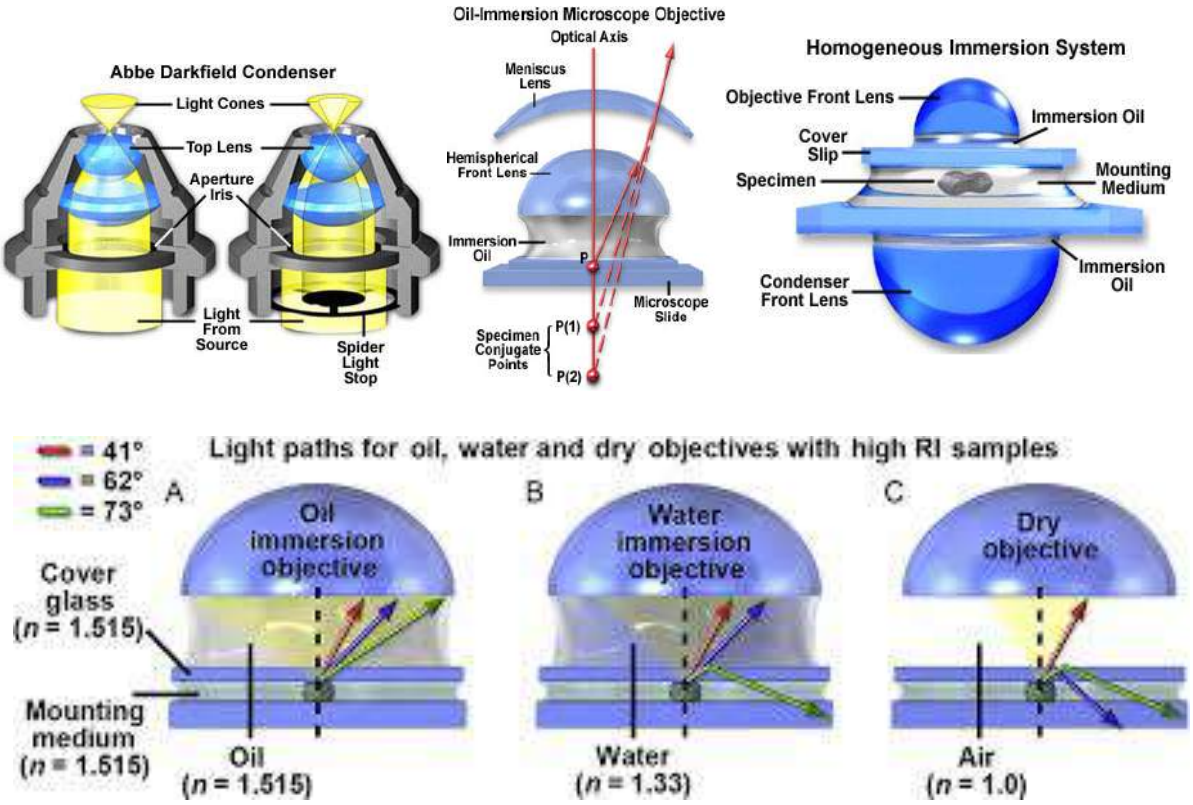

Fig. 18.10. The condenser.

Let's estimate **useful magnification** of the optical microscope.

If the subject has the size equal to the distinction limit Z, and the size of its image is Z' and if its image is located on distance of the best vision from eye, magnification of the microscope: $\Gamma = \dfrac{Z'}{Z}$.

This magnification is known as **useful magnification**.

As $Z = \dfrac{0.5\lambda}{A}$, then $\boxed{\Gamma = \dfrac{Z' \cdot A}{0.5\lambda}}$. The magnification of microscope is named *useful* because at it the eye of the person distinguishes all elements of structure of object which gives the microscope.

18.3. Some special methods of optical microscopy.

a) Measurement of the sizes of microscopic objects

Determination of size of microscopic objects is carried out with help of the eyepiece M_{eyep} and objective M_{obj} (fig. 18.11) micrometers as glass plates with scales putting on them. The eyepiece micrometer is established in the plane of

intermediate image received from the objective. In the eyepiece is observed image of the scale combined with the image of object.

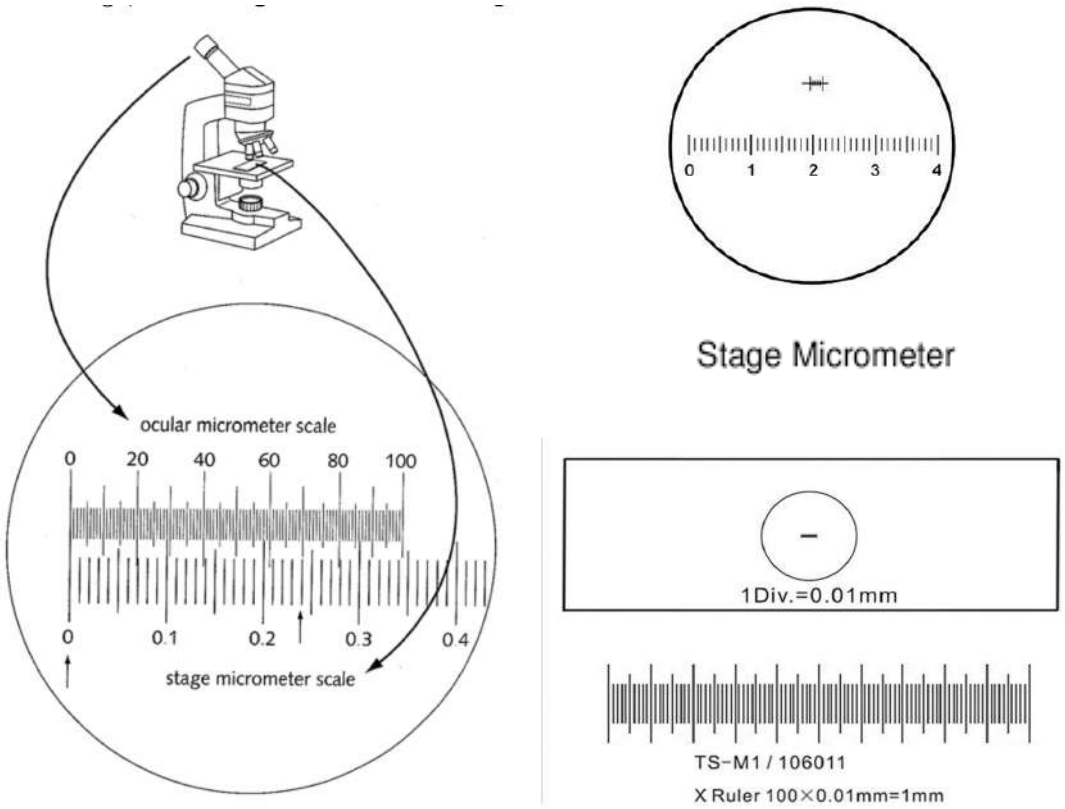

Fig. 18.11. The eyepiece micrometer and microscopic objects.

If the scale division value of a scale of the eyepiece micrometer is known it is possible to determine the size of this image given by the objective; having divided the received value on known magnification of the objective we can receive the valid sizes of object.

If the scale division value of the eyepiece micrometer is unknown, it can be determined by means of objective micrometer M_{obj} with the known scale division value (usually is 0.01 mm). The objective micrometer is placed on the place of a subject. In the eyepiece observe the combined images of both scales and determine the scale division value of the eyepiece micrometer.

b) Microprojection and microphoto

In the optical microscope the imaginary image originate due to that the intermediate valid image formed by objective settles down between forward focus F_{eyep} and the eyepiece. If to move up the eyepiece so, that the image, which gives the objective there would be before forward focus of eyepiece (fig. 18.12), last will give the valid image, which can be designed on the screen or photographic

plate. The eyepiece in this case serves as a projective lens. It is possible to remove the eyepiece and to project on the screen or a photographic plate the valid image given only by the objective, though thus magnification will be smaller.

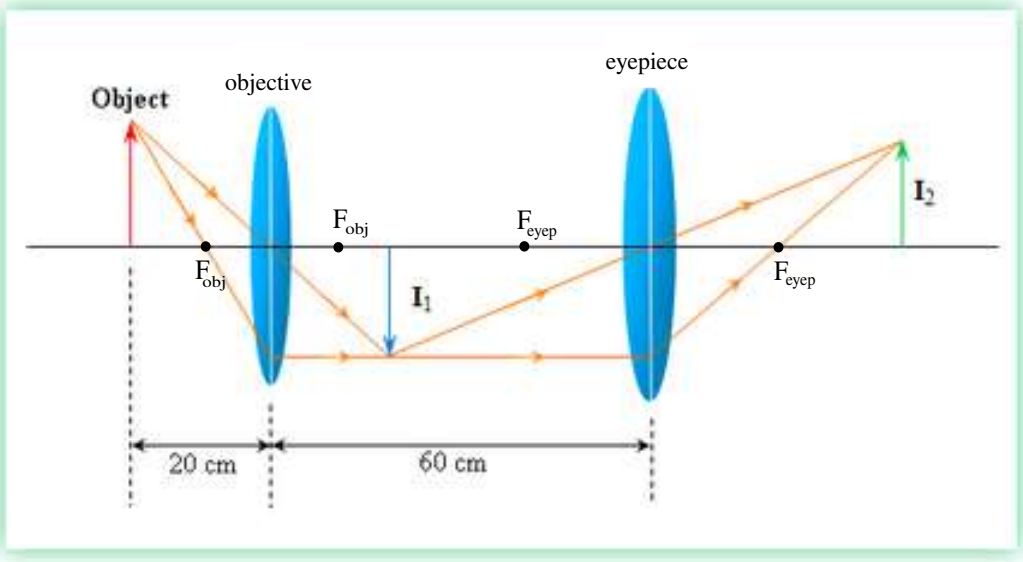

Fig. 18.12. The valid image, which can be designed on the screen or photographic plate.

Supervision on the screen of the valid image of the subjects received by one the specified ways is known as **microprojection**. Photographing of the valid image received in such way is called **microphoto**. Usually for this purpose is used the special photonozzle (photoadapter) to microscope which represents the camera, putting on the eyepiece end of the tube of microscope. The image of subject is projected on the plane of position of the photographic plate. The nozzle is supplied with visual tube for supervision over the image during shooting.

The linear magnification of photonozzle of the microscope is defined under the formula: $\Gamma_{NOZ} = n_{obj} \cdot n_{eyep}\dfrac{x}{250}$, where x is distance in mm from the eyepiece of microscope up to the photographic plate; 250 is distance of the best sight in mm. n_{obj} and n_{eyep} are magnification of the objective and eyepiece.

c) Phase-contrast method

At passage of light wave through *transparent object* intensity of light almost do not changes, but phases undergo changes, depending on thickness of object and its refractive index. To see details of such objects by usual way is practically impossible. The phase-contrast method is applied for supervision of not enough contrasting objects and is based on using of phase difference that is formed at passage of light through various structures (sites) of the researched object.

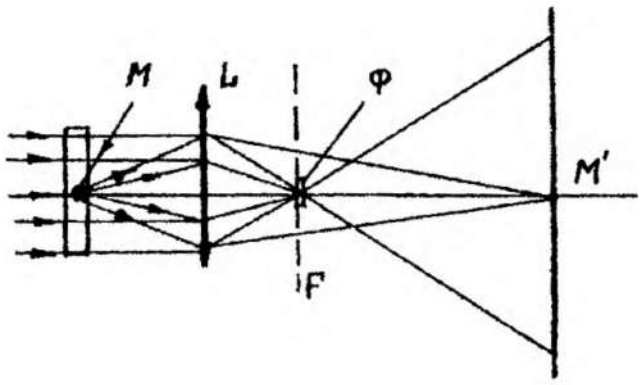

Fig. 18.13. The diffraction of light beams

We admit that in the homogeneous transparent medium of object with the refractive index of n there is the transparent inclusion M with the refractive index n_1 causing diffraction of light beams (fig. 18.13). At illumination of the object by parallel bunch of beams, part of it will pass through the medium, will come together in the small site A of focal plane F of objective and then will get on the screen by breaking up bunch. The beams formed owing to diffraction of light on heterogeneity of object fall on the objective as a divergent beam and after objective will not pass through its focus, and will be come together on the screen in the some point M', being the image of inclusion M. Between the beams, falling on the preparation in parallel and beams, diffract on heterogeneity of M be some path difference, which increases with help of the optical device (**phase plate**) up to half of the wavelength ($\lambda/2$). Therefore in the point M' direct and diffract beams interfere and mutually extinguish each other. Therefore the image of inclusion M is observed by blacked out on the light background of the medium environmental it. Phase plate Φ represents the layer of transparent substance of the certain thickness with the certain refractive index. The plate has the form of circle of very small diameter and is established in focus of the objective. Through it pass only beams, which drop on the preparation by parallel bunch. They receive thus the additional path difference in relation to beams diffracting on M. For phase-contrast microscopy applies the special objectives, containing the phase plate and special condensers, which are established in usual biological microscope.

18.4. Wave properties of particles. Electronic microscopy.

First step in creation of quantum mechanics was discovery of wave properties of microparticles.

French *physicist Louis De Broglie in 1924* has drawn the conclusion, *that any moving particle* of substance, as well as quantum of radiation, *has not only* **corpuscular** *properties, but also wave*, which can be characterized comparing to particle some wavelength, which is connected with the impulse p of particle by

the same ratio, as well as for the photon, i.e.: $\boxed{\lambda = \frac{h}{p} = \frac{h}{mv}}$, where m is mass of a particle; v is its speed; h is Planck's constant.

This wave is named the **wave of De Broglie**. It characterizes the wave properties of a moving particle. The wavelength of De Broglie is rather small. For electron at $v = 10^8\ m/s$ it has the order of 7 $\mathring{A}$ ($7 \cdot 10^{10}\ cm$), i.e. correspond to wavelength of X-ray radiation. The hypothesis of De Broglie was so unusual, that many large physics have not given to it any meaning, however several years later presence at moving particles of wave properties has been confirmed experimentally. In 1927 K. Devison and L. Dzhermer observed on the monocrystal of nickel diffraction of electrons.

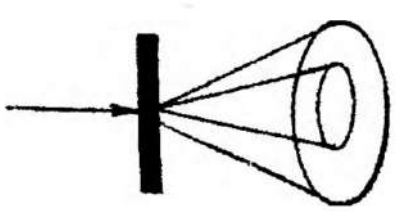

Fig. 18.14. The diffraction picture consists of lines of concentric dark and light rings.

In later experiments diffraction has been found out at passage of bunch of electrons with high energy through the metal foil (polycrystalline body). Electrons dissipate on the foil and on the photoplate or the fluorescing screen is formed the diffraction picture consisting of lines of concentric dark and light rings (fig. 18.14). The similar picture takes place at passage through the same foil of X-rays.

Wave properties of particles can be used for reception of the increased images of subjects. From told early follows that the distinction limit of optical microscope is defined by value of wavelength. To reduce the distinction limit allows the electronic microscope, in which the *data carrier about a subject is the stream of electrons* which, passage through substance will dissipate on various directions. We shall find dependence of the wavelength of De Broglie for electrons from accelerating voltage:

$\frac{mv^2}{2} = eU$, whence $v = \sqrt{\frac{2eU}{m}}$, but $\lambda = \frac{h}{mv}$, then $\boxed{\lambda = \frac{h}{\sqrt{2eUm}}}$.

From the formula for λ is possible to draw the conclusion, that the *distinction limit of the electronic microscope depends on the accelerating voltage* (other sizes are constants). Having substituted to the formula of the distinction limit of optical microscope length of the wave of De Broglie for electrons, we shall receive for an electronic microscope:

$$\boxed{Z = \frac{0.5h}{\sqrt{2emU} \cdot n \cdot \sin\alpha}}.$$

Character of dispersion of electrons depends on structure of the layer of substance through which they pass. Objects of research usually prepare as films, ultrathin cuts and are placed on special frameworks or grids from the thinnest wire, and also on the films-substrates, which are not having own structure. For work with an electronic microscope are suitable very thin objects ($5 - 100\ nm$), because electrons are strongly absorbed and dissipate by substance.

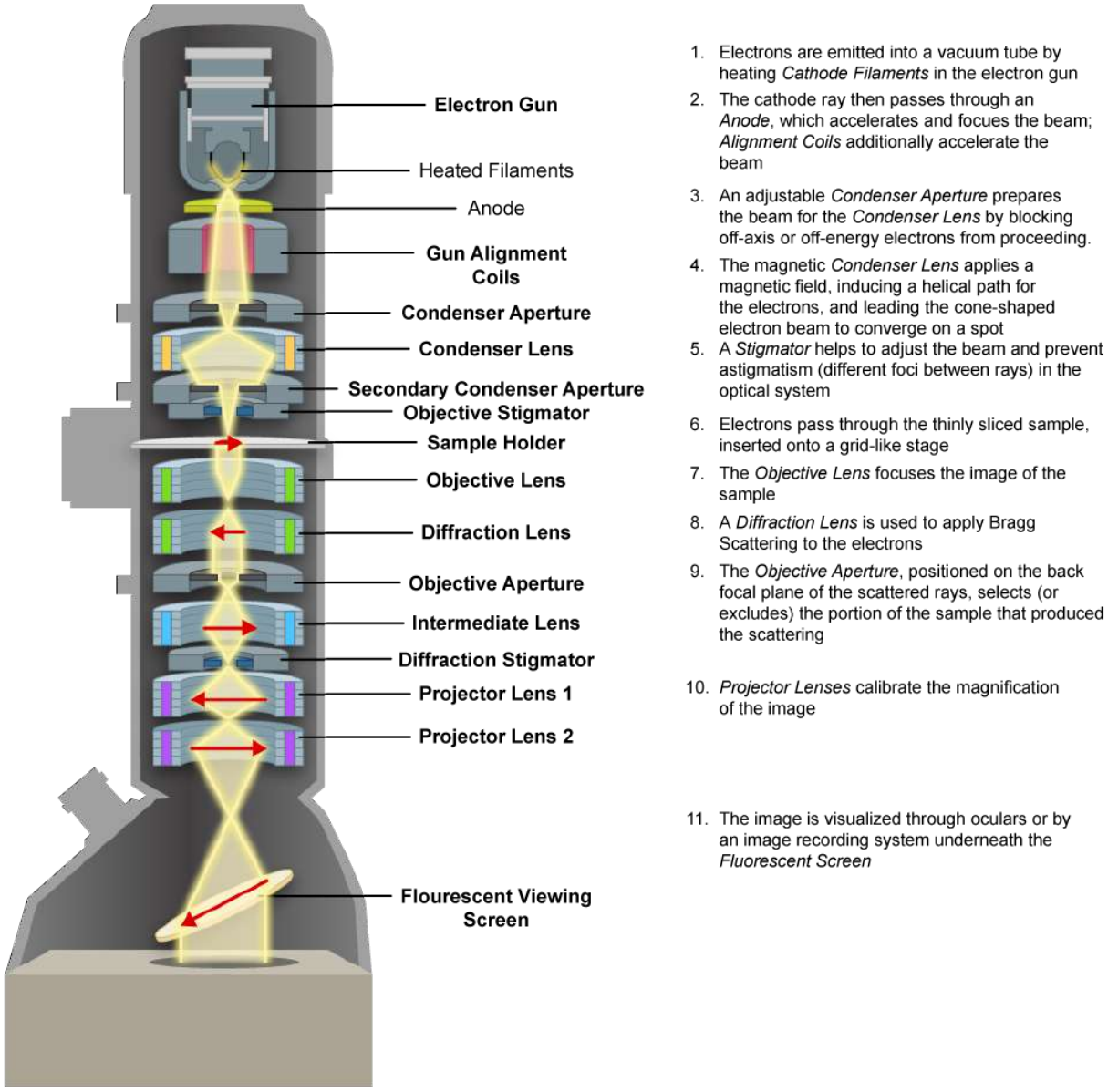

Fig. 18.15. The Transmission-Electron-Microscope.

The image which turns out on the screen or photographic plate will display structure of object. Thus it can be considerably increased in comparison with a subject. *The basic difference of an optical microscope from electronic* is that with object cooperates the bunch of electrons instead of light beams. Therefore instead of system of optical lenses in the electronic microscope by movement of

electrons operate magnetic or electric fields (magnetic or electric lenses). Such fields receive with help of the coil with a current or with the systems of charged electrodes. Magnetic lenses are more frequently used as they give less distortion. On fig.18.15 the electron-optical system of the elementary electronic microscope with magnetic lenses is schematically shown.

The bunch of fast electrons from the electronic gun (1) gets to the condenser lens $(3-4)$, directing the bunch of the necessary section on the researched object (6). Due to the different degree of dispersion of electrons by the different sites of object distinguished by thickness, density or the chemical compound, the bunch taking place through the object electrons transfers in itself the information on this object. The *objective lens* (7) gives intermediate (increased in some times) image of object. The *projective lens* (10) forms the final image (11), which is registered by the photographic way, or is observed visually on the luminescent screen in special viewing glass. All these units are connected with each other, forming column of the microscope inside which low pressure $(10^2 - 10^3\ Pa)$ is supported. The working voltage for dispersal electrons reaches of $50 - 100\ kV$.

Maximal magnification Γ of the microscope having except for the condenser and the objective only one projective lens, is defined by the focal length f_1, and f_2 of objective and projective lenses and by distance L between object and the plane of the final image: $\boxed{\Gamma = \dfrac{L^2}{4 \cdot f_1 \cdot f_2}}$. Usually sizes f_1, and f_2 make some millimeters and L is $1-2\ m$. *The useful magnification in microscopes reaches of* 10^6, and the distinction limit is $Z \approx 0.1$ nanometers, that in hundreds times is better, than at optical microscope.

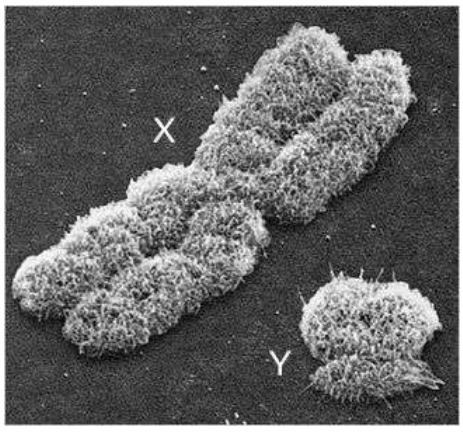

Fig. 18.16. The photo from an electron microscope.

It is necessary to note, that application of accelerating voltage greater then $100\ kV$, though raises resolution of the microscope, but it is connected with destruction of researched object by electrons having the big speed.

The image in the electronic microscope can be formed due to passage of electrons through the object and in this case the microscope is called transmission, if the image is formed by reflected from object electrons the microscope is called reflective. For biological researches is used basically transmission microscope. *With help of the electronic microscope are received unique pictures of*

various cells, subcellular structures, viruses, bacteria.

It is possible to observe large organic molecules (for example, RNA at magnification in 10^5 *times). The electronic microscopy allows to study structure of cellular membranes, nervous fibres. With help of electronic microscope the gene for the first time has been seen. The used domestic microscope of the* $EVM - 100 - LM$ *gives the maximal* 600000*-fold magnification and the distinction limit is about of* $3 \cdot 10^{-10}$ *m (*0.3 *nanometers).*

18.5. Fiber optics and its application in endoscopy.

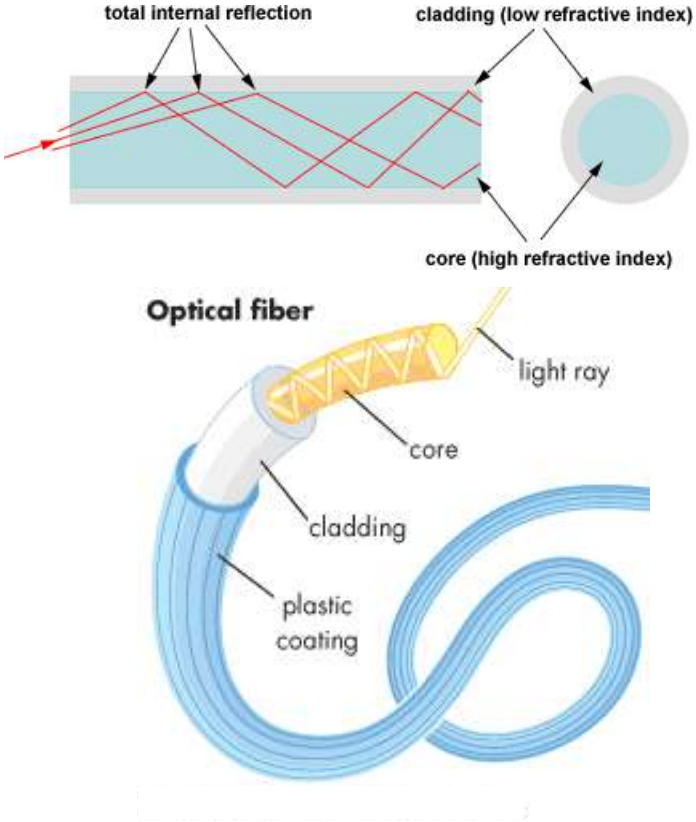

Fig. 18.17. The optical fiber structure.

In the beginning of $50 - th$ years of the last century into various branches of science and especial in medicine began to take root optical fibre elements, which are capable to transfer light on the channels, named *wavebeam guides.*

The section of optics in which is considered transportation of light and the image on wavebeam guides is known as **fiber optics**. *The fiber optics is based on phenomenon of* **total internal reflection**. Light getting inside of transparent fibre (or core) surrounded with substance with the smaller refractive index, is *repeatedly reflected* and propagated along these fibres (fig. 18.17). As at total internal reflection coefficient of reflection is rather high ($K = 0.9999$), losses of energy in a fibre basically are caused by absorption of light inside the fibre. So, for example, in the fibre of 1 m in seen area of spectrum is lost from 30 up to 70% of energy of light. For transfer of the big light streams and preservation of flexibility of wavebeam guide separate fibres gather in bunches (plaits) - optical paths, which in medicine are used for the decision of two problems: 1) transfers of light energy for illumination by cold light of internal cavities; 2) for transfer of the image.

For decision of the first problem has no value relative position of separate

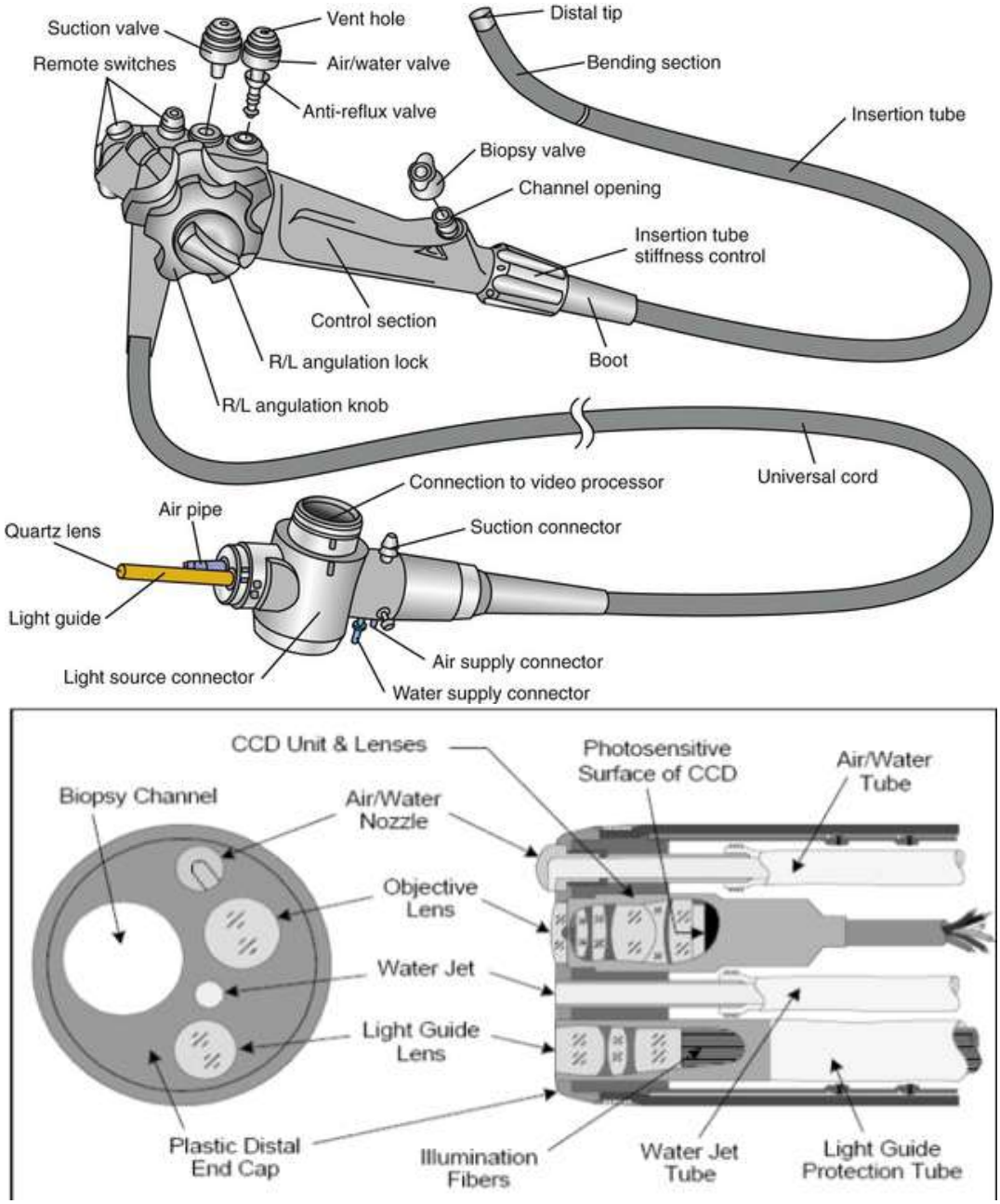

Fig. 18.18. The medical endoscope structure.

fibres. In the second case very important that the arrangement of fibres in the plait on the input and output was identical, otherwise the image will be deformed. The fiber optics has allowed tomodemize existing early the medical device **endoscope**. *Endoscope* is the special device for inspection of internal cavities (stomach, rectum, bronchial tubes etc.), which will consist of two parts: light source and the viewing part containing system of lenses (fig. 18.18). The light source (tiny bulb) is placed on the end of endoscope, which is entered inside. Using fiber optics it was possible, at first, light from the bulb to transfer inside of body on optical path and by that to avoid undesirable heating of this body, which arises at premise of light source inside of a cavity in endoscope of old design; second, flexibility of

optical fibre system of such endoscope supposes inspection of the most part of cavities of the body of person than rigid endoscopes. Fiber endoscope allows to make necessary pictures with the purpose of diagnostic. With the help of optical paths is possible to transmit laser radiation in internal bodies with the purpose of medical influence on a tumour.

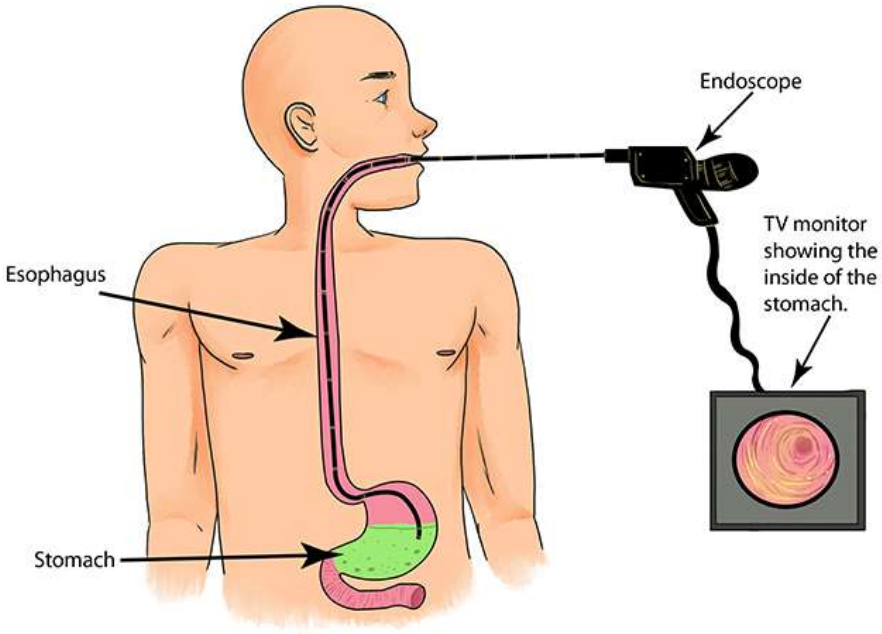

Fig. 18.19. The medical examination with an endoscope.

Test questions

1. What is an optical lens? What types of lenses do you know? What is called the main optical axis of the lens? What is called the secondary optical axis? What is it used for?

2. What rays are used to create an image using a thin lens? Sketch an image in a scattering lens and give it a characteristic.

3. Deduce of the lens formula.

4. What is called the resolution limit of the eye and what is it equal to?

5. Sketch a ray path in a magnifier (eyepiece). What is called the angular magnification of the optical device?

6. Sketch the path of the rays in a microscope according to the rules of construction.

7. What is the structure of the eyepiece micrometer; what is it used for?

8. How to determine the magnification of a microscope? List possible ways to increase the resolution of an optical microscope.

9. Which smallest object can be seen in an optical microscope? Which phenomenon fundamentally limits its resolution? What is the purpose of an optical microscope.

10. Explain the principle of the eye. What is myopia, hyperopia and astigmatism? How to get rid of them?

Lecture No 19

Heat radiation

19.1. Scale of electromagnetic waves

Table 19.1

Scale of electromagnetic waves with characteristics

Type of Wave	Approximate Frequency Range (Hz)	Approximate Wavelength Range (m)	Source
Power waves	50	$5 \cdot 10^6$	Electric currents
Radio waves			Electric currents
SW	$0.53 \cdot 10^6 - 1.70 \cdot 10^6$	$570 - 186$	
HF	$88 \cdot 10^6 - 108 \cdot 10^6$	$3.4 - 2.8$	
UHF	$54 \cdot 10^6 - 890 \cdot 10^6$	$5.6 - 0.34$	
Microwaves	$10^9 - 10^{11}$	$10^{-1} - 10^{-3}$	Special vacuum tubes
Infrared radiation	$10^{11} - 10^{14}$	$10^{-3} - 10^{-7}$	Warm and hot bodies
Visible light	$4.0 \cdot 10^{14} - 7.0 \cdot 10^{14}$	10^{-7}	Sun and lamps
Ultraviolet radiation	$10^{14} - 10^{17}$	$10^{-7} - 10^{-10}$	Very hot bodies and special lamps
X-rays	$10^{17} - 10^{19}$	$10^{-10} - 10^{-12}$	High – speed electron collisions and atomic processes
Gamma rays	above 10^{19}	below 10^{-12}	Nuclear reactions, processes in particle accelerators, and natural radioactivity

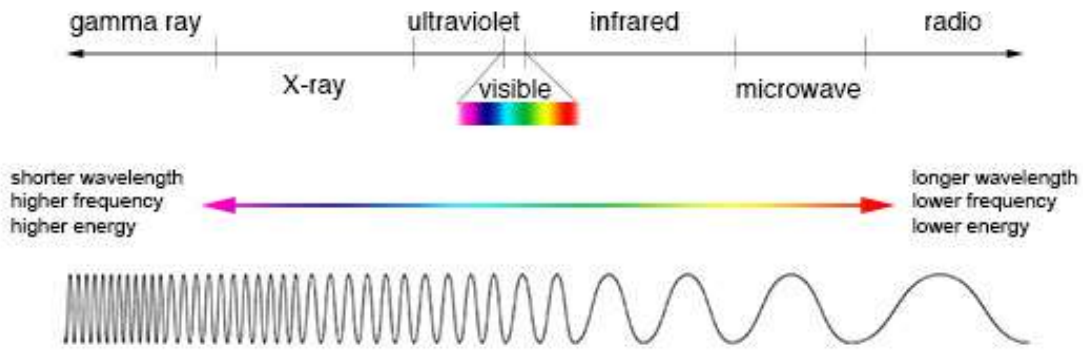

Fig. 19.1. Electromagnetic spectrum.

Electromagnetic waves are formed when an electric field couples with a magnetic field. The magnetic and electric fields of an electromagnetic wave are perpendicular to each other and to the direction of the wave (Fig. 19.2). Electromagnetic waves are classified by ranges of frequencies or wavelength in a spectrum. The electromagnetic spectrum is continuous, so the limits of the various ranges are approximate. Table 19.1 lists these frequency and wavelengths ranges for the general types of electromagnetic waves.

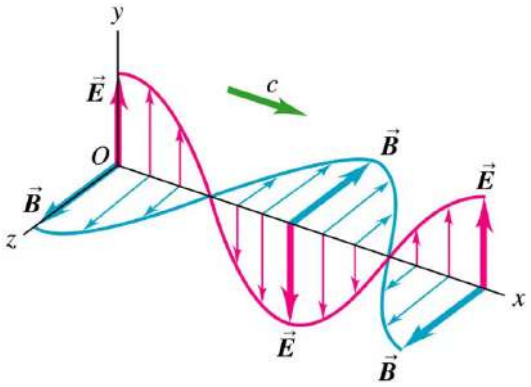

Fig. 19.2. Electromagnetic waves.

19.2. Characteristics of heat radiation

Heat (temperature, thermal) radiation is electromagnetic radiation of bodies appeared due to internal energy (energy of heat motion of atoms and molecules).

Heat radiation is specific for all bodies, the temperature of which is higher than absolute zero. Heat radiation of human body is infrared radiation. *Electromagnetic radiation within wavelength range from* 0.76 μm to 2000 μm is called infrared radiation (IR):

$$7.6 \cdot 10^{-7}\ m \leq \lambda_{IR} \leq 2 \cdot 10^{-3}\ m$$

Characteristics of heat (temperature) radiation

1. Radiation flux (Φ) is an average power of radiation per time interval, which considerably exceeds period of electromagnetic oscillations. Flux Φ is measured in Watts (W).

$$\boxed{\Phi = \frac{E}{t}}, \quad [\Phi] = W.$$

2. Radiance or radiant emittance of body (R) is the energy of the radiation at all possible wavelengths emitted by unit area of the body surface in unit time (flux of radiation from area unit of body surface):

$$\boxed{R = \frac{E}{t \cdot S}}, \quad \boxed{R = \frac{\Phi}{S}}, \quad [R] = \frac{W}{m^2}.$$

3. Spectral radiance or spectral density of energetic emittance (R_λ) is radiation emittance of the body that refers to units of spectral interval:

$$\boxed{R_\lambda = \frac{dR}{d\lambda}}, \quad [R_\lambda] = \frac{W}{m^3}.$$

where dR is the radiance of radiation within the wavelength range from λ to $\lambda + d\lambda$. From the definition of spectral radiance, it follows that:

$$\boxed{R = \int_0^\infty R_\lambda d\lambda}$$

That is why R is also called **integral radiance**.

4. Absorption coefficient (α) is a quantity that is equal to the ratio of flux of radiation absorbed by the body to the flux that falls on the body:

$$\boxed{\alpha = \frac{\Phi_{abs}}{\Phi_{fall}}},$$

5. Monochromatic absorption coefficient (α_λ) is a quantity α refers to spectral interval unit:

$$\boxed{\alpha_\lambda = \frac{\Phi_{abs(\lambda)}}{\Phi_{fall(\lambda)}}}$$

19.2.1 Absolute black body. Grey body

The body for which the monochromatic absorption coefficient is equal to one in all spectral intervals and at any temperature is called **absolute black body**, i.e.

$$\alpha = \alpha_\lambda = 1$$

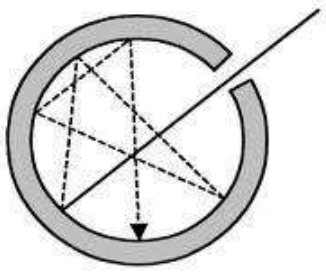

Fig. 19.3. Conceptual Black Body

The model of absolute black body may be a cavity with a very small hole (Fig. 19.3). The ray of any wavelength inside of this cavity may leave it only after multiple reflection. At each reflection from walls of the cavity a part of ray's power is absorbed and only negligible part of ray's power occurred in the opening may return back; that is why opening absorption coefficient proves to be close to one.

The notion of grey body is also introduced into the theory of heat radiation. This is the body the absorption coefficient of which is lesser than one and does not depend on wavelength. Human body may be considered as grey body in infrared part of spectrum, because its absorption coefficient in this spectral range is $\alpha_\lambda \approx 0.9$.

19.3. Laws of heat radiation

1. Kirchhoff's law

Kirchhoff's law establishes quantitative correlation between radiation and absorption property of body. This law was laid down by Kirchhoff in 1859. It states that the ratio of spectral density of radiant emittance to monochromatic absorption coefficient is equal for all bodies at the given temperature and equal to spectral density of energetic emittance of absolute black body at the same temperature

$$\boxed{\left(\frac{R_\lambda}{\alpha_\lambda}\right)_1 = \left(\frac{R_\lambda}{\alpha_\lambda}\right)_2 = \cdots = R_{\lambda\ abs\ b.\ b.}}$$

where $R_{\lambda\ abs\ b.\ b.}$ – spectral density of energetic emittance of absolute black body.

In other words the ratio of spectral density of energetic emittance of a body

to its coefficient of absorption does not depend on the nature of radiant body and is equal to spectral density of absolute black body at the given temperature.

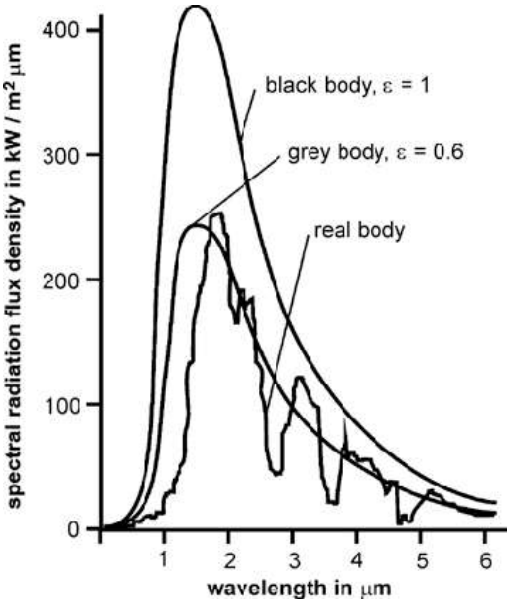

Fig. 19.4. $R_{\lambda,T}(\lambda)$ for black body, grey body, and real body

2. Stefan-Boltzmann law

Integral energetic emittance of absolute black body is in proportion to fourth power of its absolute temperature.

This is **Stefan-Boltzmann's law.**

$$\boxed{R_{abs\ b.\ b.} = \sigma T^4}$$

where $\sigma = 5.7 \cdot 10^{-8}\ \frac{W}{m^2 K^4}$ is called the Stefan-Boltzmann's constant.

First the law was laid down on the basis of experimental data given by Austrian physicist Stefan in 1879, and later it was received theoretically by Austrian physicist Boltzmann in 1884. The law proves the extremely rapid increase of heat radiation power with the rise of body temperature. Stefan-Boltzmann's law may be used for heat radiation of a man if human body is considered as "grey". In this case the suggested radiation coefficient δ is used instead of Stefan-Boltzmann's constant:

$$\boxed{R_{g.b.} = \delta T^4}$$

Where $\delta = \alpha\sigma$, i.e, the given radiation coefficient is equal to the product of Stefan-Boltzmann's constant by absorption coefficient α, which is lesser than one for grey bodies.

Stefan-Boltzmann's Law states that the total energy radiated per unit surface area.

Thus, **wavelength, at which maximum of spectral density of energetic emittance of absolute black body falls, is in inverse proportion to its absolute temperature, or frequency that corresponds to maximum of radiation property of absolute black body is in direct proportion to its absolute temperature.**

3. Wien's Displacement law

Wien's law:

$$\boxed{\lambda_{max} = \frac{b}{T}}$$

where $b = 0.29 \cdot 10^{-2}\ m \cdot K$ is called Wien constant;

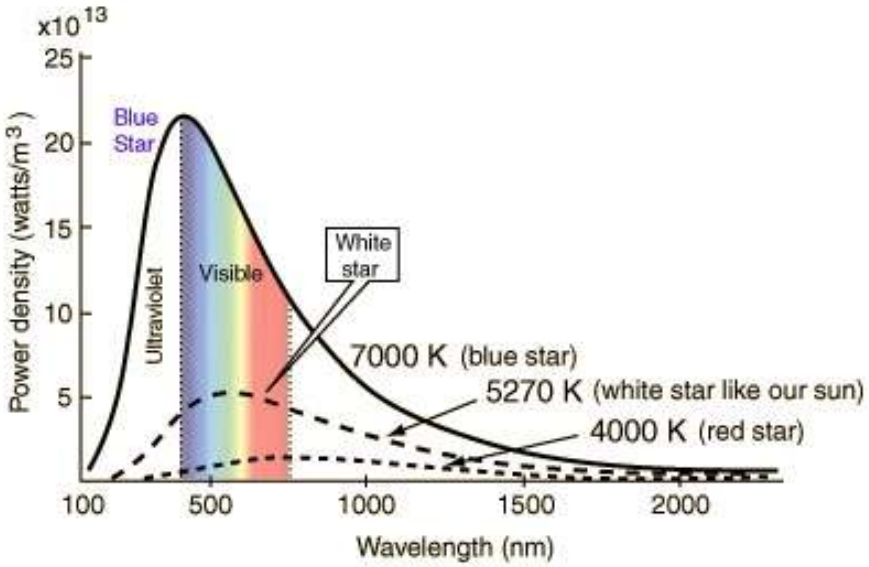

Fig. 19.5. Displacement law.

λ_{max} is the peak wavelength in meters;

T is the temperature of the blackbody in kelvins (K).

It follows from Wien law, that with an increase of temperature the wavelength, which corresponds to maximum of radiation, displaces towards lesser wavelengths (greater frequencies). That is why this law is called "displacement law".

Let's evaluate quantities λ_{max} in spectrum of heat radiation of human body. For the temperature of the skin surface $t = 32^oC (T = 273 + 32 = 305^oK)$

$$\lambda_{max} = \frac{0.29 \cdot 10^{-2}}{305} \approx 10\mu m$$

The received wavelength falls into infrared range, i.e. it is not perceived by a human eye.

19.3.1 Temperature and Temperature Scales

Temperature is a physical magnitude that characterizes the state of thermodynamic equilibrium of macroscopic system. The qualitative estimation of temperature is associated with the notions of hotness and coldness of the object while the quantitative measurement of the temperature is realized with thermometer.

All types of thermometers are based on measurement of the change in some physical properties (volume, length, pressure, resistance, color); each thermometer is supplied with the *temperature scale* calibrated in units called *degrees*. Usually a temperature scale uses two reference temperatures; the temperature difference between these two temperatures is divided into a certain number of degrees depending on the type of temperature scale. We shall discuss three the most important temperature scales that are used today.

In *Fahrenheit scale* several reference points were used. Daniel Gabriel Fahrenheit selected in 1724 as the zero point (0^oF) the temperature of a mixture of water, ice and ammonium chloride. The temperature of the same mixture but without salt was estimated as 30^oF. The temperature of the human body was used as 96^oF. The freezing point of water corresponded to 32 degrees Fahrenheit and the boiling point – to 212 degrees. The interval between the two points was divided into 180 parts – the degrees Fahrenheit (^oF).

Anders Celsius developed in 1742 the *Celsius temperature scale* (or *centigrade tem-perature scale*) which was based on 0^oC for the freezing point of water and 100^oC for the boiling point of water under a pressure of one standard atmosphere. 100 equal divisions between the freezing and boiling points was estimated as the *degrees Celsius* (^oC).

The *Kelvin temperature scale* (K) was proposed by Lord Kelvin (William Thompson) in 1848. This temperature scale has an absolute zero – the lowest possible temperature in the universe below which temperatures do not exist, and the triple point of water (the intersection on a phase diagram where the three-phase equilibrium point consists of ice, liquid, and vapour).

Zero point is defined as 0 K and corresponds to -273.15^oC, while by definition the triple point of water (273.16 K) corresponds to 0.01^oC.

Table 19.2

Conversion formulas for temperatures

From	To	Formula
Kelvin	Celsius	$^oC = K - 273.15$
Celsius	Kelvin	$K =^o C + 273.15$
Kelvin	Fahrenheit	$^oF = 9/5(K - 273.15) + 32$
Fahrenheit	Kelvin	$K = 5/9(^oF - 32) + 273.15$
Fahrenheit	Celsius	$^oC = 5/9(^oF - 32)$
Celsius	Fahrenheit	$^oF = 9/5(^oC + 32)$

19.4. Thermography

The thermal radiation created by a human body belongs, rimarily, to the infra-red spectrum. It carries information on the temperature distribution over the surface of the body, this being used in diagnostics of some diseases.

In healthy humans, the temperature distribution over the surface of the body has a distinct pattern. The presence of inflammatory processes or tumors affects this distribution. The temperature of different areas of the human body depends on blood supply. Disturbed blood supply of the limbs changes their temperature and, hence, the pattern of thermal radiation. Determining the temperature of separate areas of a human body for diagnostic purposes is known as thermography (thermovision). We can say thermovision is diagnostic method that uses human body thermal radiation for determining local temperature. The Stefan-Boltzmann law and the Wien's displacement law make up the base of thermovision, Thermography is put into practice with special devices that detect infrared radiation and convert signal to a visible image.

At present, liquid-crystal indicators are often used to determine the temperature of areas on a human body. The color of the indicators depends on their temperature, i.e. the temperature of the area on the body where they are placed.

Imaging of the temperature field of human body surface is realized by device that is called thermovision camera (thermal imager).

It is known human body temperature is changed by jumps with frequency coinciding with heart rate. Amplitude of this temperature oscillations approximately is 0.5^o with respect to average local temperature.

Stefan-Boltzmann law and Wien law lay at the basis of **thermography**, i.r. diagnostic method that uses heat radiation of human body for determination of local temperature. The considerable advantage of thermography as compared with other methods (for example, methods of radiodiagnosis) is in that at thermographic examination we use intrinsic heat radiation of human body and not external radiation which is often accompanied by heavy dose burden. Determination of temperature of body surface in thermography may be carried out in two ways. The first approach lies in application of liquid-crystal indicators, which optical properties are very sensitive to any temperature fluctuations. Placing such indicators on patient's body we may visually determine local deviation of temperature values. The second approach lies in visualisation of temperature field of body surface using devices called **thermovision cameras (infrared scan-**

ners). The main principle of this approach lies in correlation between the level of signal registered by the device and radiation property of body surface, which, in its turn, depends on temperature according to Stefan-Boltzmann law.

19.4.1. Infrared radiation

Electromagnetic radiation within wavelength range from $\lambda = 0.76\ \mu m$ to $\lambda = 2000\ \mu m$ is called **infrared radiation** (IR). This range is conditionally divided into three zones: close ($\lambda = 0.76 - 2.5\ \mu m$), medium ($\lambda = 2.5 - 50\ \mu m$) and distant ($\lambda = 50 - 2000\ \mu m$).

Therapeutic effect of IR radiation is connected with thermal effect. The best result is received using close zone of IR radiation. IR radiation penetrates the body at the depth of $\sim 2\ cm$, thus heating surface layers in the best way. As a result heat regulation, blood supply and other vital functions improve.

19.4.2. Ultraviolet radiation

Electromagnetic radiation within wavelength range from $\lambda = 400\ nm$ to $\lambda = 10\ nm$ is called **ultra-violet radiation** (UV). **Ultraviolet** (UV) light is electromagnetic radiation with a wavelength shorter than that of visible light, but longer than X-rays. It is named because the spectrum consists of refrangible electromagnetic waves with frequencies higher than those that humans identify as the colour violet.

UV light is typically found as part of the radiation received by the Earth from the Sun. Most humans are aware of the effects of UV through the painful condition of sunburn. The UV spectrum has many other effects, including both beneficial and damaging changes to human health.

This range may be conditionally divided into four zones: A ($\lambda = 400 - 315\ nm$), B ($\lambda = 315 - 280\ nm$), C ($\lambda = 280 - 200\ nm$) and vacuum ($\lambda = 200 - 10\ nm$). The later name is explained by the fact that UV radiation with $\lambda < 200\ nm$ is strongly absorbed in the air, and, hence, to study it in vacuum will be more convenient.

Therapeutic application of UV radiation is connected with its specific biological action that causes photochemical reactions. Zone A is important for creation of pigment, which colours skin. Zone B has antirachitic action, but in great doses it has carcinogenic action. Zone C is applied for bactericidal action.

Beneficial effects

The Earth's atmosphere blocks UV radiation from penetrating through the atmosphere by 98.7%. A positive effect of UVB exposure is that it induces the production of vitamin D in the skin. It has been estimated that tens of thousands of premature deaths occur in the United States annually from a range of cancers due to vitamin D deficiency. Another effect of vitamin D deficiency is osteomalacia (the adult equivalent of rickets), which can result in bone pain, difficulty in weight bearing and sometimes fractures. Other studies show most people get adequate Vitamin D through food and incidental exposure.

Many countries have fortified certain foods with Vitamin D to prevent deficiency. Eating fortified foods or taking a dietary supplement pill is usually preferred to UVB exposure, due to the increased risk of skin cancer from UV radiation.

Too little UVB radiation leads to a lack of Vitamin D. Too much UVB radiation leads to direct DNA damages and sunburn. An appropriate amount of UVB (What is appropriate depends on your skin colour) leads to a limited amount of direct DNA damage. This is recognized and repaired by the body. Then the melanin production is increased which leads to a long lasting tan. This tan occurs with a 2 day lag phase after irradiation, but it is much less harmful and long lasting than the one obtained from UVA.

Ultraviolet radiation has other medical applications, in the treatment of skin conditions such as psoriasis and vitiligo. UVA radiation can be used in conjunction with psoralens (PUVA treatment). UVB radiation is rarely used in conjunction with psoralens. In cases of psoriasis and vitiligo, UV light with wavelength of 311 nm is most effective.

Harmful effects

An overexposure to UVB radiation can cause sunburn and some forms of skin cancer. In humans, prolonged exposure to solar UV radiation may result in acute and chronic health effects on the skin, eye, and immune system. However the most deadly form – malignant melanoma – is mostly caused by theindirect DNA damage (free radicals and oxidative stress). This can be seen from the absence of a UV-signature mutation in 92% of all melanoma.

UVC rays are the highest energy, most dangerous type of ultraviolet light. Little attention has been given to UVC rays in the past since they are filtered out by the atmosphere. However, their use in equipment such as pond sterilization units may pose an exposure risk, if the lamp is switched on outside of its enclosed pond sterilization unit.

19.4.3. Light therapy

Light therapy or phototherapy, classically referred to as heliotherapy – consists either of exposure to daylight or some equivalent form of light as a treatment for *seasonal affective disorder* (SAD), or exposure of the skin to specific wavelengths of light using polychromatic polarised light to treat a skin condition.

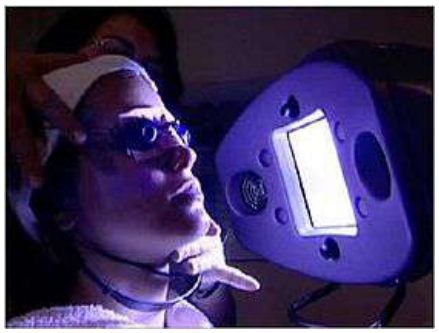

Fig. 19.6. High-intensity blue light (425 nm) used for the attempted treatment of acne.

It is used as a treatment for wintertime seasonal affective disorder and in circadian rhythm disorders, such as delayed sleep phase disorder. There is tentative evidence to support its use to treat non-seasonal psychiatric disorders, in particular major depression and depression in bipolar disorder.

As a treatment for disorders of the skin, the second kind of light therapy is meant to correct psoriasis, acne vulgaris, eczema and neonatal jaundice.

19.5. The Beer-Bouguer-Lambert law. Concepts of extinction (scattering + absorption) and emission.

As radiation struggles to make its way upwards through a stellar atmosphere, it may be weakened by absorption and scattering. The combined effect of absorption and scattering is called extinction. Scattering may simply be by reflection from dust particles. If the radiation interacts with an atom, the atom may be excited to a higher energy level and almost immediately (typically on a time-scale of nanoseconds) the atom drops down to its original level and emits a photon of the same frequency as the one it absorbed. Such a process - temporary absorption followed almost immediately by re-emission without change in wavelength - is probably best described in the present context as scattering. Individual atoms in a stellar atmosphere generally radiate dipole radiation; however, since many randomly oriented atoms take place in the process, the scattering can be regarded as isotropic. If, however, the excited atom collides with another atom before re-emission, the collision may be super-elastic; as the atom falls to a lower state, the energy it gives up, instead of being radiated as a photon, goes to kinetic energy

of the colliding atoms. The radiation has been converted to kinetic energy. This process is absorption.

Extinction and **emission** are two main types of the interaction between an electromagnetic radiation field and a medium (e.g., the atmosphere)

Extinction is a process that decreases the radiant **intensity**, while emission increases it (**extinction** = **attenuation**).

Radiation is **emitted** by **all** bodies that have a temperature above absolute zero (0 K) (called **thermal emission**).

Extinction is due to **absorption** and **scattering**.

Absorption is a process that removes the radiant energy from an electromagnetic field and transfers it to other forms of energy.

Scattering is a process that **does not** remove energy from the radiation field, but may redirect it.

Scattering can be thought of as **absorption** of radiant energy followed by **reemission** back to the electromagnetic field with negligible conversion of energy. Thus, scattering can remove radiant energy of a light beam traveling in one direction, but can be a "source" of radiant energy for the light beams traveling in other directions.

> The fundamental law of extinction is the **Beer-Bouguer-Lambert (Extinction) law**, states that the extinction process is linear in the intensity of radiation and amount radiatively active matter, provided that the physical state (i.e., T, P, composition) is held constant.

Some non-linear processes do occur as will be discussed later in the course.

Consider a small volume ΔV of infinitesimal length ds and area ΔA containing radiatively active matter. The change of intensity along the path ds is proportional to the amount of matter in the path.

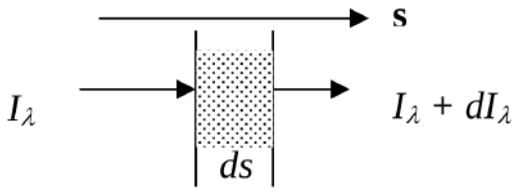

Fig. 19.7

For extinction: $\boxed{dI_\lambda = -\kappa I_\lambda ds}$

For emission: $\boxed{dI_\lambda = \kappa J_\lambda ds}$

where κ is the **volume extinction coefficient** and J_λ is the **source function**.

The **source function** J_λ has emission and scattering contributions or only scattering.

Generally, the **volume extinction coefficient** is a function of position s. **Volume extinction coefficient** is often referred to as the **extinction coefficient**.

Extinction coefficient = absorption coefficient + scattering coefficient $\boxed{\kappa = \alpha + \sigma}$

Extinction coefficient (as well as absorption and scattering coefficients) can be expressed in different forms according to the definition of the amount of matter (e.g., number concentrations, mass concentration, etc.) of matter in the path.

Volume and **mass extinction coefficients** are most often used. Mass extinction coefficient = volume extinction coefficient/density.

If ρ is the density (mass concentration) of a given type of particles (or molecules), then

$$\kappa = \rho k_e$$

$$\sigma = \rho k_s$$

$$\alpha = \rho k_a$$

where the k_e, k_s and k_a are the **mass extinction, scattering, and absorption coefficients**, respectively.

Using the mass extinction coefficient, the **Beer-Bouguer-Lambert (extinction) law** is

$$\boxed{dI_\lambda = -\rho k_e I_\lambda ds}$$

$$\boxed{dI_\lambda = \rho k_e J_\lambda ds}$$

The **extinction cross section** of a given particle (or molecule) is a parameter that measures the attenuation of electromagnetic radiation by this particle (or molecule). In the same fashion, **scattering and absorption cross sections** can be defined.

19.6 Photo-electric effect

Photo-electric effect (**photoeffect**) is called group of the phenomena arising at

interaction of light with substance, consisting in emission of electrons (**external effect**), or in change of electroconductivity of substances, or occurrence of EMF (**internal effect**).

In 1897 in G.Herts and A.G.Stoletov's experiments it has been established that under action of light metals emit electrons. This phenomenon has been named as *photoeffect*.

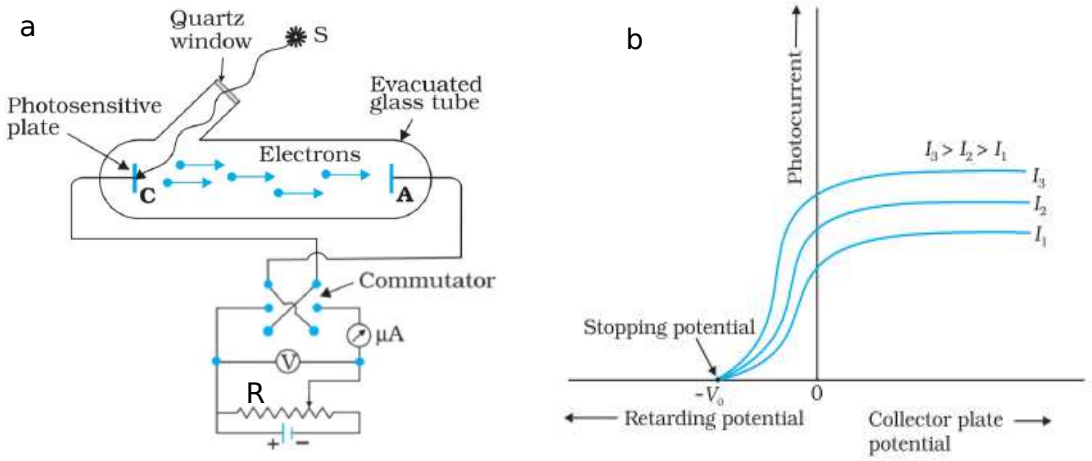

Fig. 19.8. a) external effect; b) graphs of dependence of photocurrent I from the voltage V

Detailed study of photoeffect has been carried out on installation (fig. 19.8a), in which electrodes are established in the glass vacuum chamber with quartz window for penetration of UV-beams. The photocurrent formed by the stream of electrons, knocked out by UV-beams from the cathode was fixed by the galvanometer. The voltage on electrodes changed with help of potentiometer R and was fixed with help of voltmeter V. On fig. 19.8b graphs of dependence of photocurrent I from the voltage are given at different values of light stream Φ. The current in the beginning grows and then remains the constant (photocurrent of saturation). Value of photocurrent of saturation I_H is defined by quantity of electrons n beaten out from the cathode by light per unit of time:

$$I_H = en.$$

Therefore, value of photocurrent of saturation is the measure of photo-electric action of light. If to change polarity of electrodes the electric field will brake moving of electrons, and at some value of $U = U_3$ (detaining voltage) even the fastest electrons do not achieve the anode, the photocurTent will be stopped. From experiment the following **laws of photoeffect** are established:

1. *Force of photocurrent of saturation is proportional to falling light stream*

$I_H = k\Phi$, *where* k *is coefficient of proportionality (photosensivity).*

2. The maximal energy of photoelectrons linearly grows with frequency of light and does not depend on intensity.

The photoeffect can be caused by light (irrespective of its intensity), which frequency is not lower than some minimal frequency, characteristic for the given substance of the cathode, and is known as **red border of photoeffect** ν_{red} *(3-d law of photoeffect).* The external photoeffect in metal is energetically described by **Einstein equation:**

$$h\nu = A + \frac{mv^2}{2},$$

where A is work of output of electron from metal; $h\nu$ is energy of photon; $\frac{mv^2}{2}$ is kinetic energy of electron.

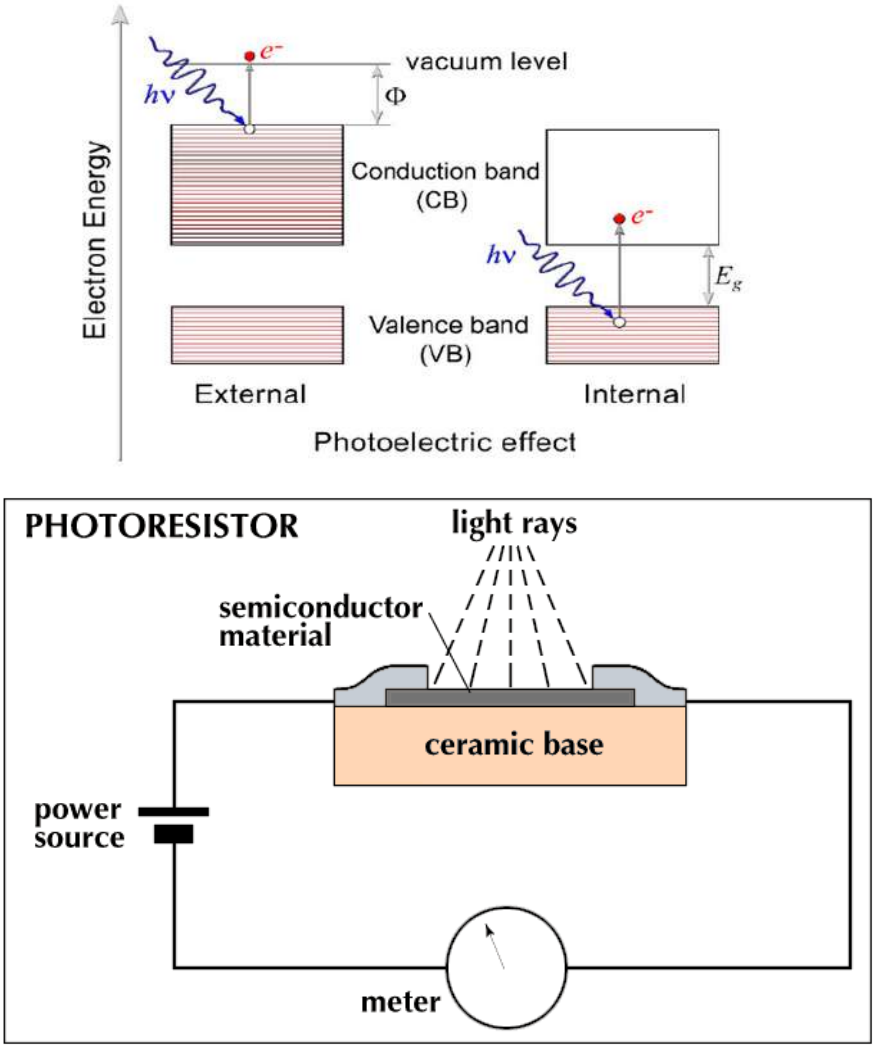

Fig. 19.9. Internal effect versus external.

According to Einstein's equation to ν_{red}, corresponds zero value of kinetic energy. In this case condition of red border of photoeffect is equal $h\nu_{red} = A$,

$\nu_{red} = \dfrac{A}{h}$ or $\lambda_{red} = \dfrac{hc}{A}$ $\left(\dfrac{mv^2}{2} = 0\right)$.

Wavelength λ_{red} and work of output for various metals will be various. These data for various metals are usually resulted in the table.

Internal effect is observed in semiconductors and dielectrics and will consist in increase of concentration of free carriers of charge inside the substance irradiated by light. Thus electroconductivity bodies increases. Due to energy of the absorbed photon connected electron is released and becomes by electron of conductivity. Differently, energy of the photon is spent for transfer of electron from valent zone to the zone of conductivity. All semiconductors are photosensitive, as energy of photon of visible region and even the IR-photon exceeds width of their forbidden zone. The internal photoeffect is easy for finding out on experiment: at switching on of a selenie plate in the circuit of direct current (fig.19.9) its conductivity sharply increases at its illumination.

19.6.1. Practical application of photoeffect

Action of receivers of radiation (**photo cells**) is based on the phenomenon of photoeffect, transforming light signal to electric. Before others the photo cell with using of external photoeffect (fig.19.9) has been created. It will consist of the cathode (source of electrons) and the anode as a loop, disk or a grid. All system is placed into the glass cylinder, from which is pumped out air.

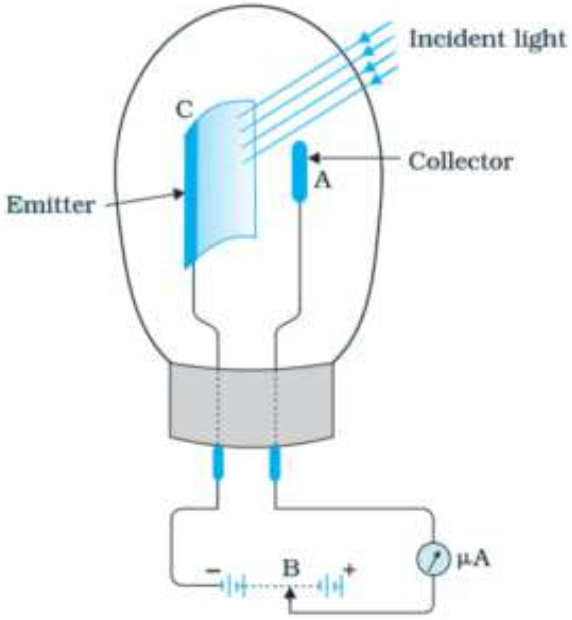

Fig. 19.10. Photo cells

The photocathode can be put on the internal surface of glass cylinder as the layer of metal. The important characteristic of the photo cell is its **sensitivity**, expressed by the ratio of photocurrent to corresponding light stream: $k = i/\Phi$. It reaches value of $k = 100\ \mu A/lm$. Low photosensivity is basic defect of vacuum photo cells.

This defect is eliminated in *photoelectronic multipliers* (PEM), in which except of external photoeffect the phenomenon of secondary electronic issue is used. PEM represents vacuum element (fig.19.11) with a number of intermediate electrodes D_1, D_2, D_3, ... Under action of light electrons

ejecting by cathode C get on dynode D_1 cause secondary emission of electrons (their number at $3-10$ time exceeds number of falling electrons). This process of multiplication repeats at the further hit of electrons on the subsequent electrodes.

The increased stream of electrons collects by anode A and forms in the circuit of loading R the current exceeding the photocurrent from cathode (initial photocurrent) in $10^5 - 10^6$ times. Sensitivity of PEM reaches of 10^3 A/lm. PEM is applied mainly for *measurement of small radiant streams*. It registers superweak *bioluminescence*. Work of the *electron-optical converter (EOC)* is based on external photoeffect (fig.19.12), intended for *transformation of the image from one area of spectrum to anothers*, and also for amplification of brightness of the image.

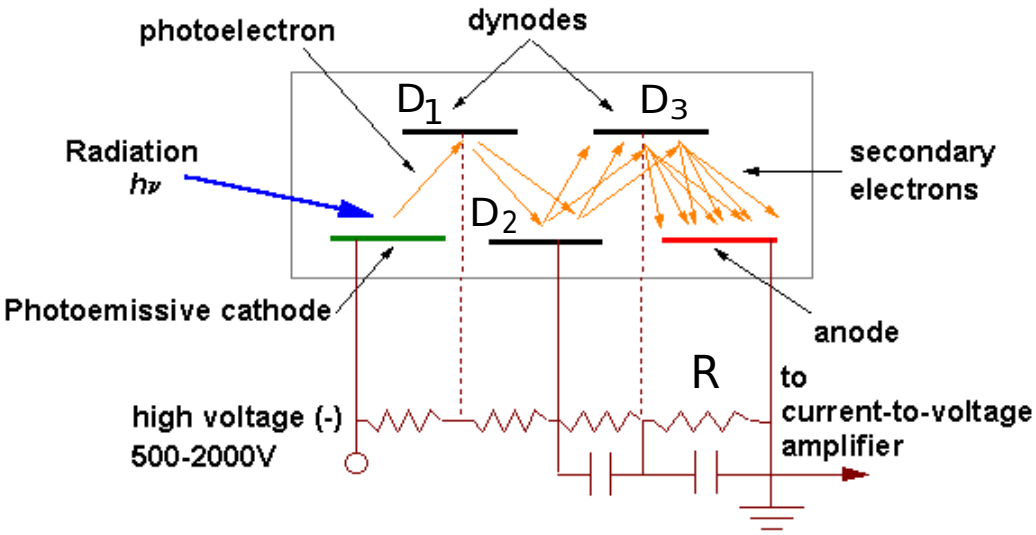

Fig. 19.11. Photoelectronic multipliers.

The light image of object (1) is projected on translucent photocathode C and transformed to the electronic image. Accelerated and focused by electric field of electrodes Э electrons get on luminescent screen L and the electronic image (2), due to cathodeluminescense again will be transformed into light image (3). *EOC are applied for amplification of brightness of the X-ray image. It allows to reduce the doze of irradiation of the person considerably.* EOC is capable to transform IR-radiation in visible region that it is possible to use for thermographical diagnostics of diseases.

The internal photoeffect in non-uniform semiconductors results to occurrence between p and n semiconductors of EMF under action of light. This phenomenon is known as **photogalvanic effect** and is used in valve photo cells (fig.19.13), which will transform light energy to energy of electric current. Valve selenie element will consist of the basic iron plate 3, on which the thin layer of selenium

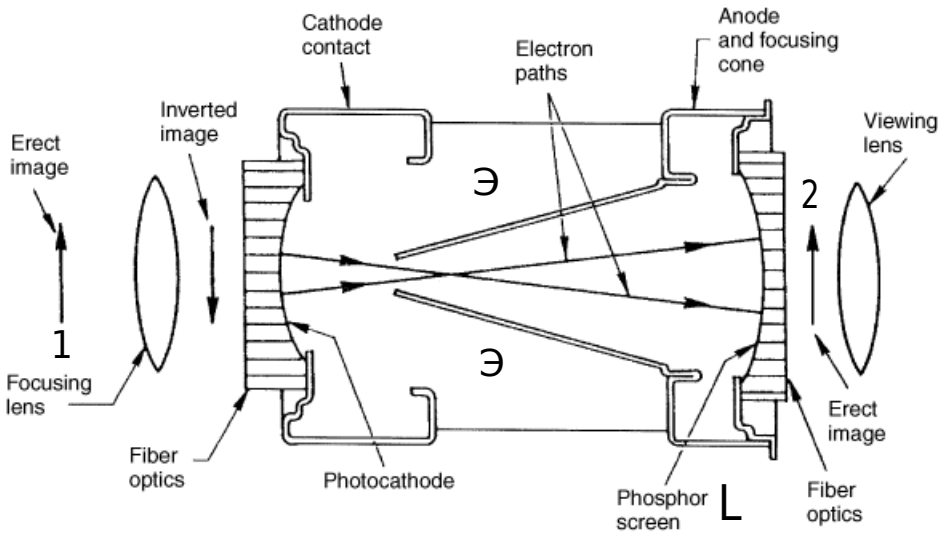

Fig. 19.12. The electron-optical converter.

4 having hole (p) conductivity is rendered. On the surface of selenium is put thin film of gold 1 transparent for light beams.

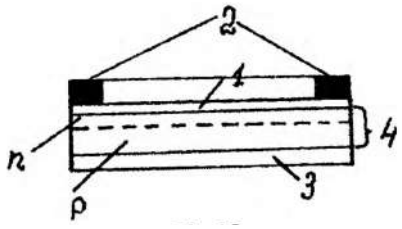

Fig. 19.13. Photogalvanic effect

Atoms of gold diffuse inside of selenium and form the connection having electronic (n) conductivity. Between semiconductors with (p) and (n) conductivity forms a locking layer (dashed fine), which interferes to penetration of electrons to the area with p-conductivity. If on the photo cell to direct a stream of light, photons will beat out electrons from atoms of selenium, the way to which aside of iron plate blocks locking layer and which move aside film of gold, charging it negatively. The layer of selenium with p-conductivity and the iron plate are charged positively. Between gold and iron plates there is a potential difference named the *photo – EMF*. If to connect gold and iron plates by a conductor on the circuit there will be photocurrent. The photocurrent is allocated with help of electrodes: the iron plate 3 and the metal ring 2.

Such photo cells are used in **luxmeters** for measurement of artificial and natural illumination.

Test questions

1. What is called the thermal radiation of bodies? What kind of bodies is such radiation inherent in?

2. Give a definition of such physical characteristics as radiation flux, radiant

emittance of body, spectral radiance and absorption coefficient.

3. Define a black body and explain what the model of this body is. What bodies are called gray?

4. Formulate Kirchhoff's law.

5. Formulate the Stefan – Boltzmann law.

6. Formulate Wien's law of displacement.

7. Explain how the body's heat exchange with the environment occurs environment?

8. What is the effect of light energy on the human body? In what the essence of light therapy?

9. What is called thermography?

10. Show the derivation of the formula expressing Bouguer's law.

Lecture No 20

Luminescence. Compelled radiation

20.1. Kinds of luminescence. Photoluminescence. Stocks rule.

All kinds of self-luminescences except of luminescence of the heated bodies are called **cold luminescence** *or* **luminescence**. *As luminescence understand own luminescence of substance arising under influence of external influence.* Example of luminescence: luminescence at electric discharge in gases, at some chemical processes (rotting of organic substances, oxidation of phosphorus), luminescence of glowworms, sea microorganisms and also some substances under action of UV-radiation. This radiation has the duration *considerably exceeding* period (10^{-15} s) of light waves. The luminescence occurs simultaneously with thermal radiation and lays in the optical range.

Depending on kind of excitation distinguish *some kinds of luminescence.*

The luminescence caused by charged particles is **ionluminescence**; by electrons is **cathodeluminescence** (luminescence of a screen of cathode-ray tube); caused by nuclear radiation is **radioluminescence**; caused by X-ray and γ-radiation is **roentgenluminescence**; caused by photons of visible range and UV-radiation is **photoluminescence**; by electric field is **electroluminescence**, which special case is the luminescence of gases at the electric discharge. The luminescence accompanying exothermal reactions (the reactions with allocation of energy) is called **hemiluminescence**. To it concern **bioluminescence**: it is the luminescence of organisms connected with processes of their viability (fungi, bacteria and insects).

Let's consider in detail the *photoluminescence which is meant as the secondary luminescence of substance under action of ultra violet or short-wave part of visible radiation.* Photoluminescence sometimes simply is named luminescence and subdivide on **fluorescence** (short-term persistence $\tau = 10^{-9} - 10^{-3}$ s) and **phos-**

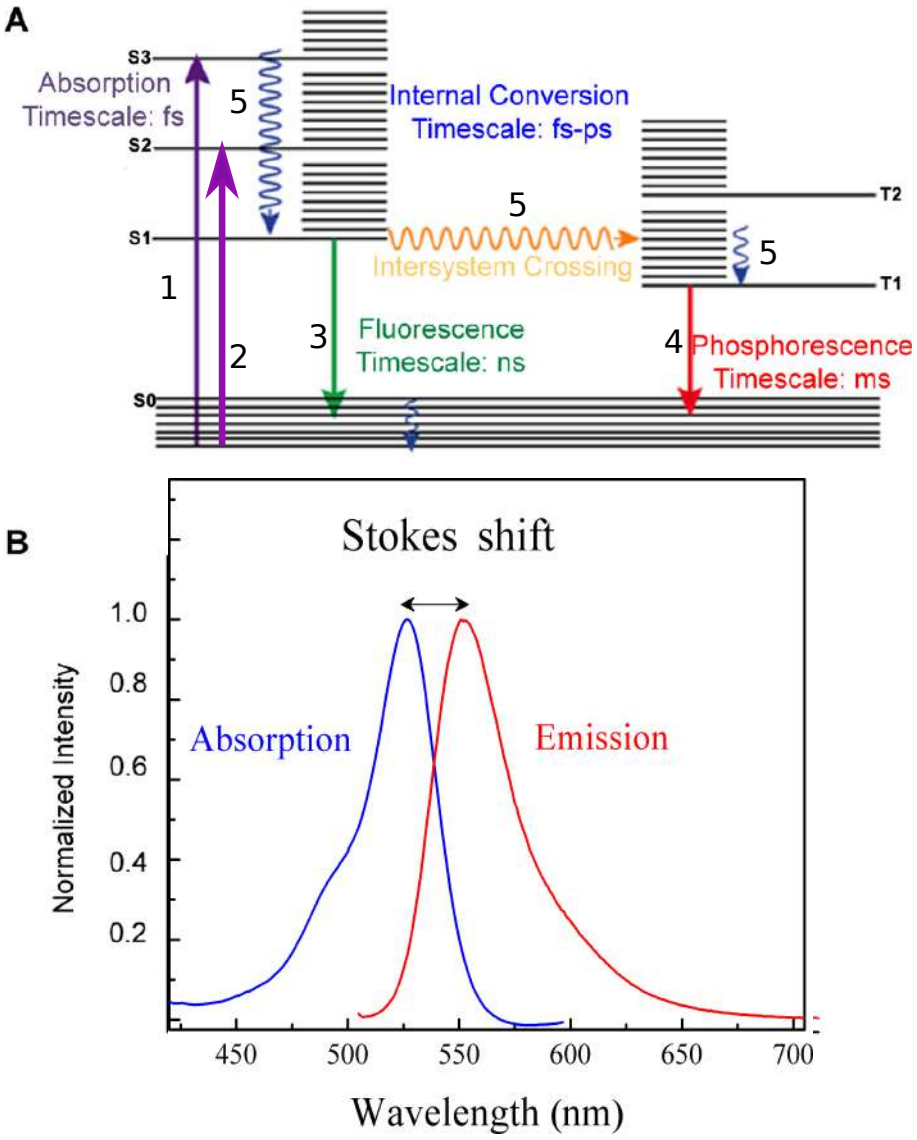

Fig. 20.1. The fluorescence and phosphorescence.

phoreccence (duration of persistence about several seconds and shares of hour). The initial act of photoluminescence is absorption of quantum of light $h\nu$ from the outside and excitation of atom or molecule. On fig.20.1a electronic levels of tirosin and electronic transitions in it are represented. If the molecule absorbs quantum of light, electrons of external shells from basic power level S_0 pass to higher power level, for example, S_2 (1) or S_3 (2). Thus electrons shells remains in the single position (all electrons are coupled, also the total spin moment is equal to zero), though the molecule becomes excited. The value of energy of the absorbed quantum is equal to difference of energy of two levels, between which is carried out electronic transition: $h\nu_{abs} = E_2 - E_0$, i.e. transition of electron from the basic single level on excited single level will correspond to absorption of light. Molecule can have some such excited single levels (S_2, S_3, S_0).

Time of presence of a molecule in the excited condition is value of the order

of $10^{-9} - 10^{-7}$ s. Electronic energy of the excited molecule can be spent as result of course of several processes:

1) it can be transferred other molecule (migration of energy),

2) it can be used for increase of thermal energy of the molecule.

In all these cases electron comes back to the basic level S_0 or on any level, laying below given excited level. Transitions, which are accompanied by transformation of energy into heat, are called *non - radiating* (5).

Besides there can be process of luminescence of the molecule, accompanying by transition of electrons from the excited levels on the basic S_0. Transition of electron from the excited levels on the basic begins with intermediate transition from the top excited levels on the lowermost excited level ($S_3 \to S_1$; $S_2 \to S_1$). Superfluous electronic energy passes thus into heat. Next step is transition from the bottom excited level S_1 on the basic level S_0 (3), thus radiates quantum of luminescence, which energy always is less than energy of the absorbed quantum on value E_{heat}, i. e. $h\nu_f = h\nu_a - E_{heat}$, i. e. $\nu_f < \nu_a$ and $\lambda_f > \lambda_a$. This dependence is known as **law of Stocks**: *wavelength of light, which is emitted at luminescence (fig.20.1b) always more of wavelength of light which has caused it (the rule o f displacement o f Stocks).* Intensity of luminescence is estimated with help of **quantum output of luminescence**: $\varphi = \dfrac{n}{N}$, where n is number of quantums of luminescence; N is number of the absorbed quantums per unit of time.

As the luminescence is always observed at transition from the bottom excited level on the basic, its intensity will not depend on what level has been thrown electron at absorption of light.

The luminescence which is observed at transition of electron from $S_1 \to S_0$ is known as **fluorescence** and is observed only directly during of illumination of object.

At some substances the luminescence is observed after deenergizing light, it is caused by transition from the triple level on the basic $T_1 \to S_0$. The triple level is such level, on which are present two not coupled electrons and their total spin moment can accept one of three values: $+1$; 0; -1. Level T_1 is located a little bit below of S_1, it is named the *forbidden level*, as here electron cannot proceed from level S_0, however, it can get here from excited single level. Its way: $S_0 \to S_1 \to T_1$.

At transition $S_1 \to T_1$ the part of its energy passes into heat. Thus spin of electron changes on opposite, therefore two electrons become not coupled, and the

molecule turns to biradical. Life expectancy of the molecule in the triple condition is from 10^3 s up to several seconds. The luminescence, which is accompanied by transition of electrons from T_1 on S_0 is called the **phosphorescence**. As $E_{T_1} < {}_{S_1}$, $\lambda_{phos} > \lambda_{fluor}$. The example: in the molecule of tirosin to transition $S_0 \to S_3$ (1) corresponds maximum in the spectrum of absorption on wavelength of $\lambda_m = 217$ nm.

To transition $S_0 \to S_2$ (2) corresponds maximum of $\lambda_m = 275$ nanometer. The maximum in spectrum of fluorescence is observed at $\lambda_m = 304$ nanometer (3). To phosphorescence corresponds transition $T_1 \to S_0$ (4), thus the quantum of $\lambda_m = 387$ nanometer is radiated.

20.2 Photoluminescent qualitative and quantitative analysis of biological systems

Phenomenon of luminescence is the basis of the method of detection and determination of the maintenance of chemical components in a mix. This method is known as **luminescent analysis**. Presence of any component (*the qualitative analysis*) determine on *colouring* of luminescent radiation, as to the maximum of spectrum of luminescence corresponds the certain color. Quantity of substance (the *quantitative analysis*) determine on *intensity* of luminescent radiation. At the luminescent analysis for excitation of molecules of substance UV-radiation is used more often.

The luminescent analysis is applied in the most various branches of science and practice. Distinctive feature of the luminescent analysis is the opportunity to find out presence of *insignificant small quantity* of substance (up to 10^{-9} g). The most part of organic connections (acid, dyes) give the characteristic luminescence at absorption of UV-radiation. For example, nicotine gives the dark-violet luminescence. The luminescent analysis is sensitive, it does not demand division of a mix, it can be carried out in biological medium, tissues and other multicomponent systems. On the basis of luminescence in sanitary-and-hygienic practice is applied the method of quality check and sorting of foodstuff (it is used for detection of initial stage of damage of products), sorting and quality check of pharmacological means, vegetative fibre (tissues), skin, detection in them of substitutes or falsifications. The photoluminescence is given tissues of alive organism, especially nails, teeth, non-pigmental (gray-haired) hair, sclera, cornea and especially crystalline

lens of eye and other tissues. The luminescent analysis is used for the control of cleanliness of reactants and water.

In criminalistics and forensic medicine the irradiation by UV-radiation allows to find out invisible traces of blood and the luminescence of blood of the person differs from luminescence of blood of animals and birds.

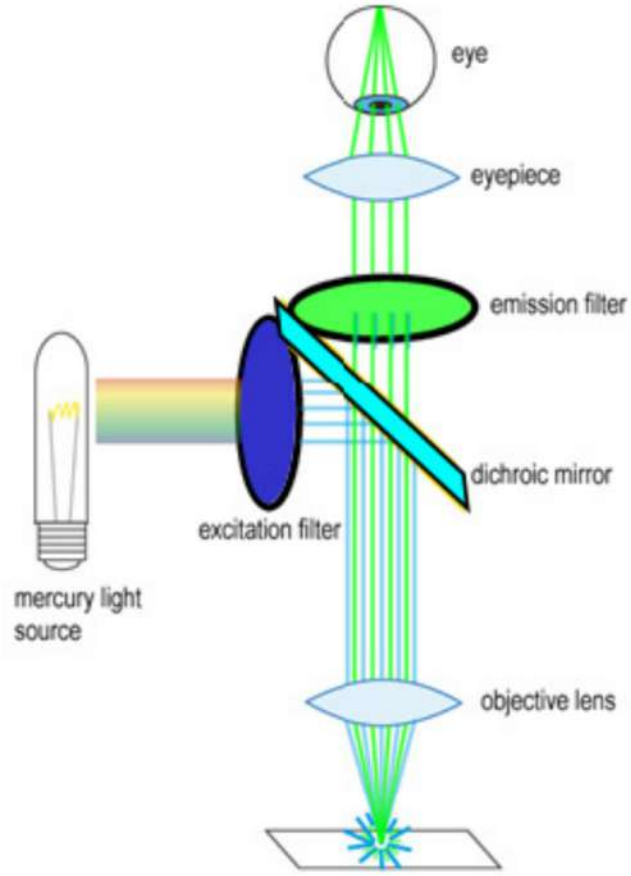

Fig. 20.2. The luminescent (fluorescent) analysis.

On color of luminescence distinguish alive and dead cells. Presence of adrenaline in blood of the person is determined on its characteristic green and yellow luminescence. The luminescent analysis is applied at diagnostics of diseases, especially skin and other illnesses. So struck by fungus hair, scale of skin under the UV-irradiation give brightly – green luminescence. In many cases as diagnostic reception is used introduction in organism of luminescent paints, which are adsorbed in some tissues. These tissues then investigate under action of UV-radiation. For example, into the vein of a person enter solution of fluorescil and in a few seconds observe the bright green of luminescence of lips and eyes raised by ultraviolet. By this method determines blood circulation in the field of the body with the lowered blood circulation. Permeability of capillaries can be determined, entering hypodermically luminescent painting substances. On fig.20.2 the circuit of the luminescent (fluorescent) analysis is shown. UV-radiation from mercury lamp (ML) goes on the object and raises its luminescence. Light of luminescence acts to receiver: an eye, a photo cell, photographic plate, photo multiplier, where it is registered. In order to visible light of a source was not imposed on light of luminescence is applied the optical filter, passing to object only UV-beams invisible for eye.

The luminescent analysis can be subdivided on the macroanalysis and the microanalysis. In the second case supervision is carried out by means of a microscope. In luminescent microscopy the preparations, capable to luminesce are

studied in microscope at UV-illumination with corresponding optical filters. By the form of luminescence of the micropreparations prepared from food stuffs, it is possible to distinguish kinds of activators of infectious diseases: tuberculosis, salmonellosis, Siberian ulcer.

It is necessary to note, that if the quantum output of luminescence is more of 1%, than such connections are easily found out by luminescent method. High quantum output has vitamins A, B_6, E, many medicinal substances. Cancerogenic hydrocarbons in air of cities, smoke of cigarettes, etc. are easily found out by luminescent method.

Some connections which are not having own fluorescence, after special chemical processing give products with high quantum output. By this method can be determined morphine, heroin and other drugs, vitamins C, D, B_{12} and others.

20.3 Induced radiation of atoms

Searches of management by radiation of atoms or molecules for reception of powerful streams of coherent radiation have resulted in creation of **masers** (or molecular amplifiers) and then **lasers** (**L**ight **A**mplification by **S**timulated **E**mission of **R**adiation). These questions are the basic in quantum electronics, which studies methods of amplification and generation of electromagnetic oscillations with use of the compelled radiation of quantum systems.

Let's familiarize with some phenomena underlying of quantum electronics. A. Einstein has proved, that except for two phenomena (absorption and emission) for atom there is one more, it is the **compelled or induced radiation**, which essence consists in the following. Photon of light, flying by the excited atom, transforms it to not excited atom (if energy of the photon coincides with energy of the excited atom), which radiates new photon. As result of the compelled quantum transition, from the atom two identical photons will be distributed: one is initial, external, and the second is secondary. Two photons, flying past by other excited atoms will transfer them also to normal condition with radiation of two photons fig.20.3.

The number of the compelled transitions, accomplished in second, will depend on number of the photons getting in substance. Besides the compelled transition will be define by filling or population of corresponding power levels. At such radiation there is the avalanche increase in number of photons, i.e. amplification of light. Such radiation also is called **induced**. *Induced radiation identically*

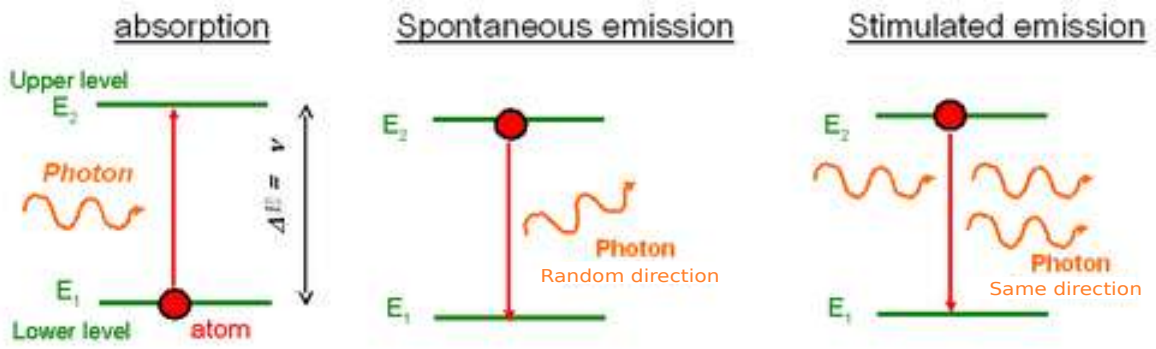

Fig. 20.3. The compelled or induced radiation.

to falling radiation in every respect including on phase, therefore it is possible to speak about coherent radiation (coherent amplification). Existence of induced radiation has been predicted by Einstein theoretically. It should be checked up experimentally.

In atom "*density of population*" (number of not exited atoms) of the bottom levels according to distribution of Boltzmann is much more than top shells. The secondary photons, arising as result of induced radiation and also many photons of external influence will be absorbed by the atoms, located at lower levels. As result absorption will be more, than radiation and amplification of light will not take place. For amplification of light it is necessary except for external influence to pick up such **active medium**, in which the number of the excited atoms would be **more** number of not excited atoms, i.e. distribution in atom of electrons should be *opposite to Boltzmann distribution* (**inversion** of density of population). As active medium can be used plasma, some gases and their mixes, crystal bodies, glasses, liquids, many semi-conductor materials. On measure of propagation of light in such medium intensity of light will grow.

20.4 Optical quantum generators (lasers)

The phenomenon of compelled radiation is used in optical quantum generators (lasers). First generator in the range of the MICROWAVE has been designed in 1955 independently from each other by the Soviet scientists N. Basov, A. Prohorov and by the American scientist I. Towns (they have been awarded for this work of the Nobel Prize). In 1969 the first generator of the visible range has been created with the ruby as working substance.

Let's consider the principle of reception of induced radiation by the example of

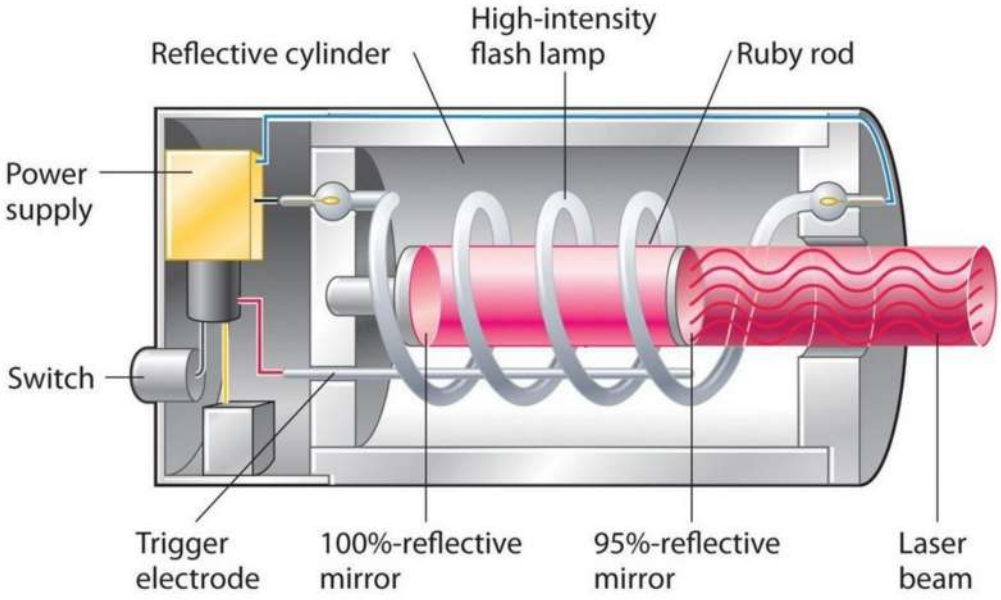

Fig. 20.4. The ruby rod laser.

the ruby laser (fig.20.4), which body is the ruby, it is crystal of oxcide of aluminium Al_2O_3 with impurity of trivalent ions of chromium Cr^{3+} $(0.03 - 0.05\%)$.

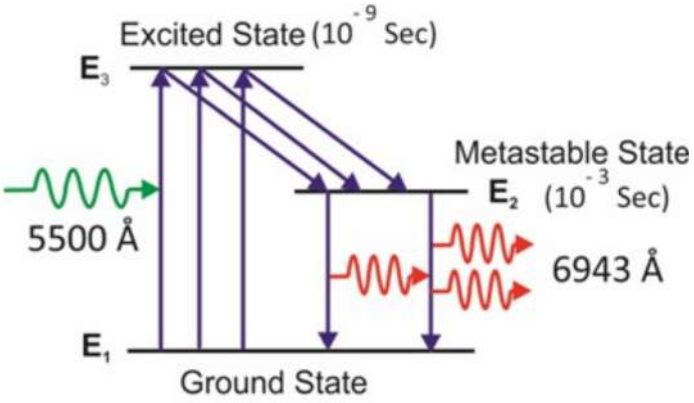

Fig. 20.5. The energy level diagram of ruby laser.

As external influence, or so-called **pump**, the Xe pulse lamp spirally located around of ruby core is used. The plasma arising as the result of the discharge in the pulse lamp radiates powerful stream of light, which acts in depth of the ruby core. From all stream of light only green beams ($\lambda = 560$ nanometers) are useful. They raise atoms of Cr (fig.20.5), throwing them from the level 1 on the level 3. At this level many atoms of Cr for a long time are not be located and pass to lower level 2 located close to level 3. This transition is non-radiating (**thermal radiation**). As the result of such transition the temperature of the crystal lattice of ruby raises. The excited atoms can pass from level 1 to level 3 in time of $t = 10^{-6}$ s; from 3 to 2 – $(t = 10^{-8}$ s); from 2 to 1 – $(t = 10^{-3}$ s).

Apparently, the biggest time is required for transition of atoms from the level 2 on the level 1, therefore level 2 will be the most filling with the excited atoms. This level is called the *metastable* (unstable or temporarily steady). If the photon

of external influence flies by the excited atom, which is taking place at level 2, the atom will pass on the level 1 having given photon of red light ($\lambda = 694.3$ nanometers). There is the coherent induced radiation.

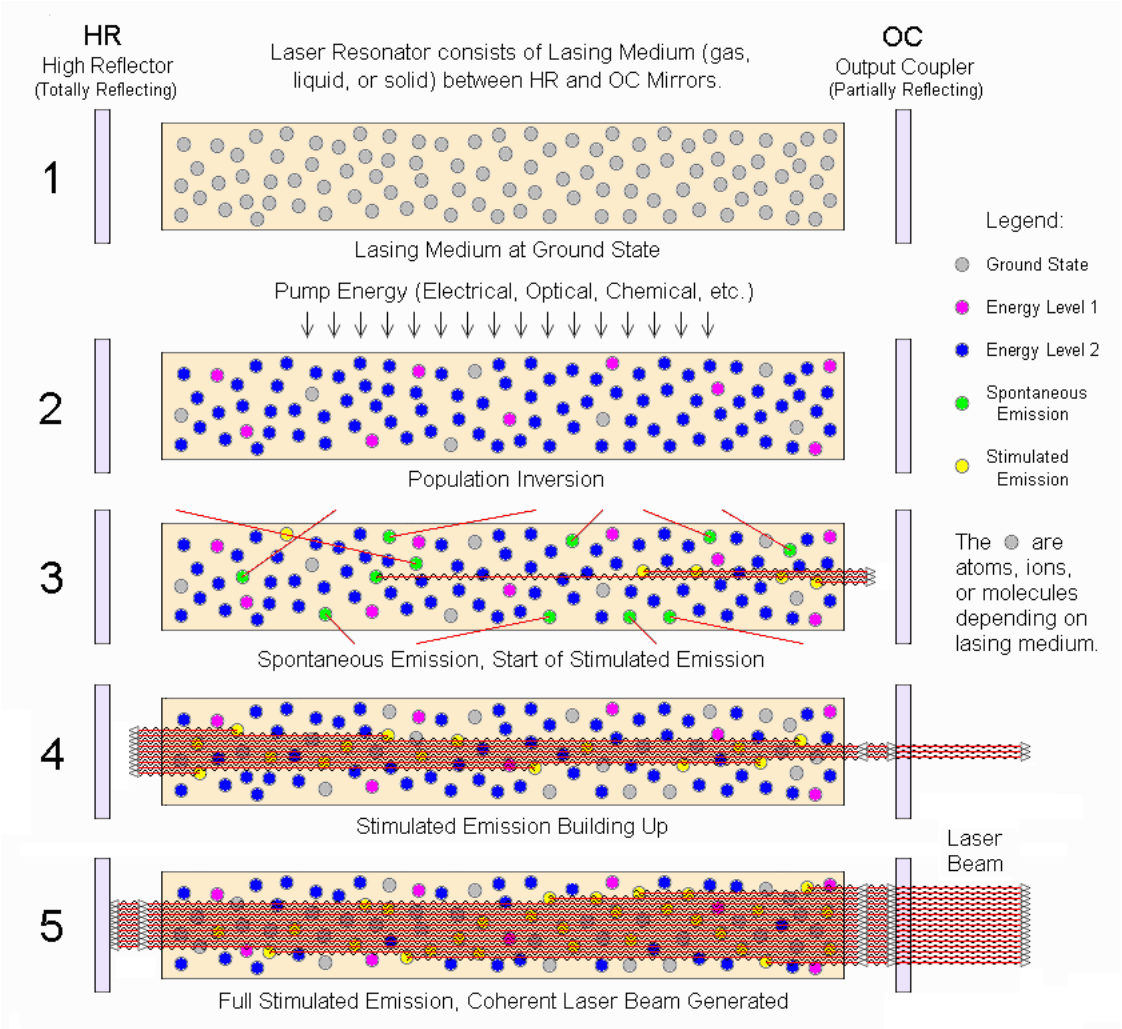

Fig. 20.6. The basic laser operation.

The crystal of the ruby has the lengthened cylindrical form, with strictly parallel ground end faces (represents **the mirror resonator**). The forward end face of it is translucent and back is not transparent (fig.20.6). The length of the ruby core is limited. Using of the core by length more than 30 cm is not obviously possible, since becomes complicated pump of atoms and focussing of radiation. Therefore for increase of the way of photons them force to be reflected repeatedly from mirror face surfaces. The stream of photons moving in parallel of axis of the crystal, leaves through the translucent end face and is focused by lens and goes on a target as sharply directed coherent beam. The optical quantum generator on ruby works in the *pulse* mode. Energy of generation during one pulse of pump reaches of 1000 J.

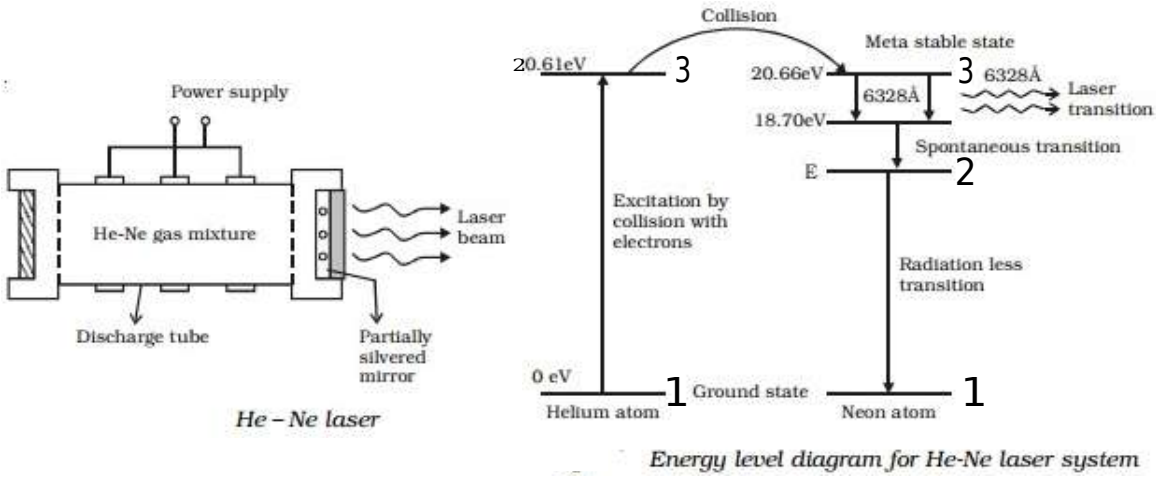

Fig. 20.7. The He-Ne laser.

Alongside with crystal lasers the wide circulation was received gas lasers (fig.20.7), in which as the active medium a gas is used. Advantage of such lasers is the *continuity of mode of radiation.* The first gas laser represented the quartz tube filled with the mix of gases of helium and neon. Excitation of gas was carried out with help of *high-frequency (HF) generator* with frequency in some tens megahertz (electrodes have been built in the tube). In it atoms of neon were radiating. Atoms of helium plays auxiliary role. On fig.20.7 the simplified circuit of power levels of atoms of helium and neon is represented. At the electric discharge the part of Ne atoms from the basic level 1 passes to the excited level 3. Time of life at this level is not enough for the pure neon and atoms pass to levels 1 and 2. For creation of inversion of "*density of population*" it is necessary to increase "density of population" of the level 3 and to reduce at the level 2. Entering into the mix of helium creates such conditions. The *first excited level of helium coincides with the level* 3 *of neon.* The excited atoms of helium at not elastic impacts with not excited Ne atoms transfer them to the excited condition and come back to initial condition. As the result it is possible to achieve primary settling of the top levels of the working gas – neon. The top levels (2, 3) have complex structure, they will consist of set of sublevels. *Therefore helium – neon lasers can work on many wavelengths* in the field of visible and infra-red radiation. So, the red helium - neon laser is radiated (transition $3 \to 2$) wavelength of 632.8 nanometers. Because at unitary passage of a beam in the active mix the beam amplifies insignificantly, it is used external reflecting plates as resonator mirrors. Conditions for self-excitation and supports of generation are created. Reflecting plates can settle down inside the gas laser also. If end faces of a tube look like the glass plates located under Bruster's angle, the leaving laser beam will be not

only *strongly monochromatic and narrowly directed, but also polarized.*

20.4.1. Basic properties of laser radiation, biophysical mechanism of its action, application in biology and medicine

Lasers for short term since time of their creation have found wide application in biology and medicine. Application of lasers is based on **properties of it radiation**: *strong monochromaticity ($\Delta\lambda \approx 0.01$ nanometers), coherency, narrow orientation (the laser beam has property of small divergence), power consumption.* Generally divergence of a beam of the optical quantum generator is defined by the phenomenon of diffraction and depends on diameter of the core of active medium: $\theta = \dfrac{1.22\lambda}{D}$, where θ is angular divergence of beam (in radians); λ is wavelength of radiation; D is diameter of the core.

High coherency of laser beam has allowed tocarry out essentially new method of photographing: reception of the three-dimensional image, which has been named **holographic** (from the Greek word *holos* = *whole*). Coherency, the narrow orientation and high concentration of energy of the laser allow to use it in the different areas of science and technics.

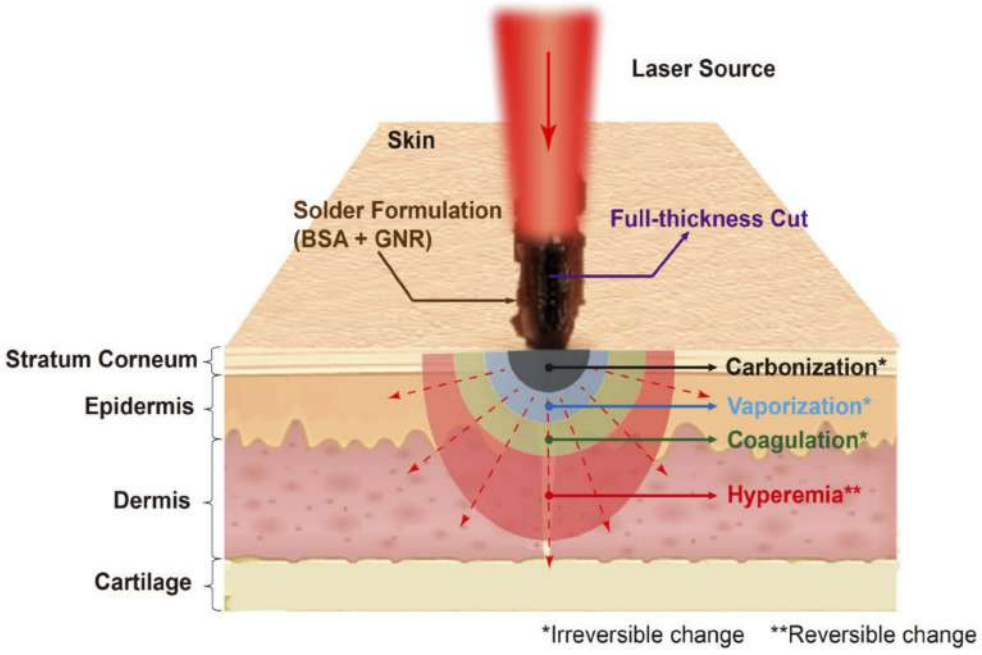

Fig. 20.8. The deep tissue laser therapy.

The mentioned above properties of laser radiation enable to focus it on rather fine biological structures and to use the laser as the research and microsurgical tool at *cellular level.* The big range of intensity of radiation allow to change character

of influence on biological objects from stimulating and therapeutic ($10^{-3}\ Wt/cm^2$) up to explosive, accompanying thermal (coagulation), electromagnetic bothacoustic processes and ionization ($10^7\ Wt/cm^2$), see fig.20.8 and fig.20.9.

The basic scopes of lasers in medicine are the **surgery, ophthalmology, oncology, therapy**. In surgery are applied CO_2-lasers with capacity of $30-100$ Wt, working in the continuous mode. Properties of laser beam to destroy the biological tissues combined with coagulation of tissue, allows to carry out some bloodless sections. The laser scalpel before a traditional scalpel has a number of advantages. The basic problems of surgery are the pain, bleeding and sterility. These problems are solved at use of the laser very simply: laser radiation, as against a usual scalpel cannot bring in an infection, it sterilizes dissected tissues, even if they are already infected with a suppuration; losses of blood do not occur, as blood vessels instantly coop up by the clotted blood. It is essential, that the laser scalpel does not render on a tissue of mechanical pressure that reduces sensation of pain. Besides with the help of modern endoscopes and flexible optical paths (fiber optics), laser radiation can be entered into internal cavities, due to what there is possible stop of internal bleeding and evaporation of suppurations without opening bodies.

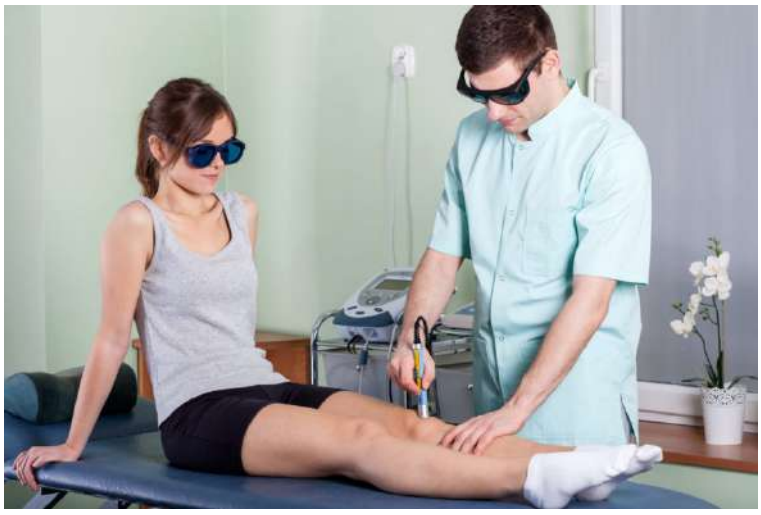

Fig. 20.9. The laser therapy.

In *ophthalmology* are used pulse ruby lasers (duration of pulses of $30-70$ nanoseconds; $E = 0.1 - 0.3$ J), which allow to carry out a number of difficult operations without infringement of integrity of eye: treatment of detachment of retina, welding it to the vascular environment; treatment of the glaucoma by means of piercing of aperture ($d = 50 - 100$ nm) by laser beam for outflow of intraocular liquid; for treatment of some kinds of cataracts.

Laser radiation is used and for destruction of cells of *malignant tumours and ulcers*. At destruction of malignant tumours is used property of non-uniform absorption of laser pulse radiation by different tissues, histologic structures or cells. For example, some people pigmental tumours absorb laser radiation much

more intensively than environmental tissues. Thus in microscopic volumes of tissue is immediately allocated heat with formation of the shock wave extending in the liquid medium with speed about of 1500 km/s. At use of the lasers working in the continuous mode, the temperature raises up to $100^o\ C$. For influence on a tumour is used the focused laser radiation ($d = 1.5 - 3$ mm) on the surface of object, thus $I = 200 - 900\ Wt/cm^2$. It is established, that laser radiation has a number of advantages before used for treatment of skin cancer of X-rays therapy, in particular, number of sessions of irradiation (up to 4 on course of treatment) is essentially removed and in expenses some times decrease. With the help of less intensive radiation it is possible to suppress growth of cancer cells (laser therapy).

Test questions

1. What are the physics of the laser oscillation? What does "inverted population" mean?

2. Explain the application of the main OQG elements (gas discharge tube, cathode, anode, mirrors).

3. What are the properties of a laser beam? Which design features they depend on?

4. Why does the laser beam have a small divergence angle?

5. What is the reason for high density of laser radiation?

6. What is the cause of the polarization of gas OQG radiation?

7. Formulate the Malus law, the Brewster condition.

8. What is the principal difference between Fresnel diffraction and Fraunhofer diffraction?

9. What is a diffraction grating? What is it designed for?

10. Tell us about the use of laser radiation in medicine.

Lecture No 21

X-rays radiation

21.1. Braking and characteristic X-ray radiation, basic properties and characteristics

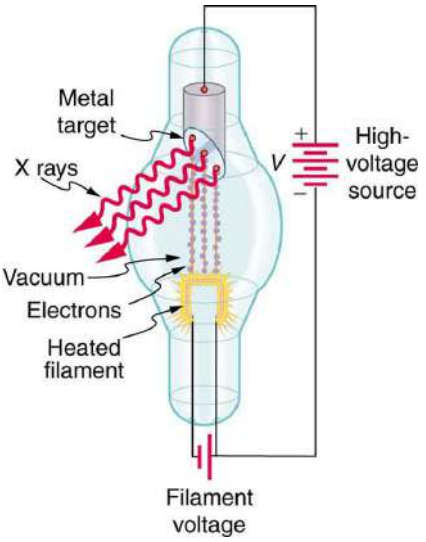

Fig. 21.1. The X-ray radiation.

In 1895 German scientist W. C. Roentgen for the first time has found out luminescence of the fluorescent screen, which has been caused by the radiation invisible to an eye going from the site of glass of the discharge tube, located opposite to the cathode (fig.21.1). This kind of radiation had ability to pass through substance, impenetrable for visible light. Roentgen has named it X-beams and has established the basic properties, allowing to apply X-rays in different branches of science and technics, including medicine.

X-ray radiation *is called radiation of wavelength* $80 \cdot 10^{-5}$ *nm.* Long-wave X-ray radiation blocks short-wave UV-radiation, short-wave is blocked long-wave γ-radiation. In medicine is used X-ray radiation of wavelength from 10 up to 0.005 nanometers. X-ray radiation is invisible for eye, therefore all supervision with it are made with help of fluorescing screens or films, as it causes luminescence and renders photochemical action. On the way of excitation X-ray radiation is subdivided on braking and characteristic radiation.

Braking X-ray radiation is caused by braking quickly moving electrons by electric field of atom (of nucleus and electrons) of substance through which they fly. It is possible to explain the mechanism of this radiation. Any moving charge represents the current around of which the magnetic field is created, which induction depends on speed of electron. *At braking of electron its magnetic induction*

decreases and according to theory of Maxwell there is an electromagnetic wave.

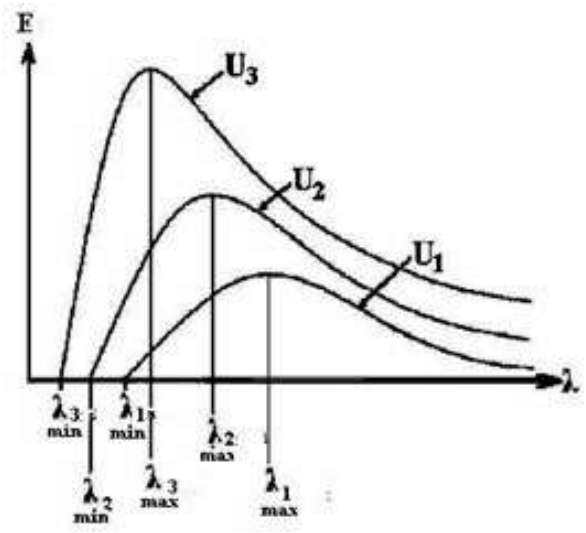

Fig. 21.2. The spectrum of brake X-ray radiation is shown for different voltage $U_1 < U_2 < U_3$

At braking of electrons only *part* of energy is spent on creation of photon of X-ray radiation, other part is spent for heating of the anode. Frequency (wavelength) of photon depends on initial kinetic energy of electron and intensity of its braking. Even if initial kinetic energy of electrons is identical in substance, condition of braking will be various, therefore emitted photons will have the diversified energy and hence wavelength, i.e. *spectrum of X-ray radiation will be continuous.* On fig.21.2 the spectrum of brake X-ray radiation is shown for different voltage $U_1 < U_2 < U_3$, where E is energy of photon of brake X-ray radiation.

In each spectrum the most short-wave radiation arises, when all energy got by electron in the accelerating field completely passes into energy of the photon:

$$eU = h\nu_k = \frac{hc}{\lambda_k}$$

If U to express in kV and to take into account the parity between other values, the formula looks like: $\lambda_k = 1.24/U$ (nm) or $\lambda_k = 1.24/U$ $\mathring{A}$ ($1\mathring{A} = 10^{-10}$ m).

From above mentioned graphs it is possible to establish that wavelength λ_{max} on which it is necessary the *maximum of radiation energy* is in the constant parity with boundary of wavelength λ_k.

$$\lambda_m \approx \frac{3}{2}\lambda_k \approx \frac{1.86}{U} \quad (nm)$$

The wavelength characterizes energy of a photon from which depends penetrating ability of radiation at its interaction with substance.

Short-wave X-ray radiation usually has the big penetrating ability and is known as *rigid* and long-wave as *soft.* Apparently from the above mentioned formula the wavelength on which is necessary the maximum of energy of radiation is inversely proportional to the voltage between the anode and the cathode of the

tube. Increasing the voltage on the anode of X-ray tube we change spectral structure of radiation and increase its rigidity.

At change of glow voltage (the temperature of reheat of the cathode changes) changes the number of electrons emitted by the cathode per unit of time or accordingly the current in the circuit of the anode of tube. Thus capacity of radiation changes proportionally to the first degree of force of current. The spectral structure of radiation will not change.

The general stream (capacity) of radiation Φ, distribution of energy on lengths of waves and also the border of the spectrum on the part of short wavelengths depends on the following three reasons: from voltage U, accelerating electrons and enclosed between the anode and the cathode of a tube; from number of electrons, participating in formation of radiation, i.e. from force of current of reheat of tube; from nuclear number Z of substance of the anode in which there is braking of electron.

The stream of brake X-ray radiation is calculated under the formula: $\Phi = KIU^2Z$, Z is serial number of atom of substance of anode (nuclear number), K is an empirical constant.

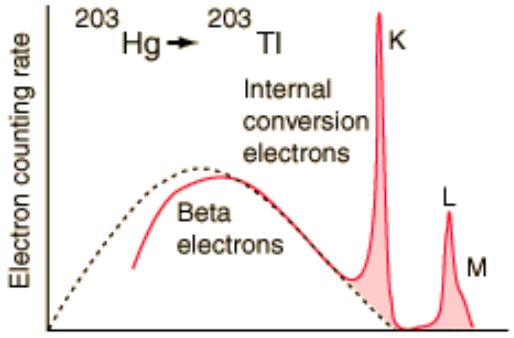

Fig. 21.3. The characteristic X-ray radiation

Increasing the voltage on the X-ray tube, it is possible to notice on the background of continuous brake X-ray radiation occurrence of separate lines (the line spectrum) that corresponds to **characteristic X-ray radiation** (fig.21.3). It arises at transition of electrons between internal shells of atom in substance (shells K, L, M). Line character of spectrum of characteristic radiation arises because accelerated electrons will penetrate deep into atoms and from their internal layers beat out electrons for limits of atom. To empty seats pass electrons (fig.21.4) from the top layers, therefore photons of X-ray radiation with the frequency corresponding to the difference of levels of energy transition are radiated. Lines in the spectrum of characteristic radiation are united in the series corresponding to transitions of electrons from more high levels to the levels K, L, M.

External influence as result of which electron is beaten out from internal layers should be strong enough. As against optical spectra, characteristic X-ray spectra of different atoms are the same. Uniformity of these spectra is caused by that

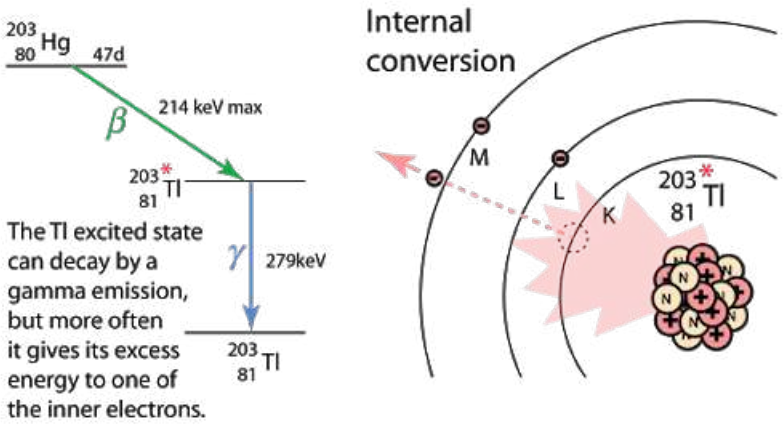

Fig. 21.4. The transitions of electrons between internal shells of atom.

internal layers at different atoms are identical and differ only energetically, since power influence on the part of nucleus increases on measure of increase of the serial number of element. It results to that characteristic spectra are shifted aside the big frequencies with increase of the charge of nucleus. Such dependence is known as *Moseley's law*: $\sqrt{\nu} = A(Z - B)$, where A and B are constants; Z is serial number of the element.

There is one more difference between X-ray and optical spectra. The characteristic spectrum of atom does not depend on the chemical compound into which the atom enters. So, for example, the X-ray spectrum of atom of oxygen is identical for O, O_2, H_2O, while optical spectra of these connections are essentially different. This feature of X-ray spectra of atoms also has formed the basis for the name of **"characteristic"**.

Characteristic radiation arises always when there are empty seats in internal layers of atom irrespective of the reasons, which have caused it. For example, it accompanies with one of the kinds of radioactive disintegration, which consists in capture by the nucleus of electron from internal layer.

21.2 Working principle of X-ray tubes and the elementary X-ray apparatus

The most widespread source of X-ray radiation is the X-ray tube: two-electrode vacuum device (Coolidge tube) (fig.21.5). It represents a glass cylinder ($p = 10^{-6} - 10^{-7}$ $mm\ Hg$) with two electrodes: the anode and the cathode between which the high voltage is created.

Heated cathode emits electrons. The anode frequently is named *anticathode*.

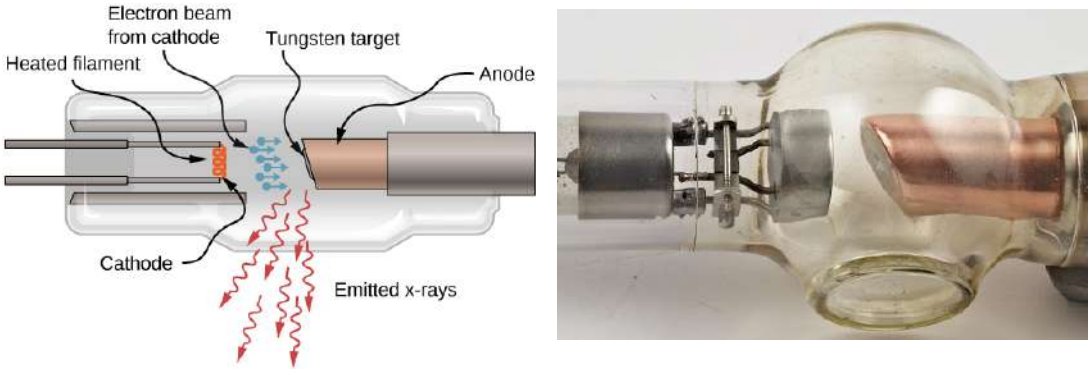

Fig. 21.5. The Coolidge tube.

It has inclined surface (fig.21.6) in order to direct arising X-ray radiation under the angle to the axis of tube.

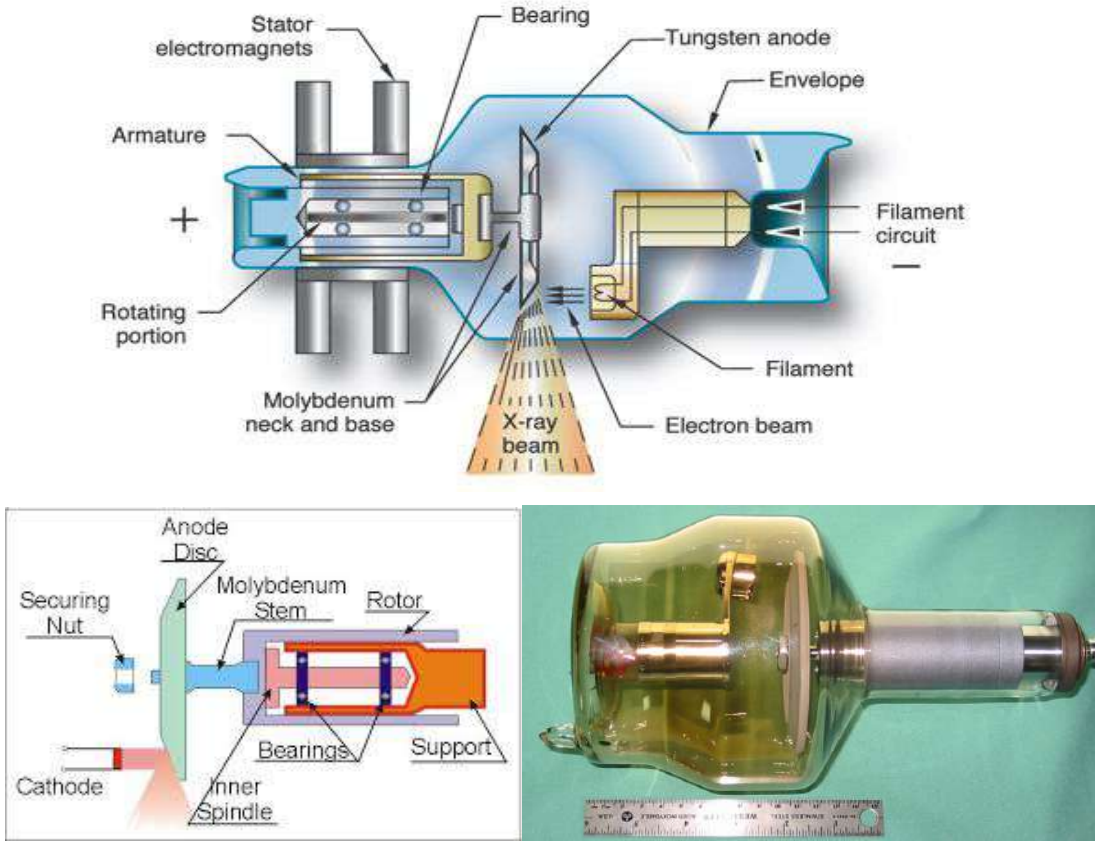

Fig. 21.6. Tube with the rotating anode.

The anode is made from metal with good heat conductivity (copper) for the heat removal formed at impact of electrons. At the oblique end face of the anode there is plate (fig.21.5) from refractory metal (tungsten) with the high nuclear number that named the *mirror of the anode*. On occasion the anode specially cool by water or oil. For diagnostic tubes is important point source of X-rays, that is possible to achieve having focused electrons on one place of the anode. Therefore structurally it is necessary to take into account two opposite problems:

on the one hand electrons should get on one place of the anode, on the other hand, to not suppose of overheating, it is desirable distribution of electrons on different sites of the anode. In this connection some X-ray tubes are made with the rotating anode (fig. 21.6).

In a tube of any construction electrons, accelerated by the voltage between the anode and the cathode get on the mirror of the anode and will penetrate deep into substance, cooperate with atoms and are braked by field of atoms. Thus there is the brake X-ray radiation. Simultaneously with brake the small amount of characteristic radiation is formed. *Only* $1-2\%$ *of electrons getting on the anode is cause of the brake radiation and other part is cause of thermal effect.* The part of the tungstic mirror on which falls basic part of electrons is known as focus of the tube.

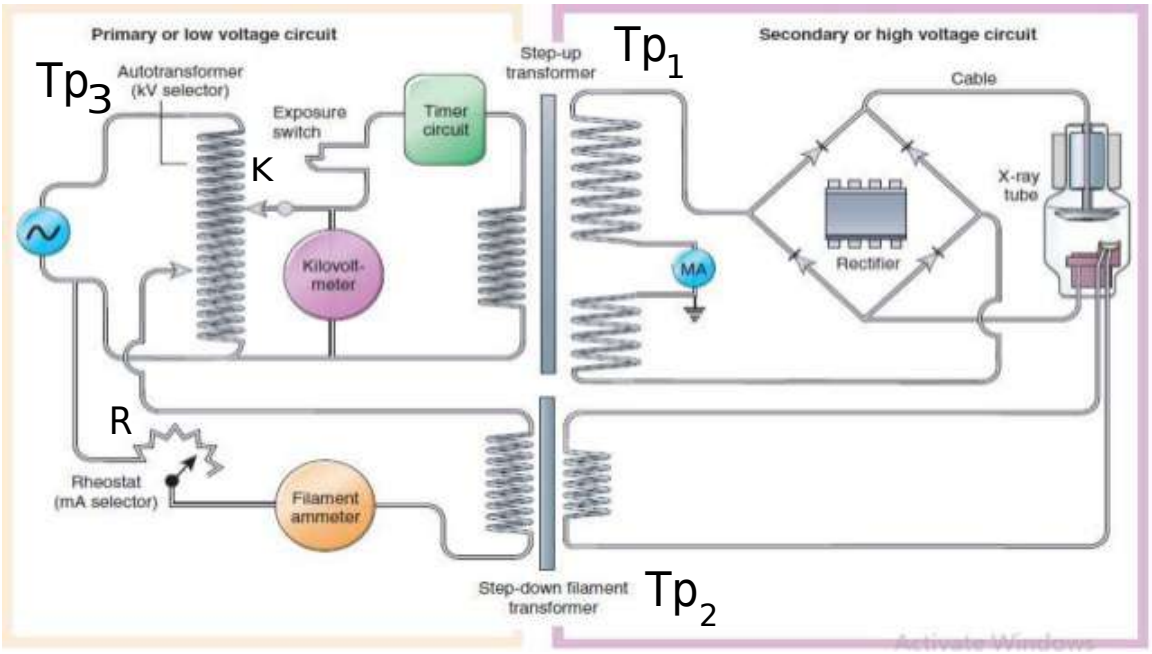

Fig. 21.7. The X-ray Machine Circuit.

For feed of a tube is required two sources: the source of high voltage for anodi circuit and low $(6-8$ V) for the circuit of heat. Both sources should have independent adjustment. *By change of anodi voltage* **rigidity** *of X-ray radiation is adjusted and by change of filament current the capacity of radiation is adjusted.*

The basic electric circuit of the elementary X-ray device is resulted on fig.21.7. In the circuit are present two transformers: $Tp1$ of high voltage and $Tp2$ for feed of heat. The high voltage on the tube is adjusted by autotransformer $Tp3$ connected to initial winding of transformer $Tp1$. The switch adjusts number of coils of winding of the autotransformer. In this connection changes the voltage of

the secondary winding of the transformer, submitted on the anode of the tube, i.e. rigidity is adjusted.

The current of reheat of the tube is adjusted by rheostat R swithed to the circuit of initial winding of transformer $Tp2$. The current of anodi circuit is measured by ammeter mA. Submitted on electrodes of the tube voltage is measured by voltmeter kV. Value of current of reheat adjustable by the rheostat is measured by the ammeter A.

21.3 Interaction of X-ray radiation with substance (simple scatter, Compton scatter, photoeffect)

At falling of x-ray radiation on any body X-rays in the small amount are reflected from it and basically passes deep into. In mass of the body radiation is in part absorbed, in part dissipates and in part passes through. Passing through the body photons of X-ray radiation cooperate basically with electrons of atoms and molecules of substance. Registration and using of X-ray radiation and also its influence on biological objects is defined by initial processes of interaction of X-ray photon with electrons. Depending on the parity of energy E of photon and energy of ionization A_i three main processes take place.

a) Coherent dispersion (or simple scatter)

Dispersion of long-wave X-ray radiation occurs basically without change of wavelength and it is named *coherent*. Interaction of photon with electrons of the internal shells strong connected with nucleus, changes only its direction, not changing its energy and so wavelength (fig.21.8). Coherent dispersion arises, if energy of the photon is less than energy of ionization: $E = h\nu < A_i$. As energy of the photon and energy of atom does not change, coherent dispersion does not cause biological action. However at creation of protection against X-ray radiation it is necessary to take into account opportunity of change of direction of the initial bunch.

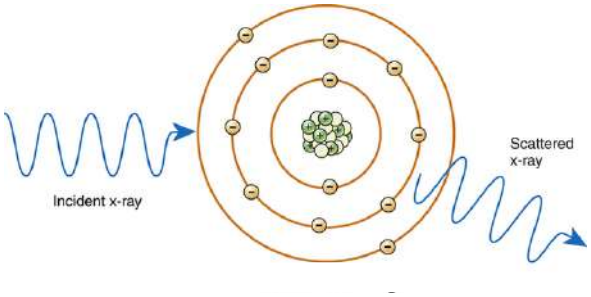

Fig. 21.8. The characteristic X-ray radiation

b) Compton scatter

In 1922 A. Compton, observing dispersion of rigid X-rays has found out reduction of penetrating ability of the dissipated bunch in comparison with incident. Dispersion of X-ray radiation with change of wavelength is called the Compton effect. It arises at interaction of the photon of anyone energy with poorly connected with nucleus electrons of external shells of atoms (fig.21.9).

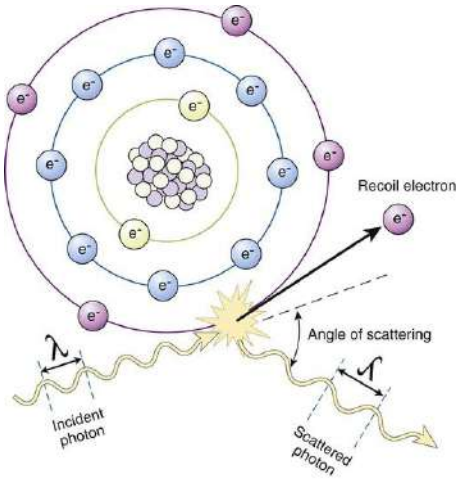

Fig. 21.9. The Compton effect.

Electron comes off atom (this electron is known as electron of recoil). Energy of the photon decreases $h\nu' = h\nu$ (wavelength accordingly increases) and also the direction of its motion changes. Compton effect arises, if energy of the photon of X-ray radiation is more than energy of ionization: $h\nu > A_i$, $h\nu = h\nu' + A_i + E_k$. Thus appear electrons of recoil with kinetic energy of E_k. Atoms and molecules become ions. If E_k is significant, electrons can ionize the next atoms by impact, forming new (secondary) electrons.

c) Photoeffect

If energy of the photon $h\nu$ is sufficient for tearing of electron, at interaction with atom the photon is absorbed and electron comes off atom. This phenomenon is known as *photoeffect*. The atom is ionized. Thus electron gets kinetic energy and if it $\frac{m_e\nu_e^2}{2} = h\nu - A_i$, is significant, it can ionize the next atoms by impact, forming new (secondary) electrons. If energy of the photon is insufficient for ionization, photoeffect can be shown in excitation of atom or molecule. At some substances it results to the subsequent radiation of photons in the visible region of radiation (*roentgenluminescence*) and in tissues it results to activation of molecules and photochemical reactions.

Photoeffect is characteristic for photons of energy about $0.5 - 1$ MeV.

Three basic processes of interaction considered above are initial; they results to the subsequent secondary, tertiary, etc. phenomena. At hit of X-ray radiation in substance there can be a lot of processes before energy of X-ray photon will turn into energy of thermal motion.

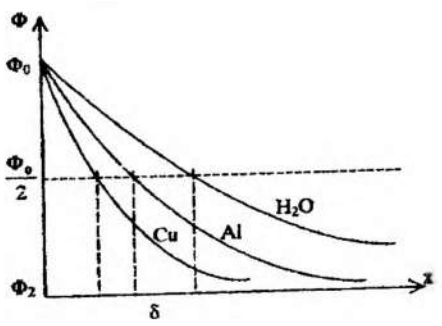

Fig. 21.10. The dependence of stream of X-ray radiation from thickness of the absorbing filter.

As result of the mentioned above processes the initial stream of X-ray radiation is attenuated. This process submits to law of Buger. We shall write down it as: $\Phi = \Phi_0 \exp -\mu x$, where μ *is the linear coefficient of attenuation*, dependent by nature of substance (mainly from density and nuclear number) and from wavelength of radiation (energy of photon). It can be presented consisting of three composed corresponding to coherent dispersion, Compton effect and photoeffect: $\mu = \mu_c + \mu_{nc} + \mu_{photoef}$

As the linear coefficient of attenuation depends on density of substance prefer to use *mass coefficient of attenuation*, which is equal to the ratio of linear coefficient of attenuation to density of the absorber and does not depend on density of substance: $\mu_m = \mu/\rho$. Dependence of stream (intensity) of X-ray radiation from thickness of the absorbing filter is submitted on fig.21.10 for H_2O, Al and Cu. Calculations show, that the layer of water by thickness of 36 mm, aluminium of 15 mm and copper of 1.6 mm reduce intensity of X-ray radiation in 2 times. This thickness is named *thickness of the half layer* δ. If the substance reduces X-ray radiation by half: $\Phi = 1/2\Phi_0(x = \delta)$, then $\Phi/\Phi_0 = \dfrac{1}{2} = e^{-\mu\delta}$ or $e^{\mu\delta} = 2\mu\delta \cdot \lg e = \lg 2$; $\mu\delta \cdot 0.4343 = 0.3010$; $\delta = 0.693/\mu$. Knowing thickness of the half layer, it is possible to determine μ. Dimension of μ is $[m^{-1}]$.

21.4 Using of X-ray radiation in medicine (roentgenoscopy, roentgenography, X-ray tomography, photoroentgenography, roentgenotherapy)

One of the most widespread applications of X-ray radiation in medicine is influence on internal bodies with the diagnostic purpose (radiodiagnosis).

For diagnostics photons of energy $60-120$ keV are used. Thus the mass coefficient of attenuation μ_m is defined basically by photoeffect. Its value is proportional to λ^3 (big penetrating ability of rigid radiation is shown) and it is proportional to the third degree of number of atoms of substance – absorber: $\mu_m = K\lambda^3 Z^3$, where is coefficient of proportionality.

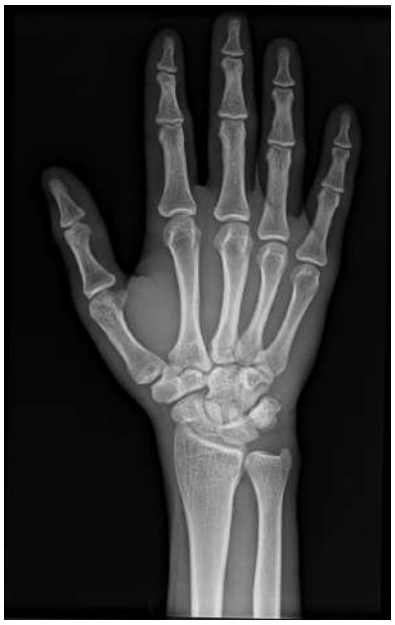

Fig. 21.11. The roentgenography.

The human body consists of tissues and the bodies having various absorbing ability in relation to X-ray radiation. Therefore at influence of X-rays the nonuniform shadow image on the screen turns out, which gives the picture of arrangement of internal bodies and tissues. The most dense absorbing radiation of tissues (heart, large vessels, bones) are visible dark, and poorly absorbing tissues (lung) are light (fig.21.11).

In many cases it is possible to judge thus their normal or pathological condition. Radiodiagnosis uses two basic methods: *roentgenoscopy* and *roentgenography* (picture). The block diagram of modern X-ray installation is shown on fig.21.12: X-ray tube, generator of feed of the tube, the control panel, place for the cartridge with the film, EOC, translucent mirror, movie camera, television camera, monitor, video.

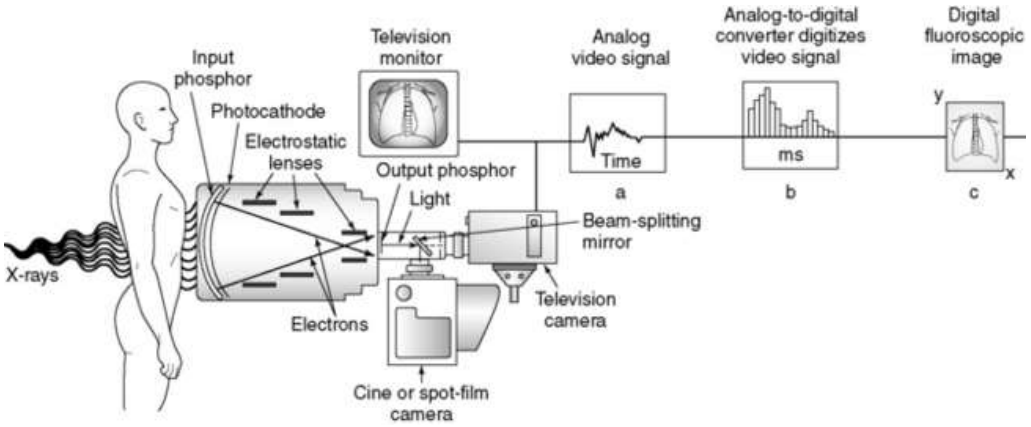

Fig. 21.12. The block diagram of modern X-ray installation

If the researched body and tissues environmental him approximately equally absorb the stream of X-ray radiation, apply special *contrast substances*. So, for example, before the X-ray research of stomach or intestine give the special kasha of barium sulfate, in this case it is possible to see it shadow image. At roentgenoscopy and roentgenography the X-ray image is the total image of all thickness of object through which pass X-rays. Those details, which are closer to the screen or film are most precisely outlined and removed become indistinct and dim. If in any body there is the pathological site, for example, destruction of lung tissues inside the extensive center of inflammation, than in some cases

this site on the roentgenogram in the sum of shadows can "be lost". To make it visible apply the special method *tomography (level-by-level record)*, which allows to receive pictures of separate layers of investigated area.

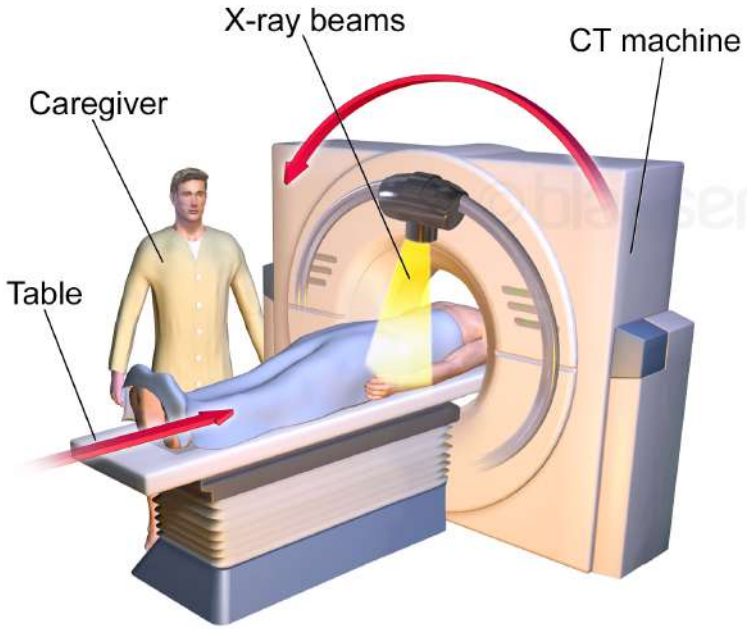

Fig. 21.13. Computerized axial tomography scan.

Such level-by-level pictures (tomograms) receive with help of the special device named the *tomograph*, in which periodically in antiphase move X-ray tube (RT) and film (F) concerning of area of research (fig.21.13). Thus X-rays at any position of RT will pass through the same point of object (the changed site) being the center concerning which periodic motion of RT and F is made. The shadow image of the site will be fixed on the film. Changing position of «the center of rocking» it is possible to receive level-by-level images of object (fig.21.13). Such modern variant of tomography is known as *computer tomography*. The tomography is widely applied at research of lungs, kidneys, bilious bubble, stomach, bones etc.

Other example is *photoroentgenography*: at it on sensitive small format film manages the image from the big roentgenluminescence screen.

The photoroentgenography combines the big opportunity of detection of reticent proceeding diseases (disease of thorax, gastroenteric path, additional bosoms of nose, etc.) with significant throughput (up to $120 - 150$ peoples at hour), in this connection it is rather effective method of mass research.

As photographing of the X-ray image at photoroentgenography is made with help of photographic optics, the image on picture in comparison with X-ray is re-

duced. In this connection resolution of photoroentgenography (i.e. discernability of fine details) is less than usual roentgenogram.

X-ray radiation is used as well for the medical purposes (*roentgenotherapy*). Biological action of radiation consists in destruction of quickly developing cells of malignant tumours. It is possible to pick up a doze of radiation, sufficient for full destruction of the tumour at rather insignificant damage of environmental healthy tissues, which owing to the subsequent regeneration are restored.

Test questions

1. What is X-ray radiation?

2. What are the known sources of X-ray radiation?

3. Explain how the X-ray tube works. What parameters of X-ray radiation are affected by the voltage and current of the filament in the X-ray tube?

4. Explain the mechanisms of the brake and the characteristic X-ray radiation.

5. Why is the spectrum of bremsstrahlung X-ray radiation continuous, and the characteristic one - linear?

6. Explain under what conditions coherent and incoherent scattering and photoelectric effect occur.

7. Formulate and explain Moseley's Law.

8. What are the known interactions of X-rays with matter?

9. Tell us about the main methods of medical X-ray diagnostics.

10. What effect does X-ray radiation have on biological objects?

11. Tell us about the methods of X-ray therapy.

Lecture No 22

Radio-activity

22.1. Basic law of radioactive decay. Activity. Units of activity

Property of unstable nucleus of some elements spontaneously (i.e. without any influences) to transform to nucleus of other elements with emission of ionizing radiation is called **radio-activity**. The phenomenon is known as **radioactive decay**. Radioactive decay is accompanied by insignificant allocation of heat. Distinguish the artificial and natural radio-activity.

The *natural* radio-activity meets at the unstable nucleus, existing naturally. Radio-activity of the nucleus formed as the result of different nuclear reactions is called the *artificial* radio-activity. Basic distinction between the natural and artificial radio-activity is not present. It is conditional division, since both kinds of radio-activity submit to the same laws.

Feature of radioactive decay is that nucleus of the same element decay not all at once and gradually in various time intervals. The moment of decay of any nucleus cannot be specified beforehand, however, the theory allows to establish *probability* of decay of one nucleus for a time unit, i.e. radioactive decay is a *statistical phenomenon*. At the big set of radioactive nucleus is possible to receive the statistical law expressing dependence of number of not broken nucleus from time. We shall receive this law.

Let for small time interval dt decay dN of nucleus. This number is proportional to the interval of time dt and also to the general number of the radioactive nucleus, which have not decay yet to the beginning of the given time interval:

$$dN = -\lambda N dt, \tag{22.1}$$

where λ is *decay constant* (characterizes probability of decay of a nucleus per unit of time and various for different radioactive nucleus). Dimension of decay

constant is s^{-1}. The mark minus specifies decrease in time of value N, i.e. $dN < 0$. Expression (22.1) represents the differential equation of 1-st order with divided variables. We shall divide variables and we shall integrate in view of that the bottom limits of integration correspond to entry conditions: at $t = 0$, $N = N_0$, where N_0 is initial number of radioactive nucleus:

$$\int_{N_0}^{N} \frac{dN}{N} = -\lambda \int_0^t dt; \quad \ln \frac{N}{N_0} = -\lambda t; \quad \ln \frac{N}{N_0} = \ln e^{-\lambda t}; \quad \frac{N}{N_0} = e^{-\lambda t}$$

$$\boxed{N = N_0 e^{-\lambda t}}, \tag{22.2}$$

i.e. number of radioactive nucleus, which have not broken up yet decrease on the exponent law. Expression (22.2) also is the **basic law of radioactive decay**. If there is the necessity to calculate quantity ΔN of nucleus, which decay to some moment of time t: $\Delta N = N_0 - N = N_0(1 - e^{-\lambda t})$.

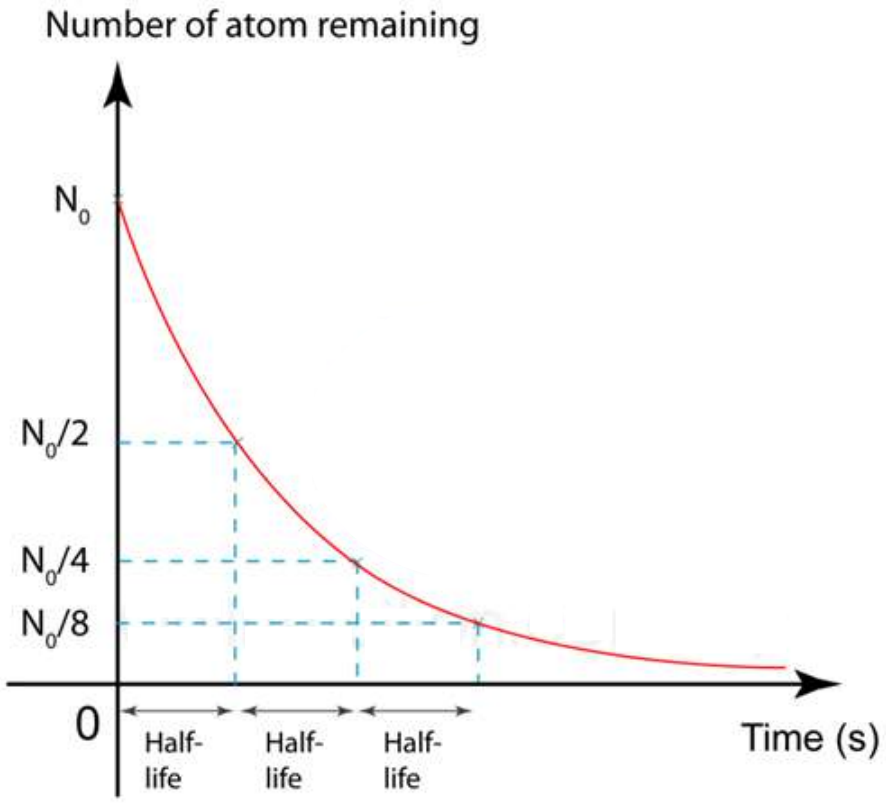

Fig. 22.1. The half-life period T.

Rate of decay of various radioactive elements characterize *half-life period* T, it is time during which breaks up half of initial number of radioactive nucleus (fig.22.1). We shall establish relationship between T and λ. The half-life period

can be determined from the following reasons: at $t = T$, $N = \frac{N_0}{2}$; $\frac{N_0}{2} = N_0 \cdot e^{-\lambda T}$; $\frac{1}{2} = e^{-\lambda T}$; $2 = e^{\lambda T}$; $\ln 2 = \lambda T$; $0.693 = \lambda T$; $T = \frac{0.693}{\lambda}$.

The half-life period for various elements matters from shares of second up to millions years. Accordingly radioactive isotopes are divided on shortly living (hours, days) and long-living (years). Examples: half-life period of Uranus $T = 4.51 \cdot 10^9$ years; Lithium $T = 0.89$ seconds. Radioactive elements of Chernobyl emission have the half-life period: ^{239}Pu is 26400 years; ^{137}Cs is 30 years; ^{90}Sr is 29 years. Tab.22.1 lists some common radioisotopes with their half-lives and illustrates, that the naturally-occuring radioisotopes in evidence today have very long half-lives.

Table 22.1

Lists some common radioisotopes with their half-lives

Radioisotope	Element	Half-Lives
Natural		
^3H (Tritium)	Hydrogen	$12 - 26$ years
^{14}C	Carbon	5760 years
^{40}K	Potassium	1300 million years
^{226}Ra	Radium	1600 years
^{238}U	Uranium	4500 million years
Artificial		
^{99}Tc	Technecium	6 hours
^{24}Na	Sodium	15 hours
^{32}P	Phosphorus	$14 - 3$ days
^{60}Co	Cobalt	$5 - 3$ years
^{125}I	Iodine	60 days
^{131}I	Iodine	8 days
^{137}Cs	Caesium	33 years

In conditions, when radioactive radiation is used for any purposes (for example, in medicine) it is necessary to know total decays per unit of time in the given quantity (mass) of radioactive element. This value is rate of decay and is known as **activity** (A). It is the essential characteristic of a radioactive preparation: $A = -\frac{dN}{dt}$, since $-\frac{dN}{dt} = \lambda N$ and $N = N_0 \cdot e^{-\lambda t}$, then $\boxed{A = \lambda N_0 \cdot e^{-\lambda t}}$. Initial activity ($t = 0$) $A_0 = \lambda N_0$. Then $A = A_0 \cdot e^{-\lambda t} = A_0 \cdot e^{\frac{0.693}{T}t}$. The activity calculated for a mass unit of an isotope is called *specific activity*. For solutions as specific activity understand activity of radioactive solution of 1 ml.

Activity of an isotope is more than is more the radio-activity of nucleus and than is less its half-life period. Activity of the preparation in time decreases on exponent law.

Unit of activity in SI is 1 Becquerel (Be), that corresponds to activity of a radioactive source in which for 1 second there is 1 act of decay.

The most used unit is *Curie* (Cur):

$$1\ Cur = 3.7 \cdot 10^{10}\ \ Bc = 3.7 \cdot 10^{10}\ \ s^{-1}.$$

Except for it there is one more stand-alone unit: *Rutherford* (R), $1\ R = 10^6\ \ Be = 10^6\ \ s^{-1}$.

22.2. Basic kinds of radioactive decay.

Under the general name of radioactive radiation are united 3 kinds of radiation, various by the nature, but having some general properties. They have been named historically as *alpha*, *beta*, and *gamma* – beams.

Alpha radiation is the stream of particles with high kinetic energy. Alpha decay will consist in spontaneous transformation of nucleus with emission of α - particles (nucleus of helium). The scheme of α - decay in the view of the rule of displacement:

$$^{A}_{Z}X \rightarrow ^{A-4}_{Z-2}Y + ^{4}_{2}\alpha,$$

where X is the symbol of initial (parent) nucleus; Y is the symbol of the nucleus – product of decay (affiliated or daughter's nucleus).

In connection with emission of α- particles the charge of the nucleus and accordingly nuclear number of the element decreases on two units and mass number on four units. Hence, the secondary element is shifted in Mendeleyev table on two numbers to the left and the nuclear mass of it becomes less on four units.

As the example of α- decay can serve decay of radium, at which radon is formed:

$$^{226}_{88}Ra \rightarrow ^{222}_{86}Rn + ^{4}_{2}\alpha.$$

Thus it is radiated γ- photon. At α- decay the affiliated nucleus can be not only in normal, but also in the excited condition and as these conditions are discrete, also values of energy of α- particles, which are taking off from different nucleus of the same radioactive substance are *discrete*. Energy of excitation of the

affiliated nucleus is allocated as γ- photon more often. For this reason α- decay is accompanied by γ- radiation. Speed of start of α- particles from a nucleus is $(1.4-2)\cdot 10^7\ km/s$, that corresponds to initial kinetic energy of $4-8.8\ MeV$. Alpha – particles, which are emitted by certain element, make some groups with close energy, therefore the spectrum of α- radiations will consist of the several close located lines.

β-decay occurs at nucleus, which instability is connected with the certain parity of number of protons and neutrons. If in any nucleus there is surplus of neutrons occurs *electronic β- decay* of nucleus, at which one neutron transforms to proton, thus in the nucleus is formed electron:

$$^1_0n \rightarrow ^1_1 p + ^0_{-1} e + \tilde{\nu},$$

Where $\tilde{\nu}$ is antineutrino (the elementary particle). Electron is thrown out from the nucleus and in it there is steadier complex of nucleons. Electronic β-decay is described by the equation:

$$^A_Z X \rightarrow ^A_{Z+1} Y + ^0_{-1} e + \tilde{\nu}.$$

Thus the charge of the nucleus and accordingly nuclear number of the element increases for unit, i.e. the secondary element is shifted in Mendeleyev table on one number to the right, its mass number remains without change. Example: $^{40}_{19}K \rightarrow ^{40}_{20} Ca + ^0_{-1} e + \tilde{\nu}$.

If in a nucleus *surplus of protons occurs positron β-decay*, at which one of protons transforms to neutron, thus in the nucleus is formed positron: $^1_1p \rightarrow ^1_0 n + ^0_{+1} e + \nu$,, where ν is neutrino.

The positron is thrown out, in the nucleus steadier complex of nucleons is formed. Positron β- decay is described by the equation:

$$^A_Z X \rightarrow ^A_{Z-1} Y + ^0_{+1} e + \nu.$$

The charge of the nucleus and accordingly nuclear number of an element changes on unit, and the secondary element is shifted in Mendeleyev table on one number to the left, the mass number of it remains constant. Example:

$$^{79}_{37}Rb \rightarrow ^{79}_{36} Kr + ^0_{+1} e + \nu.$$

Initial speed and accordingly kinetic energy of β- particles can considerably differ. The greatest initial speed has the order of $1.6 \cdot 10^8$ m/s and energy of β-particles can be in limits from tenth and 100-th shares of MeV till $10 - 12$ MeV. Power spectrum of β- particles *continuous*, i.e. their energy can accept different values. To explain distinction in energy of β- particles at decay of nucleus of the same element, V. Pauli has assumed in 1939, that at β- decay alongside with β-particle from the nucleus are thrown out neutral particles *neutrino and antineutrino* with mass equal about 1/2000 of mass of rest electron and having energy, which in the sum with energy of β- particle make some constant, characteristic for the given substance. And this energy at different nucleus divides between beta and these particles in various parities. It explains the continuous spectrum of β-particles.

At emission of β- particles, as well as at α-decay, nucleus of atoms can be in the excited condition. Their transition to the not excited condition (sometimes in steps) is accompanied by emission of γ-quantums with energy from 0.2 up to 3 MeV. Spectrum of γ- radiation is line, γ-radiation arises not only at α- and β-decays. At collision of the nucleus with a particle it can proceed to the excited condition and then coming back to the basic condition to radiate γ - photon.

There is the third kind of β - decay, that is called **electronic or e-capture**. It consists that the nucleus grasps one of internal electrons, taking place on K, L, M levels, therefore the proton of the nucleus transforms to neutron: ${}^1_1p + {}^0_{-1}e \to {}^1_0\, n + \nu$.

At electronic capture the place in electronic shell is released, therefore this kind of radio-activity is accompanied by characteristic X-ray radiation.

22.3. Methods of reception of radionuclides.

Interaction of a nucleus with an elementary particle or with other nucleus of element as. result of which this nucleus transforms to the nucleus of other element is called the *nuclear reaction.* Nuclear reactions allow toreceive from one chemical elements other elements by influence on the nucleus of atom. Effective means of such influence appeared bombardment of nucleus by particles of high energy. For the first time nuclear reaction has carried out by Rutherford in 1919. At bombardment of nucleus of nitrogen by α- particles formed at decay of radium, occured transformation of nucleus of nitrogen to nucleus of isotope of oxygen with ejection of protons: $N + \alpha \to p + O$.

Brief record of reaction: ${}^{14}N(\alpha, p)^{17}O$.

Key rule at drawing up of the equation of nuclear reaction is equality in its both parts of the sum top (mass numbers) and bottom (nuclear numbers) indexes. It is expression of laws of preservation of mass and charges of the particles participating in reactions.

The reduced record will consist of four symbols: the initial nucleus (nucleus – target), in brackets is a bombarding particle and other formed particle (or particles), out of brackets is put the symbol of nucleus – product of reaction (nuclear number of the element usually is not put).

Originally as bombarding particles were used α- particles of radioactive radiation. In 1932 by the English physicist D. Chedvik was open the *neutron.* Neutron is the stable, neutral particle, however in a free condition it for a long time does not exist. At collision with the nucleus of any element the neutron is absorbed by it and causes nuclear reaction. For example: ${}^{14}_{7}N + {}^{1}_{0}n \rightarrow {}^{11}_{5}B + {}^{4}_{2}\alpha$ or ${}^{14}N(n,\alpha){}^{11}B$.

Nuclear reactions under action of neutrons have the greatest probability. Not having electric charge, neutrons freely fly electric shells of atoms and hitting with nucleus more often cause nuclear reactions.

Further began to use and other charged particles, preliminary giving them the big speed (kinetic energy) in special accelerators, for example, in cyclotrons.

All nuclear reactions are accompanied by emission of any elementary particles (including γ- photons). Products of many nuclear reactions are radioactive and products are named *artificial radioactive isotopes (radionuclides).* Phenomenon of artificial radio-activity was open in 1934 by known physicists Frederic and Iren Jolio-Curie.

As example of reception of radioactive isotopes (radionuclides) can serve reaction of capture of neutrons by phosphorus ${}^{31}_{15}P$. At this capture is radiated γ - photon and the radioactive isotope of phosphorus is formed: ${}^{31}_{15}P + {}^{1}_{0}n \rightarrow {}^{32}_{15}P + \gamma$.

Decay of nucleus of the received isotope ${}^{32}_{15}P$ is accompanied by emission of β-particle (simultaneously is emitted antineutrino) and formation of stable isotope of silicon:

$$ {}^{32}_{15}P \rightarrow {}^{32}_{16}Si + {}^{0}_{-1}e + \tilde{\nu} $$

As well as to natural radioactive elements to artificial isotopes are peculiar α-, β- and γ- decays.

Radioactive isotopes in small amounts receive in accelerators (for example, in the cyclotron) with help of deutrons (nucleus of heavy hydrogen) d.

In industrial scale artificial radioactive isotopes receive by irradiation (mainly of neutron) of corresponding chemical elements in nuclear reactor.

Except for irradiation by neutrons radioactive isotopes receive in reactors by their allocation from fission products of nucleus of uranium, for example, radioactive iodine $^{131}_{53}I$ is widely used in medicine.

Table 22.2

Lists some radioactive isotopes with their half-lives

Isotopes	Half-Lives	Uses
Radio-carbon C^{14}	4700 years	In the study of protein metabolism and archeological dating
Radio-tritium H^3	11 years	As a tag in organic substances.
Radio-cobalt Co^{60}	$5-3$ years	Emits strong γ-rays. Used in Radiography and Radio therapy as a substitute for radium.
Radio-sulphur S^{35}	87 days	Numerous chemical and industrial applications.
Radio-phosphorus P^{32}	14 days	In the study of bone metabolism and for the treatment of blood diseases.
Radio-iodine I^{131}	8 days	Treatment ot of thyroid diseases.
Radio-sodium Na^{24}	15 hours	Its chemical property tod solubility make it useful in a number of application, i.e. in the study of circulatory disorders in blood vessels.

Already are obtained several radioactive isotopes for each chemical element, their general number exceeds of 1500. Many of them are widely applied in quality of labelled atoms in various branches of human activity, including medicine (tab.22.2).

22.4. Interaction of ionizing radiation with substance.

In connection with the general character of initial action on substance α-, β - and γ- radiation, rigid X-ray radiation and also streams of protons and neutrons are united under the general name of *ionizing radiation.* The charged particles and γ - photons being distributed in substance, cooperate with electrons and nucleus, therefore changes condition of substance and particles.

To the basic properties of radioactive radiations are *penetrating* and *ionizing abilities.*

Ionizing ability of radiation is estimated by *linear density of ionization* i: $i = dn/dl$, where dn is number of ions of one mark formed by the particle on the elementary way dl. In practice this value is estimated by number of pairs of ions formed by particle on 1 cm of run.

Ionizing ability is estimated by *linear brake ability of substance* S: $S = dE/dl$, where dE is the energy lost by charged particle at passage of elementary way dl in substance. As ionization of one molecule needs energy of 34 eV, value S can be calculated knowing linear density of ionization.

Penetrating ability of radiation is estimated by *length of free run* or *average linear run* R, it is average distance, which there passes the particle in substance while particle is capable to ionize. More charge and mass of a particle more its ability to ionize substance and the less its average run. Average linear run of α-particies in human organism is $10 - 100$ microns; of β- particles is $10 - 15$ mm; γ- radiation will penetrate on the big depth or penetrates body of the person through. Properties of ionizing particles are resulted in the tab.22.3.

Table 22.3

Properties of ionizing particles.

Kind of radiation	Average energy, MeV	Linear density of ionization i, pair/cm	Average linear run R, m	
			in air	in substance
α	$4 - 8.8$	$3 \cdot 10^4$	$(2 - 8) \cdot 10^{-2}$	–
β	$0.01 - 10$	$50 - 250$	10	$1.5 \cdot 10^{-2}$
γ	$0.2 - 3$	300	300	Near 1

Electrons displaced at ionization can beat out secondary electrons, having energy sufficient for the subsequent ionization of substance. These secondary processes can cause characteristic X-ray radiation, radioluminescense, chemical processes.

γ- photons causing insignificant initial ionization, generate secondary as result of which total ionizing effect can be rather significant.

Owing to various ionizing and penetrating abilities of radioactive radiations **ways of protection** against radiation are different: for protection from α- particles there is enough layer of a paper, clothes; from β- radiation it is possible to be protected by centimetric layer of a tree, glass or any easy metal; for protection from γ- radiation are applied thick (up to meter) layers of water, concrete, brick walls and also plates of lead by thickness up to 10 cm (fig.22.2).

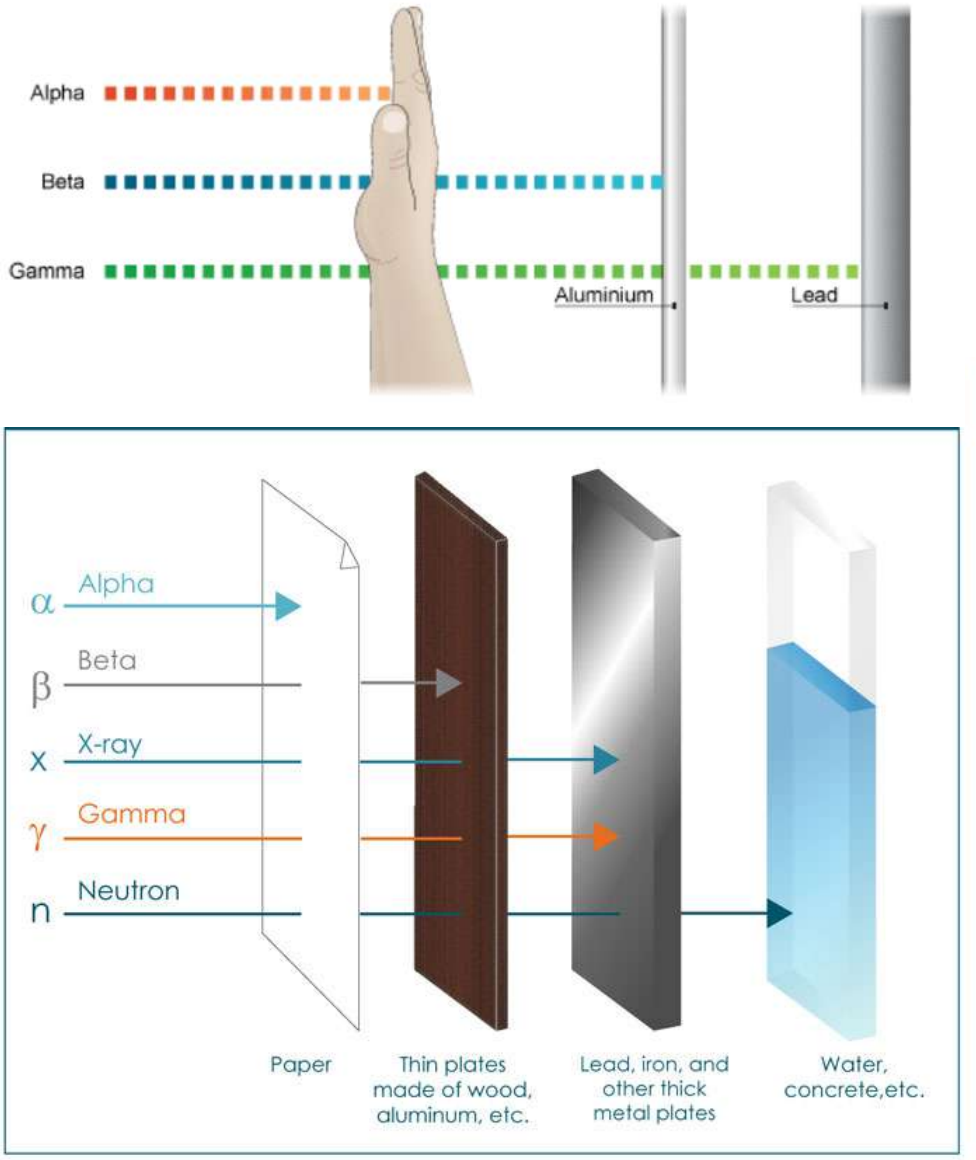

Fig. 22.2. Type of radiation and penetration.

Except for ionization the particles are capable to cause other processes (fig. 3).

α- particles can cooperate with nucleus, causing nuclear reactions, though this process more rare, than ionization.

β - particles at braking can create braking X-ray radiation.

At hit of positron in substance with high probability there is its interaction with electron, after which there are two γ - photon, which have energy not less energy of rest electron 0.51 MeV (reaction of annihilation): ${}^{0}_{-1}e + {}^{0}_{+1}e \to \gamma + \gamma$.

For α - and β- particles processes of dispersion are possible, therefore their way in substance is strongly bent.

At hit of γ- radiation in substance alongside with processes characteristic for

X-ray radiation (coherent dispersion, Compton effect (scattering), photoeffect) are possible other processes also.

At interaction of γ- photons of the big energy with nucleus is possible photonuclear reaction. For its occurrence energy of γ - photon should be not less binding energy falling on one nucleon.

At energy of γ - photon more than 1.2 MeV (not less total energy of rest electron and positron) is possible reaction of birth of pair electron-positron: $\gamma \rightarrow {}^{0}_{-1} e + {}^{0}_{+1} e$.

Attenuation of stream of γ- radiation in substance is described by the law: $\Phi = \Phi_0 \cdot e^{-\mu x}$, where μ is linear coefficient of absorption, which can be presented as the sum of corresponding coefficients of absorption, which are taking into account three processes of interaction: photoeffect (μ_{ph}), compton-effect or not coherent dispersion (μ_{nc}) and formation of pairs electron-positron (μ_p): $\mu = \mu_{ph} + \mu_{nc} + \mu_p$.

At action on substance of stream of neutrons can occur: elastic impact with a nucleus and secondary ionization, not elastic impact with a nucleus with emission of γ- quantum, capture of neutron by a nucleus with formation of the radioactive isotope. Last effect can cause formation in organism of radioactive isotopes: ${}^{1}H(n,\gamma)^{2}H$; ${}^{23}Na(n,\gamma)^{24}Na$; ${}^{31}P(n,\gamma)^{32}P$ and some other reactions.

It is necessary to note interaction of radioactive radiations with water, at which there is chemical transformation named the *radiolysis of water*. As result of interaction are possible formation of the excited molecules (H_2O^*), ions (H_2O+), radicals (for example: $\dot{H}$, $O\dot{H}$), peroxide of hydrogen (H_2O_2). These highly active connections in the chemical attitude can cooperate with other molecules of biochemical system that will result to infringement of normal functioning of membranes, cells and bodies.

22.5. Using of radionuclides in medicine.

Medical application of radionuclides can be submitted by two groups of methods: use with the *diagnostic and research purposes (labelled atoms)* and their application with the *therapeutic, medical* purpose. Bactericidal action of radiation concerns to the second group also.

The method of *labelled atoms*: in organism are entered radionuclides and are determined their location and activity in the bodies and tissues. For example, for diagnostics of disease of thyroid gland in organism enter radioactive iodine ${}^{131}_{53}I$, ${}^{125}_{53}I$, which part concentrates in gland (fig.22.3). The counter located near

to gland fixes speed of accumulation of iodine, on the basis of which it is possible to make diagnostic conclusions about condition of thyroid gland. The cancer of thyroid gland can give metastasises in different bodies that can give information about accumulation of radioactive iodine in these bodies.

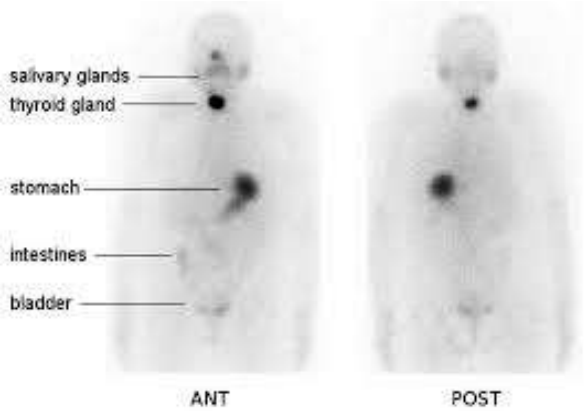

Fig. 22.3. Iodine-131 for therapy of thyroid diseases.

For detection of distribution of radionuclides in organism it is applied gammatopography, which is carried out with help of the γ- topographer. The scanning counter gradually passes the big sites above the body. Intensity of radiation of a preparation, for example, by strokes on the paper in the places of his presence is automatically fixed. The gamma - topographer gives rather rough distribution of radioactive preparation in the bodies. More exact data gives *autoradiography*. On a biological tissue apply the layer of photoemulsion. Radionuclides contained in the object give the trace on photoemulsion, as though photographing itself. Received picture is called the *autoradiogramm*.

In organism radioactive atoms enter in such small amount that neither atoms nor products of their decay do not exert harmful influence on the body. Applying radioactive isotopes it is possible to study *distribution of blood* and other biological liquids in organism. For this purpose, for example, enter the certain quantity of radioactive indicator into blood and having sustained time for its uniform distribution on blood system, it is possible to find her total amount on activity of unit of volume of blood.

Method of labelled atoms allows todiagnose *diseases of heart* and other bodies also. All researches and supervision are carried out without infringement of normal ability to live of an organism. In it value of method of labelled atoms.

Medical application of radionuclides basically is connected with use of γ- radiations (γ- therapy). Apparatus (cobalt gun) contains the protective lead container with ^{60}Co. Application of γ- radiations with big energy allows to *destroy deeply*

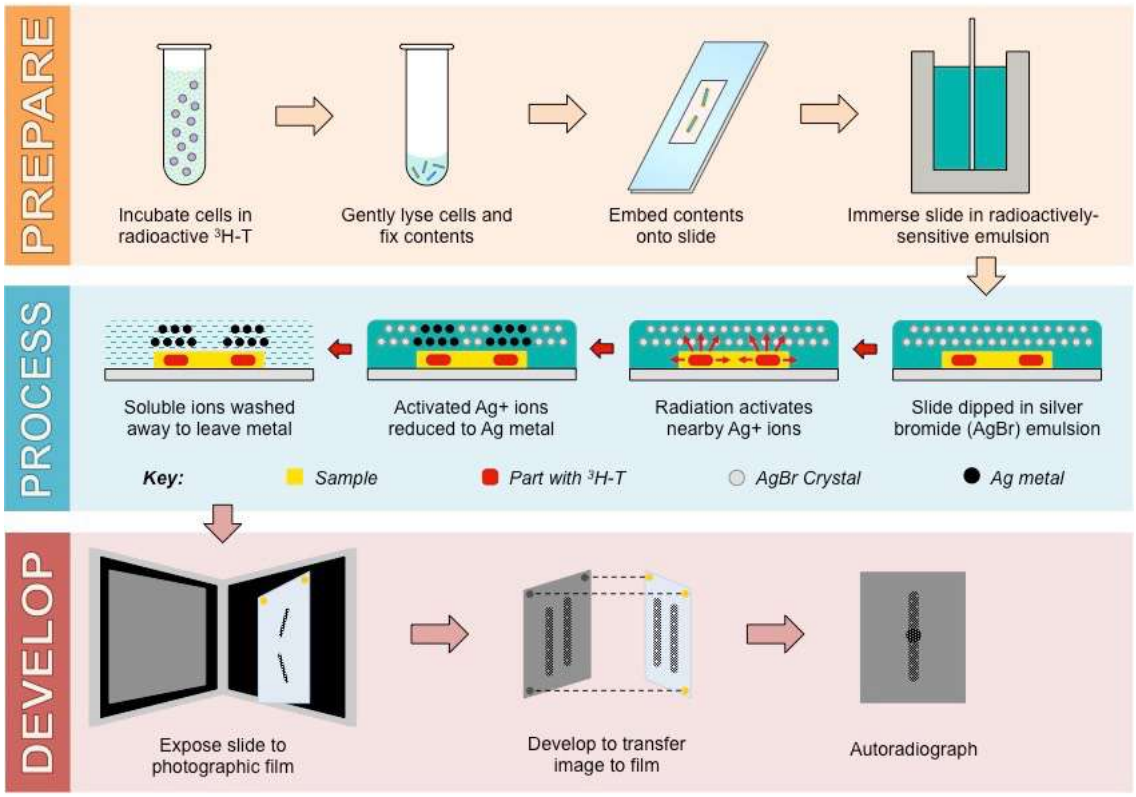

Fig. 22.4. The autoradiography.

located tumours. Superficially located bodies are exposed to smaller pernicious influence. Radioactive cobalt is applied for interstitial irradiation also. The needle containing thin pin from radioactive cobalt is stuck into the tissue.

For treatment of oncological diseases are applied α*- particles in combination with streams of neutrons.* In a tumour enter elements, which nucleus under action of the stream of neutrons cause nuclear reaction with formation of α- radiation. For example:

$$^{6}_{3}Li + ^{1}_{0}n \rightarrow ^{3}_{1}H + ^{4}_{2}\alpha.$$

Thus, α- particles and nucleus of feedback are formed in that place of the body, which it is necessary to subject to influence.

In the medical purposes (*treatment of illnesses of blood*) is used radioactive Phosphorus и $^{32}_{15}P$ (β- particles), which concentrates in the compact substance of tubular bones. $^{32}_{15}P$ irradiates marrow and thus normalizes broken at the certain diseases formation of blood. For similar purposes in relation to thyroid gland use the radioactive iodine $^{131}_{53}I$ giving electronic radiation.

It is applied radonie therapy also, at which are used the mineral waters containing ^{222}Rn and its products for influence on human skin (*radonic baths*), digestive apparatus (*drink*) and respiratory apparatus (*inhalation*).

Test questions

1. What atomic nuclei are incapable of spontaneous radioactive decay?

2. What is alpha decay? How do the charge and mass of a nucleus change during alpha decay? How does gamma radiation occur during alpha decay?

3. What is beta decay? What transformations occur in the atomic nucleus during beta decay?

4. What is meant by artificial radioactivity?

5. Formulate the law of radioactive decay and get an analytical expression of the law.

6. What is decay constant, half-life, activity of a radioactive element?

7. What determines the penetrating and ionizing ability of radioactive radiation?

8. What are the biophysical bases of the action of ionizing radiation on the body.

9. Give a definition of the main characteristics of ionizing radiation.

10. Using of radionuclides in medicine.

Lecture No 23

Nuclear Physics. Dosimetry of ionizing radiation

23.1. Some properties of nuclei.

All nuclei are composed of two types of particles: protons and neutrons. In describing the atomic nucleus are identified the following integer quantities:

• The **atomic number** Z equal the number of protons in the nucleus (the atomic number is sometimes called the charge number).

• The **neutron number** N equal the number of neutrons in the nucleus.

• The **mass number** A equal the number of nucleons (neutrons plus protons) in the nucleus.

$$\boxed{A = N + Z}.$$

The symbol used is ${}^{A}_{Z}X$, where X represents the chemical symbol for the element. *For example*, ${}^{56}_{26}Fe$ (iron) has a mass number of 56 and an atomic number of 26; therefore, it contains 26 protons and 30 neutrons.

The nuclei of all atoms of a particular element contain the same number of protons but often contain different numbers of neutrons. Nuclei that are related in this way are called **isotopes**. *The isotopes of an element have the same Z value but different N and A values.* The natural abundances of isotopes can differ substantially.

For example, ${}^{11}_{6}C$, ${}^{12}_{6}C$, ${}^{13}_{6}C$, ${}^{14}_{6}C$ are four isotopes of carbon.

The simplest element, hydrogen, has isotopes: ${}^{1}_{1}H$, the ordinary hydrogen nucleus; ${}^{2}_{1}H$, deuterium; ${}^{3}_{1}H$, tritium.

Some isotopes do not occur naturally but can be produced in the laboratory through nuclear reactions.

The proton carries a single positive charge $+e$ and the electron carries a single negative charge $-e$, where $e = 1.6 \cdot 10^{-19}\ C$. The neutron is electrically neutral,

as its name implies.

23.2. Absorbed and exposition doses. Capacity of a dose.

Quantitative estimation of action of ionizing radiation on substance of alive and abiocoen nature has resulted to occurrence of the unit of physics named *dosimetry*.

The section of nuclear physics and measuring technics in which are studied the values describing action of ionizing radiation on substance and also methods and devices for their measurement is known as **dosimetry**.

Initial development of dosimetry has been caused, first of all, by necessity of the account of action of X-ray radiation on a human organism.

Action on substance is caused not by all radiation falling on it, but only by part, that cooperates with its atoms and molecules. Part of radiation, which passes the given body through without absorption does not influence on it. Therefore the basic value describing action of ionizing radiation on substance is the *energy of radiation absorbed by mass unit of substance during irradiation.* This value is called the **absorbed dose (D)**. Various effects of ionizing radiations first of all are defined by the absorbed dose. It depends on a kind of ionizing radiation, energy of its particles, structure of irradiated substance and is proportional to time of irradiation. Unit of the absorbed dose for any kind of radiation is **"Gray" (Gy)** (L. Gray is the English radiobiologist). For 1 Gy is accepted the dose of radiation at which to the irradiated substance of mass 1 kg is transferred energy of ionizing radiation of 1 $Joule$, i.e. $1\ Gy = 1\ J/kg$.

The dose of irradiation in time unit is called the *capacity of dose* ($P = D/t$). Capacity of dose is expressed in Grays per second (Gy/s).

The stand-alone unit of dose of radiation used in radiobiology is **rad**: it is dose of any kind of ionizing radiation at which 1 g of substance absorbs the energy of radiation equal to 100 erg. $1\ Gy = 100\ rad$**;** $1\ rad = 10^{-2}\ Gy$. Capacity of dose is measured in **rad/s**.

For finding of the absorbed dose of radiation it is necessary to measure energy of ionizing radiation falling on a body, then the energy past through the body and their difference to divide on mass of the body. However practically in a human body to make it difficultly, as the body is non-uniform, *energy dissipates by a body on every possible directions.* In this connection estimate the dose absorbed by the body on ionizing action of radiation on air environmental the body. For charac-

teristic of a dose on the effect of ionization of air is used so-called **exposition dose** (or **exposure or air dose**) (X) of X-ray or γ- radiation. It is necessary to remember, *that the exposition dose is defined only for air and only for quantum radiation.* The important advantage of this dose is that for its measurement there is the simple physical method, consisting in measurement of total charge of ions formed under action of radiation. For unit of exposition dose is accepted the 1 C/kg: it is exposition dose of photon radiation at which the total charge of ions of one mark made in 1 kg of irradiated air is equal to 1 $Coulomb$.

Total ionization of air means effect both from initial action of ionizing radiation and from all secondary processes occuring at it, in particular from action of secondary electrons and nucleus of feedback. In practice is used the old standalone unit named **roentgen** (R). The exposition dose of 1 R corresponds to formation of $2.08 \cdot 10^9$ pairs of ions in 0.001293 g (in 1 cm^3) dry air under normal conditions: 1 $C/kg = 3876$ R, i.e. new unit is much larger then old. It is useful to remember the convenient rule frequently used in the practical dosimetry: **the dose of** 1 R **collects for** 1 **hour at distance of** 1 m **from the radiant of radium by mass of** 1 g**, that is activity about** 1 $Curie$.

Power unit of exposition dose is 1 A/kg and standalone unit is 1 R/s. As the absorbed dose is proportional to falling ionizig radiation between the exposition and absorbed doses, there should be proportional dependence $D = f \cdot X$, where f is the certain transitive coefficient dependent on lines of reasons and, first of all, from irradiated substance and energy of photons. It is easy to count up value of f, if irradiated substance is air. It is established, that for air for the exposition dose of 1 R corresponds the absorbed dose equal to 0.88 rad. In this case $D = 0.88 \cdot X$, $f = 0.88$. For water and soft tissues $f = 1$, hence the **absorbed dose in rads is equal to exposition dose in roentgens**.

Action of radiation on tissues of organism depends not only from the general absorbed dose, but also from capacity of radiation. For dot sources of radiations capacity of exposition dose decreases with distance under the law: $P = K_\gamma \dfrac{A}{R^2}$, where K_γ is ionizing constant or γ-constant of the radioactive isotope, dependent on its nature. Thus, *degree of influence of radiation on a human organism depends on nature of radioactive isotope (K_γ), its activity (A) and distance (R) up to the source.*

Gamma - constant is capacity of dose of radiation in R/s, created by γ-beams of the given radioactive isotope on distance of 1 cm from a dot source, if its activity is 1 mCy. The exposition dose in this case can be estimated from the

parity $X = K_\gamma \dfrac{A}{R^2} \cdot \Delta t$, where Δt is time of irradiation.

23.3. Quantitative estimation of biological action of ionizing radiation. Equivalent dose. Equivalent effective dose. Collective dose.

For protection against radioactive radiation is important to know its influence on a living tissue. For any kind of radiation biological action is usually than more, than is more absorbed dose. However experiment shows, that action of nuclear radiations on a tissue of living organism is defined not only by the dose, but also by the *nature* of ionizing radiation. Heavy particles (α - particles, protons, neutrons, fast ions) make more physiological infringements, than easy (β, γ and X-rays). Strongly penetrating streams of neutrons are especially dangerous.

In dosimetry it is accepted to compare biological effects of various kinds of radiations with the corresponding effects caused by *X-ray or γ - radiation.*

Distinctions in value of radioactive influence can be taken into account, having attributed to each radiation the *quality coefficient of radiation (K).* X-ray radiation, γ - quantums and β- particles affect an organic tissue approximately equally and for these $K = 1$. For α-particles $K = 20$, i.e. it is considered that α-particles in 20 times more dangerous at hit inside of organism, than γ- radiation. For protons and neutrons $K = 10$, etc.

In radiobiology instead of coefficient of quality is used the **relative biological efficiency** (RBE). It is equal to the ratio of the absorbed dose standard (X-ray, γ) radiation causing the certain biological effect to the absorbed dose of the investigated kind of radiation, giving the same effect. This value characterizes also quality of radiation, therefore $RBE = K$.

For account of affecting action on organism of diffeferent kinds of radiations (with their coefficients of quality) is applied the concept of **equivalent dose** (D_e), which is connected with the absorbed dose by parity: $D_e = K \cdot D_{abs}$.

As is dimensionless coefficient, the equivalent dose of radiation has the same dimension as the absorbed dose, however is called as **Sievert** (Sv) (R.Sievert is the Swedish radiobiologist). Stand-alone unit of equivalent dose is **ber** (biological equivalent of roentgen or **rem**):

$$1\ rem = 10^{-2}\ Sv; \quad 1\ Sv = 100\ rem.$$

The equivalent (biological) dose in rem is equal to the absorbed dose in rad, multiplied on K:

$$D_e(ber) = K \cdot D(rad).$$

The equivalent dose is calculated for "average" tissue of human body. But doses should be determined and for separate bodies. In particular, it is necessary in radiation therapy of tumours, when it is not required to irradiate all body.

In relation to ionizing radiations organs and biological tissues have different radiosensitivity. The marrow and genitals are most strongly affected and, for example, nervous tissue is rather steady against radiation.

Account of radiosensitivity make with help of **coefficients of radiating risk** (FR). Values of these factors for tissues and bodies of the person at uniform irradiation of all body are resulted in the table:

Red marrow – 0.12

Bone tissue – 0.03

Thyroid gland – 0.03

Mammary glands – 0.15

Lungs – 0.12

Ovary or testicle – 0.25

Other bodies – 0.30

Organism as whole – 1.00

If FR for mammary gland is 0.15, it means that the irradiation of mammary gland by the dose of 1 Sv results to the same radiation damage of organism, as irradiation by the dose of 0.15 Sv of all body.

Thus, if it is known, what bodies and with what doses are irradiated (it is especially important to know at receiption of radioisotopes with food, water, inhalation of air with the subsequent accumulation in the certain bodies), then, having multyplied equivalent doses on corresponding coefficients of risk and putting on all bodies and tissues, we shall receive the *equivalent effective dose* (EED), reflecting total effect of individual irradiation for an organism. It is measured in Sievert also.

Knowing individual doses and putting them on group of irradiated people it is possible to receive the *collective effective equivalent dose in person-Sievert.* The collective dose can be calculated for separate settlement, area and republic. Thus, collective dose is objective estimation of scale of radiating defeat. If any number of people continues to live on polluted by radionuclides territory in conditions of

long influence of radiation and laws of change of radiating influence are known, it is possible to calculate *expected collective effective equivalent dose* (it is measured in person-Sievert also) for the certain forthcoming period of time. For example, as result of failure on Chernobyl there was the pollution of significant territory by radio-activity of complex isotope structure. The estimation of expected collective dose in view of breaking up radionuclides is important for forecasting adverse consequences for living and future generations and serves as the reference point for decision-making.

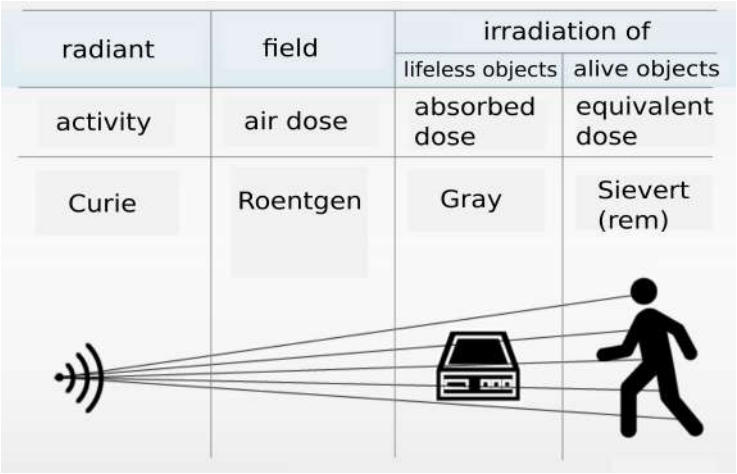

Fig. 23.1. The radiation units.

In conclusion of the paragraph we summarize sense of each concept and field of its application (fig.23.1).

Radiation hazard of used radiomaterial is convenient to estimate on **activity**, expressed in **Curie** or **Becquerels** (last unity is very small and consequently is practically inconvenient; normally it is used at measurings of samples of hazardous environment, which specific and aggregate activity insignificantly differs from background, stipulated by natural radioactive nuclides). Knowing the source activity it is possible to calculate **power of air dose (power of exposition dose)** for different distances from it and thus to spot, for example permissible residence time in this field.

The air dose (exposure) characterizes *radiation field* on its ionizing power, which is stipulated by character of radiomaterial or other radiant of ionizing radiation. For transition from air dose (characteristic of the field) to the **absorbed dose** (characteristic of interaction of the field and irradiated mediuim) it is necessary to know properties of this mediuim. At the same air dose, i.e. the same field, to water will be transferred smaller energy, than to material of the middle of the table of Mendeleev and the more so to heavy elements. The absorbed dose,

i.e. the energy absorbed in mass unit of material, on which the radiation field reacts, characterizes radiative effect for all kinds of physical and chemical bodies *except for living organisms.*

For estimation of radiation effect on living organisms, first of all for the person, it is offered and is used **the equivalent dose** of irradiation. In some practically frequently meeting events instead of D_e are used D and X. For the mixture of radiations at external and especially at internal irradiation only using of D_e allows to avoid errors in estimation of degree of radiation hazard of irradiation.

23.4. Doses of natural irradiation.

On biosphere of Earth continuously operates space radiation and also streams of α-, β- particles, γ- quantums as result of radiation of various radionuclides, situated in the earth's crust, water of underground sources, in the rivers, seas, oceans, in air. Besides, radionuclides are part of living organisms. Set of radiations of these radioactive sources is called *natural radioactive radiation.* The most widespread on the Earth from radionuclides are ^{220}Rn, ^{222}Rn and ^{40}K and also radionuclides, making the line of uranium.

Isotope of radon ^{222}Rn decays and gives α- radiation, which is accompanied by emission of γ - photon.

Mass of stable ^{40}K always contains about 0.01 % of isotope ^{40}K, which nucleuses after decay forms ^{40}Ca, β - and γ- radiation. Isotope of K contains in ground, fertilizers and also in a brain, muscles, spleen and in a marrow. So, at a person of mass 70 kg contains about 0.021 g of radionuclide ^{40}K. The half-life period of ^{40}K is $1.3 \cdot 10^9$ years. It is easy to calculate, that every second in our organism decay $5 \cdot 10^3$ atoms of ^{40}K. But it does not represent for us any danger and, apparently, is necessary for development of organism, as origin and development of life on the Earth were always accompanied by this process.

Space radiation will consist of streams of protons, α- particles, nucleus of some elements, streams of electrons, photons and neutrons. Particles of high energy, cooperating with atmosphere form as result of nuclear reactions series of radionuclides of ^3H ,^7Be ,^{22}Na and streams of neutrons and protons. This secondary radiation will penetrate into the bottom layers of atmosphere and influences biosphere.

As result of this natural both external and internal irradiation average capacity of dose makes about of 2 mSv per year (or 200 $mber$). And approximately of

2/3 of these dose (≈ 135 *mber*) the person reserves from the radioactive isotopes, which have got in organism with food, water, air (an internal irradiation) and 65 *mber* from external irradiation. It is important to note, that natural radioactive background, influencing development on life on the Earth is integral part of sphere of dwelling of the person. Infringements of radioactive background are dangerous for existence of biosphere and can result to irreparable consequences.

One of the reasons of increase of radioactive background is human activity. Creation of the large industrial enterprises, power sources, military technics, etc. can result in local changes of background. But the most dangerous reasons are emissions of radioactive particles, which can arise at nuclear explosions or at operation of atomic power stations. So, for example, at failure on the Chernobyl atomic power station there were emissions of radionuclides: ^{131}I (half-life period $T = 8$ days, it gives γ - radiation), ^{90}Sr ($T = 29$ years, gives β - radiation), ^{137}Cs ($T = 30$ years, gives β - and γ- radiations). These isotopes can collect in organism causing in him infringement of activity, both separate organs and organism as whole. So, ^{131}I collects in the thyroid gland and already 0.35 mg of radioactive iodine is dangerous to life (daily need of stable iodine about 150 mg). The isotope ^{90}Sr collects in a bone tissue and the isotope 137 Cs is in regular intervals distributed in cells of organism.

The maximum permissible biological dose for a person at professional irradiation considers 5 *ber per year. For the population is established the limiting dose in* 10 *times smaller:* 0.5 *ber per year. Minimal lethal dose is conditionally accepted* ≈ 600 *ber at irradiation of all body.*

Sometimes the radioactive background is estimated on capacity of radiation. So the *normal natural background should not exceed of* 20 $\mu R/h$. For the areas undergone to radioactive pollution as result of Chernobyl failure norms have been established: for zone of evacuation is 5 mR/h and for zone of alienation is 20 mR/h. Maximum permissible specific activity of the polluted area is considered equal to 15 Cy/km^2. Norms of specific activity of radionuclides in food stuffs are established also: grain for bakeries is $1.6 \cdot 10^{-8}$ Cy/kg; flour, groats is $1 \cdot 10^{-8}$ Cy/kg; children's feed of all kinds is $1 \cdot 10^{-8}$ Cy/kg.

23.5. Dosimetric devices.

Devices for measurement of doses of ionizing radiations or the values connected with doses are called *dosimetric devices or dosimeters*. The principle of action of

dosimeters is submitted on the circuit: detectors → the device of processing of information → indication of results.

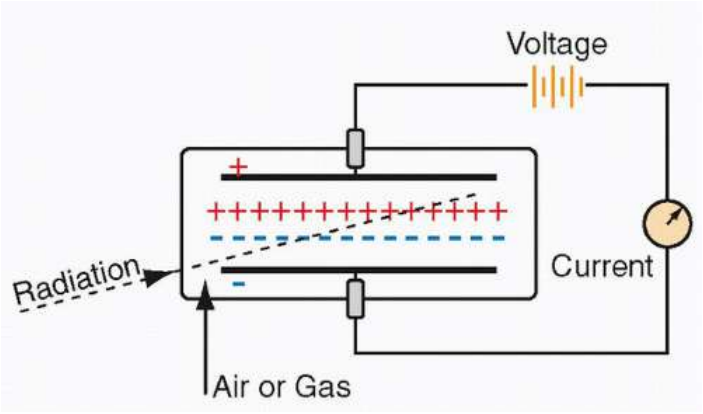

Fig. 23.2. The ionizing chamber.

In these devices for measurement of radiations always there are devices named *detectors*, in which energy of ionizing radiations will be transformed to electric signal. As in absorbing substance the particle or photon spend the energy for formation of a charge from ionization, on value of electric signal it is possible to judge about their energy and about quantity of the registered acts of decay.

There are various detectors of radiations. The most widespread detectors are: 1) **ionizing chamber**; 2) **counter of Geiger-Mulier**; 3) **semi-conductor and scintillometer detectors**.

1) In ionizing chamber as absorbing substance is served certain gas in the space between two electrodes (fig.23.2). Particles getting into the chamber and photons of radiation cause occurrence of a current. The current is proportional to number of ions formed in the chamber per second and, hence, energy flow of transiting ionizing particles. Such chambers use, in particular, in pocket dosimeters (fig.23.3).

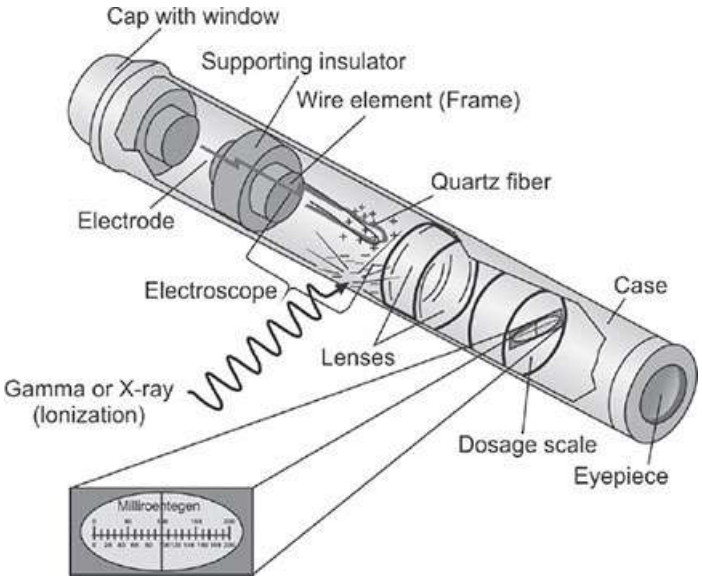

Fig. 23.3. The pocket dosimeters.

In practice of radiation monitorings the greatest distribution have received

thimble ionization chambers though frequently the sizes and the shape of such chambers reminds a thimble a little.

Such chamber can be viewed as a small vacuity filled with gas inside a solid body. As the effective atomic number of tissues of a human body $Z_{body} = 7.42$ is close to effective atomic number of air $Z_{air} = 7.64$, that it allows to determine absorbed energy for tissues of the body by results of measurings the ionization produced by explored radiation in air (walls of the chamber also produce from «tissue – equivalent» materials: polysterene, perspex, etc.).

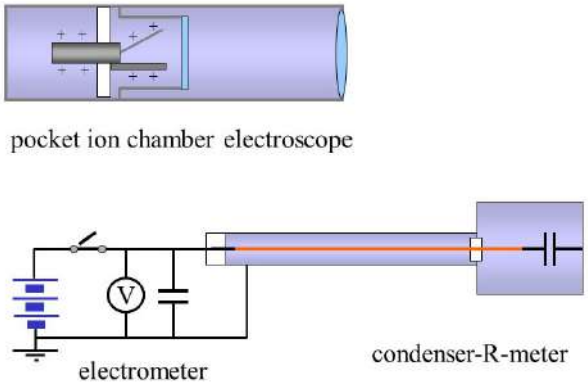

Fig. 23.4. The diagram of such chamber.

The charge gathered on electrodes of such chamber is directly proportional to the air dose. On fig.23.4 the diagram of such chamber is shown: the wall of it is one electrode and the rod included in it is the measuring electrode. If to close key S, then chamber and connected with it in parallel condenser and electrometer (electroscope) are charged up to voltage of U_1 from the source $\boxed{HT}$. Measuring begins at breaking of key S, then as the result of ionization the voltage of the measuring electrode decreases up to value U_2. Voltage $\Delta U = U_1 - U_2$ digitize on the electrometer. The charge Δq arisen owing to ionization is proportional to ΔU: $\Delta q = C \cdot \Delta U$, where C is capacity of the chamber. Corresponding air dose X:

$$X = \Delta q/V = C \cdot \Delta U/V = k \cdot (U_1 - U_2),$$

where V is volume of the chamber, k is coefficient of proportionality determined at graduation of the device. So: $X \sim \Delta q \sim \Delta U$. The main deficiency of ionization chambers is their rather low sensitivity, therefore them apply in fields of the considerable radiation intensity.

The gas-discharge counters frequently termed also as **Geiger-Muller** counters (fig.23.5), differ from ionization chambers the greater sensitivity and are capable to register individual pair of ions. By principle of the device such counter does not differ from ionizing chamber: it also is the condenser, on which the potential difference is applied, however it is so great, that in the gas gap originates new process – gas amplification that is sharp magnification of initial quantity of ions

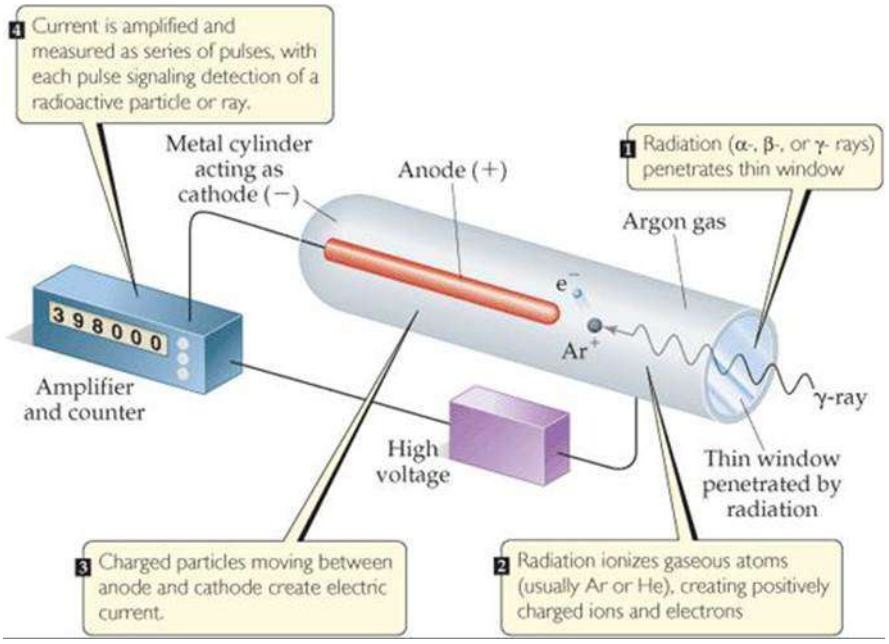

Fig. 23.5. The Geiger-Muller tube.

($\approx$ in 10^7 times). The cylindrical Geiger-Muller counter will consist of coaxially posed electrodes: the anode (the thin wire tensioned along an axis); the cathode as sprayed on the glass tube metal; contact; and insulators. Pressure of gas inside the counter is about of 100 $mm\ Hg$. To electrodes are applied a voltage at some hundreds volts. At hit in the counter of ionizing particle, in gas are formed mobile electrons which move to the anode. As the wire is thin (diameter is about of 0.5 mm), then near to wire the field is strongly nonuniform, also intensity is great. Electrons near to wire are sped up so, that start to ionize gas. As result there is a discharge and on the cicuit the current proceeds. The formed self-sustained discharge is necessary to destroy, as the counter will not react to the following particle. To discharge quenching apply: 1) inserting in series with the counter of high-resistance, on which there is the considerable voltage reduction, therefore the voltage on the counter decreases and then the discharge stops, 2) special gas filling up (argon and alcohol, halogens) giving in interruption of the discharge even at small resistances in the circuit. Limitation of a geiger: sharp dependence of efficiency of recording on energy of incident radiation («course with stiffness»).

γ- quantums in gas seldom make ionization. In this case apply semi-conductor detectors or scintillometer. In the semi-conductor detector absorption occurs in a semi-conductor material.

In scintillometer (fig.23.6) passage of γ- radiation causes fight flashes, which will be transformed to electric pulses and amplify the photoelectronic multiplier

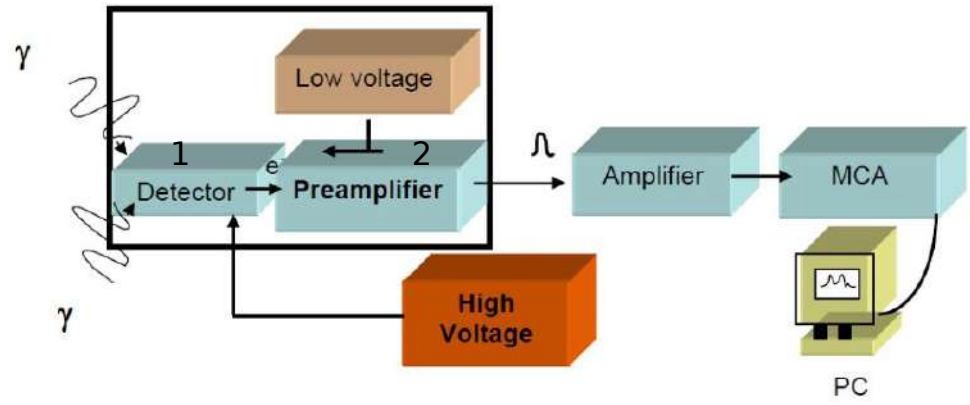

Fig. 23.6. The scintillometer.

2. For each kind of radiation select the suitable detector 1. In the capacity of detectors use monocrystals: $NaJ(Te)$, $CsJ(Na)$ – γ, β- radiation, $ZnS(Ag)$ – in α- detectors, etc.

To **destination** all dosimetric devices are subdivided on the following groups:

1. Indicators: elementary devices of radiation survey for detection and rough estimation of dose capacity.

2. Roentgenometers are intended for measurement of D_{exp}.

3. Radiometers are intended for measurement of activity or density of stream of radiation. With their help determine degree of radioactive pollution of the ground, air by α - β particles.

4. Dosimeters are intended for determination of dose and capacity of dose mainly of X-ray and γ - radiations.

5. Spectrometers are intended for reception of information on a spectrum of radiation of the person on energy of quantums or particles, to amplitude of signals, etc. parameters. Installation **SRP** (spectrometer of radiation of the person) is practically unique means of authentic estimation of doses of internal irradiation of a person.

Test questions

1. How is the charge of an atomic nucleus related to the ordinal number of a chemical element in D.I.Mendeleev's table?

2. Why does the nuclear charge determine the chemical properties of an atom?

3. What are isotopes? Why isotopes of one chemical element have the same chemical properties?

4. On the basis of what factors can we conclude that nuclear forces exist as special forces of interaction of nucleons in an atomic nucleus?

5. Why are the proton and neutron often considered as one nuclear particle - a nucleon - in two different states? Why is the mass of an atomic nucleus not equal to the sum of the masses of free protons and neutrons, of which it consists?

6. Give a definition of the main characteristics of ionizing radiation: equivalent dose, equivalent effective dose, collective dose.

7. What is dosimetry?

8. Explain what is doses of natural irradiation?

9. Where are used dosimetric devices?

Lecture No 24

Basis of mathematical analysis, probability theory and mathematical statistics

24.1. The elements of differential and an integral calculation

24.1.1 Set theory

Any set or totality of any elements joined by a common feature (sign, characteristic) is named in mathematics as a **set** or an **ensemble**.

For example, any community can be considered as a set. Let's consider people which are taking place in terrain of any city. It is **set** in our examination.

It is possible to select a **subset** of male persons, a subset of female persons, a subset of inhabitants and a subset of visitants, a subset of workers, a subset of students, a subset of teachers, etc.

The given examples are examples of **terminating sets**, but in mathematics frequently use also the **infinite sets**: a set of natural numbers (it is designated **N**), a set of integers (is designated **Z**), etc.

Sets are designated by title Roman letters **A**, **B**, **C**, **D**, **E**, ...

If **A** is a set of objects, and **x** is one of these objects, speak, that **x** there is a member of a set **A**.

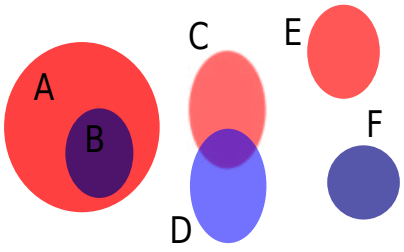

Fig. 24.1. The Sets.

Fig.24.1 displays conventional presentation of sets, where points on plane represent elements of sets. For example **B** is subset towards to set **A**; **C** and **D** are set with common elements – they are intersected sets; **E** and **F** are not intersected sets.

The mathematics explores the abstract numbers and their relations, spatial shapes in the quantitative description, and also the more composite objects giving in to the quantitative description and having regular interrelations.

In practice laws of mathematics are applied to studying objects by which any numerical values can be compared.

24.1.2 Numbers and values

Mathematical quantity (value) is relation of arbitrary value to standard same nature value (physical, geometrical) which accept for a unit.

Constant is value with definitional domain (range of definition) consist of one number.

Variable quantity is value that changes under condition of given task.

Parameter (quantity, rating, coefficient, factor) is value that constant under condition of given task.

Absolutely constant numbers. Example: $\pi = D/L = 3,14$;

Natural numbers: $1, 2, 3, \ldots, +\infty$ (integer positive). Mark as **N**.

Whole numbers, integer: $-\infty, \ldots, -2, -1, 0, 1, 2, 3, \ldots, +\infty$. Mark as **Z**.

Real number (from $-\infty$ to $+\infty$; integer positive and negative, 0, broken (fractional) numbers, rational and irrational).

Rational are periodical decimal fractions. Mark as **Q**.

Real number $\mathbf{a}$ is rational if such integer numbers $\mathbf{n_1}$ and $\mathbf{n_2}$ exist that $\mathbf{n_1} \neq \mathbf{n_2}$ and $\mathbf{a} = \mathbf{n_1}/\mathbf{n_2}$. Other real numbers are irrational (example: $\sqrt{2}$ hasn't period).

24.1.3 Function

The quantitative relations between separate numerical sets are termed as functional dependences or functions.

Let two sets D (ensemble of elements marked as X) and (ensemble of elements marked as Y) are given. And let to each set member X from set D some set member Y from set E is placed univocally in correspondence.

It means, that function $\mathbf{y} = \mathbf{f}(\mathbf{x})$ on set D with values to set is specified.

Function $\mathbf{y = f(x)}$ is correspondence between two sets where each member of set X corresponds no more than one member of set Y under the certain law or a rule.

Means of function's assignment

1. A tabular means.

For example:

Table 24.1

Tabular method of defining a function.

X	-4	-3	-2	-1	0	1	2	3	4	5	6
Y	16	9	4	1	0	1	4	9	16	25	36

2. A graphic means.

The set of all points of a coordinate plane which abscissa are equal to values of argument, and ordinates are equal to corresponding values of function refers to as the graph of function.

3. An analytical means - with the help of the formula.

For example: $\mathbf{y = 3.14 \cdot x^2}$; $\mathbf{y} = \cos(\mathbf{x})$; $\mathbf{y} = \ln(\mathbf{x})$;

$\mathbf{y} = \sin(\mathbf{x^2 + 1.237}) \cdot \cos(\mathbf{x^2 - 1.237})$.

4. A descriptive means.

Range of definition and range of values of function

Totality of all values of argument x for which function $y = f(x)$ is defined, is termed as **range of definition** $D(f)$ of this function, and a totality of all values, accepted by a variable y, is termed as a **range of values of function** $E(f)$.

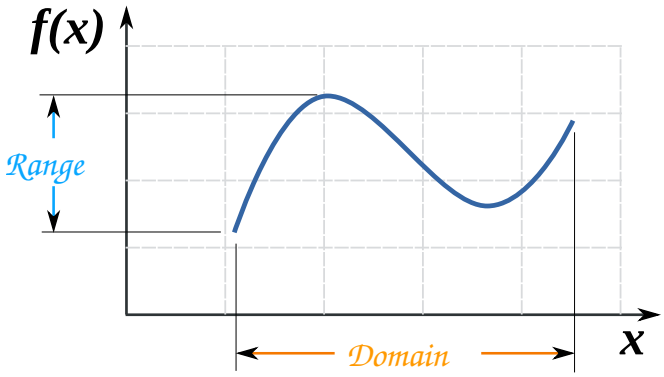

Fig. 24.2. Function.

Fig.24.2 displays function and its range of definition $[x_1; x_2]$ (domain) and range of function values $[y_1; y_2]$ (range).

24.1.4 Properties of functions

Even and odd functions.

Function $\mathbf{y = f(x)}$ is termed even if its values corresponding to any two opposite values of argument from range of function definition $\mathbf{D(f)}$, are equal, i.e.the equality $\mathbf{f(-x) = f(x)}$ is fulfilled.

Function $\mathbf{y = f(x)}$ is termed odd if its values relevant to any two opposite values of argument from range of definition $\mathbf{D(f)}$ of function, are opposite, i.e. the equality $\mathbf{f(-x) = -f(x)}$ is fulfilled.

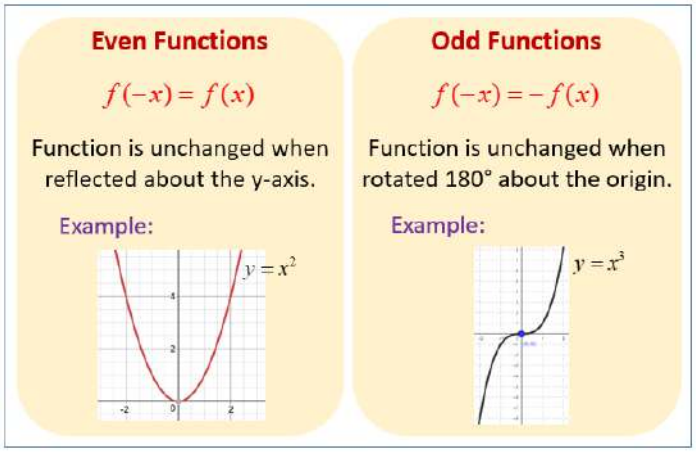

Fig. 24.3. Even and odd functions.

Periodic functions.

Function $y = f(x)$ refers to periodic if for it there is such positive number $l > 0$, that at any value of argument x numbers $x - l$ and $x + l$ belong to range of function definition $D(f)$, and next equalities are fulfilled: $f(x - l) = f(x) = f(x + l)$.

In this case the number l refers to period of function $f(x)$. If the number l is period of function $f(x)$, numbers nl will be its periods also at any whole $n \neq 0$. On fig.24.4 sine function has period $l = 2\pi$.

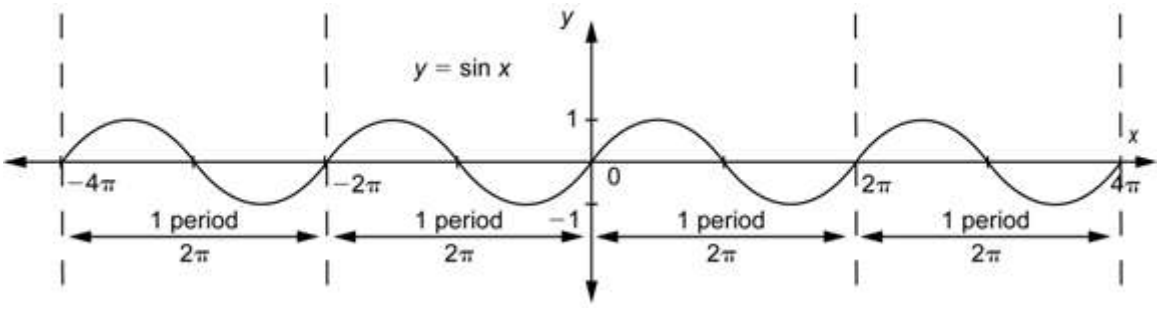

Fig. 24.4. Periodic and limited function.

Limited function.

Limited function is such: if exist number M such as $|f(x) \leq M, x \in D(f)$.

Example: $f(x) = sin(x)$, $-1 \leq f(x) \leq 1$ (see fig.24.4).

Inverse function

Inverse by function conformity can not be function.

Inverse function is conformity that inverse by function.

Graphs of function and inverse function are symmetry by straight line $y = x$.

Before to consider continuity of function, we shall remind rules of definition of limits.

Limit of function.

Limit of function $y = f(x)$ is the terminating number A at tend x to a if for each number $\varepsilon > 0$ will be such number $\delta > 0$, that $|y - A| < \varepsilon$, if $|x - a| < \delta$. Notation: $\lim\limits_{x \to a} y = a$.

For example, function $y = x^2 + 2$ at $x \to 3$ has a limit 11: $\lim\limits_{x \to 3} (x^2 + 2) = 11$.

There are functions which have no a limit in some points. For example, function $y = 1/x^2$ at $x \to 0$ grows unlimitedly and has no limit.

There are functions which have a limit at tend of argument to infinity: $\lim\limits_{x \to \infty} y = A$. For example $\lim\limits_{x \to \infty} \dfrac{1}{x} = 0$.

The basic theorems of limits.

1. The limit of a constant is equal to this constant: $\lim A = A$.

2. Limits of the sum (difference) of terminating number of functions it is equal to the total (difference) of limits of these functions: $\lim\limits_{x \to a}[f(x) + \varphi(x) + \psi(x)] = \lim\limits_{x \to a} f(x) + \lim\limits_{x \to a} \varphi(x) + \lim\limits_{x \to a} \psi(x)$.

3. The limit of product of terminating number of functions is equal to product of limits of these functions: $\lim\limits_{x \to a}[f(x) \cdot \varphi(x) \cdot \psi(x)] = \lim\limits_{x \to a} f(x) \cdot \lim\limits_{x \to a} \varphi(x) \cdot \lim\limits_{x \to a} \psi(x)$.

4. The limit of quotient of two functions is equal quotient of limits of these functions if only the limit of a denominator is not equal to zero: $\lim\limits_{x \to a} \dfrac{f(x)}{\varphi(x)} = \dfrac{\lim\limits_{x \to a} f(x)}{\lim\limits_{x \to a} \varphi(x)}$.

If limits of the functions which are in numerator and in a denominator, both are equal ∞ or both are equal 0, there is **an indeterminacy**. The estimation of such limit is termed as **disclosing of indeterminacy**.

Example:

$$\lim_{x\to\infty}\frac{x+7}{3x+8}=\lim_{x\to\infty}\frac{1+7/x}{3+8/x}=\frac{\lim\limits_{x\to\infty}(1+7/x)}{\lim\limits_{x\to\infty}(3+8/x)}=\frac{1}{3}.$$

The limit of the relation of a sine of a corner to a corner is equal 1 and is termed as a remarkable limit: $\lim\limits_{x\to\infty}\dfrac{\sin x}{x}=1$

The transcendental number $e=2,71828\ldots$ is a limit $e=\lim\limits_{x\to\infty}(1+\dfrac{1}{x})^x$. It also is termed as **a remarkable limit**.

The number e is used as a Napierian base (natural logarithm base) which are designated by a figure ln: $\log_e a=\ln a$.

Correspondingly, logarithm to the base 10 is called decimal (or common) logarithm. It is denoted $\lg x=\log_{10} x$.

Connection between natural and decimal logarithms: $\ln a\approx 0,43\lg a;\ 2,3\lg a\approx\ln a$.

Continuity of function

Function $y=f(x)$ is the continuous in a point $x=a$ if the limit of function in a point and is equal to value of function this point: $\lim\limits_{x\to a}f(x)=f(a)$.

Function, that continuous in each point of a segment $[a,b]$ is termed as the continuous on this segment.

The point in which the continuity condition of function does not satisfy, is termed as a point of discontinuity or break point.

The further problems of derivation and integration are applicable without stipulations to continuous functions.

24.1.5 Derivative of function.

Derivative $y'(x)$ of function $y(x)$ in a point is the limit of the ratio of an increment of a function to an increment of argument when the increment of argument tends to zero, i.e. $\mathbf{y'=\lim\limits_{\Delta x\to 0}\dfrac{\Delta y}{\Delta x}=\lim\limits_{\Delta x\to 0}\dfrac{f(x_0+\Delta x)-f(x_0)}{\Delta x}}$. Equivalent notation: $\mathbf{y'=\dfrac{dy}{dx}}$.

Function $y(x)$ **refers to antiderivative (or primitive, or original) in relation to a derivative** $y'(x)$. The mark «$'$» designates a derivative.

A second derivative from y on x is designated as: $\mathbf{y''=\dfrac{d^2y}{dx^2}}$ or $\mathbf{y''(x)}$.

Derivatives of the second and the superior orders are received from derivatives of the previous order by the same rules as the derivative of the first order is received out from antiderivative functions.

A third derivative from y on x is designated as: $\mathbf{y'''=\dfrac{d^3y}{dx^3}}$ or $\mathbf{y'''(x)}$.

24.1.6 Geometrical meaning of the derivative

The derivative of function in point X is equaled to a tangent of a slope angle of a tangential straight line in this point, and function $f'(x)$ – dependence of magnitude of this tangent on coordinate X.

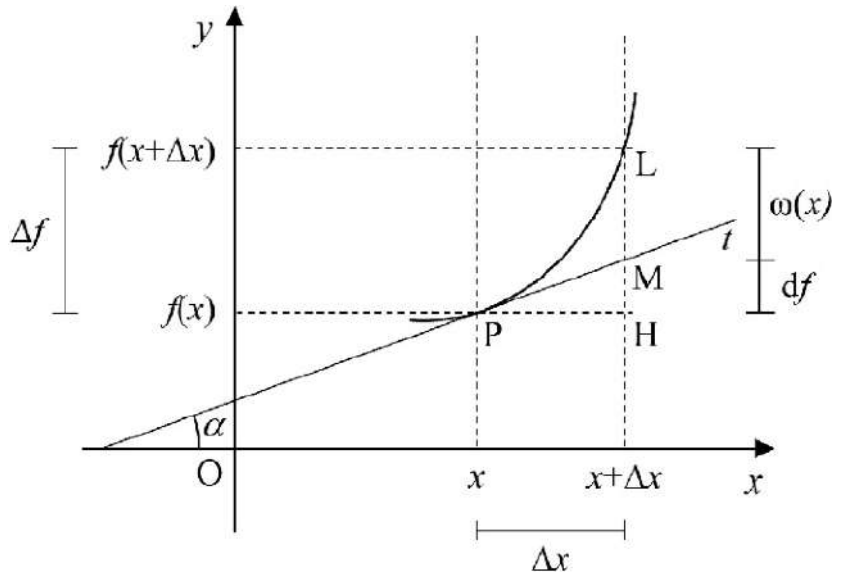

Fig. 24.5. Function and its derivative in point x.

Value of the derivative y' at every point x is equal to the tangent of the angle α between a tangent line to the graph of function $y = f(x)$, carried out at the point x, and an X-axis (on fig. 24.5 $y' = tg\alpha$). In other words, the geometrical meaning of derivative of a function is that it is equal to the slope of the tangent.

You can see on derivative graphs that maximal derivative values are observed in the time points with maximal slope of raising of corresponding graph.

Minimal derivative values are observed in the time points with maximal slope of decline of corresponding graph.

Zero derivative values correspond to the time points with horizontal tangential straight line, which are minimums and maximums of corresponding graph.

24.1.7 Physical sense of the derivative

Example of mechanical sense of the derivative

If the traversed path of a body is known as function of time t, i.e. $x = f(t)$, then the instantaneous (scalar) velocity of a body v at present time t is equal to a derivative from paths on time $\mathbf{v} = \lim\limits_{\mathbf{t_2 \to t_1}} \dfrac{\mathbf{\Delta S}}{\mathbf{\Delta t}} = \lim\limits_{\mathbf{\Delta t \to 0}} \dfrac{\mathbf{\Delta S}}{\mathbf{\Delta t}} = \dfrac{\mathbf{dS}}{\mathbf{dt}}$, and an acceleration is equal a derivative from a velocity on time $\mathbf{a} = \dfrac{\mathbf{dv}}{\mathbf{dt}}$, or a second

derivative from paths on time: $\mathbf{a} = \dfrac{\mathbf{d^2S}}{\mathbf{dt^2}}$. Here $\mathbf{S} = \mathbf{\Delta x}$ (increment of coordinate), $\mathbf{\Delta t} = \mathbf{t_2} - \mathbf{t_1}$ (increment of coordinate).

Examples of electrical sense

Derivative is the amount of an electrical charge which has passed through a cross-section of a conductor for a time unit, determines a current intensity (here Q – charge, t – time): $\mathbf{I = dQ/dt}$. Other examples are linear density of electrical charge $\tau = \mathbf{dQ/dl}$ (here Q – charge localized on linear or cylindric line, l – length of the line) or density of electrical current $\delta = \mathbf{dI/dS}$ (here I – current intensity from first example, S – area of cross-section of a conductor).

Physiological sense of the derivative

On the images there are results of simultaneous recordings electrocardiogram, rheograms from right and left parts of a head, and derivative of right-side and left-side rheograms. These records are received on different patients with various pathologies.

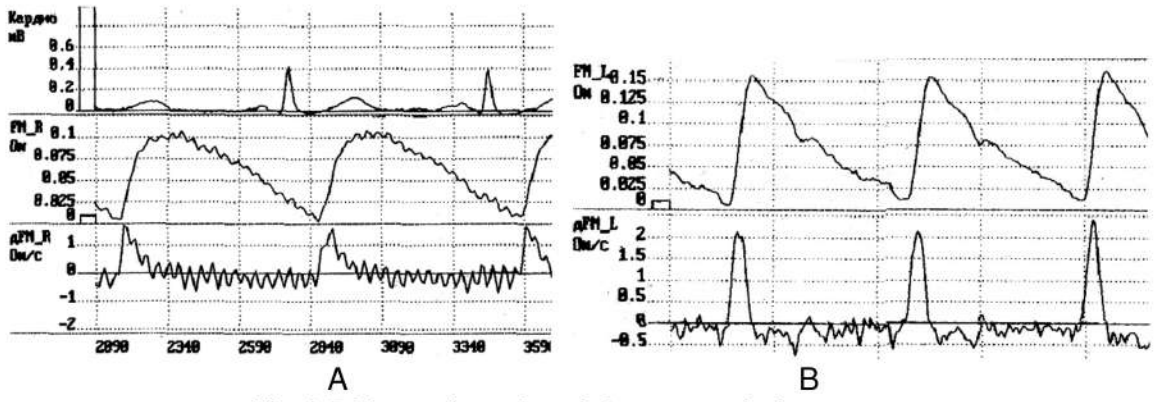

Fig. 24.6. Examples of real rheoencephalograms.

Fig.24.6 (A) displays ECG, right-side REG and its derivative; fig.24.6 (B) displays left-side REG and its derivative. Registrations were made at patients with different pathologies.

Maximal derivative values are observed in the time points with maximal slope of rising of corresponding rheogram on these graphs.

Minimal derivative values are observed in the time points with maximal slope of decline of corresponding rheogram.

Zero derivative values correspond to the time points with horizontal tangential straight line that is minimums and maximums of corresponding rheogram.

This allows to calculate easy many necessary parameters: duration of blood flow rising, time between critical points; to deduce about influence of heart work

and blood vessel wall state on the blood supply of tissue under consideration.

Derivation of functions

Operation of a determination of derivatives or differentials of functions refers to as derivation.

The common rule of a determination of a derivative coincides with its definition: an increment of a function $\Delta y = f(x + \Delta x) - f(x)$ to divide into an increment of argument Δx, and then to find a limit of a quotient at $\Delta x \to 0$.

Example. To find a derivative of function $y = x^2 + 2$.

Let's express an increment of a function $\Delta y = ((x + \Delta x)^2 + 2) - (x^2 + 2) = 2x\Delta x + \Delta x^2$. Let's divide Δy on Δx.

Will find $y' = \lim\limits_{\Delta x \to 0}(2x + \Delta x) = \lim\limits_{\Delta x \to 0} 2x + \lim\limits_{\Delta x \to 0} \Delta x = \lim\limits_{\Delta x \to 0} 2 \cdot \lim\limits_{\Delta x \to 0} x + \lim\limits_{\Delta x \to 0} \Delta x = 2x$

As the derivative is found in the given point, therefore is considered x as constant not related from variable Δx.

Table 24.2

The table of derivatives of some functions

y	y′	y	y′
C(const)	0	tgx	$\sec^2 x = \frac{1}{\cos^2 x}$
x^μ	$\mu x^{\mu-1}$	$ctgx$	$-^2 x$
x	1	$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$
$\frac{1}{x}$	$-\frac{1}{x^2}$	$\arccos x$	$-\frac{1}{\sqrt{1-x^2}}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$	$arctgx$	$\frac{1}{1+x^2}$
$\sqrt[n]{x}$	$\frac{1}{n\sqrt[n]{x^{n-1}}}$	$arcctgx$	$-\frac{1}{1+x^2}$
a^x	$a^x \ln a$	$\log_a x$	$\frac{\log a^e}{x} = \frac{1}{x \ln a}$
e^x	e^x	$\ln x$	$\frac{1}{x}$
$\sin x$	$\cos x$	$\lg x$	$\frac{\lg e}{x} \approx \frac{0.4343}{x}$
$\cos x$	$-\sin x$		

Rules of derivation

1. The derivative of the constant is equal to zero. $C' = 0$.

2. The derivative of the sum (difference) of functions is equal to the sum (difference) of derivative these functions: $y = u + v$; $y' = (u + v)' = u' + v'$.

For example, $y = e^x + x^7$, then $y' = e^x + 7x^6$.

3. The derivative of product of two functions is equal to the derivative first function multiplied on the second function, a plus a derivative of the second

functions, multiplied on the first function: $\mathbf{y = uv}$; $\mathbf{y = u'v + uv'}$.

For example, $y = 9x^7 \sin x$, then $y' = 9 \cdot 7x^6 \sin x + 9x^7 \cos x = 63x^6 \sin x + 9x^7 \cos x$.

Therefore, derivative of product of constant on functions is equal to the constant multiplied on the derivative of function: $\mathbf{y = Cu}$; $\mathbf{y' = C \cdot u'}$, where C – constant.

In other words, a constant factor can be taken outside the derivative sign.

From rule of product there follows formula of derivative of a product of several factors. If function y is product, for example, of three functions

$\mathbf{y = u(x) \cdot v(x) \cdot w(x)}$ then $\mathbf{y = uvw}$; $\mathbf{y = (uvw) = u'vw + uv'w + uvw'}$

4. The derivative of a quotient is equal to fraction which denominator is quadrate of a divisor, and the numerator is a difference to product of a divisor on a derivative of a dividend and product of a dividend on derivative of a divisor: $\mathbf{y = \frac{u}{v}}$; $\mathbf{y' = \frac{u'v - uv'}{v^2}}$.

For example, $\mathbf{y' = \frac{x^3 + 12}{e^x + 2}}$, then

$$\mathbf{y' = \frac{(x^3 + 12)'(e^x + 2) - (x^3 + 12)(e^x + 2)'}{(e^x + 2)^2} = \frac{3x^2(e^x + 2) - e^x(x^3 + 12)}{(e^x + 2)^2}}$$

5. A derivative of a composite function.

Composite function $y(x)$ is such which represents function $y = f_1(u)$ if $u = f_2(x)$. Then the derivative y'_x (on x) is equal to product of a derivative y'_u (on u) on a derivative u'_x (on x): $\mathbf{y'_x = y'_u u'_x}$.

For example, $\mathbf{y = sin(x^2 + 12)}$. Substitute $\mathbf{y = sinu}$, where $\mathbf{u = x^2 + 12}$. Then $\mathbf{y' = (sinu)'u \cdot (x^2 + 12)'x = cosu \cdot 2x = 2xcos(x^2 + 12)}$.

Application of derivatives for research of function and construction of diagrams

Increase and decrease of functions

Increasing function $y(x)$ on an interval $[x_2; x_1]$ is such which grows at increase of argument of function. Increasing function - decreases at increase of argument.

Derivative of increasing function $y'(x) > 0$; accordingly a derivative of decreasing function $y'(x) < 0$ on an interval $[x_2; x_1]$.

Determination of extreme values of function

The maximum of function $y(x)$ in a point x_{max} is observed, if in adjacent, as much as close located points on the right and at the left, function takes on smaller values.

The minimum of function $y(x)$ in a point x_{min} is observed, if in next, as much

as close located points on the right and at the left, function takes on bigger values.

Extremum is the common name of a maximum and a minimum.

Conditions of a determination of an extremum of function:

1) A derivative of function $y^{'}(x) = 0$;

2) The sign of a derivative changes at transition through a point of an extremum.

In a point of a maximum the sign of the first derivative changes from "+" on "–", in a point of a minimum the sign changes from "–" on "+".

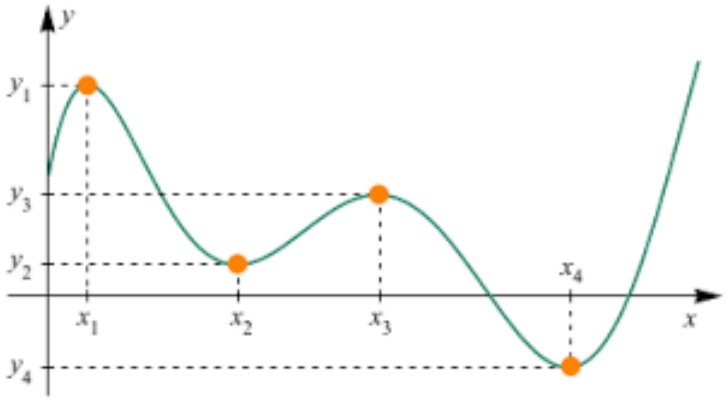

Fig. 24.7. Extremums of function.

In a point of a maximum the second derivative $y^{''}(x) = -1$; in a point of a minimum the second derivative $y^{''}(x) = 1$.

Points in which $y^{'}(x) = 0$, but the sign of the first derivative does not change at transition through this point, refer to as point of inflection. They are not extremum.

Construction of graphs of functions is facilitated if preliminary to find a derivative, then to find maxima, minima (by determination of points, in which $y^{'}(x) = 0$, and character of change of a sign $y^{'}(x)$ in these points) and intervals of increase and decrease of functions (on a signs of $y^{'}(x)$ in these points).

24.1.8 Differential of function

Initial expression for derivative making $\mathbf{y^{'} = \Delta y / \Delta x}$ can be noted as $\mathbf{\Delta y = y^{'} \cdot \Delta x}$. It refers to a differential of function and is designated as $\mathbf{dy = y^{'} \cdot \Delta x}$ or $\mathbf{dy = y^{'} \cdot dx}$. The differential of function is equal to product of a derivative on an increment of argument.

Geometrically the differential of function is an increment of ordinate of tangent Δy which corresponds to an increment of abscissa Δx.

The common rule of a determination of a differential corresponds to its definition and is spread on composite and composite functions.

1) Differential of function sum (difference):

$\mathbf{dy = (u \pm v)'dx = u'_x dx \pm v'_x dx = du \pm dv}$.

2) Differential of product of two functions:

$\mathbf{dy = d(uv) = vdu + udv}$.

3) Differential of a function quotient:

$\mathbf{dy = d(\frac{u}{v}) = \frac{vdu - udv}{v^2}}$

The concept of the **higher orders differentials** is connected to the concept of the higher orders derivatives.

Differential of the second and the superior orders are designated as:

$\mathbf{d^2y = d(dy) = d(y'dx) = d(y')dx = y''dx^2}$

Therefore $\mathbf{y''} = \dfrac{\mathbf{d^2y}}{\mathbf{dx^2}}$.

Similarly, differential of the third (and the superior) order is designated as: $\mathbf{d^3y = y'''dx^3}$; $\mathbf{d^4y = y^{(4)}dx^4}$ and so on.

Application of a differential of function for the approached evaluations

If the argument of function varies on small value dx, it is possible to find approximately an increment of a function $\mathbf{\Delta y \approx y' \cdot dx}$.

Example

Problem: How many the volume of sphere will change if as a result of heating its radius $R = 4$ m will increase on $\Delta R = 0.01$ m?

As volume of sphere $\mathbf{V = 4/3 \cdot \pi R^3}$, a differential

$\mathbf{dV = 4/3 \cdot \pi \cdot 3 \cdot R^2 dR = 4\pi R^2 dR = 2.0094\ m}$.

Application of a differential of function for the estimations of errors (the elementary theory of errors)

The approached values appear as a result of a rounding off of numbers (for example, the rounded off value $\pi \approx 3.14159\ldots$), but also as a result of measurements.

The precision of any measurements is limited, and measured value **a** is the approached value of true value **z**.

Value $\mathbf{a - z}$ refers to as **a true error of a**, the ration $\mathbf{(a - z)/z}$ – **a true relative error of a**. In most cases **z** is not known, accordingly, errors are unknown also. However, frequently it is possible to specify limits of a true error. This number – **a limit of error $\mathbf{\Delta a}$** (in other words, a **limiting absolute error**).

For it the inequality is right: $|\mathbf{a}-\mathbf{z}| \le \mathbf{\Delta a}$, i.e. $\mathbf{a}-\mathbf{\Delta a} \le \mathbf{z} \le \mathbf{a}+\mathbf{\Delta a}$.

Value $\delta\mathbf{a} = \mathbf{\Delta a}/\mathbf{a}$ refers to as a **limiting relative error**. The relative error $\delta\mathbf{a}$ is usually indicated in percentage.

For function of many variables $f(u,v,w)$ it is possible to determine a limiting absolute error Δf, if for each of variables u, v, w their limiting absolute errors Δu, Δv, Δw are known. For this purpose it is necessary to find a total differential of function df: $\mathbf{\Delta f} = |\frac{\partial \mathbf{f}}{\partial \mathbf{u}}|\mathbf{\Delta u} + |\frac{\partial \mathbf{f}}{\partial \mathbf{y}}|\mathbf{\Delta v} + |\frac{\partial \mathbf{f}}{\partial \mathbf{w}}|\mathbf{\Delta w}$.

Substitution of values Δu, Δv, Δw allows to receive required value.

Partial derivatives of the function of several variables

Total differential

If function u is function of several (any amount) independent variables – $u = f(x,y,z)$, – and one of arguments varies, whereas remaining arguments are constant, the partial difference is calculated so: $\mathbf{\Delta u} = \mathbf{f}(\mathbf{x}+\mathbf{\Delta x},\mathbf{y},\mathbf{z}) - \mathbf{f}(\mathbf{x},\mathbf{y},\mathbf{z})$.

The limit of the ratio $\mathbf{\Delta u}$ to $\mathbf{\Delta x}$ refers to as a partial derivative from u on x and is designated by a sign $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$.

$$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \lim_{\mathbf{\Delta x}\to\mathbf{0}} \frac{\mathbf{f}(\mathbf{x}+\mathbf{\Delta x},\mathbf{y},\mathbf{z}) - \mathbf{f}(\mathbf{x},\mathbf{y},\mathbf{z})}{\mathbf{\Delta x}}$$

Similarly express partial derivatives on other variables.

For a determination of a partial derivative it is necessary to differentiate function on one variable, considering other variables constants.

For example, function $\mathbf{u} = \mathbf{xy^2/z}$.

Then a partial derivative on x $\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \frac{\mathbf{y^2}}{\mathbf{z}}$, a partial derivative on y $\frac{\partial \mathbf{u}}{\partial \mathbf{y}} = \frac{\mathbf{2xy}}{\mathbf{z}}$ and a partial derivative on z $\frac{\partial \mathbf{u}}{\partial \mathbf{z}} = -\frac{\mathbf{xy^2}}{\mathbf{z^2}}$.

The partial differential of the function of several variables is a product of partial derivative of the function with respect to one of the variable multiplied by differential of the same variable.

For example, for given function three partial differentials exist: $\mathbf{du_x} = \frac{\mathbf{y^2}}{\mathbf{z}}\mathbf{dx}$, $\mathbf{du_y} = \frac{\mathbf{2xy}}{\mathbf{z}}\mathbf{dy}$, $\mathbf{du_z} = -\frac{\mathbf{xy^2}}{\mathbf{z^2}}\mathbf{dz}$.

Expression that refers to as a total differential of function $\mathbf{u} = \mathbf{f}(\mathbf{x},\mathbf{y},\mathbf{z})$:

$\mathbf{du} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{dx} + \frac{\partial \mathbf{u}}{\partial \mathbf{y}}\mathbf{dy} + \frac{\partial \mathbf{u}}{\partial \mathbf{z}}\mathbf{dz}$, where dx, dy, dz – differentials of independent variables.

The total differential also is used for the approached evaluations and an estimation of errors.

Gradient of Function

It is frequently used in physics applications of partial derivatives.

The gradient of scalar function is a vector whose projections to axes X, Y and Z are equal to partial derivatives of this function with respect to arguments x, y and z, respectively.

Thus, for scalar function $\mathbf{u}$ $\mathbf{gradu} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{i} + \frac{\partial \mathbf{u}}{\partial \mathbf{y}}\mathbf{j} + \frac{\partial \mathbf{u}}{\partial \mathbf{z}}\mathbf{k}$, where **gradu** is a gradient of function $\mathbf{u}$; and $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are unit vectors directed along the axes X, Y and Z, respectively.

In each point of space the vector of a gradient of scalar function indicates the direction of the maximal rate of increasing of the function $\mathbf{u}$, this maximal rate (related to the unit length) being equal to **gradu**. The faster the field change, the greater the modulus.

The concept of a function gradient is frequently used for the description of the various physical phenomena, for example, the electrical and magnetic fields, diffusion and osmosis, movement of viscous liquids.

24.1.9 Indefinite integral

Derivation enables to discover for given function $\mathbf{F}(\mathbf{x})$ its derivative $\mathbf{F}'(\mathbf{x})$.

Operation, inverse to derivation, is integration, that is determination $\mathbf{F}(\mathbf{x})$ on its known derivative $\mathbf{f}(\mathbf{x}) = \mathbf{F}'(\mathbf{x})$.

Function $\mathbf{F}(\mathbf{x})$ term as **antiderivative** (or **primitive**) of $\mathbf{f}(\mathbf{x})$, if for all $\mathbf{x}$ from a range of definition of function $\mathbf{F}'(\mathbf{x}) = \mathbf{f}(\mathbf{x})$.

Generally, if $\mathbf{f}(\mathbf{x})$ has an antiderivative function of $\mathbf{F}(\mathbf{x})$, totality $\mathbf{F}(\mathbf{x}) + \mathbf{C}$ also will be an antiderivative for $\mathbf{f}(\mathbf{x})$: $(\mathbf{F}(\mathbf{x}) + \mathbf{C})' = \mathbf{F}'(\mathbf{x}) = \mathbf{f}(\mathbf{x})$.

A totality of antiderivatives $\mathbf{F}(\mathbf{x}) + \mathbf{C}$ for the given function $\mathbf{f}(\mathbf{x})$ or the given differential $\mathbf{f}(\mathbf{x})\mathbf{dx}$ is named as an **indefinite integral** from function $\mathbf{f}(\mathbf{x})$ and designate: $\int \mathbf{f}(\mathbf{x})\mathbf{dx} = \mathbf{F}(\mathbf{x}) + \mathbf{C}$, where

$\mathbf{f}(\mathbf{x})\mathbf{dx}$ – **element of integration**,

$\mathbf{f}(\mathbf{x})$ – **intergrand function**,

$\mathbf{C}$ – **integration constant**.

For example, the function $4x^3$ is the derivative of x^4 , and $x^4 + C$ is an antiderivative of $4x^3$. Hence, if result of derivation is single, result of integration is infinite set of functions distinguished by arbitrary constants.

The result of computing an indefinite integral can always be verified by finding the derivative of the result. If the answer is correct the differentiation must yield the integrand. To each formula of differential calculus there corresponds a certain formula of integral calculus.

Physical sense of the integral

It follows after definitions given above.

If the body acceleration is known as function of time $\mathbf{t}$, i.e. $\mathbf{a} = \mathbf{f(t)}$ the instantaneous (scalar) body velocity $\mathbf{v}$ at present point of time $\mathbf{t}$ is equal to an integral from an acceleration $\mathbf{v} = \int \mathbf{adt}$, and a path $\mathbf{x}$ is equal to an integral from a velocity $\mathbf{x} = \int \mathbf{vdt}$, or to a double integral from an acceleration $\mathbf{x} = \int\int \mathbf{adt}$.

The table of indefinite integrals

Table 24.3

The table of simplest functions of tabular integrals

$\int \mathbf{0 \cdot dx = C}$	$\int \frac{\mathbf{dx}}{\mathbf{\cos^2 x}} = \mathbf{tg(x) + C}$
$\int \mathbf{x^\mu dx} = \frac{\mathbf{x^{\mu+1}}}{\mu+\mathbf{1}} + \mathbf{C}$, $\mu \neq -\mathbf{1}$	$\int \frac{\mathbf{dx}}{\mathbf{\sin^2 x}} = \mathbf{-ctg(x) + C}$
$\int \frac{\mathbf{dx}}{\mathbf{x}} = \mathbf{\ln x + C}$	$\int \frac{\mathbf{dx}}{\sqrt{\mathbf{1-x^2}}} = \begin{cases} arcsinx + C \\ -arccosx + C \end{cases}$
$\int \mathbf{a^x dx} = \frac{\mathbf{a^x}}{\mathbf{\ln a}} + \mathbf{C}$	$\int \frac{\mathbf{dx}}{\mathbf{1+x^2}} = \begin{cases} arctgx + C \\ -arcctgx + C \end{cases}$
$\int \mathbf{e^x dx = e^x + C}$	$\int \mathbf{sh(x)dx = ch(x) + C}$
$\int \sin \mathbf{x dx} = -\cos \mathbf{x + C}$	$\int \mathbf{ch(x)dx = sh(x) + C}$
$\int \cos \mathbf{x dx} = \sin \mathbf{x + C}$	

Some rules of integration

1. Derivative of indefinite integral is equal to element of integration $\mathbf{d}\big(\int \mathbf{f(x)dx}\big) = \mathbf{f(x)dx}$.

2. Indefinite integral of function derivative is equal to this function summarized with arbitrary constant $\int \mathbf{dF(x) = F(x) + C}$.

3. The constant factor (A) can be taken outside the integral:
$\int \mathbf{A \cdot f(x)dx = A \cdot} \int \mathbf{f(x)dx}$.

4. The indefinite integral of an algebraic sum of functions is equal to the sum of the integrals of summands: $\int [\mathbf{f(x) \pm g(x)}]\mathbf{dx} = \int \mathbf{f(x)dx} \pm \int \mathbf{g(x)dx}$.

Some methods of integration

1. The decomposition method

A given integral can often be represented in the form of a sum of tabularintegrals with the help of above formulas (3) and (4). Then is necessary to perform termwise integration and thus to obtain the answer. This is the decomposition method. Let us consider several examples. Example:
$\int \frac{\sqrt[3]{2}\cos \mathbf{x} - \sqrt[3]{x^2}\cos^3 \mathbf{x}}{\mathbf{\cos^2 x}} \mathbf{dx} = \sqrt[3]{2} \int \frac{\mathbf{dx}}{\mathbf{\cos^2 x}} - \int \mathbf{x^{\frac{2}{3}} dx} = \sqrt[3]{2}\mathbf{tg(x)} - \frac{\mathbf{3x^{\frac{5}{3}}}}{\mathbf{5}} + \mathbf{C}$.

2. The method of transformation of variable (by substitution)

It is one of the most widely spread methods of integration based on the formula of differentiation of composite function.

It is necessary to try to find a change of variable, which reduce a given integral to certain tabular integral. After performing the substitution and integration we must make the reverse substitution that is we must pass from new variable to former one.

Example 1. Integration of $\int \cos \mathbf{3xdx}$.

It is necessary use substitution $\mathbf{3x = t}$, $\mathbf{3dx = dt}$

Then $\int \cos \mathbf{3xdx} = \int \cos \mathbf{t}\frac{\mathbf{dt}}{\mathbf{3}} = \frac{\mathbf{1}}{\mathbf{3}} \int \cos \mathbf{tdt} = \frac{\mathbf{1}}{\mathbf{3}} \sin \mathbf{t} + \mathbf{C} = \frac{\mathbf{1}}{\mathbf{3}} \sin \mathbf{3x} + \mathbf{C}$

Example 2. Integration of $\int \mathbf{x^2 e^{x^3-2} dx}$.

Use substitution $\mathbf{t = x^3 - 2}$, then $\mathbf{dt = 3x^2 dx}$, $\mathbf{x^2 dx = 1/3\ dt}$.

Then $\int \mathbf{e^t} \frac{\mathbf{dt}}{\mathbf{3}} = \frac{\mathbf{1}}{\mathbf{3}} \mathbf{e^t} + \mathbf{C} = \frac{\mathbf{1}}{\mathbf{3}} \mathbf{e^{x^3-2}} + \mathbf{C}$.

3. A method of integration by parts

Unfortunately, there is no formula expressing the integral of a product of functions in terms of the integrals of the factors. As we know, the derivative of an elementary function is always an elementary function. But the integral of an elementary function may not be an elementary function, and this fact is connected with the above property.

Nevertheless if we integrate both sides of formula $\mathbf{(uv)}' = \mathbf{u}'\mathbf{v} + \mathbf{v}'\mathbf{u}$ where u and v are functions of x we obtain

$\mathbf{uv} = \int \mathbf{u} \cdot \mathbf{dv} + \int \mathbf{v} \cdot \mathbf{du}$,

that is $\int \mathbf{u} \cdot \mathbf{dv} = \mathbf{uv} - \int \mathbf{v} \cdot \mathbf{du}$.

Both equivalent formulas are called the formula of integration by parts. When applying such formula it is necessary to factorize the integrand into two factors, i.e. u and v', and then differentiate the first factor and integrate the second. Hence, we pass to an integral in which u' substitutes for u and v for v'. After such a transformation it is possible to arrive at a tabular integral or at an integral, which is simpler than the original one.

Example 1. Find $\mathbf{y} = \int \ln \mathbf{xdx}$

Suppose that $\mathbf{u} = \ln \mathbf{x}$; $\mathbf{dv} = \mathbf{dx}$; $\mathbf{du} = \mathbf{dx/x}$; $\mathbf{v} = \mathbf{x}$.

Get $\mathbf{y} = \int \ln \mathbf{xdx} = \mathbf{x} \cdot \ln \mathbf{x} - \int \mathbf{x}\frac{\mathbf{dx}}{\mathbf{x}} = \mathbf{x} \ln \mathbf{x} - \mathbf{x} + \mathbf{C}$.

Example 2. Find $\mathbf{y} = \int \mathbf{x^2} \ln \mathbf{xdx}$.

Suppose that $\mathbf{u} = \ln \mathbf{x}$; $\mathbf{dv} = \mathbf{x^2 dx}$; $\mathbf{du} = \mathbf{dx/x}$; $\mathbf{v} = \mathbf{x^3/3}$.

Get: $\mathbf{y} = \int \mathbf{x^2} \ln \mathbf{xdx} = \int \ln \mathbf{x} \mathbf{d}(\frac{\mathbf{x^3}}{\mathbf{3}}) = \frac{\mathbf{x^3}}{\mathbf{3}} \cdot \ln \mathbf{x} - \int \frac{\mathbf{x^3}}{\mathbf{3}} \mathbf{d}(\ln \mathbf{x}) =$

$= \frac{\mathbf{x^3}}{\mathbf{3}} \cdot \ln \mathbf{x} - \int \frac{\mathbf{x^2}}{\mathbf{3}} \mathbf{dx} = \frac{\mathbf{x^3}}{\mathbf{3}} \ln \mathbf{x} - \frac{\mathbf{x^3}}{\mathbf{9}} + \mathbf{C}$

Define integral

The concept of a define integral arises as outcome of a problem of square definition of a curvilinear trapezoid; this square is equal a limit of the rectangles squares summing at tendency their amount to infinity.

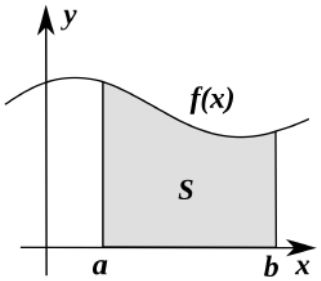

Fig. 24.8. Define integral.

$\lim\limits_{\mathbf{n}\to\infty}\sum\limits_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{f(x_i)} = \int\limits_{\mathbf{a}}^{\mathbf{b}}\mathbf{f(x)dx}$ – exact trapezoid square.

Numbers **a** and **b** are named as limits of an integration.

Physical sense of the define integral

For example, any force applied to the body move this body. And the force directs along X axis and depends on the coordinate by certain low: $\mathbf{y = f(x)}$. Problem: to find work of the force as body move from point **a** to point **b**.

As force change, to divide path on **n** short segments; and to calculate approximate work of force on every segment with number i: $\mathbf{A = f(k_i)\Delta x_i}$. Here $\mathbf{k_i}$ – is the middle point of corresponding segment. Total work is equal to sum of all elementary work values on the all segments from **a** to **b**: $\mathbf{A} = \sum_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{f(k_i)\Delta x_i}$. Total work will be exact if to calculate $\mathbf{A} = \lim\limits_{\mathbf{\Delta x_i\to 0}}\sum\limits_{\mathbf{i=1}}^{\mathbf{n}}\mathbf{f(k_i)\Delta x_i}$. If $\mathbf{\Delta x_i \to 0}$ then $\mathbf{n \to \infty}$.

Hence, $\mathbf{A} = \int\limits_{\mathbf{a}}^{\mathbf{b}}\mathbf{f(x)dx}$.

Rule of a determination of a define integral

For a determination of a definite integral use the Newton-Leibnitz formula: $\int\limits_{\mathbf{a}}^{\mathbf{b}}\mathbf{f(x)dx} = \mathbf{F(b) - F(a)}$, where $\mathbf{F(x)}$ is a primitive of function $\mathbf{f(x)}$, that is $\mathbf{F'(x) = f(x)}$.

Process of a determination of a define integral so notes in general view: $\int\limits_{\mathbf{a}}^{\mathbf{b}}\mathbf{f(x)dx} = \mathbf{F(x)}|_{\mathbf{a}}^{\mathbf{b}} = \mathbf{F(b) - F(a)}$.

Example. $\int\limits_{\mathbf{2}}^{\mathbf{4}}\mathbf{5x^3dx} = \mathbf{5}\frac{\mathbf{x^4}}{\mathbf{4}}|_{\mathbf{2}}^{\mathbf{4}} = \frac{\mathbf{5\cdot 4^4}}{\mathbf{4}} - \frac{\mathbf{5\cdot 2^4}}{\mathbf{4}} = \mathbf{300}$.

The differential equations

The ordinary differential equation is named the equation linking values of variable X, unknown function $\mathbf{y = f(x)}$ and its derivatives:

$\mathbf{F(x, y, y', y'', \ldots, y^{(n)}) = 0}$.

The degree of a differential equation is determined by the greatest order of the derivatives which are included in this equation.

For example, $\mathbf{(d^5y/dx^5)^3 - 3.14 \cdot (d^3y/dx^3)^9 + 7x^5y = 0}$, is a differential quintic (differential equation of fifth degree).

To solve the differential equation – is signifies to find function $\mathbf{y = f(x)}$, which satisfies to the given equation, i.e. at substitution of this function and its derivatives in the equation it turns to identity.

Searching all solutions of the given differential equation is major problem of the differential equations theory is. In the elementary cases this problem is reduced to an integral evaluation. Therefore a solution of the differential equation is named also as an integral, and process of searching of a solution – integration.

Many various processes in physics, chemistry, biology and medicine are described by the differential equations. The processes developing in time are usually described by such equations.

The differential equations of the first degree

To decide differential equation the first degree, apply method of a separation of variables, and then will carry out an integration by parts.

For example: $\mathbf{y' = 2xy}$ or $\mathbf{dy/dx = 2xy}$.

It is separable variables: $\mathbf{dy/y = 2xdx}$. After an integration $\int \frac{\mathbf{dy}}{\mathbf{y}} = \mathbf{2} \int \mathbf{xdx}$ we shall receive: $\ln \mathbf{y} = \mathbf{x^2 + C}$. As into the equation enters $\ln \mathbf{y}$ the constant is more convenient for expressing as the logarithm, i.e. to note $\ln \mathbf{y} = \mathbf{x^2} + \ln \mathbf{C}$, or $\ln(\mathbf{y/C}) = \mathbf{x^2}$. Exponentiating this equality, we shall receive: $\mathbf{y = Ce^{x^2}}$. It is answer.

Example of medical application.

The medical preparation entered into an organism, during defined time break up and/or to excrete from an organism. However the necessary medical effect is achieved only in the defined range of dozes. Therefore for scheduling frequency of reception the modification of preparation content in an organism or in an organ-target in time is necessary to calculate.

So, for example, it is necessary to calculate, during what time a mass of the preparation entered into an organism, will decrease twice if it is known, that in an instant $\mathbf{t_0}$ the initial mass of a preparation was $\mathbf{m_0}$. For small time $\mathbf{dt}$ the small mass of a preparation $\mathbf{dm}$ breaks up. Rate of decay and excretion of a preparation

in usual conditions is proportional to time of course of this process: $\mathbf{dm} = \lambda\mathbf{mdt}$, where λ – the constant depending on a metabolism of a preparation, and a sign «-» means a diminution of a mass in due course.

It is separable variables: $\frac{\mathbf{dm}}{\mathbf{m}} = -\lambda\mathbf{dt}$. We shall carry out an integration, having specified in as limits of an integration in the left part of the equation an initial mass $\mathbf{m_0}$ and a final mass $\mathbf{m_0/2}$, and in a right part of the equation starting time 0 and final time $\mathbf{t}$ ($\mathbf{t}$ is unknown quantity which is required to be determined): $\int_{\mathbf{m_0}}^{\mathbf{m_0/2}} \frac{\mathbf{dm}}{\mathbf{m}} = -\lambda \int_{\mathbf{0}}^{\mathbf{t}} \mathbf{dt}$. We shall receive: $\ln \frac{\mathbf{m_0}}{\mathbf{2m_0}} = -\lambda\mathbf{t}$. From here follows: $\mathbf{t} = \ln \mathbf{2}/\lambda$.

The differential equations of the second degree

Rules of them solution are similar to described above.

For example, the equation circumscribing oscillations of a pendulum.

The pendulum in length $\mathbf{l}$ and a mass $\mathbf{m}$ makes simple harmonic motions. On a pendulum operate a gravity mg and force of a tension of filament $\mathbf{F_t}$. We shall assume, that friction force is absent, and the deviation angle α a pendulum is small.

At small angles movement of a pendulum can be considered approximately rectilinear. The equation circumscribing forces operating on a pendulum:

$\mathbf{F} = -\mathbf{mg} \cdot \mathbf{tg}\alpha \approx -\mathbf{mgx/I}$, as at small angles $\mathbf{tg} \approx \mathbf{x/I}$, $\mathbf{g}$ – gravitational acceleration, sign «-» specifies opposite directions of the forces affixed on a pendulum.

Under the second law of Newton $-\frac{\mathbf{mgx}}{\mathbf{I}} = \mathbf{m}\frac{\mathbf{d^2x}}{\mathbf{dt^2}}$. We shall designate $\mathbf{g/I} = \omega$ (angular velocity), then $\frac{\mathbf{d^2x}}{\mathbf{dt^2}} = -\omega^2\mathbf{x}$. After a double integration with substitution of variables we shall receive a solution of the differential equation:

$\mathbf{x} = \mathbf{A}\cos(\omega\mathbf{t} + \varphi_\mathbf{0})$, where $\mathbf{A}$ – amplitude of oscillations, $\varphi_\mathbf{0}$ – an initial phase of oscillations (an angle of displacement during the initial moment of a time reckoning).

The differential equations with partial derivatives

The differential equations with partial derivatives contain unknown function of several variables and its partial derivatives on these variables.

For example, a wave equation: $\frac{\partial^2\mathbf{u}}{\partial\mathbf{x^2}} + \frac{\partial^2\mathbf{u}}{\partial\mathbf{y^2}} + \frac{\partial^2\mathbf{u}}{\partial\mathbf{z^2}} = \frac{\mathbf{1}}{\mathbf{a}}\frac{\partial^2\mathbf{u}}{\partial\mathbf{t^2}}$

Solution of differential equations with partial derivatives is fairly complex task

24.2. Probability theory

The probability theory studies regularities of the mass appearances having casual character. Appearances (event) which can be carried out multiply with the defined, characteristic distinctions in singularities of display which cannot be stipulated beforehand are casual; a reason of it frequently is presence of a set of the unknown factors influencing outcome of each separate trial.

The first mathematical rules of probability theory have been obtained at the analysis of regularities of gamblings, and many examples taken from this area, are convenient for illustrations.

The realization of a complex of conditions which as much as big number of times can be reproduced refers to as a **trial**.

Outcome of each trial refers to as **event**.

Certain (sure) events. Events refer to authentic if they happen inevitably as a result of each trial, and impossible if as a result of each trial they cannot take place.

Compatible events. Two events refer to joint if occurrence of one of them does not exclude occurrence of another in the same trial.

Incompatible events. Events refer to incompatible if occurrence of one of them excludes occurrence remaining. So, at tossing a playing cube one side can drop out only. Sheddings of different sides are incompatible events.

Events form a complete group if at each trial any of them can come true and all cannot come true remaining - incompatible with it. The events forming a complete group pairwise incompatible and it is equal possible (probable) events, refer to elementary. Simple events of such group refer to favorable to a realization of event A if the realization of any of simple events entails a realization of event A.

Example. At tossing a playing cube by simple events shedding sides $\mathbf{A_1}$, $\mathbf{A_2}$,. . .,$\mathbf{A_6}$ which form a complete group is. The events favorable to shedding of even number on an side, simple events $\mathbf{A_2}$, $\mathbf{A_4}$, $\mathbf{A_6}$ are.

Classical definition of probability. Probability $\mathbf{P(A)}$ of events A refers to the ration of number m the simple events favorable to event A, to number n all simple events, i.e. $\mathbf{P(A) = m/n}$. Properties of probability:

1. The probability of a certain (sure) event is equal to unit. $\mathbf{P(U) = 1}$.

2. The probability of impossible event is equal to zero. $\mathbf{P(V) = 0}$.

3. The probability of any event satisfies to an inequality $\mathbf{0 \leq P(A) \leq 1}$, as

$\mathbf{0 \leq m \leq n}$.

At carrying out of real trials their amount is limited. Ration $\mathbf{P(A) = m/n}$ refers to as a **relative frequency** of event A, occurred **m** time in the given series **n** trials.

Statistical probability of event A in the given trial is referred number $\mathbf{P(A)}$ about which values of a relative frequency are grouped at big n.

Examples. The relative frequency of shedding of one defined leg of a playing cube is equal $1/6$. The relative frequency of shedding of one defined leg of a coin is equal $0,5$, that is exhibited at the amount of repetitions exceeding $4000-5000$. The relative frequency of a birth of boys (in some country) is equal $0,514$ (however, there are the biological-social factors a little bit raising probability of a birth of girls). Relative frequencies at the big number of homogeneous experiences are steady.

24.2.1 Theorems of probability theory

The sum of two events A and **B** is referred event $\mathbf{C = A + B}$ consisting in occurrence of an event **A**, either events **B**, or events **A** and **B** simultaneously. The sum of several events name event which consists in approach even one of these events.

The probability of approach of one of two incompatible events A and B is equal to the sum of probabilities of these events: $\mathbf{P(A + B) = P(A) + P(B)}$.

The sum of probabilities of incompatible events $(\mathbf{A_1, A_2, \ldots, A_n})$, events forming a complete group, is equal to unit:

$$\mathbf{P(A_1) + P(A_2) + \ldots + P(A_n) = 1}.$$

Two events refer to **opposite** if they are incompatible and form a complete group. If event is designated as **A**, opposite event is designated as $\overline{\mathbf{A}}$. $\mathbf{P(A) + P(\overline{A}) = 1}$.

24.2.2. Independent events

Two events refer to **independent** if the probability of a event **A** realization don't depends on a event **B** realization. Some events refer to **independent in aggregate** if each of them and any combination of remaining events are independent.

Product of several events is the joint realization of all these events.

The probability of product of two independent events $\mathbf{A}$ and $\mathbf{B}$ is equal to product of probabilities of these events: $\mathbf{P(A \cdot B) = P(A) \cdot P(B)}$.

Example. The nurse serves 3 patients. The probability of a call within one hour to one patient is equal 0.3, the probability of a call to 2-d patient is equal to 0.5, and the probability of a call to 3-th patient is equal to 0.1. If to consider probabilities of calls independent that the probability of a call within one hour to all three patients is equal:

$$\mathbf{P(A \cdot B \cdot C) = P(A) \cdot P(B) \cdot P(C) = 0.3 \cdot 0.5 \cdot 0.1 = 0.015}.$$

The probability of a realization even one of events $\mathbf{A_1}$, $\mathbf{A_2}$, ..., $\mathbf{A_n}$, independent in aggregate, is equal to a difference to unit and product of probabilities of opposite events $\overline{\mathbf{A_1}}$, $\overline{\mathbf{A_2}}$, ..., $\overline{\mathbf{A_n}}$: $\mathbf{P(A) = 1 - \overline{A_1} \cdot \overline{A_2} \cdot \ldots \cdot \overline{A_n}}$.

Example. The nurse serves 3 patients. The probability of a call within one hour to one patient is equal 0.3, the probability of a call to 2 patient is equal to 0.5, and the probability of a call to 3 patient is equal to 0.1. The probability of even that the nurse within the given hour will be occupied with one of patients, is equal: $\mathbf{P(A) = 1 - (1 - 0.3) \cdot (1 - 0.5) \cdot (1 - 0.2) = 1 - 0.7 \cdot 0.5 \cdot 0.9 = 0.685}$.

24.2.3. Dependent events

Event $\mathbf{B}$ refers to dependent from event $\mathbf{A}$ if the probability of an event $\mathbf{B}$ realization depends on that, whether or not event $\mathbf{A}$ has fulfilled.

Probability of that event $\mathbf{B}$ has come true provided that event $\mathbf{A}$ has come true, is designated $\mathbf{P(B/A)}$ or $\mathbf{P_A(B)}$ and refers to as **conditional probability** of event $\mathbf{B}$ under condition of $\mathbf{A}$.

The probability of product of two dependent events A and $\mathbf{B}$ is equal to product of probability of one of them on probability of the second provided that the first event has come true: $\mathbf{P(A \cdot B) = P(A) \cdot P(B/A)}$. Such rule is fair for any other amount of dependent events too.

Example. The student has come on examination, knowing 40 problems from 50. What probability of what the student will answer all three problems of the ticket (events $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$)?

Probability of that the student knows one of problems of the ticket, is equal $\mathbf{P(A) = 40/50 = 4/5}$. Probability of that the student knows second of problems of the ticket provided that it obviously knows one of them, is equal $\mathbf{P(B) = 39/49}$. Probability of that the student knows third of problems of the ticket provided that

it obviously knows two of them, is equal $\mathbf{P(C) = 38/48 = 19/24}$. Resulting probability: $\mathbf{P(ABC) = P(A)P(\frac{B}{A})P(\frac{C}{AB}) = \frac{4}{5} \cdot \frac{39}{49} \cdot \frac{19}{24} = 0.5}$.

24.2.4. Bernoulli formula

The Bernoulli formula is used when the outcome of a trial can be the occurrence of one of two opposite events.

Let the probability of one event be equal to $\mathbf{p}$, and the probability of the opposite event be $\mathbf{q}$, and $\mathbf{q = 1 - p}$. If n trials are performed, the probability of occurrence of the first event in each successive trial being independent of which of these two events occurred in the previous trials, then the probability of the first event occurring $\mathbf{m}$ times ($\mathbf{P(m \text{ of } n)}$) can be calculated by the Bernoulli formula $\mathbf{(P(m \text{ of } n)) = \frac{n!}{m!(n! - m!)} \cdot p^m \cdot q^{n-m}}$, where $\mathbf{n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n}$. Bear in mind that $\mathbf{0! = 1}$.

24.2.5. The formula of the composite probability

This is corollary of theorems of probabilities addition for the incompatible events forming a complete group of events, and theorems of multiplication of probabilities for dependent events.

The probability of event $\mathbf{A}$ which can come true only under condition of a realization of one of incompatible events $\mathbf{A_1, A_2, \ldots, A_n}$, forming a complete group, is equal to the sum of products of probabilities of each of these events on corresponding conditional probability of event $\mathbf{A}$:

$$\mathbf{P(A) = P(A_1)P(\frac{A}{A_1}) + P(A_2)P(\frac{A}{A_2}) + \ldots + P(A_n)P(\frac{A}{A_n}) =}$$
$$\mathbf{= \sum_{i=1}^{n} P(A_i)P(\frac{A}{A_i})}.$$

Example: 40 tasks are prepared for examination on mathematics on derivation, 40 tasks on integration and 20 tasks in probability theory. What probability of what the student will solve the taken task first at random if he is able to solve 20 tasks on derivation, 16 tasks on integration and 12 tasks in probability theory?

Probability of that the student will receive a task on derivation, is equal $\mathbf{P(A_1) = 40/100 = 0.4}$. Probability of that the student knows its solution, is equal $\mathbf{P(A/A_1) = 20/40 = 0.5}$.

Probability of that the student will receive a task on an integration, is equal $\mathbf{P(A_1) = 40/100 = 0.4}$. Probability of that the student knows its solution, is equal $\mathbf{P(A/A_1) = 16/40 = 0.4}$.

Probability of that the student will receive a task in probability theory, is equal $\mathbf{P(A_1) = 20/100 = 0.2}$. Probability of that the student knows its solution, is equal $\mathbf{P(A/A_1) = 12/20 = 0.6}$.

Resulting probability: $\mathbf{P(A) = 0.4 \cdot 0.5 + 0.4 \cdot 0.4 + 0.2 \cdot 0.6 = 0.48}$.

24.3. Statistics in biology and medicine.

The questions studying of random quantities and their distributions, concern to the probability theory. But, as here they are considered from the viewpoint of practical application are included in statistics chapter.

24.3.1. Elementary concepts and main terms in statistics.

Medicine, such as biology as hole, works with human organisms that have characteristics which are various even in a health. These differences can be very much: in many times by some parameters even in one organism in various ages, seasons, conditions time ranges (for example, hormone level).

For purpose to have possibility to differ reliable norm and pathologic states in various human groups (various by age, sex, genetic, adaptation state for some conditions of environment) limits of norm must be defined for all possible situation. As minimum – for all possible situation in population that is object of following work (observation, investigation, treatment).

For example, maintenance of erythrocytes in the blood of healthy highland inhabitants don't be a comparative value for evaluation of health in plain and see coast inhabitants, since highlander organisms adapt for a life in oxygen deficiency condition, and this adaptation can be both genetically fixed and formed in the ontogenesis.

But there are many human populations which live if medium conditions. Therefore in process of determining of erythrocyte maintenance norm it is necessary to measure this characteristic in a blood of sufficiently large number of various human group representatives from various regions of planet. It is necessary to carry out measurement in ill persons with all known pathologies for determining possibility declinations from the norm.

Statistics as science investigates quantitative side of mass appearances - native or public. Objects of statistics are scientific methods of a collection of information, data processing, analysis and interpreting of data, formulation of statistical

deductions on the basis of quantitative data. Statistics has own principles, lows, methods.

Real individual objects of statistics are members of generalized groups. For example, certain person can be characterized as men (group by gender), adult (group by age), representative of north European (group by genetic population), inhabitant of South Africa (group by climatic condition of live), ill with diabetes (group by syndrome), ill with I type diabetes mellitus (group by exact diagnosis) and so on.

In all such group there are many individual objects with various characteristic by each other measuring scale. Variability of signs is inherent to all characteristic of human organism.

Variable is some characteristic of some type of some objects, which can be measured, grouped (classified) or coded with purpose of following description, abstract (generalization), analyses and comparison. Goal of such action is receiving new information about objects, lows of it's exist, activity and interaction.

Variables are things that we measure, control, or manipulate in research. They differ in many respects, most notably in the role they are given in our research and in the type of measures that can be applied to them.

In experimental research, we manipulate some variables and then measure the effects of this manipulation on other variables; for example, a researcher might artificially increase blood pressure and then record level of regulatory proteins. Data analysis in experimental research also comes down to calculating "correlations" between variables, specifically, those manipulated and those affected by the manipulation. However, experimental data may potentially provide qualitatively better information: Only experimental data can conclusively demonstrate causal relations between variables.

Independent variables are those that are manipulated whereas dependent variables are only measured or registered. The terms dependent and independent variable apply mostly to experimental research where some variables are manipulated, and in this sense they are **"independent" ("responce")** from the initial reaction patterns, features, intentions, etc. of the subjects. Some other variables are expected to be "dependent" on the manipulation or experimental conditions. That is to say, they depend on "what the subject will do" in response. Somewhat contrary to the nature of this distinction, these terms are also used in studies where we do not literally manipulate independent variables, but only assign subjects to "experimental groups" based on some pre-existing properties of

the subjects. For example, if in an experiment, males are compared with females regarding their white cell count (WCC), Gender could be called the independent variable and WCC the dependent variable.

Random variable (stochastic, accidental) is called variable x, which value is number that is result of experiment. Each value $x_1, x_2, \ldots, x_n$ has corresponding probability $p(x_1), p(x_2), \ldots, p(x_n)$. Continuous variable can take on a arbitrary (random) value in specified intervals. Other names of **random variable** are **random quantity, variate, chance variable, stochastic variable**.

24.3.2. Measurement scales

Variables differ in "how well" they can be measured, i.e., in how much measurable information their measurement scale can provide. There is obviously some measurement error involved in every measurement, which determines the "amount of information" that we can obtain. Another factor that determines the amount of information that can be provided by a variable is its "type of measurement scale." Specifically variables are classified as (a) nominal, (b) ordinal, (c) interval or (d) ratio.

(a) Nominal variables allow for only qualitative classification. That is, they can be measured only in terms of whether the individual items belong to some distinctively different categories, but we cannot quantify or even rank order those categories. For example, all we can say is that 2 individuals are different in terms of variable A (e.g., they are of different race), but we cannot say which one "has more" of the quality represented by the variable. Typical examples of nominal variables are gender, race, color, city, etc.

(b) Ordinal variables allow us to rank order the items we measure in terms of which has less and which has more of the quality represented by the variable, but still they do not allow us to say "how much more." A typical example of an ordinal variable is the socioeconomic status of families. For example, we know that upper-middle is higher than middle but we cannot say that it is, for example, 18% higher. Also this very distinction between nominal, ordinal, and interval scales itself represents a good example of an ordinal variable. For example, we can say that nominal measurement provides less information than ordinal measurement, but we cannot say "how much less" or how this difference compares to the difference between ordinal and interval scales.

(c) Interval variables allow us not only to rank order the items that are measured, but also to quantify and compare the sizes of differences between them.

For example, temperature, as measured in degrees Fahrenheit or Celsius, constitutes an interval scale. We can say that a temperature of 40 degrees is higher than a temperature of 30 degrees, and that an increase from 20 to 40 degrees is twice as much as an increase from 30 to 40 degrees.

(d) Ratio variables are very similar to interval variables; in addition to all the properties of interval variables, they feature an identifiable absolute zero point, thus they allow for statements such as x is two times more than y. Typical examples of ratio scales are measures of time or space. For example, as the Kelvin temperature scale is a ratio scale, not only can we say that a temperature of 200 degrees is higher than one of 100 degrees, we can correctly state that it is twice as high. Interval scales do not have the ratio property. Most statistical data analysis procedures do not distinguish between the interval and ratio properties of the measurement scales.

24.3.3. Sets and samplings

One from main statistical ideas is **statistical set**. This is large group of relatively homogeneous objects that are investigated at the same time or in the limited time and space ranges.

Examples of statistical set:

(a) carriers of pathological (and adaptation) gene of sickle-cell disease (drepancytic /sickle-cell /Herrick's anemia),

(b) swimmer sportsmen's,

(c) surgical ill in acute trauma department of clinic in city N.,

(d) laboratory white Wistar's rats,

(e) group of laboratory white Wistar's rats that are used by researcher K. in N. university for investigating of preparation X influence on life duration and healthy state.

Universal set (total population, universe, parent population, general totality, entire assembly) is all existing objects of investigation (observation) that can unite in one group by means of investigation purposes, for example:

(a) all carriers of pathological (and adaptation) gene of sickle-cell disease,

(b) all swimmer sportsmen's,

(c) all surgical ill in acute trauma department of all clinics.

From statistics viewpoint, all persons, for whom measurement of certain characteristic is possible, are universal set. In that time health group is one universal set, but every group with certain pathology is separate universal set.

Really in used group for characteristic determining it can include only little part of all possible objects (patients, experimental animals and other), such group is called **sampled population** or simple **sample**.

Sample (selection, sampled population, sampling) is a part of universal set, which serves as investigation (observation) object, and there are judge by it about universal set properties.

Sample must be representative that is basic variants of universal set objects must be present in its in such proportions that in universal set. Often for it sample must be sufficient large and mainly correctly selected.

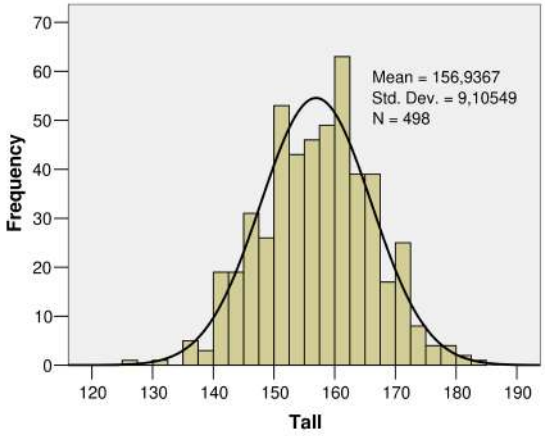

Fig. 24.9. Histogram of distribution of tall in $12 - 14$ years children.

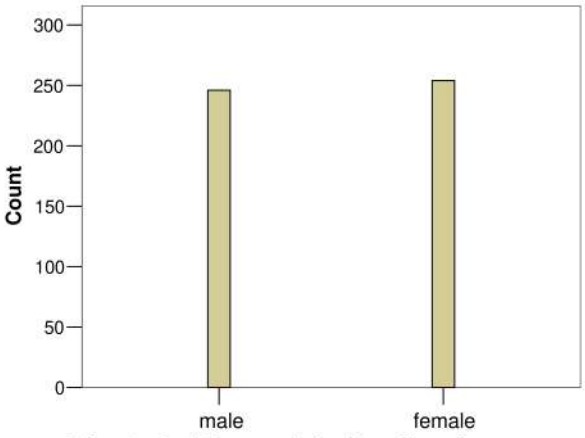

Fig. 24.10. Binomial distribution.

Representative selected rules, rules of required volume of selection determining are subject of statistics branch called **experiment planning**.

As every variable can vary in different objects of one sample, determination of sample distribution shape is important task.

24.3.4. Distributions

That is necessary to determine a frequency, with which every value of index occurs in given sample, and low (functional dependence) by which there can determine probability of given index value occurrence in arbitrary sampling object.

Distribution can be submitted graphically as histograms (fig.24.9).

There are such **main shapes of** random variable distributions (that is number variables, which value relates from result of random trial, and results of biologic-medical investigations are such).

A) For discrete random quantities:

– alternative (binomial or Bernoulli distribution) distribution (fig.24.10);

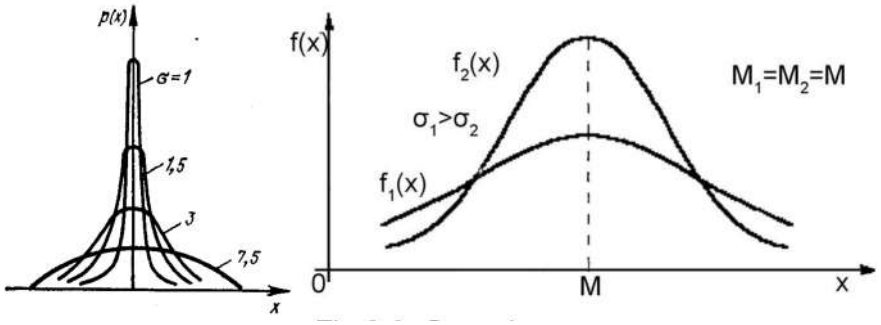

Fig. 24.11. Gaussian.

– Poisson's distribution (it dependence is enlargement of binomial distribution to the accidence of unlimited amount of sampling).

– geometric distribution;

– hypergeometric distribution.

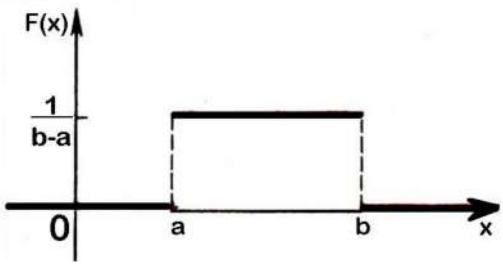

Fig. 24.12. Even distribution.

B) For continuous random quantities:

– normal distribution (Gauss distribution or Gaussian) (fig.24.11);

– χ^2 - distribution (Helmert-Pearson);

– Student 's distribution;

– Even (uniform) distribution (fig.24.12).

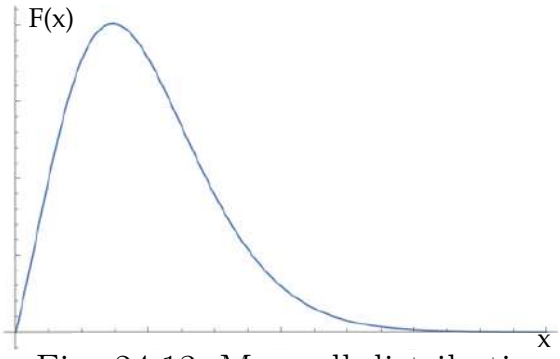

Fig. 24.13. Maxwell distribution.

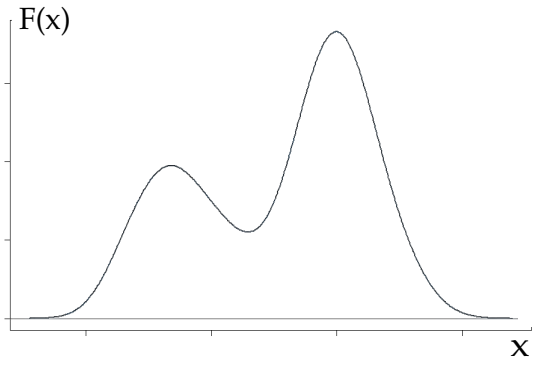

Fig. 24.14. Double-humped bimodal distribution.

A sign distribution in sampling can correspond to one from given lows, but can be asymmetry relation. For example, Charlie distribution, occurred in the case, when real distribution has asymmetry (skewness = bevel of the graph to right or to left) and kurtosis (high-summit or, turn, more similar trapezium, than bell, distribution), if some factor, which influence on random variable, greatly stronger than remainder factors. Skewed **Maxwell distribution** resembles a graph of erythrocyte distribution by its diameter value (fig.24.13).

Binomial distribution characterizes incompatible events.

Sometimes there are occur **double-humped bimodal distributions**

(fig.24.14), for example, distribution of all children by height. In age group $12-14$-years children in can be distribution with two closed maximums for girls and boys.

The **"Normal distribution"** (fig.24.11) is important because in most cases, it well approximates the function that was introduced in the previous paragraph. The distribution of many test statistics is normal or follows some form that can be derived from the normal distribution. In this sense, philosophically speaking, the Normal distribution represents one of the empirically verified elementary "truths about the general nature of reality," and its status can be compared to the one of fundamental laws of natural sciences. The exact shape of the normal distribution (the characteristic "bell curve") is defined by a function which has only two parameters: mean and standard deviation.

A characteristic property of the Normal distribution is that 68% of all of its observations fall within a range of ± 1 standard deviation (σ) from the mean, and a range of ± 2 standard deviations includes 95% of the scores. In other words, in a Normal distribution, observations that have a standardized value of less than -2σ or more than $+2\sigma$ have a relative frequency of 5% or less.

The normal distribution is extremely important that connected with the fact that many variables, which characterise certain biological and medical objects and processes, have distribution laws that are very close to the normal law. For example:

– the height and weight of adults (see fig.24.9 – distribution very close to Gaussian);

– arterial blood pressure when examining a great number of patients;

– the weight and volume of organs found when performing anatomic examinations;

– the absolute errors of readings of instruments and devices, and measurement

values;

– the enzyme activity or content in healthy people.

Not all, but most of them are either based on the normal distribution directly or on distributions that are related to, and can be derived from normal, such as t, F or Chi-square. Typically, those tests require that the variables analyzed are themselves normally distributed in the population, that is, they meet the so-called "normality assumption." Many observed variables actually are normally distributed, which is another reason why the normal distribution represents a "general feature" of empirical reality.

In such cases we can use some alternative "nonparametric" test (or so-called "distribution-free test"); but this is often inconvenient because such tests are typically less powerful and less flexible in terms of types of conclusions that they can provide.

As the sample size increases, the shape of the sampling distribution (i.e., distribution of a statistic from the sample) approaches normal shape, even if the distribution of the variable in question is not normal. This principle is called the **central limit theorem**.

24.3.5. Numerical descriptions of distributing of variants (random variable, stochastic variable)

"True" Mean and Confidence Interval

Descriptive statistics are calculated separately for each variable, and they provide such basic information as the mean, minimum and maximum values, different measures of variation, as well as data about the shape of the distribution of the variable. For the descriptive statistics procedure, you can select all variates in the current data set and a sequential table of "descriptives" for the consecutive variables will be generated (one row per variable). The measures of variation include the **standard deviation**, and the **standard error** (the standard error is the standard deviation of the sampling distribution of a mean). Numerous tests of whether the distribution of variables follows the normal distribution are also provided. For variables with normal distribution you can calculate **average value** (of mean) and it's **standard deviation** by formulas: mean $\overline{\mathbf{x}} = \frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i}$;

frequently use an entry of such aspect $\mathbf{M} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i}}{\mathbf{n}}$.

Table 24.4

Statistical measures and equations.

Measure name	Sign and design equation. Notice
Average of distribution	It characterizes center of distribution. $$\mathbf{M_x} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{x_i p(x_i)}$$ Here $\mathbf{x_i}$ – each (i-th) value of variable x; $\mathbf{p(x_i)}$ – probability of $\mathbf{x_i}$ observation. Design formula used in practical calculation: $$\mathbf{M} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \frac{x_i I_i}{\sum_{i=1}^{n} I_i}$$ Here $\mathbf{x_i}$ – each (i-th) value of variable x; I_i – quantity of x_i observation; n – quantity of observed values. In simplest case of notation this formula look such: $$\mathbf{\overline{X}} = \frac{\sum \mathbf{x}}{\mathbf{n}}$$ Here x – each of observed value, n – quantity of observed value (equal values calculate independently).
Variance of variate	It characterizes variability and has physical dimension, which differs from dimension of variable. $$\mathbf{D(x)} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{(x_i - Mx)^2 p(x_i)}$$ Design formula used in practical calculation: $$\mathbf{D} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{(x_i - M)^2 I_i}}{\sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{I_i}}$$ In simplest case of notation this formula look such: $$\mathbf{D} = \frac{\sum \mathbf{(x - \overline{X})^2}}{\mathbf{n}}$$

Measure name	Sign and design equation. Notice
Standard deviation	It characterizes variability and has physical dimension, which coincides with dimension of variable. $$\sqrt{\mathbf{D(x)}} = \sigma\mathbf{x}$$ Design formula used in practical calculation: $$\sigma = \sqrt{\mathbf{D}}$$
Standard error of mean	It is the theoretical standard deviation of all sample means of size n drawn from a population and depends on both the population variance (sigma) and the sample size (n) as indicated below. $$\sigma_{\mathbf{x-bar}} = \sqrt{\frac{\sigma^{\mathbf{2}}}{\mathbf{n}}}$$ Design formula used in practical calculation: $$\mathbf{m} = \frac{\sigma}{\sqrt{\mathbf{n-1}}}$$
Central moment with k-order	It is analog to variance by implication; it is seldom used. $$\mathbf{m_k} = \sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{x_i} - \mathbf{Mx})^{\mathbf{k}} \mathbf{p}(\mathbf{x_i})$$
Coefficient of variation	It is used for relative estimate of variability. $$\delta_{\mathbf{x}} = \frac{\sigma\mathbf{x}}{\mathbf{Mx}} \cdot \mathbf{100}\%$$
Mode	**Mo**(**x**) The most frequently occurring value. If several values share the greatest frequency of occurrence, each of them is a mode. The Frequencies procedure reports only the smallest of such multiple modes.
Amplitude of mode	**AMo**(**x**) It is value of probability for discrete distributions and value of probability density (frequency function) for continuous (uninterrupted) distributions.

Measure name	Sign and design equation. Notice
Median (Me)	It is value which divides distribution on two equal parts; it is central (medium) value of sampling, if values ranged by increasing or decreasing. If in sampling sever members, median is forth value in ordering row; if quantity of members is even, median equal average of two middle members of row.
Skewness (Index of asymmetry)	$$\mathbf{A_k} = \frac{\mathbf{m3(x)}}{\sigma^{\mathbf{3}}(\mathbf{x})}$$ A measure of the asymmetry of a distribution. The normal distribution is symmetric, and has a skewness value of zero. A distribution with a significant positive skewness has a long right tail. A distribution with a significant negative skewness has a long left tail. A skewness value greater than 1 generally indicates a distribution that differs significantly from a normal distribution.
Kurtosis (Index of kurtosis)	$$\xi = \frac{\mathbf{m4(x)}}{\sigma^{\mathbf{4}}(\mathbf{x})}$$ A measure of the extent to which observations cluster around a central point. For a normal distribution, the value of the kurtosis statistic is 0. Positive kurtosis indicates that the observations cluster more and have longer tails than those in the normal distribution and negative kurtosis indicates the observations cluster less and have shorter tails.

Standard deviation: if row is short and values are not grouped

$$\sigma = \sqrt{\frac{\mathbf{1}}{\mathbf{n-1}} \sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{x_i} - \overline{\mathbf{x}})^{\mathbf{2}}};$$

if row is long and values are grouped $\sigma = \sqrt{\frac{\mathbf{1}}{\mathbf{n-1}} \sum_{\mathbf{i=1}}^{\mathbf{n}} (\mathbf{x_i} - \overline{\mathbf{x}})^{\mathbf{2}} \mathbf{p(x_i)}}$;

Simple formula for manual computation (small data groups):

$$\sigma^{\mathbf{2}} = \frac{\mathbf{n} \sum(\mathbf{x^2}) - (\sum \mathbf{x})^{\mathbf{2}}}{\mathbf{n(n-1)}}.$$

Standard error of mean: $\mathbf{m} = \dfrac{\sigma}{\sqrt{\mathbf{n-1}}}$.

Properties of average:

1) it is generalizing, summarizing value characterizing all set without calculation of accident oscillation, difference in individual data;

2) sum of declination all variables from mean equal null $\sum\limits_{\mathbf{i=1}}^{\mathbf{n}}(\mathbf{x_i - M}) = \mathbf{0}$;

3) its place in variable's row is middle if distribution of sampling is Gaussian: $\mathbf{M = Mo = Me}$.

Standard deviation characterizes degree of scattering, dispersion of data around mean value.

Variation index is necessary for estimation and comparison data scattering in different sampling: $\mathbf{C_V} = \dfrac{\sigma}{\mathbf{M}} \cdot \mathbf{100}\%$

If $\mathbf{C_v} < \mathbf{10}\%$ usually there is telling about low dispersion of data, if $\mathbf{10}\% \leq \mathbf{C_v} \leq \mathbf{20}\%$ – about middle dispersion of data, if $\mathbf{C_v} > \mathbf{20}\%$ – about high dispersion.

The estimate of a statistical characteristic found from the sample is a random value. Therefore the interval can be determined, within which the true value falls with a certain acceptable probability α (it is usually named the **confidence probability** or **confidence coefficient**). Such an interval is known as the **confidence interval**.

The confidence interval for a statistical characteristic is a random interval whose boundaries are defined completely by the results of trials, and which covers the true value of this characteristic with the given probability α. The boundaries of the confidence interval are also random values.

The probability $\mathbf{p = 1 - \alpha}$ is known as the **significance level**.

If experimental random variable $\mathbf{X}$ have normal distribution, it is possible to calculate the **confidence interval limits**. Amount of data influences on choosing of computational expression.

1. If volume of sample large enough ($n > 30$), the sample variance estimate can be accepted as the true variance value, i.e. consider quantity $D(x)$ as known.

In this case, the confidence interval for $M(x)$ is computed by the formula:

$$\overline{\mathbf{x}} - \mathbf{t}(\alpha)\sqrt{\frac{\mathbf{D(x)}}{\mathbf{n}}} \leq \mathbf{M(x)} \leq \overline{\mathbf{x}} + \mathbf{t}(\alpha)\sqrt{\frac{\mathbf{D(x)}}{\mathbf{n}}}$$

where $\overline{\mathbf{x}}$ is the sample mean; n is the sample size; and value $t(\alpha)$ is found in tables for the Laplace function values with account of the set α value. α value is confidence probability (probability that assertion is correct).

In medical-biological studies, it is usually taken that $\alpha = 0.95(p = 0.05)$. In this case $t(\alpha) = 1.96$.

2. If the variance is unknown and $n < 30$ (typical case), to determine the confidence interval for $M(x)$, the formula is used:

$$\overline{\mathbf{x}} - \mathbf{t}(\alpha, \mathbf{k})\frac{\sigma}{\sqrt{\mathbf{n}}} \leq \mathbf{M}(\mathbf{x}) \leq \overline{\mathbf{x}} + \mathbf{t}(\alpha, \mathbf{k})\frac{\sigma}{\sqrt{\mathbf{n}}}$$

where σ is the sample estimate of the standard deviation; k is the **number of degrees of freedom** ($k = n - 1$); and $\mathbf{t}(\alpha, \mathbf{k})$ is **Student's coefficient** found in the table with the values of Student's coefficient.

For example, at $p = 0.05$ and $k = 20$ $t(\alpha, k) = 2.09$.

Comparison of two groups of experimental data

One of problems of the statistician theory is comparison of variables of one nature, but obtained in different experimental conditions, for example, to compare indexes of healthy people and patients with certain disease, to find out, whether there are distinctions on the given index. There are many methods of comparison, each of which has the defined observations to application. In case of the experimental data distributed under the normal law, apply Student's method which concerns to group of parametrical methods. Nonparametric methods of the analysis are applied to the data distributed under other laws, for example, binomial.

On Student's method comparison of average values of two groups will spend in view of variability of data around of average. Conditions of application of the given method: the data are continuous (not discrete), are normally distributed, variances are equal two populations, two samples are independent: associations between separate representatives of different samplings are not present; both samples are simple casual samples of corresponding populations.

If groups are close in volume, use such formula: $\mathbf{t} = \dfrac{\mathbf{M_1} - \mathbf{M_2}}{\sqrt{\mathbf{m_1^2} - \mathbf{m_2^2}}}$. Here M_1 and M_2 – average values of 1-t and 2-d groups accordingly, and m_1 and m_2 - errors of average of 1-t and 2-d groups accordingly.

If amounts of groups differ significantly in volume, use such formula:

$$\mathbf{t} = \frac{(\mathbf{M_1} - \mathbf{M_2})\sqrt{\dfrac{\mathbf{n_1 n_2}}{\mathbf{n_1} + \mathbf{n_2}}}}{\sqrt{\dfrac{\sigma_1^2\dfrac{(\mathbf{n_1} - \mathbf{1})}{\mathbf{n_1}} + \sigma_2^2\dfrac{(\mathbf{n_2} - \mathbf{1})}{\mathbf{n_2}}}{\mathbf{n_1} + \mathbf{n_2} - \mathbf{2}}}}$$

Here M_1 and M_2 – average values of 1-t and 2-d groups accordingly, and s_1 and s_2 – errors of average of 1-t and 2-d groups accordingly, and n_1 and n_2 – amount of groups.

The obtained value compares to the tabulated data $\mathbf{t}(\alpha, \mathbf{k})$ in view of sizes of both experimental groups. You can find this table in handbooks with standard mathematical tables. $\mathbf{t}(\alpha, \mathbf{k})$ is **Student's coefficient** found in the table with the values of Student's coefficient; and $\mathbf{k}$ is the **number of degrees of freedom** ($k = n_1 + n_2 - 2$).

24.3.6. Relations between variables

Regardless of their type, two or more variables are related if in a sample of observations, the values of those variables are distributed in a consistent manner.

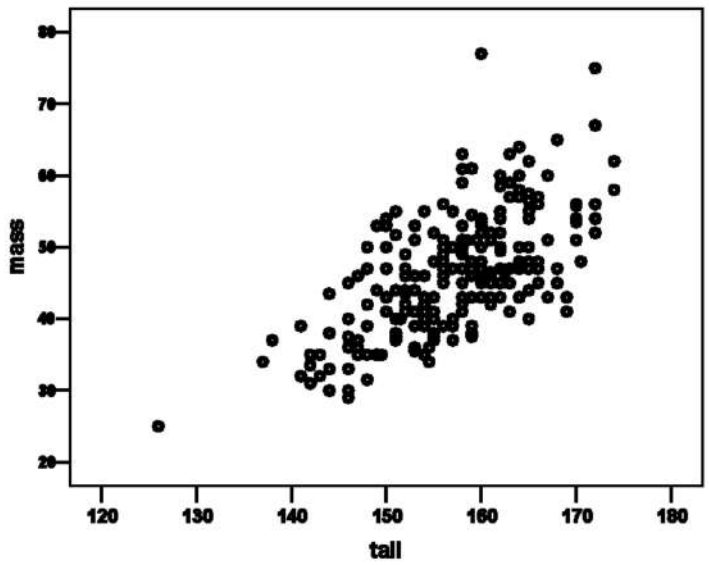

Fig. 24.15. Relation between tall and body mass in $12 - 14$ years girls.

In other words, variables are related if their values systematically correspond to each other for these observations. For example, **Gender** and **Tall** would be considered to be related if most males had high **Tall** and most females low **Tall**, or vice versa; **Height** is related to **Weight** because typically tall individuals are heavier than short ones; **IQ** is related to the **Number of Errors** in a test, if people with higher **IQ's** make fewer errors.

Ultimate goal of every research or scientific analysis is finding relations between variables. No other way of representing "meaning" except in terms of relations between some quantities or qualities; either way involves relations between variables. Thus, the advancement of science must always involve finding new relations

between variables. Correlation research involves measuring such relations in the most straightforward manner. However, experimental research is not any different in this respect. For example, the above mentioned experiment comparing **Tall** in males and females can be described as looking for a correlation between two variables: **Gender** and **Tall**. Statistics does nothing else but help us evaluate relations between variables. Actually, all of the hundreds of procedures that are described in this manual can be interpreted in terms of evaluating various kinds of inter-variable relations.

Correlation is a measure of the relation between two or more variables. **Correlation coefficients** can range from −1.00 to +1.00. The value of −1.00 represents a perfect negative correlation while a value of +1.00 represents a perfect positive correlation. A value of 0.00 represents a lack of correlation.

The relationship between two variables is such that as one variable's values tend to increase, the other variable's values tend to decrease. This is represented by a negative correlation coefficient. The relationship between two variables is such that as one variable's values tend to increase, the other variable's values also tend to increase. This is represented by a positive correlation coefficient.

Correlation coefficient marked **r** in cases of linear correlation.

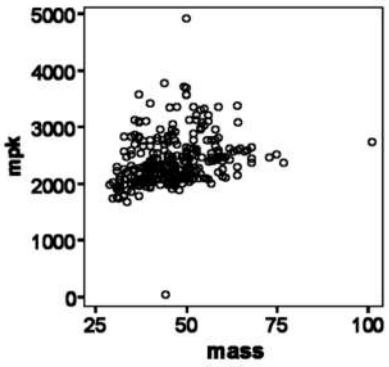

Fig. 24.16. Positive correlation.

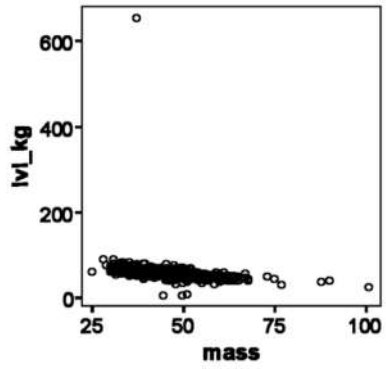

Fig. 24.17. No relation.

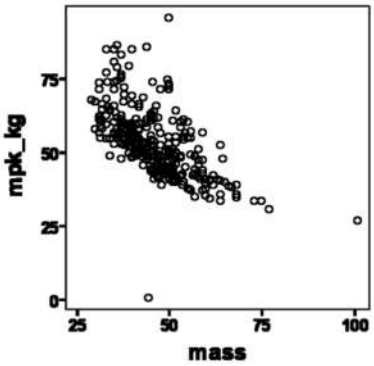

Fig. 24.18. Negative correlation.

Two basic features of every relation between variables

The two most elementary formal properties of every relation between variables are the relation's (a) **magnitude** (or "size") and (b) its **reliability** (or "truthfulness").

(a) **Magnitude (or "size")**. It measure is index of correlation (correlation coefficient).

(b) **Reliability (or "truthfulness")**. The reliability of a relation is a much

less intuitive concept, but still extremely important. It certain to the "representativeness" of the result found in our specific sample for the entire population. In other words, it says how probable it is that a similar relation would be found if the experiment was replicated with other samples drawn from the same population. *Remember that we are almost never "ultimately" interested only in what is going on in our sample; we are interested in the sample only to the extent it can provide information about the population.* If our study meets some specific criteria (to be mentioned later), then the reliability of a relation between variables observed in our sample can be quantitatively estimated and represented using a standard measure (technically called **p-level** or **statistical significance level**).

The statistical significance (p-level) of a result is an estimated measure of the degree to which it is "true" (in the sense of "representative of the population"). The value of the p-level represents a decreasing index of the reliability of a result. The higher the p-level, the less we can believe that the observed relation between variables in the sample is a reliable indicator of the relation between the respective variables in the population. Specifically, the p-level represents the probability of error that is involved in accepting our observed result as valid, that is, as "representative of the population." For example, a p-level of 0.05 (i.e.,1/20) indicates that there is a 5% probability that the relation between the variables found in our sample is a "fluke". In other words, assuming that in the population there was no relation between those variables whatsoever, and we were repeating experiments like ours one after another, we could expect that approximately in every 20 replications of the experiment there would be one in which the relation between the variables in question would be equal or stronger than in ours. In many areas of research, the p-level of 0.05 is customarily treated as a "border-line acceptable" error level.

24.3.7. How to determine that a result is "really" significant

There is no way to avoid arbitrariness in the final decision as to what level of significance will be treated as really "significant." That is, the selection of some level of significance, up to which the results will be rejected as invalid, is arbitrary. In practice, the final decision usually depends on whether the outcome was predicted a priori or only found post hoc in the course of many analyses and comparisons performed on the data set, on the total amount of consistent supportive evidence in the entire data set, and on "traditions" existing in the particular area of research. Typically, in many sciences, results that yield $p < 0.05$ are considered

borderline statistically significant but remember that this level of significance still involves a pretty high probability of error (5%). Results that are significant at the $p < 0.01$ level are commonly considered statistically significant, and $p < 0.005$ or $p < 0.001$ levels are often called "highly" significant. But remember that those classifications represent nothing else but arbitrary conventions that are only informally based on general research experience.

Statistical significance and the number of analyses performed.

Needless to say, the more analyses you perform on a data set, the more results will meet "by chance" the conventional significance level. For example, if you calculate correlations between ten variables (i.e., 45 different correlation coefficients), then you should expect to find by chance that about two (i.e., one in every 20) correlation coefficients are significant at the $p < 0.05$ level, even if the values of the variables were totally random and those variables do not correlate in the population.

Interrelation beyween strength and reliability of a relation between variables. Strength and reliability are two different features of relationships between variables. However, they are not totally independent. In general, in a sample of a particular size, the larger the magnitude of the relation between variables, the more reliable the relation.

The stronger the relation found in the sample, the less likely it is that there is no corresponding relation in the population.

Significance of a relation between variables depends on the size of the sample. If there are very few observations, then there are also respectively few possible combinations of the values of the variables, and thus the probability of obtaining by chance a combination of those values indicative of a strong relation is relatively high.

Small relations can be proven significant only in large samples. Analogously, if a relation in question is "objectively" very large (i.e., in the population), then it can be found to be highly significant even in a study based on a very small sample.

For example, if you work with data with normal distribution and want to check presence of pair linear correlation between two variables (for example, human tall and body weight), you can use such formula for calculation Pearson correlation index: $\mathbf{r_{xy}} = \dfrac{\mathbf{M(xy)} - \mathbf{M(x)M(y)}}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}$. Here $\mathbf{M(xy)} = \sum(\mathbf{X_i} - \mathbf{M_x})(\mathbf{Y_i} - \mathbf{M_y})$, $M(x) = \sum(X_i - M_x)^2$, $M(y) = \sum(Y_i - M_y)^2$.

Or, other form $\mathbf{r} = \dfrac{\sum_{\mathbf{i=1}}^{\mathbf{n}}(\overline{\mathbf{X}} - \mathbf{x_i})(\overline{\mathbf{Y}} - \mathbf{y_i})}{\sqrt{\sum_{\mathbf{i=1}}^{\mathbf{n}}(\overline{\mathbf{X}} - \mathbf{x_i})^{\mathbf{2}} \cdot \sum_{\mathbf{i=1}}^{\mathbf{n}}(\overline{\mathbf{Y}} - \mathbf{y_i})^{\mathbf{2}}}}$

Here x_i – members of first variable row, X – it mean value; y_i – corresponding members of second variable row, Y – mean value.

In some branches there are used such estimation of correlation force, but it can be used only in cases of very large selections (hundreds or thousands of objects) and statistical significance is present.

Table 24.5

Statistical significance

Correlation coefficient	**Statistics in biology and medicine**
$0 < \|r\| < 0.3$	weak correlation
$0.3 < \|r\| < 0.5$	insignificant relation
$0.5 < \|r\| < 0.7$	significant relation
$0.7 < \|r\| < 0.9$	strong correlation
$0.9 < \|r\| < 1.0$	very strong correlation

In the majority of cases with rather short selections (for, example, $10 - 100$ objects) you can evaluate only sign of correlation: positive or negative. Example of positive correlation is fig.24.16; example of negative correlation is fig.24.18. Lack of correlation is observed on fig.24.17 and 24.18. Can "no relation" be a significant result? When the magnitude of the effect approaches 0, the necessary sample size to conclusively prove it approaches infinity.

24.3.8. Calculation the level of statistical significance

In order to determine the level of statistical significance, we need a function that represents the relationship between "magnitude" and "significance" of relations between two variables, depending on the sample size. The function we need would tell us exactly "how likely it is to obtain a relation of a given magnitude (or larger) from a sample of a given size, assuming that there is no such relation between those variables in the population." In other words, that function would give us the significance (p) level, and it would tell us the probability of error involved in rejecting the idea that the relation in question does not exist in the population. This "alternative" hypothesis (that there is no relation in the population)

is usually called the null hypothesis. It would be ideal if the probability function was linear, and for example, only had different slopes for different sample sizes. Real function is more complex. Most of those functions are related to a general type of function which is called normal.

Test questions

1. Tell about physiological sense of the derivative?

2. Write down and explain rules of derivation?

3. Tell about physical sense of the define integral?

4. Write down and explain the differential equations of the first degree for example of medical application?

5. Explain the difference between independent events and dependent events?

6. Write down and explain Bernoulli's formula?

7. Write down and explain the formula of the composite probability?

8. Explain the classification of variables in statistics?

9. Draw and explain what a normal distribution is?

10. Tell us about statistical measures and equations?

Bibliography

[**1**.] Remizov A.N. Medical and biological physics: textbook / Remizov A.N. – Москва : ГЭОТАР-Медиа, 2021. – 576 с.

[**2**.] Касюк С.Т. Physics : textbook / С.Т. Касюк. – Челябинск : ТЕТА, 2021. – 276 с.

[**3**.] Клюшин В. Л. Mathematics : Textbook For Students Majoring In Non Mathematical Subjects / В. Л. Клюшин. - М. : Издательство РУДН, 2019. - 669 с.

[**4**.] Eliseeva S.V Laboratory works in physics. Guidelines for medical students of the first year: teaching aid / S.V. Eliseeva, Yu.N. Zubkov – Ulyanovsk: USU, 2022. – 189 p.

[**5**.] Tsurganov A.G. Medical and biological physics lectures course (for internal using only) for foreign students of the first year: The manual / A.G. Tsurganov, – second edit., – Vitebsk: VSMU Press, – 2014, – 315 pp.

[**6**.] Stadnichenko S.N. Medical and Biological Physics: a Course of Lectures / S.N. Stadnichenko. – Dnipropetrovsk, SE "Dnipropetrovsk Medical Academy of Health Ministry of Ukraine", 2014. – Part 1. – 67 pp.

[**7**.] Stadnichenko S.N. Medical and Biological Physics: a Course of Lectures / S.N. Stadnichenko. – Dnipropetrovsk, SE "Dnipropetrovsk Medical Academy of Health Ministry of Ukraine", 2014. – Part 2. – 67 pp.

[**8**.] Resnick R, Walker J, Halliday D. Fundamentals of physics. Hoboken: John Wiley, – 1988, – 1330 pp.

[**9**.] Korovina L. D. Bases of the biological and medical physics: Electron tutorial. Module 2. Bases of medical physics. – Poltava – 2019, – 113 pp.

Appendix A

Tables

Greek alphabet

Alpha	A	α	Nu	N	ν
Beta	B	β	Xi	Ξ	ξ
Gamma	Γ	γ	Omicron	O	o
Delta	Δ	δ	Pi	Π	π
Epsilon	E	ε	Rho	P	ρ
Zeta	Z	ζ	Sigma	Σ	σ
Eta	H	η	Tau	T	τ
Theta	Θ	θ	Upsilon	Υ	υ
Iota	I	ι	Phi	Φ	φ
Kappa	K	κ	Chi	X	χ
Lambda	Λ	λ	Psi	Ψ	ψ
Mu	M	μ	Omega	Ω	ω

SI Prefixes

Factor	Prefix	Symbol	Factor	Prefix	Symbol
10^{24}	yotta	Y	10^{-1}	deci	d
10^{21}	zetta	Z	10^{-2}	centi	c
10^{18}	exa	E	10^{-3}	milli	m
10^{15}	peta	P	10^{-6}	micro	μ
10^{12}	tera	T	10^{-9}	nano	n
10^{9}	giga	G	10^{-12}	pico	p
10^{6}	mega	M	10^{-15}	femto	f
10^{3}	kilo	k	10^{-18}	atto	a
10^{2}	hecto	h	10^{-21}	zepto	z
10^{1}	deka	da	10^{-24}	yocto	y

Some physical constants

Velocity of light in vacuum	$c = 3,00 \cdot 10^8$ m/s
Permeability of free space	$\mu_0 = 4 \cdot \pi \cdot 10^{-7}$ H $\cdot$ m^{-1}
Permittivity of free space	$\varepsilon_0 = 8,85 \cdot 10^{-12}$ F $\cdot$ m^{-1}
Universal constant of gravitation	$G = 6,67 \cdot 10^{-11}$ Nm$^2 \cdot$ kg^{-2}
Planck constant	$h = 6,63 \cdot 10^{-34}$ J $\cdot$ s
Rest mass of electron	$m_e = 9,1 \cdot 10^{-28}$ g $= 9,1 \cdot 10^{-31}$ kg
Rest mass of proton	$m_p = 1,673 \cdot 10^{-27}$ kg
Rest mass of neutron	$m_n = 1,674 \cdot 10^{-27}$ kg
Electron charge	$e = 1,60 \cdot 10^{-19}$ C
Specific charge of electron	$e/m = 1,76 \cdot 10^{11}$ C $\cdot$ kg^{-1}
Atomic mass unit	$u = 1,66 \cdot 10^{-27}$ kg
Avogadro constant	$N_A = 6,02 \cdot 10^{23}$ mol^{-1}
Faraday constant	$F = 9,65 \cdot 10^4$ C/mol
Molar gas constant	$R = 8,31$ J/(mol $\cdot$ K)
Boltzmann constant	$k = R/N_A = 1,38 \cdot 10^{-23}$ J $\cdot$ K^{-1} = $= 8,62 \cdot 10^{-5}$ eV $\cdot$ K^{-1}
Free-fall acceleration	$g = 9,81$ m/s^2
Stefan constant	$\sigma = 5,7 \cdot 10^{-8}$ W/$(m^2 \cdot K^4)$
Bohr magneton	$\mu_B = 9,27 \cdot 10^{-24}$ J/T
Coefficient in the law of Coulomb	$k = 1/(4\pi\varepsilon_0) = 9,00 \cdot 10^9$ m/F
Absolute zero	$t = -273,16$ ^{o}C

Учебное издание

Eliseeva S.V. Zubkov Yu.N.

Physics, Mathematics

A COURSE OF LECTURES
for foreign medical students of the first year:
educational book

Оригинал макет подготовлен *С. В. Елисеевой*

Подписано в печать 14.12.2023. Формат 60×84/16.
Гарнитура Times New Roman. Усл. печ. л. 22,7
Тираж 100 экз. Заказ № 55

Отпечатано с оригинал-макета в лаборатории оперативной полиграфии
Ульяновского государственного университета
432017, г. Ульяновск, ул. Л. Толстого, 42