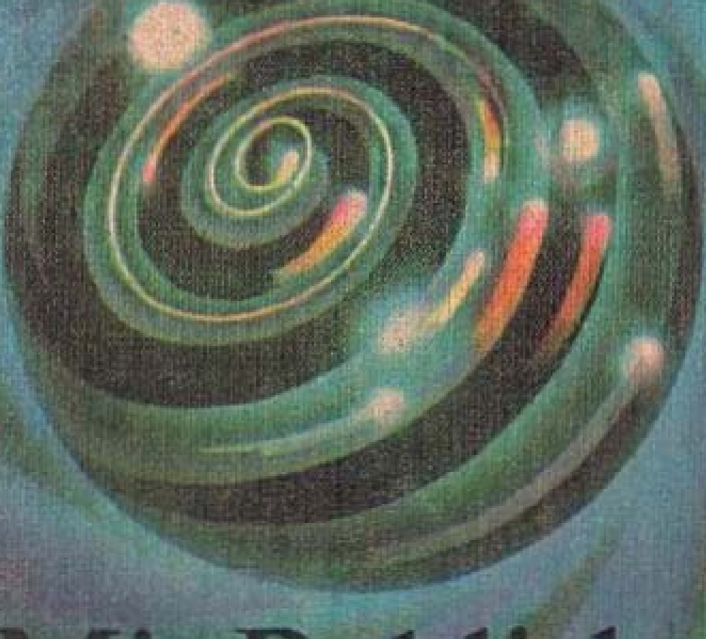
A. N. Matveev

Mechanics and Theory of Relativity



Mir Publishers
Moscow

Mechanics and Theory of Relativity

А. Н. Матвеев

Механика и теория относительности

Издательство «Высшая школа»
Москва

A. N. Matveev

Mechanics and Theory of Relativity

Mir Publishers Moscow

Translated from Russian by Ram Wadhwa

First published 1989
Revised from the 1986 Russian edition

На английском языке

Printed in the Union of Soviet Socialist Republics

ISBN 5-03-000267-7

Preface

This is the first volume (second Russian edition) of a course on general physics (the second, third and fourth volumes were published in 1985 (*Molecular Physics*), 1986 (*Electricity and Magnetism*) and 1988 (*Optics*) respectively; the fifth volume (*Atomic Physics*) is under preparation).

The task of a course on physics can be summed up as follows. Firstly, it should contain the basic principles and laws of physics and their mathematical formulation, introduce the basic physical phenomena, methods of their investigation and experimental studies, a proper form of expression for physical ideas, a quantitative formulation and solution of physical problems, estimates of the order of physical quantities, and a clear idea about the limits of applicability of physical models and theories. Secondly, it should inculcate in the studies a skill in experimental work, indicate the main methods of exact measurement of physical quantities and simple ways of analyzing the experimental results and basic physical instruments. Thirdly, it should provide an insight into the philosophical and methodological problems of modern physics and describe the various stages of evolution of science. Finally, it should point out the true role of physics in the scientific and technical progress and arouse the student's curiosity, interest and ability to solve scientific, engineering and other applied problems.

These problems can be solved only through a proper combination of experimental and theoretical instruction. Experimental skill is acquired in laboratories with the help of appropriate practical guides for laboratory work. This book provides the theoretical background. Of course, it also contains a description and analysis of physical phenomena, measurement of physical quantities, experimental methods of investigation, and other allied problems, but only from the point of view of theoretical understanding.

The curriculum of physics education in colleges at present aims at strengthening the basic level of knowledge. Physics is a leading discipline among fundamental sciences. Hence this book contains a detailed material on the measurement and determination of physical quantities, the role of abstractions, and the methods of physical investigation. Kinematics is treated not as a mathematical theory, but from a physical point of view. This allows the introduction of relativistic concepts of space and time, as well as Lorentz transformations, right at the beginning of the book. Consequently, the concepts of space and time, motion and material are linked inseparably in kinematics. The physical content of Newton's laws is described in detail, different methods of substantiation of mechanics are reviewed critically, and the connection between the conservation laws and symmetry of space and time is established in a comprehensible form.

A modern specialist should not only acquire the basic skills, but also learn to effectively apply the results of physical studies to accelerate the pace of scientific progress. In this connection, we have also considered in this book problems like motion in noninertial reference frames, inertial navigational systems, gyroscopic phenomena, motion of the artificial Earth's satellites, dynamics of bodies of variable mass, motion in electromagnetic fields, relation between mass and energy.

The same methodological approach has been used in writing all the volumes of this course. Each chapter contains a resumé of the basic ideas, and each section contains a formulation of the crux of the problems discussed in it. Examples have been chosen in such a way that they illustrate the methods of solving the most important problems. Problems for independent work-out are included at the end of each chapter and answers are also provided. Brief formulations of the most important statements and formulas are provided throughout the book, and questions for testing the level of understanding of the material are also given in each section. The material is supplemented by a large number of diagrams. The appendices contain the necessary material for reference.

The author is grateful to Prof. S. P. Kapitza and to the staff of the department chaired by him for a careful review of the manuscript and for valuable comments.

A. N. Matveev

Contents

1. Introduction

2. Kinematics of a Point and a Rigid Body

3. Coordinate Transformations

4. Corollaries of Lorentz Transformations

5. Dynamics of a Point Mass

Chapter 1

Introduction

Basic idea:
Physical models are mathematical, but mathematics is not what they are based on. The quantitative relations between physical quantities are clarified through measurements, observations and experimental investigations. Mathematics only serves as the language in which these relations are expressed. There is no other language for constructing physical theories.

Sec. 1. PROBLEMS AND EXPERIMENTAL METHODS IN PHYSICS

The model nature of physical concepts is analyzed and experimental methods in physics are described.

PROBLEMS OF PHYSICS. When faced with physical objects, phenomena, situations and the connections between them in life, everyone creates in his consciousness a model consisting of the images of these objects, phenomena, situations and the connections between them. He also establishes certain rules for handling this information. Models were first formed right from the origin of human consciousness itself. Hence it is not surprising that some of the elements of these models (like the concepts of space and time) have made such deep inroads into our consciousness that some philosophers have called them "forms of consciousness" instead of calling them reflections of the external world in our consciousness. *While studying physics as a science, we must always keep in mind that physics is based on the concept of models.*

The task of physics is to create in our consciousness such a picture of the physical world that faithfully reflects the properties of the world and that ensures the same relations between the elements of the model as exist between the elements of the outer world.

ABSTRACTION AND LIMITEDNESS OF MODELS. The relations between phenomena and objects are so diverse in the real world that it is impossible in principle to comprehend all of them either in practice or in theory. This is due to the inexhaustibility of the properties of matter. *Hence in order to* *construct a model, we need to consider only those properties and connections that are significant for the phenomena being simulated.* Only such a restriction makes it possible to create a model that can be grasped by our imagination. The rejection of all that is not significant for a given phenomenon is one of the most important elements of physical investigation. For example, while considering the laws governing planetary motion around the Sun, we need not take into account the pressure of solar rays or the solar wind on the planets. However, the solar wind is significant when the behaviour of comet tails is being investigated. On many occasions a scientific study has come to naught because the investigators tried to take into account factors that were of no consequence.

Considering just the significant factors involves an abstraction from the real situation and the creation of a model within the framework of the abstractions.

Physical models are only approximate and their validity can only be guaranteed for the range of applicability of the abstractions used to construct the model. Beyond this range, a model may become inapplicable or even meaningless.

Therefore, it is important during an investigation to understand at each stage why the model being used is applicable in the particular situation. It must be emphasized here that the same physical object may be represented by different models under different conditions. For example, the motion of the Earth around the Sun can be considered in terms of a point-mass model, the point having the Earth's mass concentrated at its centre. The same model can also be used as the first approximation for the Earth when satellite trajectories are considered around the Earth at large distances. However, the model is not applicable to generate a more accurate description of a satellite trajectory since the Earth is not a perfect sphere and its mass is not uniformly distributed over its volume. Because of this, the force of gravity acting on the satellite cannot be approximated by the gravitational force exerted by a point mass at the centre of the Earth. Moreover, the model of a gravitating point mass permits satellite trajectories to pass through points within one radius of the Earth, which is impossible, since the satellite would collide with the Earth's surface.

EXPERIMENTAL METHODS IN PHYSICS. At birth, human beings possess neither any elements of the models of the world surrounding them nor any rules for operating with the models. We acquire these elements and rules as we grow up. An individual human being makes them elements of his consciousness in the course of his personal activity and during the process of learning.

Our physical model of the world is continuously being broadened and extended as a result of scientific activity. This is only possible through experiment and observation. Hence physics is an experimental discipline. Its models must adequately reflect the properties that are discovered by observation and experiment. On the other hand, the limits of applicability of a model are also set by experiment.

Consequently, the experimental method in physics can be described as follows. On the basis of experiment and observation, a model is created and used to predict phenomena that are in turn verified by experiments and observations. In the light of these experiments and observations, appropriate corrections are made to the model which is then used to make fresh predictions, and so on.

There are two cases in which significant advances are made in physics. In the first case, the predictions of a model are not confirmed by experiment, while in the second case, a new set of physical phenomena is discovered for which there are no models at all. In the first case, the model has to be refined or replaced by a new one. If the replacement of the model is accompanied by a radical change in the basic concepts, a revolution is said to have occurred in physics. In the second case, a new branch of physics is opened.

The creation of the special theory of relativity, which led to a complete reconsideration of the basic concepts in the Newtonian model of space and time, is an example of the first kind. The creation of quantum mechanics as a new branch of physics is an example of the second kind. *In both these cases, we are talking not of the rejection of the existing models, but simply of the establishment of the limits of their applicability and the creation of new models which can be applied in cases where the old models cannot.*

The experimental method in physics requires a unique interpretation of all ideas, concepts and other elements of a physical model. The model must not contain elements which cannot be uniquely interpreted through relations either with the objects, processes, situations, etc. in the real world or with other elements of the model which have already been established.

This is most significant in the study of physics. It must be ensured that each element of the model under consideration has a well-defined meaning and a clearly formulated relation with the appropriate element of the real physical world.

It is only under this condition that physical concepts have any real meaning reflecting the objective laws of the material world and the objective properties of physical bodies and processes.

!

The same physical object can be represented by different models in different situations. Physical models are approximate, and their validity can be guaranteed only within the limits of applicability of the abstractions used. Beyond these limits, a model may become unsuitable or even meaningless.

Sec. 2. PHYSICAL QUANTITIES AND THEIR MEASUREMENT

The idea of measurement as a quantitative comparison of the common in the diverse is discussed.

DIFFERENCE AND COMPARISON. The first step to knowledge lies in establishing the difference between physical objects. This enables us to identify the objects of investigation. The next problem is that of comparison. However, a comparison is only possible among like quantities. Hence we must find out common features in diverse objects. The common and the diverse appear here in their dialectical unity. A comparison of different objects is only possible on the basis of something that is common in all of them. For example, all objects occupy a certain volume in space, and hence a comparison can be made between different objects on the basis of this criterion.

COMPARISON AND MEASUREMENT. The meaning of the statement "a melon is bigger than an apple" is intuitively quite clear: the volume occupied by an apple is a fraction of the volume occupied by a melon. Such a comparison is qualitative in nature. For example, even if it is known that the particular melon is bigger than another apple, no information can be drawn from here as to which of the two apples is bigger.

Thus it becomes imperative to express the result of comparison between the melon and each of the apples in such a form that information can be obtained about the result of comparison between the apples. This is made possible by measurement as a result of which the property under consideration is expressed in terms of a number.

!

Measurement of physical properties is the procedure of assignment of certain numbers to these properties in such a way that a comparison of properties is reduced to a comparison of numbers. Each physical quantity can be expressed only in its own units. Hence the number of units is equal to the number of physical quantities. However, using the relations between physical quantities, we can express one quantity in terms of another and thus restrict the number of physical quantities whose units can be used to express all the remaining quantities.

MEASUREMENT. It was mentioned above that measurement involves a comparison of identical properties or qualities in different objects, phenomena or processes. For example, the most general property of objects is their extent (size), while the most general characteristic of processes is their duration. Let us consider one of these properties, viz. the extent of an object. For the sake of simplicity, we shall consider this property only in one dimension, i.e. we shall only consider length. The objects whose lengths we measure are called lines.

The measurement of physical properties is the procedure of assigning certain numbers to these properties in such a way that the comparison of properties can be reduced to a comparison of these numbers.

In the problem under consideration, each line is assigned a certain number which uniquely characterizes its length. Conversely, each number must make it possible to choose from among all the existing lines only those whose length is uniquely determined by this number. The property which can be characterized in this way is called a physical quantity and the procedure which is used for determining the number characterizing a physical quantity is called measurement.

The measurement of a length is reduced to the comparison of lengths with a certain length which is taken as a unit. The procedure of comparison and the determination of the corresponding number is the essence of measurement, and may turn out to be quite complicated. In this definition, the length of a line is given by the formula $l = ml_0$, where m is a dimensionless number indicating the number of times the unit length lies within the length being measured. The quantity denoted by the symbol l_0 is the unit of length which is often given a name, for example, a centimetre, a metre, etc.

UNITS OF PHYSICAL QUANTITIES. Thus, in order to measure a physical property, we must choose a unit for the quantity, i.e. the physical property which can be assigned the number 1. For example, to measure the extent of a body, we choose a body (ruler scale) whose length is taken as unity and is denoted by the number 1. Measurement is then reduced to a comparison of the property being measured with the property taken as unity. The properties, qualities, etc. which are used in physics are called physical quantities. In this respect, the task of measurement is reduced to finding numerical value of a physical quantity.

NUMBER OF UNITS OF PHYSICAL QUANTITIES. Many physical quantities are used in physics. Each of these can only be expressed in its own particular units. *Hence the number of units of physical quantities is equal to the number of physical quantities themselves.* Such a large number of different units is very inconvenient and can be reduced. As a matter of fact, physical quantities are not independent of one another, but are rather related through the laws studied in physics. *These laws can be used to express one physical quantity in terms of another, and restrict the number of units of physical quantities in terms of which all the remaining physical quantities can be described. These units are called base units and the aggregate of units is called a system of units.*

Sec. 3. ON THE DEFINITION OF CONCEPTS AND QUANTITIES IN PHYSICS

The necessity of defining physical quantities in two ways is emphasized.

TWO CATEGORIES OF CONCEPTS USED IN PHYSICS. The first category of concepts with which physics deals could be called the category of physical concepts. This category includes concepts like force, velocity, acceleration, capacitance and viscosity. People who have not specially studied physics have very vague ideas about these concepts. Physicists, on the other hand, have a clear understanding of these concepts, usually backed by a quantitative definition, i.e. by their definition as physical quantities. However, a physicist cannot confine himself to these concepts in his work. He also needs concepts which are not restricted to physics but are more general in nature. This category of concepts could be termed the category of general concepts. This category includes concepts like existence, annihilation, verity, causality, determinism and objectivity. The impressions a physicist might have about these concepts do not differ in practice from those possessed by nonphysicists. They may be vague or clear, depending on the circumstances. Many of these concepts are considered in philosophy. However, general concepts by themselves are still inadequate, and must be assigned a concrete form so that they can be applied in physics. The physical analysis of most of these concepts has enriched our knowledge and the creation of physical theories.

TWO WAYS OF DEFINING PHYSICAL QUANTITIES. Each physical concept must have a clear and unique definition. This statement requires a further clarification of the term "definition" itself. What does the "definition" of a physical quantity mean? It means that we must indicate the property which makes this physical quantity specific, and also indicate the common factor that makes it an element of the general physical relation between phenomena. *A physical quantity is mathematically defined as a relation between other physical quantities.* This is the first way in which a physical quantity can be defined. For example, if we assume the definition of velocity to be known, we can define acceleration as the rate of change of velocity. In this case, acceleration can be expressed in the form $a = \mathrm{d}v/\mathrm{d}t$, which is just the definition of the quantity a. To emphasize that a certain formula is a definition, the equality sign "$=$" is sometimes replaced by the identity symbol "$\equiv$".

Obviously, not all physical quantities can be defined in this manner. As we go over from one set of quantities to another, we encounter quantities which must be defined in some new way. For example, acceleration is expressed in terms of velocity and the change in velocity (with time). Hence we must define velocity and time intervals. Velocity is expressed in terms of path length and time intervals which were also used in the definition of acceleration. These must be determined by a different method, without indicating their relation with known physical quantities. (It should be emphasized that we are now referring to space and time not as philosophical concepts, but as metres and seconds, and to the measurement of lengths and time intervals as physical quantities.) This method is called operational: *we indicate a physical object whose property is taken as unity, and define a measuring* *procedure which can be used to compare the properties of the object being measured and the unit object.* For example, when we measure length, we must indicate the standard taken as the unit length and the procedure for measuring other lengths in terms of this standard. Similarly, to measure a time interval, we must indicate the time interval which is taken as unity. This question will be considered in greater detail at a later stage.

These two methods of defining physical quantities exist side by side and complement each other.

ON GENERAL CONCEPTS. Most of the concepts now termed general do not have any special definitions in physics. It is assumed that no additional explanation is required for the meaning of the words in terms of which these concepts are described. Sometimes, references can be made to philosophical literature where these concepts are clarified. However, it has become clear with the passage of time that physics as a science cannot do without an analysis and interpretation of these concepts. Gnosiological, methodological and philosophical questions in physics have been considered in many works. These questions cannot be by-passed while studying physics, and their elaboration is stimulated by the development of physical concepts. For example, considerable progress has been made towards the understanding of causality and determinism with the development of quantum mechanics. The development of the theory of relativity has led to a better understanding of the relation between matter, space and time. These and other similar examples show that the interpretation and application of general concepts in physics are closely related to the progress of physics as a discipline. Many advancements in the evolution of physics have been linked to some extent with the progress in the interpretation of the general concepts.

Sec. 4. SYSTEMS OF UNITS OF PHYSICAL QUANTITIES

The arbitrariness in the choice of a system of units is emphasized and the International System of Units (SI) is introduced.

BASE AND DERIVED UNITS. We noted above that the number of different units must be equal to the number of different physical quantities. However, some physical quantities are described in terms of some other physical quantities. This allows us to decrease the number of base units which are defined without any reference to other units.

DIMENSIONS OF A PHYSICAL QUANTITY. As mentioned above, a physical quantity is usually (though not always) defined so that it can be represented by a formula of the type

$$a = m_a e_a. \tag{4.1}$$

Here, the symbol e_a stands for a unit quantity, i.e. a physical quantity of the same type as the quantity a being measured and with numerical value set equal to unity. Thus, the symbol e_a describes the nature of the quantity being measured, as well as the scale of measurement. The number m_a is a dimensionless number showing the number of units e_a comprising the quantity a being measured. Moreover, it follows from formula (4.1) that the summation of two quantities a_1 and a_2 involves the addition of two numbers m_{a1} and m_{a2}:

$$a_1 + a_2 = (m_{a1} + m_{a2})\, e_a. \tag{4.2}$$

It should be noted that this formula expresses the summation rule for physical quantities if they possess the property of additivity in the particular situation under which their summation is carried out. For example, a conductor can be characterized by its resistance R or conductance $\gamma = 1/R$. When two conductors are connected in series, their resistances are added, and when connected in parallel, their conductances are added. In both these cases, (4.2) is satisfied.

The nature of the quantity being measured is determined by its dimensions. Usually, the dimensions of a physical quantity are indicated by the same letters enclosed in brackets. For example, if the quantity a under consideration is length, its dimensions will be length, denoted by L. This is mathematically expressed by the equality $[a] = L$. The dimensions of the unit are the same, i.e. $[e_a] = L$. *For example, when we state that a quantity has the dimensions of length, we simply describe the nature of the quantity without saying anything about the scale of the unit used for measuring this quantity.* This could be, say, a metre, or a centimetre, or some other length taken as unity. All these units have the same dimensions L.

Let us consider two other physical quantities which are described by formulas .

$$b = m_b e_b \quad \text{and} \quad c = m_c e_c. \tag{4.3}$$

Suppose that the quantities a, b and c are related by a physical law.

It must be borne in mind that the law is established in the form of a relation not between the physical quantities a, b and c, but between the numbers m_a, m_b and m_c, through which these quantities are measured. For example, let us suppose that this law has the form

$$m_c = A m_a^p m_b^q. \tag{4.4}$$

Here, the numbers A, m_a, m_b and m_c are also dimensionless, while p and q are the powers to which the numbers m_a and m_b are raised. Using (4.2) and (4.3), we can write this relation formally as follows:

$$\frac{c}{e_a} = A\frac{a^p}{e_a^p}\frac{b^q}{e_b^q}, \tag{4.4a}$$

or

$$c = \left(A\frac{e_c}{e_a^p e_b^q}\right)a^p b^q. \tag{4.4b}$$

By definition, the numerical values of the quantities e_a, e_b and e_c are equal to 1. Hence the quantity within the parentheses on the right-hand side of (4.4b) is numerically equal to A, but is dimensional. Denoting it by A', we can write the physical law (4.4b) in the form

$$c = A' a^p b^q. \tag{4.4c}$$

Physical laws are usually described in this form rather than in the nondimensional form (4.4).

The following two rules are used to determine the dimensions of a complex expression:

$$\left[\frac{1}{a}\right] = \frac{1}{[a]}, \quad [ab] = [a]\,[b]. \tag{4.5}$$

Hence the dimensions of the quantity A' in (4.4c) can be given by the expression

$$[A'] = A\left[\frac{e_c}{e_a^p e_b^q}\right] = [A]\frac{[e_c]}{[e_a]^p \cdot [e_b]^q} = \frac{[c]}{[a]^p [b]^q}, \tag{4.6}$$

where we have considered the fact that A is a dimensionless number. This ensures that the left- and right-hand sides of Eq. (4.4c) have the same dimensions.

Mathematical equalities can be established only between quantities that have the same dimensions. The dimensions of physical quantities usually vary with the system of units. A reliable and rapid way of avoiding gross errors in computational formulas is to verify that the left- and right-hand sides of the equalities have the same dimensions, as well as the various terms in a sum or a difference since addition and subtraction is only possible between physical quantities with the same dimensions. Hence if the dimensions of the left- and right-hand sides of the equalities do not coincide or if two quantities with different dimensions are added or subtracted in a formula, it can confidently be stated that an error has been committed. The error is detected most easily if a dimensionless number is added to, or subtracted from, a dimensional quantity.

SELECTION OF BASE UNITS. The choice of physical quantities whose units are taken as the base units is a matter of convention. In principle, it is impossible to say why one physical quantity is preferred over another.

From a practical point of view, however, not all the units are equally suitable as base units. As a matter of fact, the base unit must be defined by a direct indication of the material object and the physical procedures leading to this unit. This raises questions concerning the invariance of the material object, reproducibility of the procedures, convenience of operation, etc. When these things are taken into consideration, the arbitrariness in the choice of base units is considerably reduced. Hence it is not surprising that the units of length, mass and time are invariably chosen together with other units as base units in many systems of units.

NUMBER OF BASE UNITS. The maximum number of base units is equal to the number of all measurable physical quantities, so that each physical quantity has its own unit. For example, quantities like velocity v, length l and time t have their own particular units. The dimensions of the units are the same as those of the physical quantity being measured. In the present case, these dimensions are the dimensions of velocity $[v] = V$, length $[l] = L$ and time $[t] = T$.

A study of uniform motion leads to the establishment of the following law:

$$l = Avt, \tag{4.7}$$

where A is a dimensional constant. Its numerical value depends on the choice of the units for velocity, length or time, while its dimensions are given by the formula

$$[A] = LT^{-1}V^{-1}. \tag{4.8}$$

For a given choice of the system of units, Eq. (4.7) is a universal relation between l, v and t, while the constant A is a universal constant. In view of this, we can choose any two quantities (say, L and T) as base ones, and select the dimensions and magnitude of the third unit (i.e. V) in such a way that A becomes a dimensionless constant equal to unity. For this purpose, we should choose unit velocity such that a unit distance is traversed in a unit time, while the dimensions of the velocity should be chosen in such a way that the quantity A in (4.8) is dimensionless, i.e.

$$[v] = LT^{-1}. \tag{4.9}$$

As a result, Eq. (4.7) assumes the form $l = vt$, while the unit of velocity is no longer a base unit, but becomes a derived unit with dimensions LT^{-1}. This leaves only two units, viz. length and time, as the base units.

Let us further reduce the number of units. For this purpose, we use the fundamental law concerning the constancy of the velocity of light, which will be described in detail at a later stage. A ray of light propagating with a velocity c will cover a distance

$$l = ct \tag{4.10}$$

in time t. In this relation, the velocity of light c is a universal dimensional constant which is independent of both the system of coordinates and the velocity of the source or the observer. As before, we choose, say, time as the base unit. The other unit is then a derived unit and is defined such that c becomes a dimensionless quantity equal to unity. For this purpose, the dimensions of length and time must be the same, i.e. $[l] = T$.

If we choose 1 s as the unit of time, a length l can be measured as the number of seconds it takes light to traverse it. For example, the length of a writing-table is about 0.5×10^{-8} s (corresponding to a distance of about 1.5 m), while the Earth's equator is 0.13 s long. Sometimes, though not always, it is convenient to use such units. In astronomy, the measurement of distance in light years is a very clear and widely used example of this kind. It is the same system of units that was considered above, the only difference being that one year is now chosen as the unit of time.

ARBITRARINESS IN THE CHOICE OF THE SYSTEM OF UNITS. We have thus shown that *the choice of base units and their number cannot be dictated by any basic or general philosophical considerations.*

From the point of view of principle, all systems of units are identical. They differ only in expedience and convenience both as regards their application and in meeting the above-mentioned requirements imposed on base units.

THE INTERNATIONAL SYSTEM OF UNITS (SI). After about one hundred years of discussion, the scientific and engineering communities throughout the world arrived at the International System of Units (SI) as the most expedient one. An agreement was reached by various international organizations and ratified by the member countries. This system was also adopted by the USSR.

The following units are chosen as SI base units: length–metre, time–second, mass–kilogram, electric current–ampere, temperature–kelvin, luminous intensity–candela. The definition of the first three units is vital in mechanics.

SECOND, THE UNIT OF TIME. For a very long time, the unit of time was defined in terms of the visible motion of the stars and the Sun in the sky due to the rotation of the Earth about its axis and its motion around the Sun. More and more accurate studies of these motions gradually refined the unit of time. A solar day is the interval of time between two successive passages of the Sun across the meridional plane which is imagined to pass through the axis of rotation of the Earth and the point of observation. A sidereal day is the interval of time between two successive passages of a fixed star through the meridian. The Earth moves around the Sun in an elliptic orbit and at the same time rotates about its axis in the same direction as it rotates around the Sun. This axis is not perpendicular to the plane of its orbit. In view of these and many other factors, the solar day is about four minutes shorter than the sidereal day. The duration of both solar and sidereal days varies during the year. Hence the tropical year, defined as the time elapsed between two successive passages of the Sun through the vernal equinox, was chosen as the standard of time. However, the duration of a tropical year also varies with time. Hence in 1956 the International Commission on Weights and Measures fixed the tropical year as the standard and gave the following definition of the second:

$$1 \text{ second} = \frac{1}{31{,}556{,}925.9747} \text{ of the tropical year 1900.}$$

!

The choice of system of units is a matter of convention. There are no basic arguments favouring one system of units over another. However, from a practical point of view, not all systems of units are equivalent. The decisive factors for a practical choice of a system of units are the invariance of the objects in terms of which a unit is expressed, convenience, reproducibility, practical convenience of the scale of the unit and the basic nature of the quantities selected as units.

The International System of Units (SI) has been chosen as the basic system by international convention as well as by the governments of most countries. Upon a transition from one system of units to another, the dimensions of physical quantities generally vary.

The Commission also indicated the starting and terminating moments of this year, which we shall not describe here. This definition of a second provided a fairly constant and reliable unit of time. The unit of time (second) can be measured with a relative error not exceeding 10^{-8} by means of the existing experimental techniques.

However, it became clear by the beginning of the sixties that atomic processes are more suitable for defining the unit of time. Atoms emit light whose frequency can be determined very accurately. The XIII General Conference on Weights and Measures decided in 1967 to adopt a certain radiation under definite circumstances as the means for finding the standard of time. The frequency of this radiation was assigned such a numerical value that the value of a second determined with its help was in fairly good agreement with the existing standard.

A second is defined as the duration of 9,192,631,770 periods of the radiation corresponding to the quantum transition between the levels $F = 4$, $m_F = 0$ and $F = 3$, $m_F = 0$ in the hyperfine structure of the ground state $^2S_{1/2}$ of the ^{133}Cs atom.

Using the methods worked out in radiophysics and optics, it is possible to measure very accurately the number of oscillations of the normal radiation emitted by cesium atoms in this interval of time. This enables us to graduate secondary time standards (say, a clock) which in turn can be used to measure intervals of time directly in seconds, hours, etc.

METRE, THE UNIT OF LENGTH. Initially, small lengths were measured in terms of the body. We see traces of this in units like the pace, cubit and foot which are still in use today. There was a huge multitude of such units in history, but they could not come up to the requirements of practice. Hence it was found necessary to create a single and stable scale. The decisive step towards the creation of a unified system was taken after the French revolution. The Earth's meridian was chosen as the physical object forming the basis for the definition of the unit of length. The unit of length was called a metre and was described as exactly 10^{-7} of a quarter the meridian circle. Accordingly, a platinum bar was specially prepared to serve as the standard metre. The metric system of weights and measures was introduced in France in 1799 and gradually extended to other countries. The use of nonmetric units was banned in France in 1840. Following the developments in measurement techniques, it was found that the standard metre was not sufficiently precise (as compared to its definition) and that a platinum-tipped measure of length was not effective. Besides, the invariability of the material of the standard was admitted to be inadequate. Hence the standard with platinum tips was replaced by a platinum-iridium bar (containing 90% platinum and 10% iridium) with two marks engraved on it. After this, the metre was defined in terms of the standard without any mention of the length of the Earth's meridian (according to the measurements carried out in the sixties, for example, one quarter of the Earth's meridian was found to be 10,001,954.5 m long). However, even the new standard changed in length due to recrystallization processes in the alloy. For example, it is believed that the standard became 0.5×10^{-6} m shorter between 1899 and 1957. Hence an incessant search was made to find a more reliable standard of length. The standard adopted at present is based on the constancy of the velocity of light in vacuum outside a gravitational field. This value is assumed to be 299,792,458 m/s. In 1975 the General Conference on Weights and Measures adopted this value of the velocity of light as a universal constant. As a result, the following definition was given to the unit of length:

A metre is the distance traversed by a plane electromagnetic wave in vacuum during 1/299,792,458 of a second.

KILOGRAM, THE UNIT OF MASS. This is the only base unit which is related to a material prototype. The prototype is chosen so that it should be easily reproducible and well preserved.

At first, it was assumed that the prototype was to have a mass equal to the mass of 1 dm^3 of water having its highest density (at a temperature of 3.98°C) under a pressure of 1 atmosphere (101.325 Pa). However, the prototype was found to be 28×10^{-6} kg heavier than 1 dm^3 of water under these conditions.

At present, the international prototype of 1 kg is a cylinder of an alloy of platinum (90%) and iridium (10%), 39 mm in diameter and 39 mm long, which is carefully preserved at the International Bureau of Weights and Measures at Sèvres near Paris.

It has been established that the prototype of mass ensures a constant value of the mass 1 kg with a relative error not exceeding 10^{-8} over several thousand years.

OUT-OF-SYSTEM UNITS. The Committee on Standards, Measures and Measuring Instruments under the USSR Council of Ministers approved and adopted on November 18, 1961 the State All-Union Standard 9867-61 on the International System of Units, which came into effect on January 1, 1963. The abbreviation SI was adopted for the system. During a transition period, use of systems of units other than SI and individual units not pertaining to any system (out-of-system units) was allowed. The USSR State Committee on Standards introduced the State All-Union Standard 8.417-81 from January 1, 1982 for the "Units of Physical Quantities". This was aimed at the total implementation of the International System of Units in the USSR. Most of the out-of-system units will become obsolete, although the deadline of use for all such units has not yet been set. Some out-of-system units, however, may be used for an unlimited period of time together with the SI units. These include the following units which have been retained in mechanics (the corresponding symbols are indicated in parentheses): ton (t), minute (min), hour (h), day (D), litre (l), hectare (ha). The unit electron volt (eV) is allowed for measuring energy. For measuring astronomical distances, astronomical units (a. u.), a light year and a parsec are allowed. Plane angles can be measured in degrees (°), minutes (′) and seconds (″).

PREFIXES FOR FRACTIONAL AND MULTIPLE UNITS. In addition to the SI base units, multiple and fractional units can also be used. These units are obtained by multiplying the base units by 10^n, where n is a positive or negative integer. The multiple and fractional units are named by adding prefixes to the SI base units according to the table on p. 27.

The use of these prefixes simplifies both the notation and pronunciation of the quantities. For example, it is simpler to write 10^{-9} m as 1 nm and call it a nanometre. Two prefixes cannot be used simultaneously. For example, it is not permissible to write 10^{-9} m as $10^{-6} \times 10^{-3} = 1$ μmm and call it a micromillimetre, since prefixes can be assigned only to SI units,

Power	Prefix	
	name	symbol
10^{18}	hexa	H
10^{15}	peta	P
10^{12}	tera	T
10^{9}	giga	G
10^{6}	mega	M
10^{3}	kilo	k
10^{2}	hecto	h
10	deka	da
10^{-1}	deci	d
10^{-2}	centi	c
10^{-3}	milli	m
10^{-6}	micro	μ
10^{-9}	nano	n
10^{-12}	pico	p
10^{-15}	femto	f
10^{-18}	atto	a

and a millimetre is not an SI unit. In general, fractional and multiple units are not SI units.

DIMENSIONAL ANALYSIS. Equalities are possible only between quantities of similar dimensions. If we know the physical quantities involved in a process, dimensional analysis can be used in many cases to establish the nature of functional dependence between them. For example, suppose that we have to establish the functional dependence of distance on time in the case of a uniformly accelerated motion. It is logical to assume that the distance will be expressed by a formula of the type $l = Aa^n t^m$, where l is the distance, a is the acceleration, t is the time, A is a dimensionless constant and n, m are unknown numbers. Since $[l] = L$, $[a] = LT^{-2}$, $[t] = T$ and $[A] = 1$, we obtain $L = L^n T^{m-2n}$. Consequently, $n = 1$, $m - 2n = 0$ and $m = 2$, i.e. $l = Aat^2$. The numerical value of A cannot be determined by dimensional analysis, but this is not important from the point of view of functional dependence.

Many problems can be solved by dimensional analysis. However, we must take into account that dimensional constants may exist. For a more detailed account of this subject, the reader is referred to the literature on dimensional analysis.

Chapter 2

Kinematics of a Point and a Rigid Body

Basic idea:
Kinematics describes specific mechanical motions without going into the cause and the realizability of such motions in nature. Kinematics concerns itself only with the physical substantiation and mathematical rigour within the framework of the accepted models.

Sec. 5. COORDINATE SYSTEMS

The experimental content of geometrical statements is discussed. Important coordinate systems and coordinate transformations are described.

SPACE AND GEOMETRY. All bodies have a length, occupy a certain position and are arranged in a definite manner with respect to one another. *As a result of practical experience, these general properties of bodies were imprinted on the consciousness of human beings as the concept of space. The mathematical formulation of these properties assumed the form of geometrical concepts and relations between them.* Geometry was formulated as a science some two and a half thousand years ago by Euclid.

The concept of space, which arose in our consciousness as a reflection of the properties of bodies, later acquired a relative independence in the minds of some philosophers who interpreted it as an entity capable of existence without bodies. Thus, geometry became a discipline concerning the properties of space capable of existing independently of bodies rather than a branch of science dealing with the properties of bodies. Another group of philosophers did not accept the isolation of the concept of space from the properties of bodies. These two conflicting opinions have existed throughout the development of science.

There is no need for us to follow the whole long and winding road along which these conflicting views were developed. Suffice it to say that Newton's views on space reflect a synthesis of both viewpoints. According to Newton, space can exist independently of bodies as absolute space which in essence remains fixed and unchanged irrespective of all

Nikolai Ivanovich Lobachevski (1793-1856)
Russian mathematician, the founder of non-Euclidean geometry. This discovery has revolutionized the ideas of space based for more than 2,000,000 years on the Euclidean teaching and strongly influenced the evolution of mathematics and physics. He was the author of many scientific works in mathematics, mechanics, physics and astronomy.

external factors. However, there also exists a relative space which is a part of the bounded space associated in our consciousness with certain bodies.

The next important step towards an understanding of the relation between space and bodies was taken by the creators of non-Euclidean geometry. N. I. Lobachevski explained the problem as follows: "Strictly speaking, we only comprehend motion in space for without motion sensory perception is not possible. All other concepts, for example, geometry, are artificially derived by our mind from the properties of motion. Hence space does not exist on its own."

The statement that space and matter are inseparably linked was subsequently developed in the theory of relativity. Philosophically, the development of these ideas culminated in the teachings of dialectical materialism on space and time. According to dialectical materialism, space and time are forms of existence of matter, and hence cannot be conceived without matter.

GEOMETRY AND EXPERIMENT. Geometrical concepts are abstractions of real relations between bodies. Hence geometry is an experimental science. The "building blocks" in geometry are idealized forms of the properties of bodies in the real world, such as a point, line, surface or volume. These forms are used to create the geometrical model of the real world. It was believed for a long time that there was no need to think about a relation between geometry and the real world, since Euclidean geometry was the only conceivable model of the real world. Later, it was shown that there is an infinite number of other models (non-Euclidean geometries) which do not have any intrinsic contradictions. Hence the question as to which model or geometry correctly reflects the real world can be answered only experimentally by comparing all the conclusions drawn from the model with the situation prevailing in the real world.

For example, Euclidean geometry states that the sum of the angles of a triangle is equal to π. In principle, this statement can and must be verified experimentally. Indeed, a straight line is defined as the shortest distance between two points. Hence, by taking three points associated with an object, we can in principle construct a triangle with vertices at these three points. The question that arises now concerns the invariance (rigidity) of the scale of measurement as we go from one point to another, and the invariance of the body with which the three points under consideration are associated. This question can also be answered experimentally, and not just by a particular experiment, but by experimental knowledge. For example, the measurement of length is a comparison of the length of an object with the length of another object taken as the standard. But is there any reason to doubt the invariance of the length of the body taken as the standard? Indeed, there is a very definite reason for doubting it. As a matter of fact, measurement is a comparison of two bodies, both arranged in an identical manner. Hence each individual act of measurement of a body by means of the body taken as the standard is at the same time an act of measurement of this standard with respect to the first body.

Having adopted the length of a body as unity and measured the lengths of all the other bodies with respect to this length, we can draw a conclusion about the unit of length itself. Indeed, let us assume that at a certain instant of time the lengths of all the bodies have changed, say, increasing by 10%. In other words, the numbers denoting the lengths of these bodies have increased by 10%. By definition, the length of the body taken as the standard has remained equal to unity. However, we can look at this event from a different point of view. All the bodies can be adopted as the scale of measurement in turn. Each time we arrive at the conclusion that the lengths of all the other bodies have remained unchanged with the exception of the body which was first adopted as the standard, and whose length has now decreased by 10%. The complete set of results shows that the event under consideration involves not a 10% increase in the length of all the bodies, but rather a 10% decrease in the length of the body taken as the standard. This example shows that it is meaningful to question the invariance of the standard.

The question concerning the invariance of perfectly rigid bodies is also important. The invariance of the scales or standards for physical quantities is gradually being attained, and their suitability is being verified by using all the experience at the disposal of man. In accordance with the results of these investigations, the standards that were once chosen as the base units have gradually been changed. For a long time, it was assumed that the length of the Earth's meridian is constant. This quantity was therefore chosen as the base for the standard of length. However, we now believe that a more suitable quantity from the point of view of invariance and constancy is the distance traversed by light in a vacuum within a fixed period of time in the absence of a gravitational field. This is the kind of definition used to select the SI base unit of length.

Let us now return to the verification of the verity of Euclidean geometry. According to what has been mentioned above, we can state that it is indeed possible to construct a triangle whose sides are uniquely defined. Further, all the angles of the triangle can be measured using a suitable

!

The concept of space is a reflection in human consciousness of the properties of bodies to have a length, to occupy a certain position and to arrange themselves in a definite order relative to one another.

Geometry is an experimental science by origin. Among all conceivable geometries, experience helps in the selection of the one which correctly describes the relation between the geometrical objects of the real world.

The right-handed system of rectangular Cartesian coordinates cannot be made to coincide with the left-handed system by any type of motion in space.

technique. The sum of these results will be either equal to π or not. If the sum is not equal to π, it can confidently be stated that Euclidean geometry is unsuitable as a model of the real world, and some other model is called for. The Pythagorean theorem can be verified in a similar manner. Experimentally, it involves the construction of a right triangle and the measurement of its hypotenuse and legs.

An analysis of many physical phenomena does yield a conclusion concerning the limits of applicability of Euclidean geometry, namely: Euclidean geometry provides a fairly accurate description of the geometry of the real world, starting from distances of the order of one-tenth of the size of the nucleus, i.e. from distances of about 10^{-16} m, to distances close to the "size of the Universe", i.e. about 10^{26} m $\simeq 10^{10}$ light years. At such large distances (of the order of 10 billion light years), however, the non-Euclidean properties of space must start manifesting themselves if the predictions of the theory of relativity are correct. There is every reason to believe that Euclidean geometry remains valid at distances below 10^{-16} m, although the lower limit of its applicability has not yet been established.

However, the statement that Euclidean geometry is applicable near the Earth's surface is valid only with the reservation "to a very high degree of accuracy". The statement is not absolutely true. The general theory of relativity states that geometry in a gravitational field is not Euclidean.

It has recently been verified to a very high precision that on the scale of solar system the geometry of space is non-Euclidean by measuring experimentally the way radio waves are bent in the gravitational field of the Sun and from the delay in the radiosignals.

POINT MASS. The concept of a point mass is the most important abstraction for constructing the models of mechanical systems. *A point mass is defined as a physical object which is mathematically a point in the geometrical sense which has a certain mass.* The phrase "in the geometrical sense" means that a point mass does not have any internal geometrical structure, shape or size.

Whether or not a concrete object can be treated as a point mass depends on the conditions of the problem and the questions which have to be answered. *The same body can be considered a point mass under some conditions, while it is impossible to do so under some other conditions.* For example, the Earth can be treated as a point mass when its motion around the Sun is being considered. However, it cannot be considered a point mass when we are dealing with the Earth's satellites. *A point mass used for simulating a body can be* *ascribed certain physical (but not geometrical!) properties of the body*. For example, it can be assumed that a point mass has an intrinsic angular momentum, charge, etc. This can be done when the corresponding physical properties of the body can be assumed under the conditions of the problem to be localized at a point mass.

BODY. The model of a body in mechanics does not take into consideration its atomic and molecular structure. It is assumed that the mass of the body is uniformly "smeared" over its volume with a certain density ρ. A volume element dV contains a mass $dm = \rho\, dV$. The volume elements are assumed to be different and define different points of the body in the limit $dV \to 0$. This allows us to refer to a body as a set of point masses for the sake of brevity and to use such a concept for simplification in mathematical calculations. *It must, however, be borne in mind that point masses have nothing in common with the real atoms and molecules, and are simply auxiliary entities.* Experience shows us that in some bodies, their different parts have a freedom of movement relative to one another (for example, liquids, free-flowing dry substances). For some other bodies, however, the different parts preserve their relative position, and hence their shape remains unchanged. Such bodies are called solids, or rigid bodies. The relative invariance of the mutual arrangement of the different parts of a solid ensures that its extent in space is relatively constant. As a result, the comparison of the extents of solids acquires a clear physical meaning, and it becomes possible to define the length of a solid and the operation of measurement, and to give a quantitative measure of the relative invariance of the length of a solid with respect to a body taken as a standard or a unit scale. However, even if the ratio of the lengths of two bodies is constant, we cannot, for the time being, obtain a quantitative characteristic for the important concept of a "perfectly riqid body". We must study the mutual relation between different bodies and analyze their stability. After this, we can select those bodies which turn out to be the most stable and invariable, and use them as standards for measurements. It was mentioned above that the use of a standard of measurement can lead to information about its invariance and thus help in its refinement.

Thus, man's activities over many centuries have shown what materials, processes and conditions can be used to define an invariable length and to choose length for measuring the extents of bodies. *This choice was historical in nature and varied over time, since new experience drawn from practice led to new conclusions concerning the relative invariance of the objects of the material world surrounding us.*

DISTANCE BETWEEN POINTS. Having chosen the unit of length, we can measure one-dimensional extents, i. e. the length of a line passing through two points of a body. Any two points of a body can be joined by an infinite number of lines, and the length of each line can be measured. An analysis of all these lengths shows that the largest among these cannot be defined, but there is the smallest length. *The smallest length is called the distance between the two points, and the corresponding line is called a straight line.*

It is possible to measure the distance not only between two point masses belonging to the same object, but also between two points that are isolated in space. For this purpose, it is sufficient to choose in the simplest case a long solid standard metre and arrange it in such a way that the two point masses coincide with certain divisions on the standard metre. The distance between the two divisions will then be equal to the distance between the two point masses. Point masses are situated in space. Hence we sometimes refer to spatial points and to the distance between them. *This, however, does not mean that spatial points can be labelled in some way so that they can be tracked in the same way as point masses.* A spatial point can be defined only with respect to a body. *Hence any statement concerning a spatial point is only meaningful if its position is specified relative to a body.* In turn, the position of a spatial point relative to a body is characterized by the position of an imaginary point mass located at the spatial point. Hence in order to describe a space, we must indicate a reference body relative to which the position of spatial points is defined.

PERFECTLY RIGID BODY. Generally speaking, any object can be chosen as the reference body. However, the concept of a perfectly rigid body was of prime importance in the development of Euclidean geometry. *A perfectly rigid body is the one the distance between any of whose points remains unchanged.* We have already discussed the invariance of the scale used for measuring distances.

The geometrical images and concepts of Euclidean geometry are inseparably linked with the images and concepts formed in our consciousness by the ideas of perfectly rigid bodies and their motion. Consequently, it is desirable to choose a perfectly rigid body as a reference body. This, however, is not always possible. Nevertheless, a perfectly rigid body can always be chosen as a reference body in classical mechanics and in the special theory of relativity. This is what we shall do throughout this course.

REFERENCE FRAME. *The set of all spatial points and the body relative to which the position of the spatial points is determined is called a reference frame.* If a reference frame is

?

What is the meaning of the statement on the geometrical properties of space? What is the meaning of the question whether a certain geometry is authentic or false?
What is a perfectly rigid body and what role does it play in the development of geometrical concepts?
What is the idea behind the concept of invariance of a scale taken as a unit when this property of invariance is valid by definition?

given, the position of a point in space is characterized by the spatial point with which the point mass coincides. Thus the space acquires an "independent" existence, a point mass may move from one spatial point to another, and so on. Our task is to indicate the manner in which the position of spatial points can be described in a reference frame. This is done by introducing a coordinate system.

COORDINATE SYSTEMS. The concepts of distance, line, straight line, angle, etc. are defined in a given reference frame. The problem of establishing relations between them is of experimental nature. Some of the relations seem so obvious that there is a temptation to assign them the status of self-evident propositions which do not require any proof. Such assumptions are called axioms. The building of the entire geometrical structure from the axioms forming the basis of geometry requires simple logical thinking and is not directly associated with experiment. Different systems of axioms lead to different kinds of geometry, which are all valid by themselves without any reference to the real world. Each such geometry is a model of the relations which could generally exist in the real world. *Only experiment can determine which of the feasible geometries is the model of the real physical world.*

Since it has experimentally been established that Euclidean geometry is valid with a very high degree of accuracy over a very wide range of distances in the real physical world, we shall assume that the axioms of Euclidean geometry are valid for all the reference frames throughout this book.

In order to describe the motion of a point mass or of a rigid body, we must agree how to specify the position of a point. It was mentioned above that the "address" of point mass is determined by the "address" of the imaginary point in the reference frame with which the point mass coincides. Hence our task is to assign "addresses" to all the points in a reference frame in such a way that each point has its own individual "address", and each "address" leads to just one point. For this purpose, we introduce a coordinate system. *The introduction of a coordinate system is tantamount to an agreement as to how to assign "addresses" to different points in the reference frame.* For example, it is agreed that the "addresses" of all the points on the Earth's surface are given in the form of numbers having the dimensions of degrees, and called latitude and longitude. Each point on the Earth's surface lies at the intersection of a meridian and a parallel, and its "address" is specified by two numbers corresponding to the meridian and parallel. The assignment of a number to each meridian and parallel is arbitrary, and the only important point is to ensure unambiguity in the assignment of the number: each meridian

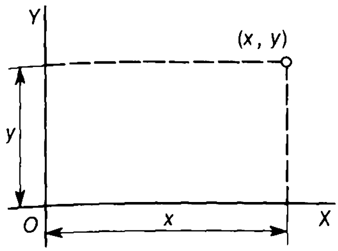


Fig. 1. Rectangular Cartesian coordinate system in a plane.

The two numbers characterizing the position of a point are the distances x and y from the origin to the projections of the point on the coordinate axes.

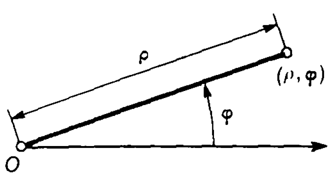


Fig. 2. Polar coordinate system.

The two numbers characterizing the position of a point in a plane are the distance ρ from the origin to the point and the angle φ between the ray drawn from the origin and the line segment connecting the origin and the point.

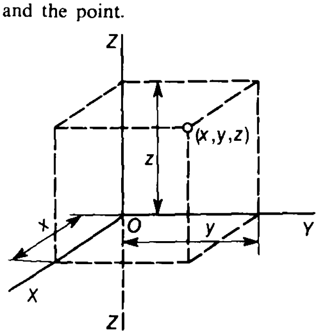


Fig. 3. Rectangular Cartesian coordinate system in space.

The three numbers characterizing the position of a point are the distances x, y and z from the origin to its projections on the coordinate axes.

must be ascribed a definite number and each number must correspond to a definite meridian. For example, instead of characterizing longitude by the angle between the plane of a meridian and the plane of another meridian taken as the starting point, we could characterize it by the distance along the equator from the point where the equator intersects the reference meridian and the point where the equator intersects the meridional plane passing through the point under consideration. In this case, we would have to state that a certain point lies with so many kilometres of longitude and so many degrees of latitude. Of course, *there is in principle no advantage or drawback to the different methods of introducing a particular coordinate system. However, in practice, different coordinate systems have varying significance.* Quite often, each success in solving a problem depends on the correct choice of the coordinate system.

DIMENSIONALITY OF SPACE. It can be seen from the above example that the position of each point is characterized by two numbers, regardless of what these numbers are. The only significant point is that the method for specifying the numbers must ensure a continuity and unambiguity in the "addresses". It is also important that there are two such numbers. This is so because we are considering the surface of the Earth. The position of a point on a surface is defined by two numbers. In other words, we say that a surface has two dimensions.

The space in which we live is three-dimensional. This means that the position of any point in space is characterized by three numbers. The choice of the numbers depends on the coordinate system used for defining the positions of points in space.

The procedure for assigning numbers to a spatial point is called the arithmetization of a space. If the space is arithmetized, there is no need to remember the reference body or the coordinate system since all the information about them is contained in the arithmetization. In this case, it is not necessary (in many cases, it is not even permissible) to associate a reference frame with an imaginary perfectly rigid body. Geometry is based entirely on the arithmetized specification of a space and the square of its infinitesimal linear element. However, such a general approach is not necessary for this book. Within the framework of the special theory of relativity and classical mechanics, the geometry is Euclidean and the reference frames can conveniently be associated with imaginary perfectly rigid bodies. This is because the spatial relations in Euclidean geometry are simply a generalized formulation of the relations between the geometrical characteristics of perfectly rigid bodies, their motion and mutual arrangement.

IMPORTANT COORDINATE SYSTEMS. Only a few of the infinitely large number of possible coordinate systems are

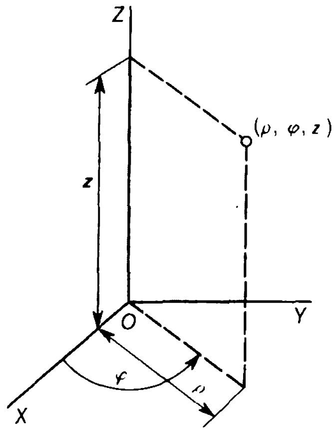


Fig. 4. Cylindrical coordinate system.

The three numbers characterizing the position of a point are the distances ρ and z from the origin and the angle φ between the segment ρ and the X-axis.

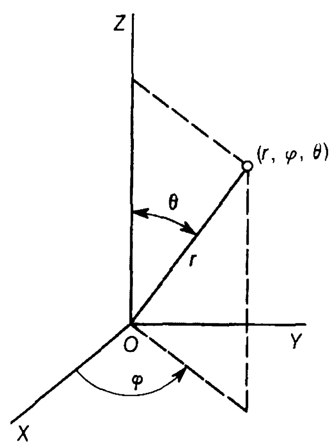


Fig. 5. Spherical coordinate system.

The three numbers characterizing the position of a point are the distance r from the origin and the angles φ and θ.

simple, important and frequently used in practice. Detailed information on these coordinate systems can be found in other textbooks. The following systems should be memorized:

(1) *in a plane*:

(1a) the rectangular Cartesian coordinate system (Fig. 1) in which the two numbers (x, y) characterizing the position of a point are the distances x and y;

(1b) the polar coordinate system (Fig. 2) in which the two numbers (ρ, φ) characterizing the position of a point are the distance ρ and the angle φ;

(2) *in space*:

(2a) the rectangular Cartesian coordinate system (Fig. 3) in which the three numbers (x, y, z) characterizing the position of a point are the distances x, y and z.

It should be remarked that *it is possible to have two rectangular Cartesian coordinate systems in space, which cannot be made to coincide by any translation or rotation. These are called the right-handed and left-handed Cartesian coordinate systems, and their axes are differently oriented.*

If, looking at the XY-plane in the positive direction of the Z-axis, a clockwise rotation of the X-axis is the shortest way to make it coincide with the Y-axis, the system is right-handed. If, on the other hand, an anticlockwise rotation of the X-axis is the shortest way to make it coincide with the Y-axis, the system is left-handed.

Figure 3 shows a right-handed coordinate system. The dashed line shows the Z-direction in a left-handed coordinate system, the X- and Y-directions remaining unchanged. It can easily be seen that a right-handed system cannot be made to coincide with a left-handed system by any translation or rotation.

We must always bear in mind the coordinate system being used, since a transition from the right-handed to the left-handed system involves a sign reversal in some formulas. In most cases, as well as in this book, the right-handed system is used.

(2b) The cylindrical coordinate system (Fig. 4) in which the three numbers (ρ, φ, z) characterizing the position of a point are the distance ρ, the angle φ and the distance z.

(2c) The spherical coordinate system (Fig. 5) in which the three numbers (r, φ, θ) characterizing the position of a point are the distance r, and the angles φ and θ.

The numbers defining the position of a point in a coordinate system are called the coordinates of the point. For the sake of convenience, the coordinates of a point are often described by the same letter but by different subscripts, say, x_1, x_2, x_3. These numbers denote the following coordinates: $x_1 = x$, $x_2 = y$, $x_3 = z$ (in the Cartesian system); $x_1 = \rho$, $x_2 = \varphi$,

$x_3 = z$ (in the cylindrical system) and $x_1 = r, x_2 = \varphi, x_3 = \theta$ (in the spherical system).

COORDINATE TRANSFORMATIONS. The formulas connecting the coordinates of a point in one system with its coordinates in another system are called coordinate transformations. We shall describe here the transformation formulas between cylindrical, spherical and Cartesian coordinates. These formulas can be obtained by visual inspection from Figs. 4 and 5.

Transformation from cylindrical coordinates to rectangular Cartesian coordinates:

$$x = \rho \cos \varphi, \quad y = \rho \sin \varphi, \quad z = z. \tag{5.1}$$

Transformation from spherical to Cartesian coordinates:

$$\begin{aligned} x &= r \sin \theta \cos \varphi, \\ y &= r \sin \theta \sin \varphi, \\ z &= r \cos \theta. \end{aligned} \tag{5.2}$$

Transformation formulas are of great practical importance for going over from one Cartesian coordinate system to another when their origins and the directions of their axes do not coincide. However, it is more convenient to analyze this situation using vector concepts.

Sec. 6. VECTORS

Basic concepts of vector algebra are described.

DEFINITION OF A VECTOR. Many physical quantities are characterized by just one number. These include, for example, temperature, which is expressed in degrees of a particular scale, and mass, which is expressed in kilograms. Such quantities are called scalars. In order to characterize other physical quantities, however, it is necessary to specify several numbers. For example, velocity has both magnitude and direction. For example, it can easily be seen in Fig. 5 that a direction in space is completely defined by two numbers, viz. the angles φ and θ. Hence velocity is described by three numbers in all. Such quantities are called vectors.

While considering vector quantities, we must clearly demarcate two aspects of the problem: firstly, the definition and properties of vectors as mathematical quantities; secondly, an analysis of the properties of a physical quantity and whether it can be represented by a vector.

The mathematical definitions of a vector and vector operations are not associated with the point of application of the

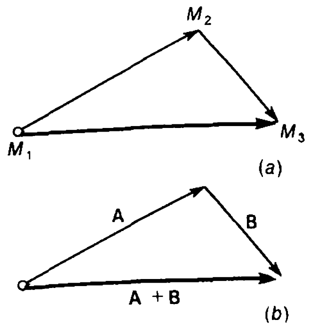


Fig. 6. Addition of vectors.
The vector addition rule is a natural generalization of the obvious rule for displacement addition.

vector, since they are described solely in terms of the three numbers defining the vector. The point of application is only significant for a geometrical interpretation of the vector. It is clear that *in the mathematical definition of a vector, the point of its application is arbitrary and is identified with a specific point only for the sake of convenience or visual interpretation.*

A vector is depicted by an arrow whose length is proportional to the quantity being represented by it, and the arrow points in the direction of the vector. In this book, vectors will be denoted by blue letters, for example, vector A, while their absolute numerical value, called the magnitude of the vector, will be represented either by a blue letter enclosed within two vertical bars: $|\mathrm{A}|$ or by the letter of an ordinary type: A.

ADDITION OF VECTORS AND MULTIPLICATION OF A VECTOR BY A SCALAR. Displacement is one of the most important concepts associated with a vector. The displacement of a point mass from position M_1 to position M_2 (Fig. 6*a*) is described by vector $\overrightarrow{M_1M_2}$, which is represented by the segment joining points M_1 and M_2, and directed from M_1 to M_2. If the point then moves from M_2 to M_3, this sequence of two displacements, i.e. the sum of two displacements, is equivalent to a single displacement $\overrightarrow{M_1M_3}$, which can be written in the form of the vector displacement equation

$$\overrightarrow{M_1M_2} + \overrightarrow{M_2M_3} = \overrightarrow{M_1M_3}. \tag{6.1}$$

This formula expresses the addition rule for vectors. Sometimes, it is called the parallelogram law, since the sum of the vectors is equal to the diagonal of the parallelogram whose sides are formed by the two vectors being added. *By definition, the addition formula is applicable to any two vectors.* Figure 6*b* shows the addition of two arbitrary vectors A and B.

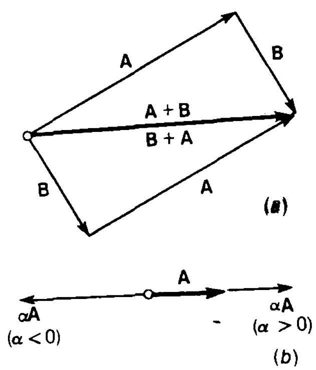


Fig. 7. Commutativity of vector addition (*a*) and multiplication of a vector by a scalar (*b*).
The sum of two vectors does not depend on the order of addends. When a vector is multiplied by a negative scalar, its direction is reversed.

From the example on the addition of displacements, it is clear that *the sum of vectors does not depend on the order in which the displacements take place, i.e. on the order in which vectors are added* (Fig. 7*a*):

$$\mathrm{A} + \mathrm{B} = \mathrm{B} + \mathrm{A}. \tag{6.2}$$

This rule can be extended to the addition of all types of vectors.

Multiplication of a vector by a scalar boils down to the multiplication of the magnitude of the vector by the scalar without changing the vector's direction if the number is positive, but reversing the vector's direction if the number is negative (Fig. 7*b*).

SCALAR PRODUCT. The scalar product $\mathrm{A} \cdot \mathrm{B}$ of two vectors A and B is the scalar which is the product of the magnitudes of

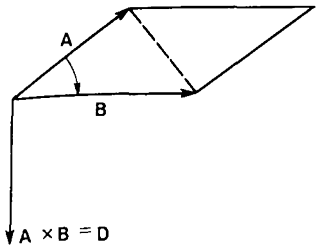


Fig. 8. Vector product A × B = D.

This vector is normal to the plane in which the vectors to be multiplied lie.

!

The vector addition rule is a definition whose expedience is proved by the properties of a number of simple physical quantities.
In order to present a physical quantity in the form of a vector, it is essential that it be the vector sum of its components, and the vector projections corresponding to the components be transformed upon a transition from one coordinate system to another in accordance with the transformation rules for the projections of mathematical vectors. In other words, the projections of the components must be transformed as the projections of the radius vector under the transformation (6.20).

the vectors and the cosine of the angle between them:

$$\mathbf{A}\cdot\mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos(\widehat{\mathbf{A},\mathbf{B}}). \tag{6.3}$$

It can easily be verified that the following rules are valid for the scalar products of vectors:

$$\begin{aligned} &\mathbf{A}\cdot\mathbf{B} = \mathbf{B}\cdot\mathbf{A}, \\ &\mathbf{A}\cdot(\mathbf{B}+\mathbf{C}) = \mathbf{A}\cdot\mathbf{B} + \mathbf{A}\cdot\mathbf{C}, \\ &\mathbf{A}\cdot\alpha\mathbf{B} = \alpha\mathbf{A}\cdot\mathbf{B} = \alpha(\mathbf{A}\cdot\mathbf{B}), \end{aligned} \tag{6.4}$$

where α is an arbitrary number.

VECTOR PRODUCT. The vector product A × B of two vectors A and B is the vector D = A × B which is defined in the following way (Fig. 8):

(1) it is perpendicular to the plane containing vectors A and B being multiplied, and is directed towards the side in which a right-handed screw would move if rotated in the direction vector A must be turned to meet vector B through the shortest path. In other words, vectors A, B and A × B are oriented in the same way as the positive X-, Y- and Z-directions in a right-handed coordinate system;

(2) the magnitude of the vector product is equal to the product of the magnitudes of the vectors being multiplied and the sine of the angle between them:

$$|\mathbf{D}| = |\mathbf{A}\times\mathbf{B}| = |\mathbf{A}||\mathbf{B}|\sin(\widehat{\mathbf{A},\mathbf{B}}). \tag{6.5}$$

Here it is important that the angle between vectors A and B is measured from the first factor A to the second factor B along the smallest arc, i.e. the angle between A and B is less than or equal to π so that the sine in (6.5) cannot be negative. It follows from (6.5) that the magnitude of the vector product is equal to the area of the parallelogram formed by the vectors being multiplied (see Fig. 8).

The following properties of the vector product can easily be verified:

$$\begin{aligned} &\mathbf{A}\times\mathbf{B} = -\mathbf{B}\times\mathbf{A}, \\ &\mathbf{A}\times(\mathbf{B}+\mathbf{C}) = \mathbf{A}\times\mathbf{B} + \mathbf{A}\times\mathbf{C}, \\ &\mathbf{A}\times\alpha\mathbf{B} = \alpha\mathbf{A}\times\mathbf{B} = \alpha(\mathbf{A}\times\mathbf{B}). \end{aligned} \tag{6.6}$$

REPRESENTATION OF VECTORS IN TERMS OF A UNIT VECTOR. The direction of a vector can be indicated in terms of a dimensionless unit vector. Any vector A can be represented

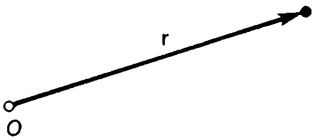


Fig. 9. Radius vector.
The position of any point in space relative to point O taken as the reference point is completely characterized by its radius vector.

in the following form:

$$A = \frac{A}{|A|}|A| = n|A| = nA, \tag{6.7}$$

where $n = A/|A|$ is a dimensionless unit vector that gives the direction of vector A.

ADVANTAGES OF VECTOR NOTATION. The concept of a vector and all the operations associated with it are introduced independently of any coordinate system. This makes it possible to use physical quantities without resorting to their expression in any particular coordinate system. Certain relations between physical quantities are more simply expressed in vector form than in coordinate form. These are significant advantages of vector notation which ensure its wide application. On the other hand, it is often more convenient to carry out numerical computations in coordinate form in which all the quantities are purely algebraic entities.

Hence it is important that we should be able to write all vector expressions and operations in coordinate form. Above all, we should know how to do this in Cartesian coordinates.

RADIUS VECTOR. It is convenient to describe the position of spatial points in terms of their radius vectors. *The radius vector of a point is the vector which begins at the origin of the coordinate system and terminates at the point under consideration* (Fig. 9). If the position of a point is specified in terms of a radius vector, there is no need to use any coordinate system. The positions of points can be described by radius vectors in incoordinate form.

!

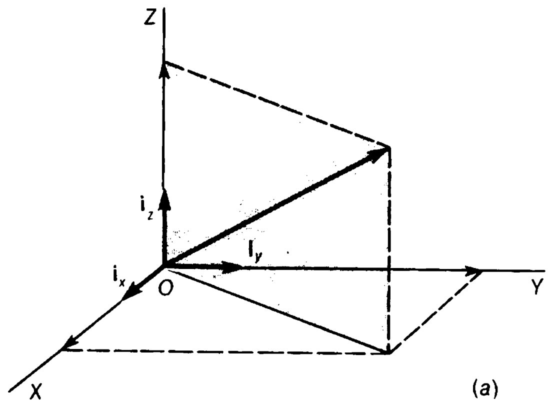
A direction in space is determined by two numbers.
By definition, the radius vector emanates from the origin of coordinates. Other vectors, generally, originate at other points. The relation between the positions of a point relative to different origins can easily be established in terms of radius vectors.
The relation is expressed through the magnitudes of quantities in different coordinate systems with the help of the coordinate transformations and has a more complicated form.

VECTOR PROJECTIONS IN THE CARTESIAN COORDINATE SYSTEM. Suppose that a point O is taken as the reference point. We choose a Cartesian coordinate system with its origin at O. The position of any point can be characterized either by its radius vector r or by the three numbers (x, y, z) which are the Cartesian coordinates of the point. Let us establish a relation between r and the numbers x, y, z. We first introduce dimensionless unit vectors in the positive X-, Y- and Z-directions and denote these unit vectors by i_x, i_y and i_z. Taking into consideration the addition vector rule (6.1) and formula (6.7), we can see from Fig 10*a* that the radius vector r can be represented as the sum of three vectors $i_x x$, $i_y y$ and $i_z z$, which are also directed along the coordinate axes:

$$r = i_x x + i_y y + i_z z. \tag{6.8}$$

The numbers x, y and z are called the projections of the radius vector r. They coincide with the coordinates of the point characterized by r.

In fact, any vector can be represented as the sum of vectors



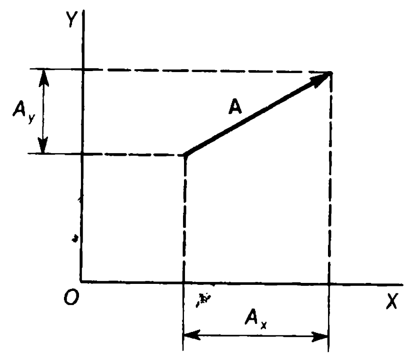


Fig. 10. The projection of a radius vector in a spatial Cartesian coordinate system (*a*) and of an arbitrary vector A in the same system in a plane (*b*).

The projections of a vector on coordinate axes are algebraic quantities, their sign being determined by the sign of the cosine of the angle between the directions of the vector and the unit vector of the corresponding axis. The projections of the radius vector are coordinates of the point characterized by it.

directed along the coordinate axes (Fig. 10*b*):

$$\mathrm{A} = \mathrm{i}_x A_x + \mathrm{i}_y A_y + \mathrm{i}_z A_z. \tag{6.9a}$$

The numbers A_x, A_y and A_z are the projections of vector A onto the *X*-, *Y*- and *Z*-axes. To be able to find the projections of a vector and to express all vector operations in coordinate form, we must know certain relations between the unit vectors i_x, i_y and i_z.

Equation (6.9a) can also be written in the form

$$\mathrm{A} = \mathrm{A}_x + \mathrm{A}_y + \mathrm{A}_z, \tag{6.9b}$$

where vectors $\mathrm{A}_x = \mathrm{i}_x A_x$, $\mathrm{A}_y = \mathrm{i}_y A_y$ and $\mathrm{A}_z = \mathrm{i}_z A_z$ are called the components of vector A in the *X*-, *Y*- and *Z*-directions.

RELATION BETWEEN VECTORS i_x, i_y AND i_z. Since these vectors are unit vectors and are orthogonal, we obtain

$$\begin{gathered} \mathrm{i}_x^2 = \mathrm{i}_y^2 = \mathrm{i}_z^2 = 1, \\ \mathrm{i}_x \cdot \mathrm{i}_y = 0, \quad \mathrm{i}_x \cdot \mathrm{i}_z = 0, \quad \mathrm{i}_y \cdot \mathrm{i}_z = 0. \end{gathered} \tag{6.10}$$

By the definition of the vector product, we arrive at

$$\mathrm{i}_x \times \mathrm{i}_y = \mathrm{i}_z, \quad \mathrm{i}_y \times \mathrm{i}_z = \mathrm{i}_x, \quad \mathrm{i}_z \times \mathrm{i}_x = \mathrm{i}_y,$$
$$\mathrm{i}_x \times \mathrm{i}_x = 0, \quad \mathrm{i}_y \times \mathrm{i}_y = 0, \quad \mathrm{i}_z \times \mathrm{i}_z = 0. \tag{6.11}$$

COMPUTATION OF VECTOR PROJECTIONS. The scalar multiplication of both sides of Eq. (6.9) successively by i_x, i_y and i_z and a comparison of the result with (6.10) lead to

$$A_x = \mathrm{A} \cdot \mathrm{i}_x, \quad A_y = \mathrm{A} \cdot \mathrm{i}_y, \quad A_z = \mathrm{A} \cdot \mathrm{i}_z. \tag{6.12}$$

It can easily be seen that the projections of the vectors on the axes of a rectangular Cartesian coordinate system are nothing but the projections of the vectors on the axes, calculated using the sign rule. For example,

$$A_x = \mathrm{A} \cdot \mathrm{i}_x = |\mathrm{A}|\,|\mathrm{i}_x| \cos(\widehat{\mathrm{A}, \mathrm{i}_x}) = |\mathrm{A}| \cos(\widehat{\mathrm{A}, \mathrm{i}_x}),$$

where $(\widehat{\mathrm{A}, \mathrm{i}_x})$ is the angle between vector A and the positive X-direction. This proves the above statement. A similar situation prevails for the other projections.

EXPRESSION OF VECTOR OPERATIONS IN COORDINATE FORM. In order to obtain these expressions, we must represent vectors in the form (6.9a) and use the formulas obtained earlier for unit vectors. Suppose that we are given

$$\mathrm{A} = \mathrm{i}_x A_x + \mathrm{i}_y A_y + \mathrm{i}_z A_z,$$
$$\mathrm{B} = \mathrm{i}_x B_x + \mathrm{i}_y B_y + \mathrm{i}_z B_z. \tag{6.13}$$

Adding vectors A and B, we obtain

$$\mathrm{C} = \mathrm{A} + \mathrm{B} = \mathrm{i}_x(A_x + B_x) + \mathrm{i}_y(A_y + B_y) + \mathrm{i}_z(A_z + B_z). \tag{6.14}$$

Thus, the projection of the sum of two vectors is equal to the sum of the projections of the components:

$$C_x = A_x + B_x,$$
$$C_y = A_y + B_y, \tag{6.14a}$$
$$C_z = A_z + B_z.$$

?

What are the projections of a radius vector?
Describe the method for finding the coordinate expressions for vector operations.
Under what conditions can a physical quantity be called a vector?
Describe two methods for geometrical determination of the vector sums.

Similarly, it can be seen that the multiplication of a vector by a scalar boils down to the multiplication of each of its projections by the scalar. For a scalar product, we obtain the following expression from (6.10):

$$\mathrm{A} \cdot \mathrm{B} = A_x B_x + A_y B_y + A_z B_z. \tag{6.15}$$

Directly computing the vector product by means of (6.11), we obtain:

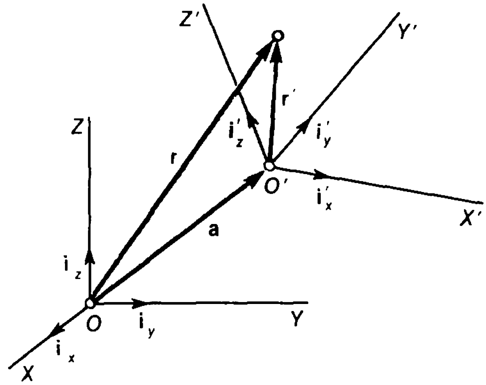


Fig. 11. Coordinate transformations.

Vector a characterizes the position of the origin of the primed coordinate system relative to the unprimed system, while the cosines of the angles between the unit vectors of the two systems determine their mutual orientation in space.

$$\mathrm{A} \times \mathrm{B} = \mathrm{i}_x(A_yB_z - A_zB_y) + \mathrm{i}_y(A_zB_x - A_xB_z) + \mathrm{i}_z(A_xB_y - A_yB_x). \tag{6.16}$$

TRANSFORMATION OF CARTESIAN COORDINATES. Using vector notation, we can find the formulas for a coordinate transformation from one Cartesian system to another. In the general case, neither the origins of coordinates nor the axes of the two systems coincide (Fig. 11). The position of the origin of the K' coordinate system relative to the origin of the K system is described by vector a. It can be seen from Fig. 11 that the radius vectors r and r′, which define the position of a point in the K and K' systems, are related thus:

$$\mathrm{r} = \mathrm{a} + \mathrm{r}'. \tag{6.17}$$

If we express r and r′ in terms of their components along the coordinate axes, we can write

$$\mathrm{i}_x x + \mathrm{i}_y y + \mathrm{i}_z z = \mathrm{a} + \mathrm{i}'_x x' + \mathrm{i}'_y y' + \mathrm{i}'_z z'. \tag{6.18}$$

In order to determine the relation between the coordinates of a point, we must form the scalar product of both sides of this equation and the corresponding unit vector. For example, the scalar product of both sides of (6.18) and i_x will lead to the x-coordinate. Thus we get

$$x = \mathrm{a}\cdot\mathrm{i}_x + \mathrm{i}'_x\cdot\mathrm{i}_x x' + \mathrm{i}'_y\cdot\mathrm{i}_x y' + \mathrm{i}'_z\cdot\mathrm{i}_x z',$$

or, which is the same,

$$x = a_x + \cos(\widehat{\mathrm{i}'_x, \mathrm{i}_x})x' + \cos(\widehat{\mathrm{i}'_y, \mathrm{i}_x})y' + \cos(\widehat{\mathrm{i}'_z, \mathrm{i}_x})z'. \tag{6.18a}$$

Thus, in order to effect a transformation, we must know the angles between the coordinate axes and the mutual arrangement of the origins of coordinates.

The expression for the y- and z-coordinates are obtained in a similar manner. In order to derive the inverse transformations for x', y' and z', we must form the scalar product with the unit vectors i'_x, i'_y and i'_z respectively. For example, the scalar product of both sides of (6.18) and i'_x is

$$i_x \cdot i'_x x + i_y \cdot i'_x y + i_z \cdot i'_x z = a \cdot i'_x + x',$$

or

$$x' = -a'_x + \cos(\widehat{i_x, i'_x})x + \cos(\widehat{i_y, i'_x})y + \cos(\widehat{i_z, i'_x})z. \tag{6.19}$$

Here, $a'_x = a \cdot i'_x$ is the x-projection of vector a in the K' coordinate system. This vector points towards the origin of the K' coordinate system. If we change its direction so that it originates at the point O' in the K' system and terminates at the point O in the K system, the sign of the first term on the right-hand side of (6.19) will be reversed and this relation will become the same as (6.18a). If the origins of the two coordinate systems coincide, vector a becomes equal to zero.

In order to simplify the transformations, we introduce the notation

$$x = x_1, \quad y = x_2, \quad z = x_3;$$
$$x' = x'_1, \quad y' = x'_2, \quad z' = x'_3;$$
$$i_x = e_1, \quad i_y = e_2, \quad i_z = e_3;$$
$$i'_x = e'_1, \quad i'_y = e'_2, \quad i'_z = e'_3;$$

$\cos(\widehat{e_m, e_n}) = \alpha_{mn}$ $(m = 1, 2, 3; \quad n = 1, 2, 3)$. Using this notation, we can write the transformations (6.18a) for $a = 0$ as

$$\begin{aligned} x_1 &= \alpha_{11}x'_1 + \alpha_{12}x'_2 + \alpha_{13}x'_3, \\ x_2 &= \alpha_{21}x'_1 + \alpha_{22}x'_2 + \alpha_{23}x'_3, \\ x_3 &= \alpha_{31}x'_1 + \alpha_{32}x'_2 + \alpha_{33}x'_3. \end{aligned} \tag{6.20}$$

?

Does the scalar product of two quantities depend on the order of the factors? Substantiate your answer.
What is the dependence of a vector product on the order of the factors?
Will the definition of a vector product change if we replace the right-handed Cartesian coordinate system by the left-handed system?
What is meant by the projections of a vector? How is their sign determined?

These transformations are carried out by a rotation of the rectangular Cartesian coordinate systems having a common origin.

Let us consider the application of (6.20) to the two-dimensional case ($x_3 = 0$, $x'_3 = 0$) shown in Fig. 12:

$$\alpha_{11} = \cos(\widehat{e_1, e'_1}) = \cos\varphi,$$
$$\alpha_{12} = \cos(\widehat{e_1, e'_2}) = -\sin\varphi,$$
$$\alpha_{21} = \cos(\widehat{e_2, e'_1}) = \sin\varphi,$$
$$\alpha_{22} = \cos(\widehat{e_2, e'_2}) = \cos\varphi.$$

Hence the transformations in (6.20) assume the form

$$\begin{aligned} x_1 &= \cos\varphi \cdot x'_1 - \sin\varphi \cdot x'_2, \\ x_2 &= \sin\varphi \cdot x'_1 + \cos\varphi \cdot x'_2. \end{aligned} \tag{6.21}$$

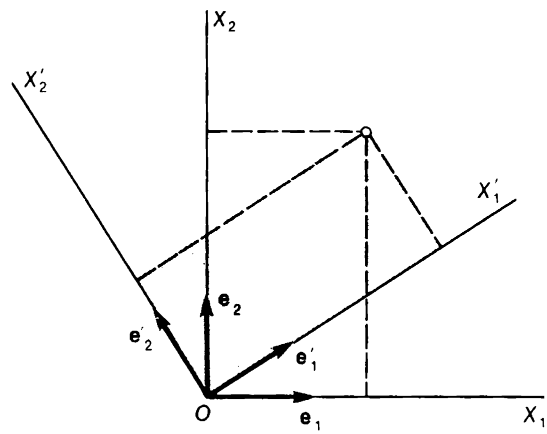


Fig. 12. Rotation of a coordinate system.

If the origins coincide in the two-dimensional case, the mutual orientation of the coordinate axes is completely characterized by the angle of rotation between the X_1- and X_1'-axes.

TRANSFORMATION OF VECTOR PROJECTIONS. It has already been mentioned above that not every quantity characterized by three numbers is a vector.

A quantity described by three numbers is a vector only if upon a transition from one coordinate system to another the vector components behave like the projections of the radius vector under the transformations (6.20).

PHYSICAL VECTOR. It is not easy to say whether a physical quantity having a numerical value and a direction can be represented as a vector. In the first place, it must be emphasized that *the properties of a physical vector are not always identical to those of a mathematical vector.* For example, the point of application is quite significant for many physical vectors and cannot be moved anywhere. For example, the vector of a force acting on a point mass must be applied to the point. This does not impose any restriction on the representation of a physical quantity by a vector since, by definition, the point of application of a mathematical vector is arbitrary and may be located wherever needed by the physical requirements. However, this means that in order to completely characterize a physical vector, we must not only indicate its projections, but also the point of its application.

In order to represent a physical quantity in the form of a vector, it is essential that it must be the vector sum of its components, and that the vector projections corresponding to these components be transformed from one coordinate system to another in accordance with the rules for transforming the projections of mathematical vectors, i.e. like the projections of the radius vector under the transformations (6.20).

Thus, the representation of a physical quantity in terms of a vector depends on its physical properties and not on whether a directed segment of a straight line can formally be projected onto the coordinate axes. *Consequently, the definition of a physical vector does not include the addition rule at all, but just the rule for forming a physical vector from its own components as vectors, and the rule for transforming vector projections from one coordinate system to another.*

In classical physics, we consider three-dimensional vectors whose components point along the axes of a spatial Cartesian coordinate system since time is independent of the coordinate system. Hence in the definition of a vector, the transformation of vector projections means a transformation in accordance with (6.20). A direct inspection shows that a three-dimensional displacement of a point satisfies this definition, i.e. displacement is a vector. Since the transformations (6.20) are independent of intervals Δt of time, the velocity of a point and the difference between its velocities over an interval Δt of time are also vectors. Hence the acceleration of a point is also a vector. The vector properties of other quantities in classical physics can be analyzed in a similar manner.

In relativistic physics, the analysis of the vector properties of quantities cannot be confined to three spatial dimensions since in this case time depends on the coordinate system and appears in the transformations on an equal footing with the spatial coordinates. In other words, when defining vectors, we must consider transformations of the type (6.20), but for four independent projections of the radius vector. This means that the vectors must be characterized in this case by four projections, i.e. in relativistic theory we consider four-dimensional vectors rather than three-dimensional ones. These vectors are defined in the same way as the three-dimensional vectors in classical physics, but on the basis of relativistic transformations of the spatial coordinates and time (see Sec. 13).

Sec. 7. TIME

The philosophical and physical meanings of the concepts of time are discussed and the measurement of time intervals is described.

CONCEPT OF TIME. The world around us is continuously changing. Processes take place in a definite sequence, and each process has a certain duration. The world is in a perpetual state of evolution. These general properties of the changing and growing world are reflected in human consciousness as the concept of time.

By the time we mean the property of processes to have a certain duration, to follow one another in a certain sequence and to develop in steps and stages.

Thus, time cannot be separated from matter and its motion, and is a form of matter. Just as there is no sense in speaking of space as an entity, it is meaningless to speak of time as such. *The concept of passage of time without any relation to processes is devoid of meaning.*

Only the investigation of the processes and the relations between them provide a physical meaning to the concept of time.

PERIODIC PROCESSES. Among the vast multitude of processes occurring in nature, those that repeat, attract the most attention. These include the cycles of day and night, the seasons, the movement of stars in the sky, the heartbeat and breathing. Their study and comparison lead to the idea of duration, while a comparison of their durations leads to the idea of their measurement.

Analyzing all kinds of processes, we can single out those processes that have the most stable durations, and so we can choose a process which serves as the standard. The situation is identical to the one described in Sec. 5 for length, and hence there is no need to repeat the arguments we put forth there.

A periodic process used as a standard is the one that regulates clocks. It is obvious that all clocks must have the same speed in various reference frames.

Let us discuss the physical meaning of this requirement.

Suppose that a physical process can carry information from one point to another. Such a process is called a signal. A signal can be in the form of a flash of light or a bullet fired from one point at another. There is no need to know by what law signals propagate, and it is sufficient to know that the transmission, propagation and reception of the signals take place under invariable, and hence identical, conditions.

Let us transmit signals from one point to another at regular intervals of time measured by means of a clock located at the first point. *If the signals arrive at the other point after the same intervals of time according to a clock located there, it can be stated that the clocks at the two points are at the same speed.*

This verification can be made in principle for all possible pairs of points. The following condition is satisfied in this case: if the speed of a clock at point A is the same as that at point B, and if the speed of the clock at point B is the same as that at point C, then the speed of the clocks at points A and C will also be the same. Of course, we need not confine ourselves to just one type of signal. That clocks at different points in the reference frame operate at the same speed must be verified using every kind of signal at the disposal of the experimenter.

In principle, these experiments may yield two results, viz. the clocks at different points in the reference frame all work at the same speed or the clocks at different points work at different

!

The concept of time is the reflection in human consciousness of the feature of processes to have a certain duration, a certain order of occurrence, and their evolution in steps and stages.
The concept of the passage of time is meaningless unless it is associated with a process. There is no sense to speak about time as such. Only an investigation of the physical processes and their interaction gives a physical meaning to the concept of time.

speeds. Both these results, which were fictitious in principle, have experimentally been shown to be possible. For example, let us consider an intra-atomic process as a standard process (clocks) since it is independent of pressure, temperature effects, etc. We shall try to verify whether the process occurs at the same speed using the method described above. Suppose that at the beginning of the process, a signal is sent from a point above the Earth's surface to a point on the surface, where the same process is occurring. We assume that the signal arrives at the second point at the instant the process begins there. The next signal is sent from the first point at the instant the process at the first point ends. The law governing the signal's propagation from the first point to the second is of no consequence. The only important thing is that the propagation should occur in exactly the same way for both signals, in other words, that the conditions of transmission, propagation and reception of signals must be the same for any sequence of signals. Experiment shows that the second signal will arrive at the Earth not at the instant the process at the second point ends, but somewhat earlier.

These experiments have indeed been carried out recently and will be described in greater detail at a later stage. Here, it is important just to note that the situation is possible in principle, namely, when the rates of physical processes are different at different points in the reference frame. This possible experimental situation is expressed in the form of a statement that time is not the same at each point in the reference frame, and that the rate at which time passes is different at different points. Strictly speaking, this situation will always prevail in a reference frame fixed to the Earth. But the difference between the speeds of clocks at different points near the Earth's surface is quite insignificant. For example, if the difference between the heights of two points above the Earth's surface is about 10 m, the duration of a process at the points will differ by about 10^{-15} of the entire duration. Such an absolutely tiny difference was first established experimentally in 1960. If such small disparities are neglected, we can state with a high degree of precision that a unified time exists in the reference frame fixed to the Earth.

Henceforth, we shall consider only such reference frames in which a unique time can be introduced if not in the absolute sense, then at least with a high degree of accuracy. It should be observed that the impossibility of introducing a unique time near the Earth's surface is due in principle to the gravitational field. However, this field is not strong, and we can speak of a unique time to a very close approximation.

However, gravitation is not the only factor hindering the introduction of a unique time. Let us suppose, for example, that the reference frame is rotating relative to fixed stars or is moving with an acceleration relative to them. In this reference frame too, we cannot introduce a unique time. Such reference frames are called noninertial. A unique time can be introduced in these frames only with a certain error. This will be used in Chap. 7.

Thus, all reference frames can be divided into two classes. In one class of frames, time is unique and Euclidean geometry is applicable. In the other, there is no unique time and the geometry is non-Euclidean.

How can we in practice determine the class to which a particular reference frame belongs? A direct investigation of the properties of space and time seems to be the most natural approach. But this method is ineffective since in most cases the difference between the quantitative characteristics of space and time for the two classes of frames is much smaller than the error involved in measuring the properties themselves. Hence the two types of reference frame in kinematics are practically indistinguishable. The difference between the two classes becomes significant only in dynamics. *The first class contains reference frames in which no gravitational forces are present and Newton's first law of motion is applicable. Hence these frames are termed* inertial frames. *Reference frames in which gravitational forces act and Newton's first law is not obeyed are called* noninertial frames.

A large number of experiments and observations have shown that all inertial frames move without any perceptible acceleration and rotation relative to fixed stars. Their motion relative to one another is also without any acceleration and rotation. Hence in many cases the most convenient way of verifying whether a frame is inertial is to establish that its motion relative to fixed stars is free from acceleration and rotation.

Another reference frame which is convenient in many respects and is equivalent to fixed stars has been developed in recent decades. This is the reference frame associated with relict radiation. According to modern concepts, the Universe was formed about 10-15 billion years ago by an explosion of a superdense state of matter. This explosion generated matter and electromagnetic radiation. For a certain period of time, the matter and radiation could be transformed into each other. However, this interaction came to a stop after some time, and the electromagnetic radiation that remained continued to exist independently. As a result of the expansion of the Universe, the temperature of the radiation dropped to its present value of about 2.7 K. This radiation is isotropic. The reference frame in which relict radiation is isotropic is assumed to be at rest relative to the radiation. In principle, this reference frame is equivalent to the one of fixed stars. The movement of any other frame relative to this one can be determined from the anisotropy of the relict radiation in the frame.

SYNCHRONIZATION OF CLOCKS. The duration of a physical process at a point is measured by means of clock located at the same point. The measurement of a duration of time boils down to the determination of the beginning and end of the measuring process on the scale of another process which is taken as the standard. This is done by reading the clock at the beginning and end of the process, although it is not related in any way to the location of the clock (process) at the point where measurements are made. The results of measurements enable us to compare the durations of processes at different points under the condition that each process occurs at the same point from beginning to end. But what about a physical process which begins at one point and ends at another? What do we mean by the duration of this process? How to measure the duration of this process? Obviously, it cannot be measured by a single clock. We can only record the beginning and end of the process by means of the clocks located at the different points. However, these readings will not be any good since no reference point for time has been fixed for the different clocks or, in other words, the clocks have not been synchronized.

A simple method to be adopted to synchronize clocks would be to adjust the hands of all the clocks "simultaneously" to the same reading. However, this statement has no meaning either since it is not clear as to what we mean by "simultaneously". Hence we must define the synchronization of clocks in terms of the physical procedures with which this synchronization is associated rather than in terms of some other known concepts. In the first place, we must establish a physical relation between the clocks at the points where they are located, i.e. we must again turn to signals. However, in this case, not only must the signals be propagated under invariable physical conditions, but the law governing their propagation assumes significance. However, the law governing the propagation of a signal is not known and cannot be established without synchronized clocks.

From a logical and historical point of view, the synchronization of clocks and the study of the laws governing the propagation of various physical signals were developed simultaneously, both development processes often completing and refining each other. The very large value of the velocity of light played an extremely important role in the development. As a matter of fact, light was from the very beginning a natural signal to be used to synchronize clocks, and its velocity was considered to be almost infinite in comparison with all other known velocities. This gave rise to the idea of synchronizing clocks using a signal which propagates at an infinite velocity. The idea is realized as follows: the hands of the clocks at all points are fixed at the same position. Then signals are emitted from a certain point in all directions, and each clock is started at the instant the signal passes through the point the clock is located. This synchronization has a very important property: if clock A is synchronized with clock B, and clock B is synchronized with clock C, then clock A is synchronized with clock C whatever the arrangement of the clocks A, B and C.

Light signals can be used for signals that propagate at an infinite velocity. Naturally, this will be only an approximate synchronization with an error appoximately equal to the time it takes for the light to propagate between the two most distant points in the region under consideration. For example, this synchronization is satisfactory for most situations in everyday life. It was also satisfactory for quite a long time for scientific investigations in laboratories. In particular, it enabled a study of mechanical motion at low velocities and led to the concept of constant velocity. After this, it became possible to synchronize clocks using a signal that propagates at a finite velocity. Essentially, this is done using the definition of constant velocity, namely, that if we transmit a signal at a constant velocity v from a point the clock shows a time t_0, the clock must show a time $t = t_0 + s/v$ when the signal arrives at a point a distance s from its starting point. It can easily be seen that this synchronization will agree with the synchronization achieved using a signal that propagates at an infinite velocity.

An increase in the accuracy of measurement of the time intervals and an extension of the region over which measurements are carried out not only showed that velocity of light is finite but that the velocity can be measured. Then light was considered a signal carrier that has a finite velocity of propagation. So now when clocks are synchronized using a light signal, the formula $t = t_0 + s/c$ is involved, where c is the velocity of light.

When the synchronization is achieved using a signal that propagates at a constant velocity v, we must know the value of v and whether it is indeed constant. In particular, this question was also posed for the velocity of light. Investigations were carried out to find whether the velocity of light depends on the direction of propagation, the velocity of the light source, the velocity of the receiver or other physical factors. These investigations led to the following fundamental result:

In inertial reference frames, the velocity of light is independent of the velocities of the source and receiver, and has the same value for all directions in space, namely, equal to the universal constant c.

This universal constant, viz. the velocity of light in vacuum, is taken to be $c = 2{,}997{,}924{,}58 \times 10^8$ m/s (exactly). We shall describe later and in detail how this value was arrived at. For the present, we shall confine ourselves to the result as it concerns the synchronization of clocks. According to the synchronization rule, any two clocks located at two points a distance s apart are synchronized if the difference between the readings of the clocks at the instant a light signal arrives at one point and the instant it departs from the other point is s/c. The physical meaning of this statement is not so much that two clocks can be synchronized in this way, but that such a synchronization is possible and is noncontradictory for all pairs of points in the reference frame.

In practice, synchronization is done as follows. At some point, taken as the starting point, the clock is set at 0. A light signal is emitted from this point in the form of a spherical wave. When the wave front arrives at a point a distance r from the starting point, the clock located at this point must show a time r/c.

Thus, when we state that an event occurred at a point at an instant t, we mean that the hands of the clock located at this point show a time t at the instant. Of course, it is not necessary to have a clock at each point. This is simply a brief version of the statement that if a clock synchronized according to the above process were to be located at that point, it would show a time t.

We now possess all the prerequisites for analyzing the motion in reference frames in which the positions of points and the times of occurrences of events can be described by coordinates and time whose exact meaning was explained in preceding sections.

Sec. 8. DISPLACEMENT, VELOCITY AND ACCELERATION OF A POINT

Basic physical quantities are defined for the kinematics of a point.

METHODS OF DESCRIBING MOTION. At this stage, we are neither interested in the agency responsible for the motion of a point mass nor in finding out why a particle moves in a certain way and not otherwise nor in the causes of its motion. Our task is just to describe its motion. To describe the motion of a point mass means to indicate its position at any instant of time. A moving particle continuously passes through a sequence of points in the reference frame, forming a trajectory.

Since the position of points in a reference frame can be characterized in different ways, its motion can be described accordingly.

COORDINATE FORM OF MOTION. Let us choose a coordinate system in which the position of a point is characterized by three coordinates. Let us denote these coordinates by x_1, x_2 and x_3 in the general case. As we mentioned in Sec. 5, this means that $x_1 = x$, $x_2 = y$, $x_3 = z$ (see Fig. 3) for a Cartesian coordinate system; $x_1 = \rho$, $x_2 = \varphi$, $x_3 = z$ (see Fig. 4) for a cylindrical coordinate system; $x_1 = r$, $x_2 = \varphi$, $x_3 = \theta$ (see Fig. 5) for a spherical coordinate system. When a point moves, these coordinates vary with time or, in other words, are functions of time. To describe the motion of a point means to give these functions:

$$x_1 = x_1(t),\ x_2 = x_2(t),\ x_3 = x_3(t). \tag{8.1}$$

It should be recalled that a function describes a rule by which each value of variable is assigned a numerical value of another quantity. The rule is tentatively described by a certain letter, for example, $y = f(x)$. Here, f stands for the function by which each value of the variable x is assigned a definite value of the quantity y. However, in order to avoid the introduction of too many letters, the same functional dependence is frequently written in the form $y = y(x)$. The symbol y on the right-hand side of this equality is analogous to f, while the symbol y on the left-hand side indicates the numerical value of the quantity y corresponding to this situation. This method of describing functional dependences is more economical and is widely used. The same notation was employed to write formulas (8.1).

Let us consider some examples in which motion is described by using this method. Suppose that a point begins to move at an instant $t = 0$ away from the initial position along a straight line in such a way that its distance s from the initial point along its trajectory is proportional to time: $s = At$, A being the proportionality constant. The formulas describing this motion depend on the coordinate system that is chosen and on the way the system is oriented. Let us consider the Cartesian coordinate system whose origin coincides with the initial position of the point and one of the axes, say, the y-axis, is directed along the trajectory. In this case, (8.1) becomes

$$x_1 = x = 0,\ \ x_2 = y = At,\ \ x_3 = z = 0. \tag{8.2a}$$

If, however, the coordinate axes are oriented such that the trajectory of the particle lies in the XY-plane and coincides with the bisector of the angle between positive X- and

Y-directions, (8.1) can be written as:

$$x_1 = x = \frac{At}{\sqrt{2}},$$
$$x_2 = y = \frac{At}{\sqrt{2}}, \tag{8.2b}$$
$$x_3 = z = 0.$$

For a spherical coordinate system oriented as in Fig. 5 relative to the Cartesian axes, which, in turn, are oriented relative to the trajectory under consideration in the same way as indicated by (8.2b), (8.1) assumes the following form:

$$x_1 = r = At,$$
$$x_2 = \varphi = \frac{\pi}{4}, \tag{8.2c}$$
$$x_3 = \theta = \frac{\pi}{2}.$$

If the origin of the coordinate system is not chosen to coincide with the initial position of the point, all formulas become more complicated, especially in the spherical coordinate system. We leave it to the reader to verify this.

Suppose that a point moves uniformly in a circle of radius R. We assume its position at an instant of time $t = 0$ to be the reference point. The distance s traversed by the point along its circular trajectory is proportional to the time, i.e. $s = At$, where A is the proportionality constant. We orient the Cartesian coordinate system so that the circle lies in the XY-plane, the origin of the coordinate system coincides with the centre of the circle, and the Z-axis is directed so that the motion appears to be anticlockwise to an observer watching it from the positive Z-direction. Moreover, we assume that the positive X-direction passes through the reference point from which the motion of the particle started. In this case, (8.1) assumes the following form for the motion in a circle:

$$x_1 = x = R\cos\left(\frac{At}{R}\right),$$
$$x_2 = y = R\sin\left(\frac{At}{R}\right), \tag{8.3a}$$
$$x_3 = z = 0.$$

In a spherical coordinate system, (8.1) is transformed as follows to describe the same motion:

$$x_1 = r = R,$$
$$x_2 = \varphi = \frac{At}{R}, \qquad (8.3b)$$
$$x_3 = \theta = \frac{\pi}{2}.$$

For a cylindrical coordinate system oriented as in Fig. 4 relative to the Cartesian axes, which in turn are oriented relative to the trajectory under consideration in the same way as indicated by (8.3a), (8.1) assumes the following form:

$$x_1 = \rho = R,$$
$$x_2 = \varphi = \frac{At}{R}, \qquad (8.3c)$$
$$x_3 = z = 0.$$

All these formulas become much more complicated if the origin of coordinates does not coincide with the centre of the circle or if the coordinate axes are oriented in some other manner.

VECTOR NOTATION OF MOTION. The position of a point can be described using a radius vector $\mathbf{r}$ relative to a point taken as the origin. It was mentioned in Sec. 5 that this method of describing the position of a point does not presuppose the introduction of any coordinate system, but only assumes the presence of a reference body. The radius vector $\mathbf{r}$ is considered to be a directly specified quantity. As a point moves, its radius vector continuously varies, its tip following the trajectory. The motion is described in a coordinate-free form as follows:

$$\mathbf{r} = \mathbf{r}(t). \qquad (8.4)$$

Formulas of this type describe the vector function of a scalar argument. The vector function of a scalar argument is a rule according to which each numerical value of the argument (t in the present case) has a corresponding vector ($\mathbf{r}$ in the present case). In (8.4), this rule is denoted by $\mathbf{r}$ on the right-hand side, while the vector obtained from the rule is denoted by $\mathbf{r}$ on the left-hand side. As in (8.1), no confusion is caused by such a utilization of the same symbol in two different senses.

Formulas (8.2a)-(8.2c), which have different forms, describe the same motion. In order to present this motion in the form (8.4), let us use $\boldsymbol{\tau}$ to denote a dimensionless unit vector in the direction of motion and consider that the radius vector starts from the initial point of the trajectory. The motion is then described by a formula which is independent of the coordinate system:

$$\mathbf{r} = \boldsymbol{\tau} A t. \tag{8.5}$$

It should be emphasized once again that (8.4) is not an abbreviation for three scalar equalities of the type (8.1); it is rather a starting form which, if necessary, can be split into three scalar equalities, but exists irrespective of whether such a notation is possible or not.

DESCRIPTION OF MOTION WITH THE HELP OF TRAJECTORY PARAMETERS. If a point's trajectory is given, the problem reduces to the specification of the equation of motion along the trajectory. An arbitrary point along the trajectory is then taken as the starting point, and any other point can be characterized by the distance s from the initial point along the trajectory. In this case, the motion is described by the following formula:

$$s = s(t). \tag{8.6}$$

For example, the equation of motion in a circle, described by (8.3a), has the form

$$s = At. \tag{8.7}$$

The known parameters in this case are the circle and the starting point of the motion. The positive values of s correspond to the direction of motion of the point in the circle.

DISPLACEMENT VECTOR. The displacement vector $\Delta \mathbf{r} = \mathbf{r}(t + \Delta t) - \mathbf{r}(t)$ is numerically equal to the distance between the terminal and initial points and is directed from the initial point to the terminal point (Fig. 13). This vector joins the points on the trajectory where a point mass is situated at the instants t and $t + \Delta t$.

VELOCITY. The average velocity vector $\mathbf{v}_{av}$ for motion between two points is defined as the vector coinciding in direction with the displacement and equal in magnitude to the displacement vector divided by the time taken for the displacement (see Fig. 13):

$$\mathbf{v}_{av}(t,\ t + \Delta t) = \frac{\Delta \mathbf{r}}{|\Delta \mathbf{r}|} \frac{|\Delta \mathbf{r}|}{\Delta t} = \frac{\Delta \mathbf{r}}{\Delta t}. \tag{8.8}$$

The parentheses after $\mathbf{v}_{av}$ indicate the interval of time over which the velocity is averaged. As Δt tends to zero, the average velocity tends to its limiting value which is termed the instantaneous velocity $\mathbf{v}$:

$$\mathbf{v}(t) = \lim_{\Delta t \to 0} \frac{\Delta \mathbf{r}}{\Delta t} = \frac{d\mathbf{r}}{dt} = \mathbf{r}. \tag{8.9}$$

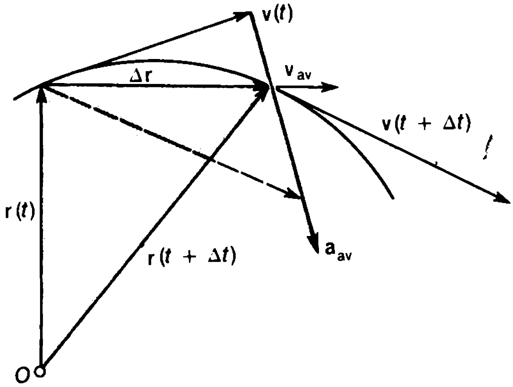


Fig. 13. Displacement, velocity and acceleration.

In a motion between two points of the trajectory, the average velocity coincides in direction with the displacement vector. Generally, it is not directed tangentially to the trajectory either at the initial or at the terminal point. Point O is the origin.

In the Cartesian coordinate system, we have

$$\mathrm{r}(t) = \mathrm{i}_x x(t) + \mathrm{i}_y y(t) + \mathrm{i}_z z(t),$$
$$\mathrm{v} = \frac{\mathrm{d}\mathrm{r}}{\mathrm{d}t} = \mathrm{i}_x \frac{\mathrm{d}x}{\mathrm{d}t} + \mathrm{i}_y \frac{\mathrm{d}y}{\mathrm{d}t} + \mathrm{i}_z \frac{\mathrm{d}z}{\mathrm{d}t}. \tag{8.10}$$

Consequently, the projections of the velocity are given by the formulas

$$v_x = \frac{\mathrm{d}x}{\mathrm{d}t}, \quad v_y = \frac{\mathrm{d}y}{\mathrm{d}t}, \quad v_z = \frac{\mathrm{d}z}{\mathrm{d}t}. \tag{8.11}$$

If the motion is described in terms of trajectory parameters, the trajectory and the time dependence of the path are specified. The path length is measured from the point on the trajectory taken as the starting point. Each point on the trajectory is characterized by its own value of s. Hence its radius vector is a function of s, and the trajectory may be defined by the equation

$$\mathrm{r} = \mathrm{r}(s). \tag{8.12}$$

Consequently, $\mathrm{r}(t)$ in (8.9) can be considered a compound function $\mathrm{r}[s(t)]$, and its derivatives can be found by differentiating a compound function:

$$\mathrm{v} = \frac{\mathrm{d}\mathrm{r}}{\mathrm{d}t} = \frac{\mathrm{d}\mathrm{r}}{\mathrm{d}s}\frac{\mathrm{d}s}{\mathrm{d}t}. \tag{8.13}$$

The quantity Δs is the distance between two points along the trajectory, while $|\Delta \mathrm{r}|$ is the distance between them along a straight line. It is obvious that as the two points approach each other, the difference between these two quantities decreases. In view of this, we can write

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = \lim_{\Delta s \to 0} \frac{\Delta \mathbf{r}}{\Delta s} = \lim_{\Delta s \to 0} \frac{\Delta \mathbf{r}}{|\Delta \mathbf{r}|} \frac{|\Delta \mathbf{r}|}{\Delta s} = \boldsymbol{\tau},$$

where $\boldsymbol{\tau}$ is a unit vector tangential to the trajectory. Moreover, by definition, $\mathrm{d}s/\mathrm{d}t = v$ is the projection of the velocity on the direction of the tangent. Hence (8.13) assumes the form

$$\mathbf{v} = \boldsymbol{\tau} v. \tag{8.14}$$

It follows hence that the velocity is directed along the tangent to the trajectory.

In our discussion of velocity, the statement that a point mass can occupy two infinitesimally close points in space at two infinitesimally close instants of time was assumed to be objective. The validity of this statement is obvious for a point mass in Newtonian mechanics and is simply a corollary of the definition of a point mass which must exist continuously in time and space in an invariant form. When macroscopic objects are modelled as point masses, it is not difficult to describe the geometrical and physical meaning of its coordinates at each instant of time. However, the situation is quite different if the motion of atomic and subatomic particles is to be simulated. At first glance, it would seem natural to model them as point masses in view of their small size. However, this is not correct. *The absolute geometrical size of an object is not important when modelling it as a point mass; what is important is whether its physical properties can be modelled.* An experimental investigation of the laws of motion of atomic and subatomic particles shows that their motion cannot be described on the basis of the statement that each particle has definite coordinates at each instant of time. Hence the concept of instantaneous velocity is meaningless for these particles, and hence the concept of a particle's motion along a trajectory loses its meaning as do the other concepts used to describe the motion of a point mass in Newtonian mechanics.

The laws of motion of atomic and subatomic particles are studied in quantum mechanics, where these particles are modelled using quantum-mechanical laws rather than the classical ones.

ACCELERATION. Acceleration is the rate of change of velocity. Suppose that the velocity at instants t and $t + \Delta t$ is respectively equal to $\mathbf{v}(t)$ and $\mathbf{v}(t + \Delta t)$. This means that the velocity has changed by $\Delta \mathbf{v} = \mathbf{v}(t + \Delta t) - \mathbf{v}(t)$ over a time interval Δt. The average acceleration $\mathbf{a}_{av}$ during this interval is (see Fig. 13)

$$\mathbf{a}_{av}(t,\ t + \Delta t) = \frac{\Delta \mathbf{v}}{\Delta t}. \tag{8.15}$$

We shall use the same initial point to represent vector $\mathbf{v}(t)$ at different instants of time. The tip of vector $\mathbf{v}(t)$ describes a curve called the hodograph of motion (Fig. 14). Indefinitely decreasing the time interval Δt over which the average velocity

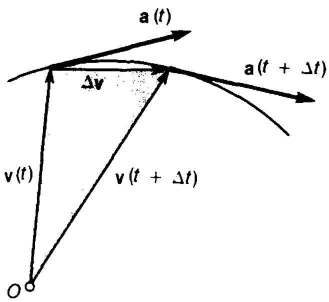


Fig. 14. Hodograph of motion.
It is a curve described by the tip of the velocity vector drawn from the fixed origin (point O).

is measured, we obtain acceleration in the limit, i.e.

$$\mathrm{a} = \lim_{\Delta t \to 0} \frac{\Delta \mathrm{v}}{\Delta t} = \frac{\mathrm{d}\mathrm{v}}{\mathrm{d}t}. \tag{8.16}$$

Since $\mathrm{v} = \mathrm{d}\mathrm{r}/\mathrm{d}t$ and $\mathrm{r} = \mathrm{i}_x x + \mathrm{i}_y y + \mathrm{i}_z z$, the acceleration can be represented in the form $\mathrm{a} = \mathrm{d}^2\mathrm{r}/\mathrm{d}t^2$, or

$$\mathrm{a} = \mathrm{i}_x \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + \mathrm{i}_y \frac{\mathrm{d}^2 y}{\mathrm{d}t^2} + \mathrm{i}_z \frac{\mathrm{d}^2 z}{\mathrm{d}t^2}. \tag{8.17}$$

Consequently, the projections of acceleration in the Cartesian coordinate system are given by

$$a_x = \frac{\mathrm{d}^2 x}{\mathrm{d}t^2}, \quad a_y = \frac{\mathrm{d}^2 y}{\mathrm{d}t^2}, \quad a_z = \frac{\mathrm{d}^2 z}{\mathrm{d}t^2}. \tag{8.18}$$

Next, we must consider the orientation of acceleration relative to velocity and the trajectory. It is obvious that acceleration is always tangential to the hodograph of motion and may be inclined at any angle to velocity. This means that acceleration may be directed at any angle to the tangent to the trajectory. In order to find the factors on which the direction of acceleration depends, we compute the acceleration on the basis of (8.14):

$$\mathrm{a} = \frac{\mathrm{d}\mathrm{v}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\tau \quad) = \frac{\mathrm{d}\tau}{\mathrm{d}t} v + \tau \frac{\mathrm{d}v}{\mathrm{d}t}. \tag{8.19}$$

The unit tangential vector τ is completely defined by a point on the trajectory, while the point on the trajectory is uniquely characterized by its distance s from the initial point. Hence vector τ is a function of s, i.e. $\tau = \tau(s)$, while s is a function of time. Hence we can write $\mathrm{d}\tau/\mathrm{d}t = (\mathrm{d}\tau/\mathrm{d}s)(\mathrm{d}s/\mathrm{d}t)$. Vector τ is invariant in magnitude. Consequently, $\mathrm{d}\tau/\mathrm{d}s$ is perpendicular to τ. This can be verified by simply differentiating the equation $\tau^2 = 1$, which indicates the invariance of the magnitude of vector τ: $[\mathrm{d}(\tau^2)/\mathrm{d}s] = 2(\tau\, \mathrm{d}\tau/\mathrm{d}s)$. But if the scalar product of two vectors is zero, and neither vector is itself zero, the vectors must be perpendicular. Thus τ and $\mathrm{d}\tau/\mathrm{d}s$ are indeed perpendicular. Vector τ is directed along the tangent to the trajectory. Hence vector $\mathrm{d}\tau/\mathrm{d}s$ is perpendicular to the tangent or, in other words, $\mathrm{d}\tau/\mathrm{d}s$ is directed along the normal which is called the principal normal. The unit vector in the direction of the principal normal is denoted by n. The *magnitude of vector* $\mathrm{d}\tau/\mathrm{d}s$ is $1/R$, where R is the radius of curvature of the trajectory. The planes in which vectors τ and n lie are called osculating planes.

!

Velocity is always directed tangentially to the trajectory. Acceleration may form any angle with velocity, i.e. may be inclined at any angle to the trajectory.
The normal component of acceleration does not change the magnitude of velocity, but only changes its direction.
A change in the magnitude of velocity is due only to the tangential component of acceleration.

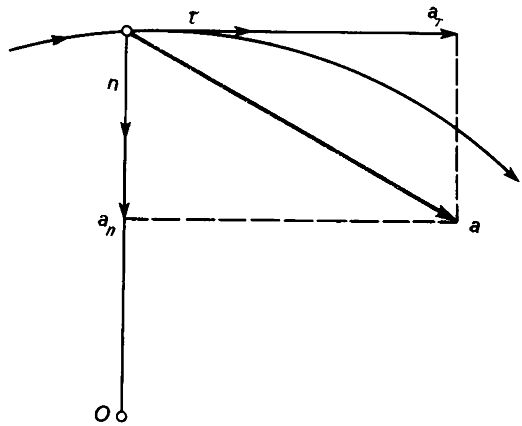


Fig. 15. Decomposition of the total acceleration vector $\mathbf{a}$ into the tangential $\mathbf{a}_\tau$ and normal $\mathbf{a}_n$ accelerations.

Point O is the centre of curvature of the trajectory, τ is the unit tangential vector, and $\mathbf{n}$ is the unit vector in the direction of the principal normal.

The point a distance R from the trajectory in the direction of the principal normal $\mathbf{n}$ is called the centre of curvature of the trajectory. Thus, we can write

$$\frac{d\tau}{ds} = \frac{\mathbf{n}}{R}. \tag{8.20}$$

Since $ds/dt = v$ is the magnitude of velocity, we can rewrite (8.19) in its final form using (8.20), i.e.

$$\mathbf{a} = \mathbf{n}\frac{v^2}{R} + \tau\frac{dv}{dt}. \tag{8.21}$$

Total acceleration is the resultant of two perpendicular vectors, viz. the acceleration $\tau(dv/dt) = \mathbf{a}_\tau$ which is directed along the trajectory and is called the tangential acceleration, and the acceleration $\mathbf{n}v^2/R = \mathbf{a}_n$ which is directed perpendicular to the trajectory along the principal normal, i.e. towards the centre of curvature of the trajectory (Fig. 15), and is called the normal acceleration. Squaring both sides of (8.21) and considering that $\mathbf{n}\cdot\tau = 0$, we obtain the magnitude of the total acceleration:

$$a = \sqrt{\mathbf{a}^2} = \sqrt{\left(\frac{v^2}{R}\right)^2 + \left(\frac{dv}{dt}\right)^2}. \tag{8.22}$$

When a point moves in a circle, the normal acceleration is called centripetal since the centres of curvature at all the points are the same, namely, the centre of the circle.

Example 8.1. A cat is chasing a mouse (Fig. 16) which is running along a straight line at a constant velocity $\mathbf{u} = \text{const}$.

?

What methods do you know to describe a motion?
What are the advantages of the vector notation and the vector form to describe a motion?
What is the instantaneous velocity and how is it oriented relative to the trajectory?
What are the directions of the normal and tangential accelerations relative to the trajectory and how is their absolute magnitude determined?
Why do we state that the angular velocity is a vector?
Is a finite angular displacement a vector?
What is the angular acceleration vector? What is its direction if the direction of the angular velocity does not change?

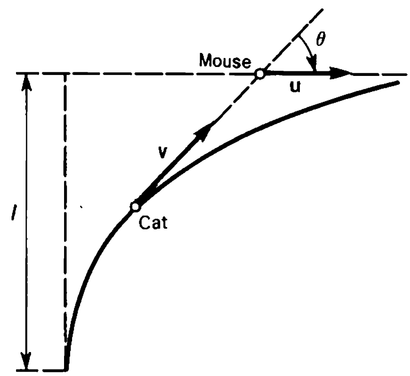


Fig. 16. Cat catches mouse.

As long as the cat catches the mouse, the velocity vector v of the cat is directed to the mouse.

The velocity of the cat has a constant value $v > |\mathbf{u}|$ and is directed towards the mouse. At the initial moment of time, the velocities of the cat and the mouse are perpendicular, and the distance between them is equal to l. Find how long it will take the cat to catch the mouse.

Let r and θ denote the distance between the cat and the mouse and the angle between their velocities (measured from the direction of the cat's velocity). Using $\mathbf{v}_{rel} = -\mathbf{v} + \mathbf{u}$ to denote the relative velocity of the two animals in the polar coordinate system centred on the cat, we obtain

$$\dot{r} = -v + u\cos\theta, \qquad r\dot{\theta} = -u\sin\theta. \tag{8.23}$$

Hence

$$\dot{r}(u\cos\theta + v) - r\dot{\theta}u\sin\theta = u^2 - v^2 \tag{8.24}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}[r(u\cos\theta + v)] = u^2 - v^2. \tag{8.25}$$

Integrating both sides of this equation between 0 and t and considering that $r(0) = l$ and $\theta(0) = \pi/2$, we obtain

$$r(u\cos\theta + v) - vl = (u^2 - v^2)t. \tag{8.26}$$

The cat catches the mouse when $r = 0$. Hence the time it takes to do so is $vl/(v^2 - u^2)$.

Example 8.2. A stone is thrown at a velocity u from the origin of coordinates at an angle α to the horizontal X-axis. Neglecting the resistance of the air, find the angle α_0 at which the stone hits the point (x_0, y_0) (Fig. 17).

The radius vector of the stone at an instant t is

$$\mathbf{r} = \mathbf{u}t + \frac{\mathbf{g}t^2}{2}, \tag{8.27}$$

where vector $\mathbf{g}$ is the acceleration due to gravity and is always

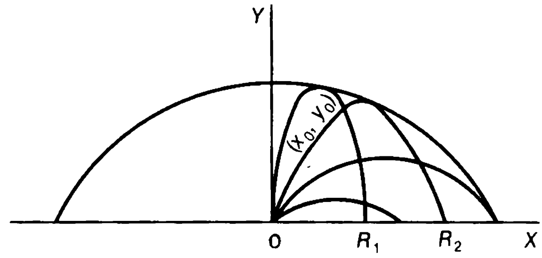


Fig. 17. Safety parabola.

directed downwards. If the Y-direction is upwards, we can rewrite (8.27) in coordinate form as

$$x = ut\cos\alpha, \qquad y = ut\sin\alpha - \frac{gt^2}{2}. \tag{8.28}$$

Eliminating t, we can write the equation for the trajectory as

$$y = x\tan\alpha - \frac{gx^2}{2u^2\cos^2\alpha}. \tag{8.29}$$

For $y = y_0$ and $x = x_0$, this equation gives

$$\tan\alpha_0 = \left\{1 \pm \left[1 - \frac{2g}{u^2}\left(y_0 + \frac{gx_0^2}{2u^2}\right)\right]^{1/2}\right\}\frac{u^2}{gx_0}. \tag{8.30}$$

Real values of α_0 corresponding to the possible trajectories can only be obtained if the radicand is positive, i.e. if

$$1 \geqslant \frac{2g}{u^2}\left(y_0 + \frac{gx_0^2}{2u^2}\right),$$

or

$$y_0 \leqslant \frac{u^2}{2g}\left(1 - \frac{g^2x_0^2}{u^4}\right). \tag{8.31}$$

Thus, a stone thrown at the initial velocity u can only reach points lying below the parabola

$$y = \frac{u^2}{2g}\left(1 - \frac{g^2x^2}{u^4}\right), \tag{8.32}$$

which is called the safety parabola. The points lying beyond the region enclosed by the parabola cannot be reached by the stone.

It is sufficient to confine ourselves to the region $x_0 \geqslant 0$, $y_0 \geqslant 0$. It follows from (8.32) that $u^2/(gx_0) \geqslant 1$. Together with (8.30), this means that at least one value of α_0 is equal to or greater than $\pi/4$. A point lying on the safety parabola can only be reached for one value of $\alpha_0 \geqslant \pi/4$, while the points lying below the parabola are attainable for two values, say, α_{01} and

α_{02} (see Fig. 17). The distances at which the corresponding trajectories intersect the horizontal plane are respectively $R_1 = u^2 \sin 2\alpha_{01}/g$ and $R_2 = u^2 \sin 2\alpha_{02}/g$.

It is obvious that the points of intersection of the parabolas are reached at different instants of time.

Sec 9. KINEMATICS OF A RIGID BODY

The basic physical quantities in the kinematics of a rigid body are defined.

DEGREES OF FREEDOM. In order to describe the motion of a point mass, we must specify the three functions characterizing the time dependence of its coordinates. In order to describe a system of N point masses moving independently, we must specify $3N$ functions characterizing the time dependences of their coordinates. *The number of independent functions (or parameters, as they are often called) describing the motion of a system of point masses is called the* degree of freedom *of the system.* A point mass has three degrees of freedom, while a system of two independent point masses has six degrees of freedom.

If, however, two point masses are rigidly connected through a rod of constant length l, the six coordinates of the two points are no longer independent since they satisfy the equation $l^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2$ in which (x_1, y_1, z_1) and (x_2, y_2, z_2) are the Cartesian coordinates of the points. From this equation, we can express one of the six coordinates in terms of the length l and the remaining five coordinates. Thus, we are left with only five independent parameters to describe the motion of two rigidly connected point masses. Hence this system has five degrees of freedom.

DEGREES OF FREEDOM OF A RIGID BODY. The position of a rigid body can be fixed in space by specifying any three of its points so long as they do not lie on the same straight line. These three points are defined by nine coordinates which are related by three equations that reflect the invariance of the distances between different points of the rigid body. Hence the position of a rigid body is characterized by six independent parameters or, in other words, *the rigid body has six degrees of freedom.* These six independent parameters can be specified in different ways depending on the circumstances.

DECOMPOSITION OF THE MOTION OF A RIGID BODY INTO COMPONENTS. It is convenient to use three independent parameters to describe the motion of any point of a rigid body. This point is taken as the origin of the rectangular Cartesian coordinate system whose axes move parallel to themselves, i.e. without rotating, as the origin moves. The position of the rigid body relative to the axes is characterized

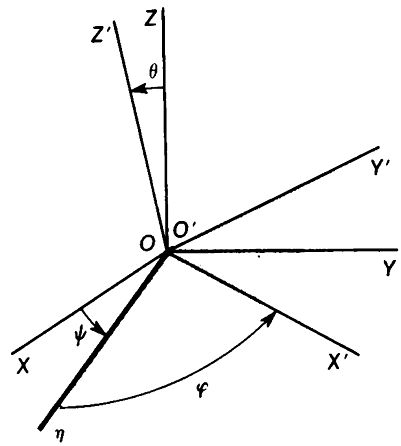


Fig. 18. The Euler angles characterize the mutual arrangement of two rectangular Cartesian coordinate systems.

The $O'X'Y'$-plane intersects the OXY-plane along the line $O\eta$.

by the remaining three independent parameters. The kinematics of a point was considered in Sec. 8. Hence in order to describe the kinematics of a rigid body, we simply have to describe the motion of the rigid body fixed at the origin of the coordinate axes. This description is carried out in terms of the Euler angles.

EULER ANGLES. We associate a coordinate system (X', Y', Z') with a rigid body, the system being characterized by the unit vectors i'_x, i'_y and i'_z. The origin of this coordinate system, as well as the origin of the coordinate system (X, Y, Z) in which the body is moving, coincide with the point at which the rigid body is fastened (Fig. 18). The position of the rigid body is completely defined by the position of the X'-, Y'- and Z'-axes relative to the X-, Y- and Z-axes.

The $O'X'Y'$- and OXY-planes intersect along the line $O\eta$, which is called the nodal line. The positive direction along this line is defined by the vector $\tau = \mathrm{i}_z \times \mathrm{i}'_z$. The Euler angles are defined as the angles

$$\begin{aligned} \varphi &= \angle \eta O X' \quad (0 \leqslant \varphi \leqslant 2\pi), \\ \psi &= \angle X O \eta \quad (0 \leqslant \psi \leqslant 2\pi), \\ \theta &= \angle Z O Z' \quad (0 \leqslant \theta \leqslant \pi). \end{aligned} \tag{9.1}$$

The angles φ, ψ and θ are respectively called the angles of intrinsic rotation, precession and nutation (see Sec. 26).

It can be seen from the definition of these angles that they are independent variables and characterize the position of a rigid body fastened at one point. Any motion of a body fastened at a point can be described by specifying three

!

The position of a system with six degrees of freedom is completely defined by means of six numbers called coordinates. These numbers are arbitrary, only their independence must be verified. One possible choice of coordinates is the Euler angles that have a number of advantages.

The angular velocity is a vector since it is determined in terms of a fundamental infinitesimal angular displacement which is a vector. The average angular velocity in the case of a rotation by a finite angle is not a vector, although it may possess both a magnitude and a direction.

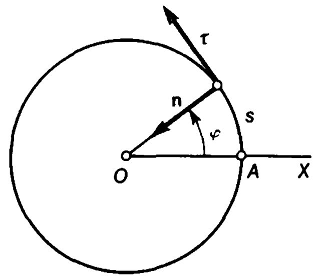


Fig. 19. Motion of a point in a circle.

The position of a point in a circle is completely characterized by the path length s traversed by it from point A taken as the reference point. The centre of the circle is the centre of curvature of the trajectory.

functions:

$$\varphi = \varphi(t), \quad \psi = \psi(t), \quad \theta = \theta(t). \tag{9.2}$$

TRANSLATIONAL MOTION. The translational motion of a rigid body occurs when all the points of the body move along similar trajectories. *This means that all the points of the body have the same velocity at each instant of time. Any straight line joining two points on the body moves parallel to itself. In the case of translational motion, the Euler angles are constant.* Hence this motion is described by specifying the motion of any point of the body. In other words, a body moving by translation has three degrees of freedom. In the kinematic sense, this motion is identical to the motion of a point mass.

PLANE MOTION. In plane motion, the trajectories of all points lie in parallel planes. *In this case, the motion of a body is described by the motion of one of its sections in any of the parallel planes, while the position of the section is defined by that of any two of its points.* The position of two points in a plane is described by four parameters (coordinates). These parameters are connected by one relation because the distance between the two points is constant. This leaves only three independent parameters, and hence the plane motion is characterized by three degrees of freedom.

ROTATIONAL MOTION. In the case of rotational motion, at least two points of the body remain fixed all the time. The straight line passing through these two points is called the axis of rotation. All the points of a rigid body lying on the axis of rotation are fixed, while all the remaining points move in circles in planes perpendicular to the axis. The centres of these circles lie on the axis of rotation.

Thus, the rotational motion of a rigid body is a plane motion.

!

Any motion of a rigid body can be represented as a combination of the motion of a point and the rotation of the body at an instantaneous velocity passing through the point.
The axis of rotation whose points have a zero translational velocity is called the instantaneous axis of rotation. The velocity of all the points of a body at a given instant of time is represented as the velocity of rotational motion about the instantaneous axis.

ANGULAR VELOCITY VECTOR. Let us consider the motion of any point of a rigid body in a circle (Fig. 19) whose radius is R. Taking point A as the reference point for measuring distance along the trajectory, we can write $s = R\varphi$. The velocity is given by $v = ds/dt = R\, d\varphi/dt$. The rate of variation of the angle φ, i.e. $d\varphi/dt = \omega$, is called the angular velocity. It has the same value for all the points of a rigid body and is called the angular velocity of rotation of the body. If this velocity is constant, it is called the cyclic frequency ω of rotation of the rigid body about an axis. The cyclic frequency is related to the period T of rotation of the rigid body about an axis thus: $\omega = 2\pi/T$. The rotation of a rigid body is described by its angular velocity. All these properties of rotation of a rigid body are combined by the concept of the angular velocity vector ω of rotation. It is equal in magnitude to the scalar quantity $d\varphi/dt$ and is directed along the axis of rotation in

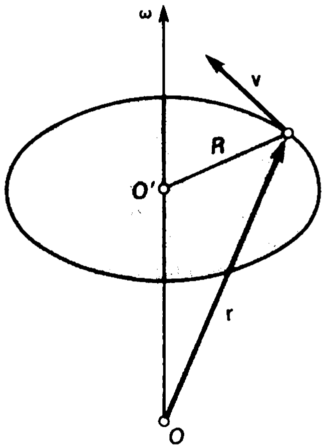


Fig. 20. Angular velocity of a rigid body.

The velocity is parallel to the axis of rotation and is related with the direction of rotation through the right-hand screw rule.

such a way that the linear velocity $\mathbf{v}$ of the points of the rigid body (Fig. 20) is expressed by the formula

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}. \tag{9.3}$$

The origin of the radius vectors $\mathbf{r}$ of the points of the rigid body is assumed to lie on the axis of rotation. We have to prove that the quantity $\boldsymbol{\omega}$ defined in this way is indeed a vector.

FUNDAMENTAL ANGULAR DISPLACEMENT VECTOR. The points of a rotating rigid body are displaced in time dt about the axis of rotation by an angle $d\varphi = \omega\, dt$ (see Fig. 19).

The fundamental angular displacement $d\boldsymbol{\varphi}$ *is characterized by both its magnitude and the plane in which it takes place.* In order to fix the plane, we must treat $d\boldsymbol{\varphi}$ as a vector perpendicular to the plane. The direction of the vector is determined from the right-hand screw rule, viz. if the screw is turned in the φ-direction, the longitudinal motion of the screw will coincide with the vector $d\boldsymbol{\varphi}$. However, in order to verify that the quantity $d\boldsymbol{\varphi}$ defined in this way is indeed a vector, we must prove that it possesses vector properties.

Let $d\boldsymbol{\varphi}_1$ and $d\boldsymbol{\varphi}_2$ be two angular displacements (Fig. 21). We shall prove that *these two quantities are added like vectors.* If we describe a sphere of unit radius about point O, these angles will have two infinitesimal arcs $d\mathbf{l}_1$ and $d\mathbf{l}_2$ corresponding to them on the surface of the sphere. The infinitesimal arc $d\mathbf{l}_3$ constitutes the third side of a triangle. We can treat this infinitesimal triangle on the surface of the sphere as a plane. The vectors $d\boldsymbol{\varphi}_1$, $d\boldsymbol{\varphi}_2$ and $d\boldsymbol{\varphi}_3$ are perpendicular to the sides of the triangle and lie in its plane. Obviously, the following vector equality is satisfied in this case:

$$d\boldsymbol{\varphi}_3 = d\boldsymbol{\varphi}_1 + d\boldsymbol{\varphi}_2 \tag{9.4}$$

Q. E. D. These vectors can be decomposed into components along the coordinate axes. In view of (9.4), these components behave like vector components, and hence *the fundamental angular displacement is indeed a vector.*

It should be noted that only fundamental infinitesimal angular displacements possess the properties of vectors. *Displacements by finite angles are not vectors* since if the displacements are represented by segments of straight lines directed perpendicular to the plane in which displacement occurs, the segments cannot be added using the parallelogram law (9.4).

An infinitesimal angular displacement $d\varphi$ of a point mass takes place in an infinitesimal time dt. Hence the *angular velocity*

$$\boldsymbol{\omega} = \frac{d\boldsymbol{\varphi}}{dt} \tag{9.5}$$

?

What determines the number of degrees of freedom of a mechanical system?
How many degrees of freedom has a rigid body in different cases of motion?
Give the geometrical definition of the Euler angles.
Prove that it is possible to present the velocity of plane motion of a rigid body as the sum of its rotational and translational velocities.

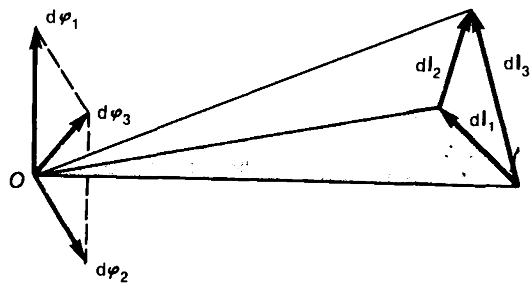


Fig. 21. Proving the vector nature of fundamental angular displacements.

is a vector in view of the fact that $d\varphi$ is a vector, while dt is a scalar. The directions of ω and $d\varphi$ coincide and are determined by the right-hand screw rule.

Consequently, the linear velocity of the points of a rotating rigid body is indeed expressed by (9.3).

ANGULAR ACCELERATION. The time derivative $d\omega/dt$ of the angular velocity is called the angular acceleration. It is used to describe the acceleration of the points of a rigid body. The magnitudes of the linear velocity, normal and tangential accelerations of the points of a rigid body are respectively equal to $v = R d\varphi/dt = R\omega$, $a_n = v^2/R = \omega^2 R$ and $a_\tau = dv/dt = R d\omega/dt$. Hence the total acceleration of the points is expressed by the formula

$$a = \sqrt{a_n^2 + a_\tau^2} = R\sqrt{\omega^4 + \dot{\omega}^2},$$

where $\dot{\omega} = d\omega/dt$ is the angular acceleration.

It can be seen from these formulas that the vectors of the total acceleration of the points of a rigid body lying on the same radius that is perpendicular to the axis of rotation are parallel and increase in proportion to the distance from the axis of rotation (Fig. 22). As can be seen from Fig. 22, the angle α characterizing the direction of acceleration relative to the radius is defined by the relation $\tan\alpha = a_\tau/a_n = \dot{\omega}/\omega^2$ and is thus independent of R.

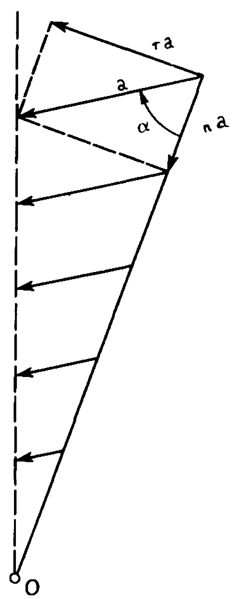


Fig. 22. As a body moves away from the axis of rotation, the total acceleration remains constant in direction, but increases in magnitude.

The axis of rotation (point O) is perpendicular to the plane of the figure.

INSTANTANEOUS AXIS OF ROTATION. In plane motion, the position of a rigid body is completely defined by the position of a segment of a line rigidly fastened to the points on one of the sections. Let us study the displacement of the segment over a certain interval of time from position A_0B_0 to position AB (Fig. 23). This displacement can be decomposed into two parts: (1) the translation from A_0B_0 to $A'B'$, during which the straight line moves parallel to itself, and (2) the rotation of the body through an angle α about an axis passing through point O' and perpendicular to the plane in which the body moves. The decomposition of the displacement is not unique: for example, we could translate the straight line from

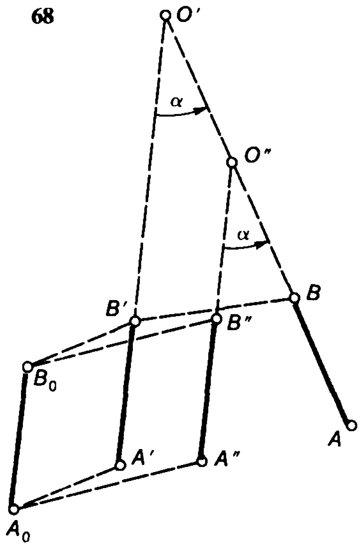


Fig. 23. Decomposition of displacement into translation and rotation.

The decomposition is ambiguous and can be made by an infinite number of ways, but the angle of rotation remains the same in all cases.

A_0B_0 to $A''B''$ and then rotate the body through an angle α about an axis passing through O''.

Thus, the decomposition of a displacement into translation and rotation is ambiguous, but the angle of rotation α is always the same. Over a time interval dt, all the points of the body are translated by $d\mathbf{l}$ and simultaneously rotated by $d\alpha$ about O' through an angle α. Hence the velocity of all the points of the body is composed of two parts: (1) the translational part $\mathbf{v}_0 = d\mathbf{l}/dt$, and (2) the rotational part $\mathbf{v}' = \boldsymbol{\omega} \times \mathbf{r}$, where $\boldsymbol{\omega} = d\boldsymbol{\alpha}/dt$, and the origin of the radius vector $\mathbf{r}$ is point O' through which the axis of rotation of the body passes. Since this point is one of the points of the rigid body, it has a translational velocity $\mathbf{v}_0$. Consequently,

$$\mathbf{v} = \mathbf{v}_0 + \boldsymbol{\omega} \times \mathbf{r}. \tag{9.6}$$

Since the decomposition of a displacement into translation and rotation is not unique, the decomposition of a velocity into translational and rotational velocities is also not unique. This is illustrated in Fig. 24 in the form of a symbolic equality: the motion on the left-hand side is composed of a translation at a velocity $\mathbf{u}$ and a rotation about the O-axis, while the motion on the right-hand side is composed of a translation at a velocity $\mathbf{u}'$ less than $\mathbf{u}$ and a rotation about the O'-axis.

Varying the translational velocity of a body, we are at the same time varying the position of the axis of its rotation. It can be stated that any axis perpendicular to the plane of motion is an axis of rotation.

Moreover, the translational velocity of a body will depend on the axis chosen as the axis of rotation. The axis of rotation for which the translational velocity is zero is called the instantaneous axis of rotation.

The velocity of all the points of a rigid body at any instant of time can be represented as the velocity of rotational motion about the instantaneous axis.

The velocities of all the points of a rigid body that lie on the instantaneous axis are zero. If the body is finite in size, the instantaneous axis may lie outside the body, but its definition and properties remain the same.

Figure 25 shows the construction for finding an axis such that the plane motion of a rigid body around it can be represented as a pure rotation. Point O through which this axis passes is the point of intersection of perpendiculars to AA' and BB'. This is clear from the fact that the triangle ABO and $A'B'O$ are congruent since a side OB is equal to OB' and OA is equal to OA' because they are drawn from a point on the right bisector to the ends of the segment, while a side AB is equal to

?

What is the instantaneous axis of rotation? Give examples of the instantaneous axis of rotation in the simplest cases of motion.
Describe a method for finding the instantaneous axis of rotation.
Give the proof of Euler's theorem.
What are the velocities constituting the velocity of the points of a rigid body in the case of an arbitrary motion?
If a body has a translational motion, where is the instantaneous axis of rotation situated?

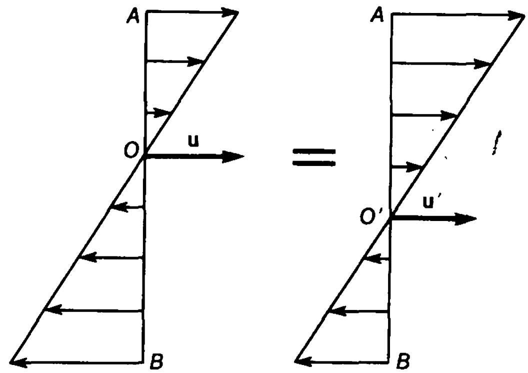


Fig. 24. Decomposition of the velocity of points of a rigid body into translational and rotational velocities.

The decomposition is ambiguous. In the two cases connected through the equality sign, the total velocity of any point along AB, which is equal to the sum of the translational and rotational velocities, is the same.

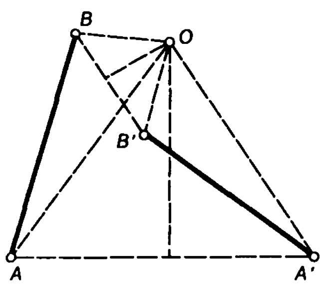


Fig. 25. Construction for finding an axis such that the displacement of a rigid body in a plane can be represented as a pure rotation about this axis.

$A'B'$ because they represent different positions of the same segment. For an infinitesimal displacement, this construction gives point O through which the instantaneous axis of rotation passes. *With the passage of time, the position of the instantaneous axis varies relative to the body and the system of coordinates in which the motion of the body is considered.*

Let us illustrate this by considering the example of a wheel rolling along a straight line. Its instantaneous axis of rotation is a straight line parallel to the axle of the wheel and passing through the point where the wheel touches the ground (Fig. 26). This axis keeps on changing its position relative to the ground and is displaced in time t to a point whose distance from the starting point is vt, v being the velocity of the wheel's axle. At different instants of time, the instantaneous axis passes through different points of the wheel along the rim. The instantaneous axis is an imaginary axis which does not have a material analogue, and *hence there is no physical meaning to the term "the velocity of the instantaneous axis".*

A physical meaning lies instead in the fact that the points of the wheel on the instantaneous axis are at rest at each particular instant of time, and the motion of the wheel boils down to a rotation about the axis.

All this refers to the plane motion of a body. Let us now consider a body which is fastened at one point, and find out if the instantaneous motion of the body can be described as the rotation about an axis passing through the point at which the body is fastened. This question is answered by Euler's theorem which states:

A rigid body having only one fixed point can be transferred from one position to another by a single rotation through a certain angle about a fixed axis passing through the fixed point.

Euler's theorem is valid for both infinitesimal and finite

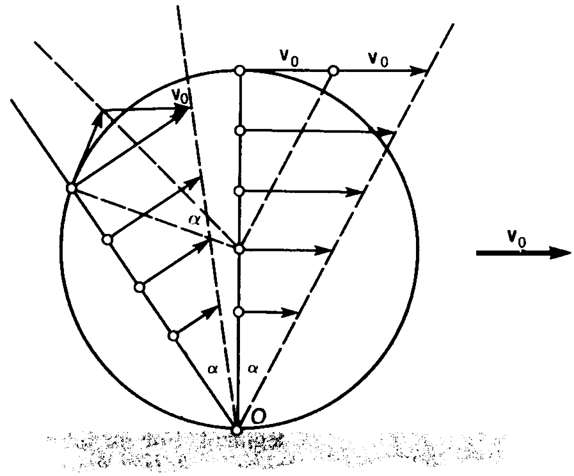


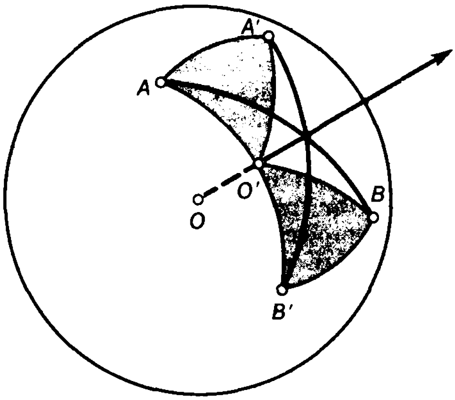
Fig. 26. The instantaneous axis of rotation of a wheel rolling over the ground is the straight line parallel to the axle of the wheel and passing through the point of contact between the wheel and the ground.

displacements. In order to prove this, we circumscribe a sphere of unit radius about the fixed point in a rigid body and draw an arc AB on the sphere. The position of the arc specifies the position of the body. As the body moves, the position of the arc changes in space so that it always remains on the surface of the sphere of unit radius. Euler's theorem then reduces to a statement that the arc AB can be transferred to any other position by rotating the sphere about an axis passing through its centre. Let us consider two positions AB and $A'B'$ of the arc on the sphere (Fig. 27). We join A to A' and B to B' along arcs of the sphere's great circles. We then draw great-circle arcs perpendicular to the previous arcs through their centres until they intersect at O'. It can be seen from the construction that the spherical triangle $AO'B$ is equal to the spherical triangle $A'O'B'$. Hence they can be made to coincide by rotating through an axis passing through O' and the centre of the sphere. This proves Euler's theorem.

It follows directly from Euler's theorem that *the motion of a rigid body fastened at a point is considered at each instant of time as a rotation about an instantaneous axis passing through the point at which the body is fastened.* The position of the instantaneous axis varies with time relative both to the body and to the stationary coordinate system in which the body is fastened at one point. The velocity of a point of the body can be represented in the form

$$v = \omega_i \times r, \tag{9.7}$$

where ω_i is the instantaneous angular velocity, and r is the radius vector relative to the fixed point. Since the angular velocity ω_i is a vector, we can represent it as the sum of two



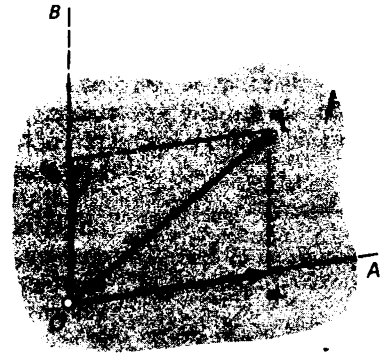


Fig. 27. To the proof of Euler's theorem.

The arc AB of the sphere can be made coincident with the arc $A'B'$ by a single rotation about an axis passing through the centre of the sphere and point O'.

Fig. 28. Decomposition of the instantaneous angular velocity vector ω_i of rotation of a rigid body into the components ω_0 and ω'.

The direction of the angular velocity ω_0 is constant relative to the stationary coordinate system, while the direction of the angular velocity ω' is constant relative to the body, but varies relative to the stationary coordinate system.

vectors (Fig. 28), one directed along a line OA' which is fixed relative to the body, and the other directed along a line OB which is stationary in the coordinate system in which the motion of the body is being considered, i.e.

$$\omega_i = \omega_0 + \omega', \quad \mathrm{v} = \omega_0 \times \mathrm{r} + \omega' \times \mathrm{r}. \tag{9.8}$$

Having represented the velocity of a body in this form, we can state that its motion is composed of two parts: a rotation at an angular velocity ω' about an axis whose position does not change relative to the body, and a rotation at an angular velocity ω_0 relative to an axis whose direction does not change in space. During the course of motion, the angular velocity ω_0 changes only in magnitude, but its direction remains unaltered, while the angular velocity ω' changes both in magnitude and direction.

Example 9.1. A rigorous mathematical description of a kinematic problem does not indicate its physical reality. It may so happen that a mathematically rigorous solution of a kinematic problem is devoid of any physical meaning. Let us clarify this point.

A ladder of length l stands vertically by a vertical wall (Y-axis). At $t = 0$, its lower end begins to slide along the floor (X-axis) at a velocity u. What is the trajectory described by the midpoint of the ladder, and what is the velocity at which it moves?

The solution seems to be obvious. Denoting the coordinates of the ends of the ladder by x and y, we have $x^2 + y^2 = l^2$. Since the coordinates of the midpoint of the ladder are $x_1 = x/2$ and $y_1 = y/2$, we get $x_1^2 + y_1^2 = l^2/4$, i.e. the midpoint of the ladder describes an arc of a circle, and the velocity of the centre is $v = [(\mathrm{d}x_1/\mathrm{d}t)^2 + (\mathrm{d}y_1/\mathrm{d}t)^2]^{1/2} = ul/[2\sqrt{l^2 - u^2t^2}]$. When the ladder is lying on the floor, $ut = l$ and $v = \infty$, which is physically meaningless, although the problem has been solved correctly in the kinematic sense. The reason behind the error will be discussed in Example 34.3.

Example 9.2. A wheel of radius a rolls along a horizontal plane without slipping. The centre of the wheel moves at a velocity u. Find the magnitude of the velocity of a point on the rim of the wheel if the line joining it to the centre of the wheel forms an angle α with the vertical.

Let us first solve the problem without using the concept of an instantaneous centre of rotation. Since the wheel rolls without slipping at a velocity u, the time taken by the wheel to complete one revolution is $T = 2\pi a/u$, and hence the angular velocity of rotation of the wheel is given by the formula $\omega = 2\pi/T = u/a$. The motion of any point of the wheel, which is a rigid body, is the result of the translational motion of its centre at the velocity u and its rotational motion about an axis passing through the centre of the wheel and perpendicular to its plane. The magnitude of the velocity of rotational motion is $\omega a = u$, the velocity being directed along the tangent to the rim. This means that the velocity of a point on the rim of the wheel is composed of the velocity u directed horizontally and the velocity tangential to the rim, also equal to u in magnitude. The angle between these velocities is α since the velocities are perpendicular to the vertical and the straight line joining the point on the rim to the centre of the wheel (angles between mutually perpendicular sides). Hence the magnitude of the velocity of the point on the rim is equal to the length of the diagonal of the parallelogram whose sides are equal to u and form an angle α. Consequently, $v = 2u\cos(\alpha/2)$.

The same problem can be solved by using an instantaneous centre of rotation. For a wheel rolling without slipping, the instantaneous fixed point is the point of contact between the rim and the surface on which the wheel rolls. Hence this point is the instantaneous centre of rotation. The motion is plane, and therefore we can use the concept of the centre of rotation instead of the axis of rotation, and assume that the axis of rotation is perpendicular to the plane of motion. It can immediately be seen that the angle between the vertical and the straight line joining the instantaneous centre of rotation to the point on the rim is $\alpha/2$. From the triangle formed by the instantaneous centre of rotation, the point on the rim and the upper tip of the vertical diameter of the wheel, we can determine the distance between the instantaneous point of rotation and the point on the rim. This distance is $R = 2a\cos(\alpha/2)$ and is equal to the radius of rotation of the wheel about the instantaneous centre of rotation at an angular velocity $\omega = u/a$. Consequently, the required magnitude of velocity is $v = \omega R = 2u\cos(\alpha/2)$, which is the same as the result obtained above.

For points that do not lie on the rim of the wheel, the calculations are carried out in exactly the same way, although the formulas become more cumbersome since the radius of rotation about the instantaneous centre is not the base of an isosceles triangle in this case. The sides of the triangle can be obtained by using methods of geometry.

PROBLEMS

2.1. The radius vectors of three consecutive vertices of a regular hexagon are $\mathbf{r}_1 = 0$, $\mathbf{r}_2$ and $\mathbf{r}_3$ respectively. Find the radius vectors of the remaining vertices.

2.2. Find the angles between two nonzero vectors $\mathbf{A}$ and $\mathbf{B}$ which satisfy the conditions:
(a) $3\mathbf{A} + 5\mathbf{B} = 0$; (b) $|\mathbf{A}| = |\mathbf{B}| = |\mathbf{A} + \mathbf{B}|$; (c) $\mathbf{A}\cdot(\mathbf{A} \times \mathbf{B}) = 0$, $|\mathbf{B}| = 2|\mathbf{A}|$; (d) $(\mathbf{A} \times \mathbf{B}) \times \mathbf{A} = \mathbf{B} \times (\mathbf{B} \times \mathbf{A})$.

2.3. Points A, B and C lie on the surface of a sphere of unit radius with its centre at O. The angles BOC, COA and AOB are respectively denoted by α, β and γ. Find the angle between the planes OAB and OAC.

2.4. Let $\mathbf{r}(t)$ be the radius vector of a moving point, $\mathbf{r}_0 = \mathbf{r}/r$. Prove that $\dot{\mathbf{r}}_0 = (\mathbf{r} \times \dot{\mathbf{r}}) \times \mathbf{r}/r^3$.

2.5. A stone must fly over two walls of height h_1 and h_2 ($h_2 > h_1$) from the side of the lower wall. The distance between the upper points of the two walls near which the stone's trajectory lies is l. Find the minimum initial velocity of the stone.

2.6. A rod of length l slides so that its ends move along lines which are at right angles to each other. What is the shape of the curve described by a point on the rod at a distance αl ($\alpha < 1$) from one of its ends?

2.7. A stone is thrown from a point on the ground at a distance l from an obstacle of height h at the lowest admissible velocity. At what distance from the obstacle will the stone fall to the ground on the other side?

2.8. A cannonball is fired at a velocity u at a target lying in the same horizontal plane as the cannon. Errors committed when the cannonball is fired amount to ε radians in the inclination of the barrel, and 2ε radians in the direction at the target in the horizontal plane. At what distance from the target will the ball land?

2.9. Lumps of mud fly from the car's wheel of radius a, the car moving at a velocity u ($u^2 \geqslant ga$). Find the maximum height attained by the lumps.

2.10. A stone is thrown at a velocity u and at an angle α to the Earth's surface. Taking the point from which the stone is cast as the origin of a polar coordinate system (r, θ), where r is the distance between the origin O and the stone, and θ is the angle between the radius vector of the stone and the Earth's surface, find the equation for the trajectory of the stone.

2.11. A rod of length l slides with its ends moving along lines which are at right angles to each other. At a certain instant of time, the velocity of one end on the rod is v, while the angle between the rod and the line on which the other end is moving is θ. What is the distance between the first end of the rod and the point having the lowest velocity? What is the value of this velocity?

2.12. A cylinder of radius $(r_2 - r_1)/2$ is placed in the space between two circular coaxial cylinders of radii r_1 and r_2 $(r_2 > r_1)$. The inner and outer cylinders rotate about their axis at angular velocities ω_1 and ω_2 respectively. Assuming that the surface of contact between the middle cylinder and the coaxial cylinders does not slip, find the angular velocity of the middle cylinder about its axis and the angular velocity of the points on its axis about the axis of the coaxial cylinders.

2.13. A rigid body which is fixed at one point to the origin of a Cartesian coordinate system is moved to a new position by successive rotations around the X-axis by $\pi/2$ and around the Y-axis by $\pi/2$. How should the axis of rotation be oriented and through what angle must the body be rotated about the axis so that the same position is attained by a single rotation?

2.14. A rigid body is turned through an angle φ about an axis passing through the origin of coordinates. The unit vector $\mathbf{n}$ is directed along the axis of rotation. The directions of rotation and of vector $\mathbf{n}$ are related by the right-hand screw rule. Find the radius vector of a point of the body after a rotation if the radius vector of the point before the rotation was $\mathbf{r}$.

2.15. A particle moves in a plane with a constant radial acceleration a directed away from the centre, and a normal acceleration $2v\omega$, where v is the particle's velocity and ω is a positive constant. Taking the direction of acceleration at the origin of coordinates as the polar axis, derive the equation for the trajectory of the particle.

2.16. Two wheels of radius r_0 each are mounted on an axle of length l. The wheels on the axle can rotate independently. They roll along a horizontal surface without slipping, and their centres have velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively. Find the magnitudes of the angular velocities of the wheels.

2.17. A ball of radius $(r_2 - r_1)/2$ is placed between two concentric spheres of radii r_1 and $r_2 (r_2 > r_1)$. The concentric spheres rotate at angular velocities ω_1 and ω_2 respectively, while the ball rolls between the spheres without slipping. What is the trajectory described by the centre of the ball and what is its angular velocity?

2.18. An object on the Earth is observed from an aeroplane flying horizontally along a straight line at a constant velocity v. Two observations of the object are made from the aeroplane at a time interval t. The object is known to be situated in a vertical plane passing through the trajectory of flight of the aeroplane, and is stationary relative to the Earth. The angles between the vertical and the direction of the telescope are α_1 and α_2 respectively for the two observations. Find the height at which the aeroplane is flying.

2.19. A linear segment AB moves in a plane. At a certain instant of time, the directions of the velocities of the ends of the segment form angles α and β respectively with the straight line AB. The magnitude of the velocity of point A is v. Find the magnitude of the velocity of point B.

2.20. A ring of radius R_2 moves at an angular velocity ω and without slipping around a stationary ring of radius R_1. The centres of the rings lie on opposite sides of the point of their contact. Find the velocity of displacement of the point of contact around the stationary ring.

2.21. Solve Problem 2.20 for the case when the centres of the rings lie on the same side of the point of their contact and $R_2 < R_1$.

2.22. A cylinder of radius R rotates without slipping between two parallel boards moving in the same direction perpendicular to the generators of the cylinder at velocities v_1 and $v_2\,(v_2 > v_1)$ along their long sides. Find the angular velocity of the cylinder and the velocity of its axis.

ANSWERS

2.1. $\mathbf{r} = 2\mathbf{r}_3 - 2\mathbf{r}_2$, $\mathbf{r}_5 = 2\mathbf{r}_3 - 3\mathbf{r}_2$, $\mathbf{r}_6 = \mathbf{r}_3 - 2\mathbf{r}_2$. **2.2.** (a) π; (b) $2\pi/3$; (c) $2\pi/3$; (d) 0 or π. **2.3.** $\arccos[(\cos\alpha - \cos\beta\cos\gamma)/(\sin\beta\sin\gamma)]$. **2.5.** $[g(h_1 + h_2 + l)]^{1/2}$. **2.6.** The arc of an ellipse with semiaxes αl and $(1 - \alpha)l$. **2.7.** $lh/\sqrt{l^2 + h^2}$. **2.8.** $2u^2\varepsilon/g$. **2.9.** $a + a^2g/(2u^2) + u^2/(2g)$. **2.10.** $r = 2u^2\cos\alpha\cdot\sin(\alpha - \theta)/(g\cos^2\theta)$. **2.11.** $l\cos^2\theta$, $v\sin\theta$. **2.12.** $(\omega_2 r_2 - \omega_1 r_1)/(r_2 - r_1)$, $(\omega_1 r_1 + \omega_2 r_2)/(r_1 + r_2)$. **2.13.** $\mathbf{i}_x + \mathbf{i} - \mathbf{i}_z$, $2\pi/3$. **2.14.** $\mathbf{r}\cos\varphi + \mathbf{n}\times\mathbf{r}\cos\varphi + \mathbf{n}(\mathbf{n}\cdot\mathbf{r})(1 - \cos\varphi)$. **2.15.** $r = a(1 - \cos\varphi)/\omega^2$. **2.16.** $[(v_1 - v_2)^2/l^2 + v_1^2/r_0^2]^{1/2}$, $[(v_1 - v_2)^2/l^2 + v_2^2/r_0^2]^{1/2}$. **2.17.** Circle $(r_1\omega_1 + r_2\omega_2)/(r_1 + r_2)$. **2.18.** $vl/(\tan\alpha_2 - \tan\alpha_1)$. **2.19.** $v\cos\alpha/\cos\beta$. **2.20.** $\omega R_1 R_2/(R_1 + R_2)$. **2.21.** $\omega R_1 R_2/(R_1 - R_2)$. **2.22.** $(v_2 - v_1)/(2R)$, $(v_1 + v_2)/2$.

Chapter 3

Coordinate Transformations

Basic idea:
The transformation of coordinates belonging to the same inertial reference frame is a purely mathematical question, while the transformation of coordinates belonging to different inertial reference frames is a problem falling in the purview of physics. This problem can be solved only by means of experiments.

Sec. 10. RELATIVITY PRINCIPLE

The difference between a geometrical coordinate transformation and a coordinate transformation due to the motion of the reference frames is analyzed.

GEOMETRICAL COORDINATE TRANSFORMATIONS. The coordinate transformations considered in Secs. 5 and 6 involve points in the same reference frame. They are derived from the definition of coordinate systems as a result of geometrical constructions. These transformations do not include time. They are purely geometrical coordinate transformations, and their physical significance does not extend beyond a change in the values of physical quantities.

PHYSICAL COORDINATE TRANSFORMATIONS. Different bodies for which different reference frames exist may be in motion relative to one another. Each reference frame has its own coordinate system, and time is measured at different points of the system using clocks which are at rest at the points and are synchronized as described in Sec. 7. Let us find the connection between the coordinates and time in various reference frames if the frames are in relative motion. This problem cannot be solved merely from geometrical considerations. It is essentially a physical problem which is transformed into a geometrical problem only when the relative velocity of various reference frames is zero so that there is no physical difference between the reference frames, and they can be treated as a single reference frame.

INERTIAL REFERENCE FRAMES AND THE RELATIVITY PRINCIPLE. The simplest motion of a rigid body is a uniform translation in a straight line. Correspondingly, the simplest relative motion of a reference frame is its uniform translation

Galileo Galilei (1564–1642)
Italian scientist, the founder of modern mechanics. He formulated the relativity principle and established the laws of inertia, of free fall, of composition of motions, and so on. He constructed the first telescope and made a number of astronomical discoveries. He was an active supporter of the heliocentric system and had to face an inquisition court in 1633 for his convictions.

in a straight line. One reference frame is conditionally called stationary, while the other is called moving. We introduce a Cartesian coordinate system in each reference frame. We denote the coordinates in the stationary K system by (x, y, z), and in the moving K' system by (x', y', z'). We assume that each quantity in the moving coordinate system is denoted by the same letter as in the stationary coordinate system but primed. The axes of the coordinate systems are directed as indicated in Fig. 29. Instead of stating "the reference body to which the K' coordinate system is attached moves at a velocity v", we shall use the simplified form "the K' coordinate system moves at a velocity v relative to the K coordinate system". This does not cause confusion since each coordinate system has a meaning only when the reference body to which it is attached is indicated. We shall use similar expressions when speaking of time measurements in various coordinate systems or of synchronization of clocks, etc. since all the operations are carried out in appropriate reference systems.

The first question of prime importance that arises in this connection can be stated as follows. The measurement of coordinates and time was considered in Secs. 5 and 7 by assuming the validity of Euclidean geometry, a unified time and the possibility of synchronization of clocks described therein. It was mentioned that the existence of such frames had been confirmed by the experiment. We must now indicate how to find such reference frames. This can only be done by studying the course of various physical processes in reference frames moving relative to one another.

It follows from the results of a multitude of experiments that in all reference frames moving uniformly in a straight line relative to fixed stars, and hence relative to one another, all mechanical phenomena occur identically.

It is assumed that gravitational fields are negligibly small. Such reference frames are called inertial reference frames because Newton's law of inertia is obeyed in them, namely, that a body unattached by other bodies moves uniformly in a straight line relative to the reference frame.

!

The relativity principle states that physical laws are identical for all inertial reference frames. The inertial nature of the reference frames and the validity of the relativity principle in them are due to the properties of space and time. There exists an infinite number of inertial reference frames. All these frames have a uniform motion in a staight line relative to one another.

The hypothesis (first put forth by Galileo) that mechanical phenomena occur identically in all inertial frames is called Galileo's relativity principle. Later, as a result of investigations of other phenomena, including electromagnetic phenomena, the validity of this hypothesis was extended to all phenomena. In such a general form, it is called the relativity principle in the special theory of relativity, or simply the relativity principle.

At present, this principle has experimentally been verified to a very high degree of accuracy for mechanical and electromagnetic phenomena. In spite of this, the relativity principle remains a postulate, i.e. a basic assumption which is beyond the scope of experimental verification. This is due to two reasons.

Firstly, within the range of investigated physical phenomena, a statement can experimentally be verified only to the accuracy permitted by contemporary measurement techniques. A statement, however, is absolute in nature, i.e. whatever the accuracy of the experiment, the results will always be in agreement with the statement. Clearly, this cannot experimentally be verified, no matter how developed a science, experiments can only be carried out with a finite accuracy. Secondly, there are some physical phenomena that have not yet been discovered. The statement that all phenomena which will be discovered in future will obey the relativity principle is beyond the realms of experimental verification. Hence the relativity principle is a postulate and will always remain so. But this does not lower its significance in any way. All scientific concepts, laws and theories are worked out for a certain class of physical phenomena and are only valid within certain limits. Transgressing the limits of their applicability does not render these concepts, laws, or theories invalid. It simply indicates the limits and conditions within which they can be applied, as well as their accuracy. Progress in science simply means an extension of the range of applicability of the existing theories.

Sec. 11. GALILEAN TRANSFORMATIONS

The role of the invariants of transformations in physics is discussed and the invariants of Galilean transformations are analyzed.

GALILEAN TRANSFORMATIONS. A moving coordinate system (see Fig. 29) occupies a single position relative to the stationary coordinate system at all instants of time. If the origins of both coordinate systems coincide at $t = 0$, the origin of the moving system will be a distance $x = vt$ from the stationary system at the instant t. Galilean transformations assume that at each instant of time the same relation exists between the coordinates and time of the systems (x, y, z) and (x', y', z'), which would exist between them if the systems were at rest relative to each other. In other words, coordinate transformations can be reduced to geometrical transformations that were considered earlier, while the time remains the same in the two systems:

$$x' = x - vt, \quad y' = y, \quad z' = z, \quad t' = t. \tag{11.1}$$

These formulas are called Galilean transformations.

It is obvious that we could have chosen the K' system as the

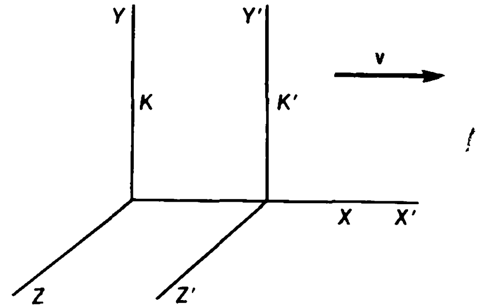


Fig. 29. Relative motion of the K' and K coordinate systems.

By spatial rotation of the coordinate systems and by displacing the origin, one can always attain the situation that the X- and X'-axes of these systems coincide, and the motion occurs along the X-axis. With such a mutual arrangement of the systems, the coordinate transformations have the simplest form.

stationary system. In this coordinate system, K system moves at a velocity v in the negative x'-direction, i.e. at a negative velocity. Hence the transformation formulas would be obtained in this case from (11.1) by replacing primed quantities by unprimed ones, and v by $-v$. These relations have the following form:

$$x = x' + vt', \quad y = y', \quad z = z', \quad t = t'. \tag{11.2}$$

It is worth noting that these formulas were derived from (11.1) by applying the relativity principle to the transformations (11.1) and not by computation, i.e. not by solving (11.1) for the unprimed quantities. Of course, the same formulas (11.2) can be obtained from (11.1) simply by solving them as a system of equations in unprimed quantities. The fact that both results are the same means that (11.1) and (11.2) do not contradict the relativity principle.

!

Invariants of transformations reflect essential features of the objects under investigation, the features which do not depend on the choice of a coordinate system and so actually reflect the properties of the objects.

INVARIANTS OF TRANSFORMATIONS. The numerical values of various geometrical and physical quantities generally change as a result of a coordinate transformation. For example, the position of a point is characterized by three numbers (x_1, y_1, z_1). These numbers change when a different coordinate system is used. Obviously, the numbers depend not on any objective property of the point, but simply on the geometrical position of the point relative to a particular coordinate system.

?

What is the difference between physical and purely geometrical coordinate transformations?
If there exist various reference frames, under what conditions will the transformation of the coordinates fixed to a reference frame become a geometrical transformation?

If the value of a quantity does not change as a result of a coordinate transformation, then the quantity has an objective value which is independent of the choice of the coordinate system. Such quantities reflect the properties of the very phenomena and objects under investigation, and not the relations of these phenomena or objects to the coordinate system in which they are being considered. *Quantities whose numerical values are not changed by a coordinate transformation are called* invariants of transformations. They are of prime importance to physics. Hence we must study the invariants of Galilean transformations.

INVARIANCE OF LENGTH. Let us consider a rod whose ends have coordinates (x'_1, y'_1, z'_1) and (x'_2, y'_2, z'_2) in a K' coordinate system. This means that the length of the rod in this system is

$$l = \sqrt{(x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2}.$$

The rod has a translational motion in a K coordinate system, and all its points have a velocity v. By definition, *the length of the moving rod is the distance between the coordinates of its ends at a certain instant of time.* Thus, in order to measure the length of the moving rod, we must simultaneously fix the position of its ends for the same reading of the clocks located at the ends in a stationary coordinate system. Suppose that the ends of the moving rod are marked in the stationary coordinate system at an instant t_0 and have the coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2). According to the transformation formulas (11.1), the coordinates and time in the moving and stationary systems are connected through the following relations:

$$\begin{aligned} x'_1 &= x_1 - vt_0, & x'_2 &= x_2 - vt_0, \\ y'_1 &= y_1, & y'_2 &= y_2, \\ z'_1 &= z_1, & z'_2 &= z_2, \\ t'_1 &= t_0, & t'_2 &= t_0. \end{aligned} \tag{11.3}$$

Hence

$$\begin{aligned} x'_2 - x'_1 &= x_2 - x_1, \\ y'_2 - y'_1 &= y_2 - y_1, \\ z'_2 - z'_1 &= z_2 - z_1. \end{aligned}$$

Consequently, we obtain

$$\begin{aligned} l &= \sqrt{(x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2} \\ &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} = l'. \end{aligned} \tag{11.4}$$

This means that the rod has the same length in both coordinate systems. We can therefore state that *length is an invariant of Galilean transformations.*

UNIVERSALITY OF THE CONCEPT OF SIMULTANEITY. Let us consider the last row in (11.3). These equalities show that when the ends of the moving rod are described in the stationary coordinate system, the clocks have the same readings at the points in the moving coordinate system which coincides with the ends of the rod. This is a consequence of the transformation formula for time from one coordinate system to another, i.e. the relation $t' = t$. Accordingly, *the events that* *occur simultaneously in one coordinate system are also simultaneous in the other system.* In other words, *a statement that two events are simultaneous is universal in nature and independent of the coordinate system.*

INVARIANCE OF TIME INTERVAL. The invariance of time intervals is proved in terms of the transformation formula $t' = t$. Suppose that two events occurred at instants t'_1 and t'_2 in the moving coordinate system. The time interval between these two events is

$$\Delta t' = t'_2 - t'_1. \tag{11.5}$$

In accordance with (11.2), these events occurred in the stationary coordinate system at the instants $t_1 = t'_1$ and $t_2 = t'_2$ respectively, and hence the time interval between them is

$$\Delta t = t_2 - t_1 = t'_2 - t'_1 = \Delta t'. \tag{11.6}$$

Thus, it can be stated that *a time interval is an invariant of Galilean transformations.*

VELOCITY SUMMATION. Suppose that the coordinates of a point mass moving in the K' coordinate system are given as a function of time:

$$x' = x'(t'), \quad y' = y'(t'), \quad z' = z'(t'), \tag{11.7}$$

while the velocity projections are

$$u'_x = \frac{dx'}{dt'}, \quad u'_y = \frac{dy'}{dt'}, \quad u'_z = \frac{dz'}{dt'}. \tag{11.8}$$

The coordinates of the point mass vary with time in the stationary coordinate system in accordance with (11.2) as follows:

$$\begin{aligned} x(t) &= x'(t') + vt', \\ y(t) &= y'(t'), \\ z(t) &= z'(t'), \\ t &= t'. \end{aligned} \tag{11.9}$$

The velocity projections of the point mass in the stationary system are

$$\begin{aligned} u_x &= \frac{dx}{dt} = \frac{dx'}{dt} + v\frac{dt'}{dt} = \frac{dx'}{dt'} + v\frac{dt'}{dt'} = u'_x + v, \\ u_y &= \frac{dy}{dt} = \frac{dy'}{dt} = \frac{dy'}{dt'} = u'_y, \\ u_z &= \frac{dz}{dt} = \frac{dz'}{dt} = \frac{dz'}{dt'} = u'_z. \end{aligned} \tag{11.10}$$

These relations are the formulas for velocity summation in classical nonrelativistic mechanics.

INVARIANCE OF ACCELERATION. Differentiating (11.10) and substituting $dt = dt'$, we obtain

$$\frac{d^2x}{dt^2} = \frac{d^2x'}{dt'^2}, \quad \frac{d^2y}{dt^2} = \frac{d^2y'}{dt'^2}, \quad \frac{d^2z}{dt^2} = \frac{d^2z'}{dt'^2}. \tag{11.11}$$

These formulas show that acceleration is an invariant of Galilean transformations.

Sec. 12. CONSTANCY OF THE VELOCITY OF LIGHT

Various stages in the development of our ideas concerning the velocity of light and resulting in the statement that it is constant are described.

EXPERIMENTAL VERIFICATION OF THE VALIDITY OF GALILEAN TRANSFORMATIONS. The validity of Galilean transformations can be verified by comparing their corollaries with experimental results. The most important conclusion that can be drawn from Galilean transformations is the summation formulas (11.10). It was the verification of these formulas that revealed their approximate nature. The higher the velocity, the more serious the deviations from these formulas. The discrepancies are especially significant at velocities close to the velocity of light. They were first discovered during investigations of the velocity of light whose behaviour was not just strange from the point of view of classical concepts, it was inexplicable. Hence it is important first to consider the velocity of light.

EVOLUTION OF VIEWS ABOUT THE VELOCITY OF LIGHT. Ancient thinkers had two different ideas about the nature of light. Plato (427-347 B.C.) adhered to the theory that certain optic rays emanate from the eye and "feel" the object around us. Democritus (460-370 B.C.) supported the theory that atoms flow from the object to the eye, a theory Aristotle (384-322 B.C.) also supported. However, both theories practically merged into one after Euclid (*circa* 300 B.C.) had provided a geometrical description of optics and established the concept that light rays propagate in straight lines and formulated the laws of reflection. Atomistic ideas acquired predominance, and it was assumed that light propagates at a very high velocity, or even instantaneously. This conviction was based on an analogy with the flight of an arrow from a bow: the greater the velocity of the arrow, the closer its trajectory to a straight line.

Galileo, the founder of classical physics, believed that light has a finite velocity. However, he did not have any real idea about its magnitude and tried to measure it by using unsuitable methods. Descartes (1596-1650) hypothesized that

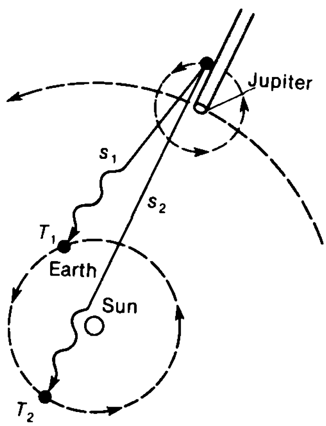


Fig. 30. Determining the velocity of light by Roemer.

light is a form of pressure transmitted through a medium at an infinite velocity. Thus, Descartes clearly suggested that the transmission of light would only be possible through a medium. Grimaldi (1618-1663) and Hooke (1635-1703) proposed the wave theory of light, according to which light is a wave motion in a homogeneous medium. However, the real founder of the wave theory was Huygens (1629-1695) who presented his ideas to the Paris Academy of Sciences in 1678. Newton (1643-1727) was reluctant to deliver his opinion about the nature of light and did not want to "invent a hypothesis". However, he obviously believed in the corpuscular theory of light, although he did not insist on its unconditional acceptance. In 1675, Newton wrote that in his opinion, light should not be defined as an ether or its oscillatory motion, but as something propagating from a luminous body. This "something" may be a group of peripatetic properties or, still better, an aggregate of extremely small and fast corpuscles.

DETERMINATION OF THE VELOCITY OF LIGHT BY ROEMER. The velocity of light was first measured by Roemer in 1676. Observations of eclipses of Jupiter's satellites showed that the apparent period of their revolution decreases when the Earth approaches Jupiter in the course of its yearly movement, and increases when the Earth moves away from Jupiter. Roemer interpreted this effect to be related to the finite velocity of light and calculated it from the results of measurements. Figure 30 shows the position of a satellite of Jupiter just after the eclipse. Since the orbital period of Jupiter around the Sun is much longer than the orbital period of the Earth, we can assume for the calculations that Jupiter is stationary relative to the Earth. Suppose that a satellite of Jupiter emerges from its shadow at an instant t_1. This event will be observed on the Earth at the instant

$$T_1 = t_1 + \frac{s_1}{c}, \tag{12.1}$$

where s_1 is the distance from the Earth to the point where the satellite appears when it is observed, and c is the velocity of light. After the satellite has completed one revolution around Jupiter, it will emerge from the shadow at the instant t_2. This event will be observed on the Earth at the instant

$$T_2 = t_2 + \frac{s_2}{c}. \tag{12.2}$$

Thus, according to the observations carried out on the Earth's surface, the period of revolution of the satellite is

$$T_{obs} = T_2 - T_1 = T_{tr} + \frac{s_2 - s_1}{c}, \tag{12.3}$$

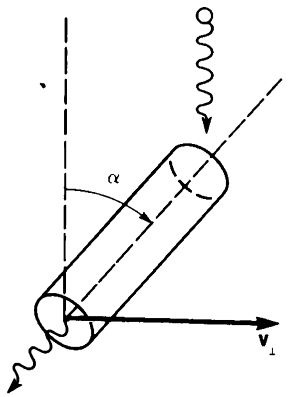


Fig. 31. Aberration of light.
While watching a star located at right angles to the Earth's velocity, the axis of the telescope must form an angle α with the true direction towards the star because of the aberration of light.

where $T_{tr} = t_2 - t_1$ is the true period of revolution of the satellite. Thus, the observed period of revolution of the satellite differs from the true period due to the difference $s_2 - s_1$ in the distance between the Earth and Jupiter. Making a large number of measurements of this period both as the Earth approaches and moves away from Jupiter, the average value will be equal to the true period since the terms $(s_2 - s_1)/c$ will have opposite signs in the two cases and cancel out upon averaging.

Knowing the value of T_{tr}, we can determine the velocity of light from (12.3):

$$c = \frac{s_2 - s_1}{T_{obs} - T_{tr}}. \tag{12.4}$$

The values of s_2 and s_1 are known from astronomical observations because the motions of Jupiter and the Earth have been studied thoroughly. The motion of Jupiter can easily be taken into consideration in calculations. Roemer obtained a value of $c = 214{,}300$ km/s for the velocity of light from such calculations.

This was the first reliable measurement of the velocity of light, which was quite accurate for that time.

ABERRATION OF LIGHT (Bradley, 1727). When the weather is calm, raindrops fall vertically to the ground. However, they form inclined tracks on the glassplane of a moving train. This is due to the superposition of the vertical velocity of the drops and the horizontal velocity of the train. A similar effect, called aberration, is also observed for light. As a result of the aberration of light, the apparent direction of a star differs from the true direction by an angle α, which is called the aberration angle (Fig. 31). It can be seen from the figure that

$$\tan\alpha = \frac{v_\perp}{c}, \tag{12.5}$$

where $v_\perp$ is the component of the Earth's velocity perpendicular to the direction of the star, and c is the velocity of light.

The aberration caused by the velocity of the Earth's motion around the Sun is called the annual aberration of light. The maximum annual aberration, which is the same for all stars, is equal to $\alpha_{max} = \arctan(v_\perp/c) = 20''.47$, since $v_\perp/c \simeq 10^{-4}$. The linear velocity of points on the Earth's surface due to its revolution around its own axis is responsible for the diurnal aberration of light. It is proportional to the cosine of the geographical latitude of the point of observation, and varies from $0''.13$ at the equator to $0''.00$ at the poles.

In order to observe aberration, we must know how to orient the telescope along a fixed direction in space. We cannot use the light from the stars since the light emitted by all stars undergoes the same kind of aberration. The orientation of the telescope in a fixed direction is calculated from the laws governing the Earth's revolution around the Sun and about its own axis. Bradley was able to measure the value of α. Given the value of $v_{\perp}$, he calculated the velocity of light and confirmed the results obtained by Roemer to within the same accuracy.

VARIOUS INTERPRETATIONS OF THE VELOCITY OF LIGHT. After the velocity of light had been measured, it was then possible to ask whether it depended on other factors. At that time, the answer depended on what views one held concerning the nature of light.

If light moved like a wave in a homogeneous medium, its velocity relative to the medium would be a constant which should depend on the properties of the medium. But the velocity of light relative to the source and the observer would be a variable quantity which should depend on the velocity of the source and the observer relative to the medium and would be given by (11.10) for the velocity summation.

If light were a flow of particles emitted by the source at a very high velocity, naturally the particle's velocity relative to the source would be a constant value, while the velocity relative to the observer would be the vector sum of the particle's velocity and the observer's velocity relative to the source, again calculated by (11.10).

ABSOLUTE ETHER AND ABSOLUTE VELOCITY. Newton's reputation guaranteed the success of the corpuscular theory of light. Although Huygens' wave theory did have some supporters, it was pushed into background for more than a century. However, new discoveries in optics at the beginning of the 19th century completely altered the pattern. In 1801, Young formulated the interference principle and used it to explain the colour of thin films. The ideas put forth by Young, however, were qualitative and were not widely recognized. The final blow to the corpuscular theory was dealt by Fresnel in 1818 when he explained diffraction on the basis of the wave theory. Until Fresnel, every attempt to explain it by using the corpuscular theory had proved fruitless. The basic idea underlying Fresnel's work was a combination of Huygens' principle of elementary waves and Young's interference principle. For many years after this, the corpuscular theory was completely rejected, and the concept of light as a wave process in a medium was universally accepted. This medium, which fills the entire Universe, was called absolute ether. The theory of light was reduced to a theory of oscillations in ether. The role of ether was subsequently extended, and it made responsible for other phenomena as well (gravitation, magnetism, electricity). Many famous scientists helped create the theory of absolute ether in the last century. However, these investigations are only of historic importance now, and there is no need to dwell upon them here. We have mentioned the term "absolute ether" simply to explain the concept of absolute velocity and the methods for determining it.

Absolute ether and absolute velocity were exceptionally important concepts in a general picture of the Universe in prerelativistic times.

Ether was supposed to fill the entire space in which bodies move, and remains stationary in space. The velocity of light relative to ether was supposed to be a constant whose value depended on the properties of ether. Bodies moved relative to the stationary ether. Obviously, the motion of bodies relative to ether was absolute and differed from the motion of bodies relative to one another. Indeed, if a body A moved relative to a body B at a velocity , the velocity could be varied by applying a force to either A or B. However, the motion of the body A relative to ether could only be changed by applying a force to the latter and not to any other body. The velocity of a body relative to ether was called "absolute". The absolute velocity of a body was supposed to be independent of the motions of other bodies. It was assumed that it would be a meaningful value even if no other bodies existed. The only question that arose was how to measure this absolute velocity.

MEASUREMENTS OF "ABSOLUTE" VELOCITY. Since the velocity of light relative to ether is constant, it is variable relative to other bodies moving in the ether and depends on their velocity relative to ether. By measuring the velocity of a body relative to light or, which is the same, the velocity of light relative to a body, we can determine its velocity relative to ether (the velocity of light relative to ether is assumed to be known). The situation here is analogous to the one in which an oarsman in a boat can determine his velocity relative to water by measuring the velocity of the boat relative to the waves and knowing the velocity of the waves relative to stationary water.

The first attempts to determine the absolute velocity of the Earth by using this method were made by Michelson and Morley in 1881 and 1887.

THE MICHELSON-MORLEY EXPERIMENT. The idea behind the experiment was to compare the propagation of light along two paths, one in the direction of the body's motion in ether and the other perpendicular to this direction. A diagram of the experiment is represented in Fig. 32.

A ray of monochromatic light, i.e. light with a single

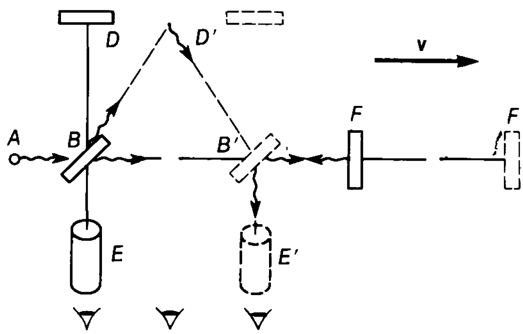


Fig. 32. Diagram of the Michelson-Morley experiment in a reference frame fixed to ether.
The figure shows the successive positions of the interferometer relative to ether.

frequency, from a source A falls upon a translucent plate B set at 45° to the path of the ray. The ray is split by the plate into two. Since these two rays are generated by the same incident ray, the waves in them are not independent and are in phase. By analogy with ripples on water surface, we can state that the oscillations of both waves are perfectly synchronized where the ray is split. We can also imagine a case when both waves oscillate with the same period, but one wave oscillates slightly behind the other. In other words, there is a constant phase difference between the two waves, and if the oscillation at one point is described by the function $\sin \omega t$, the oscillation at another point is given by the function $\sin(\omega t + \varphi)$, where $\varphi = \text{const}$ is the phase difference between the two waves. Such waves are called coherent. Thus, we can state that the ray is divided at B into two coherent rays, one of which is reflected by the plate and is incident on a mirror D, while another passes through the plate and is incident on a mirror F. The rays are reflected by the mirrors D and F, and return to the plate B. The ray partially reflected by the mirror D and passing through the translucent plate B encounters, in an interferometer E, the ray reflected by the mirror F and the plate B. Thus, two coherent rays which have traversed different paths after splitting meet in the interferometer. Obviously, if the paths are traversed by the rays in the same period of time, the phase difference between the waves at the point where they meet will have the same value φ as at the point where they parted. For example, suppose that the phase difference was $\varphi = 0$ where they were split, i.e. the oscillations in both rays are in phase. In this case, the rays will meet in the interferometer in the same phase and hence interfere constructively, i.e. the crest of one wave coincides with that of the other wave. If, however, the rays take different times to traverse their paths, a phase difference will appear between the oscillations at the point where the rays meet. For example, it may happen that the crest of one wave coincides with the trough of another wave and so they will cancel each other out.

A change in intensity as a function of the phase difference of the components of light oscillations is called interference. Observing interference, we can draw conclusions concerning the phase difference between the coherent waves arriving at the interferometer and thus calculate the time one wave lags relative to the other. This is what was done by Michelson and Morley. The optical part of the experiment and the construction of the Michelson interferometer are described in *Optics* *.

CALCULATION OF THE PATH DIFFERENCE BETWEEN RAYS. Suppose that the instrument is moving in the direction of $BF = l_1$ at a velocity v relative to ether (see Fig. 32). The velocity of light relative to ether is denoted by c. As the ray is moving from B to F, the directions of the velocity of light and the velocity of the instrument coincide. Consequently, the velocity of light relative to the instrument should be $c - v$, and the time taken by the ray to traverse the distance from B to F would be

$$t_{BF}^{(1)} = \frac{l_1}{c - v}. \tag{12.6}$$

This is what would arise if an "ether wind" is blowing past the instrument at a velocity $c - v$. The time during which the path from F to B is traversed after reflection is

$$t_{FB}^{(1)} = \frac{l_1}{c + v} \tag{12.7}$$

since the light moves against an advancing instrument, and the velocities are added. Thus, the total time spent in traversing a path up to the mirror F and back is

$$t_{\parallel}^{(1)} = t_{BF}^{(1)} + t_{FB}^{(1)} = \frac{2l_1}{c}\frac{1}{1 - v^2/c^2}. \tag{12.8}$$

Superscript (1) indicates the time intervals during which different paths are traversed if the instrument is oriented relative to the velocity direction as shown in Fig. 32. If the velocity direction coincides with the arm BD, superscript (2) is employed.

In order to determine the time required for light to traverse the path $BD'B'$, we consider that for a ray to fall on D after reflection by B, the velocity of light must be decomposed into two components, viz. a velocity v in the direction of motion of the instrument and $c_{\perp}$, the perpendicular component directed

* A. N. Matveev, *Optics*, Mir Publishers, Moscow, 1988.

from B to D. Hence we can write

$$c^2 = c_\perp^2 + v^2. \tag{12.9}$$

The ray will traverse a path $BD = l_2$ in the time

$$t_{BD}^{(1)} = \frac{l_2}{c_\perp} = \frac{l_2}{\sqrt{c^2 - v^2}} = \frac{l_2}{c}\frac{1}{\sqrt{1 - v^2/c^2}}. \tag{12.10}$$

When the ray moves in the opposite direction, its velocity is again $c_\perp$, and hence the time required to traverse the path DB remains the same. Hence the total time required for light to traverse a distance to the mirror D and back is

$$t_\perp^{(1)} = \frac{2l_2}{c}\frac{1}{\sqrt{1 - v^2/c^2}}. \tag{12.11}$$

The orbital velocity of the Earth around the Sun is about 30 km/s. The linear velocity of its rotation is about 60 times slower (approximately 500 m/s) and may be neglected in comparison with its orbital velocity. Hence if an instrument is situated on the Earth, the quantity $(v/c)^2$ is of the order of 10^{-8}. Since $(v/c)^2$ is negligibly small, we can expand the expressions (12.8) and (12.11) into a series in this quantity. Truncating the expansion to the first terms, we obtain

$$t_\parallel^{(1)} \approx \frac{2l_1}{c}\left(1 + \frac{v^2}{c^2}\right),$$
$$t_\perp^{(1)} \approx \frac{2l_2}{c}\left(1 + \frac{1}{2}\frac{v^2}{c^2}\right). \tag{12.12}$$

!

If two shots are fired at an interval of, say, 1 s in a moving train, an observer standing on the track on which the train is moving towards him will hear these shots as fired at an interval of less than 1 s. To an observer from which the train is moving away, the shots will appear to have been fired at an interval of more than 1 s.
If there is no wind accompanying rainfall, the umbrella must be held in a vertical position to protect from rain. If, however, we have to run, we must incline the umbrella in the direction of motion.

Consequently, the time difference between the rays in traversing their paths is

$$\Delta t^{(1)} = t_\parallel^{(1)}(l_1) - t_\perp^{(1)}(l_2) = \frac{2}{c}\frac{v^2}{c^2}\left(l_1 - \frac{l_2}{2}\right) + \frac{2}{c}(l_1 - l_2). \tag{12.13}$$

Let us now turn the instrument by 90° so that BD is collinear with the velocity, while BF is perpendicular to it. The path difference between the rays in this case is calculated in the same way as before, but l_1 is now replaced by l_2, and vice versa.

Since l_1 is perpendicular to the velocity in the second case, while l_2 is parallel to it, we obtain

$$\Delta t^{(2)} = t_\perp^{(2)}(l_1) - t_\parallel^{(2)}(l_2) = -\frac{2}{c}\frac{v^2}{c^2}\left(l_2 - \frac{l_1}{2}\right) - \frac{2}{c}(l_2 - l_1). \tag{12.14}$$

Thus, the total change in the time difference for the rays upon a rotation of the instrument by 90° is

$$\Delta t = \Delta t^{(2)} - \Delta t^{(1)} = -\frac{l_1 + l_2}{c}\frac{v^2}{c^2}. \tag{12.15}$$

RESULTS OF THE MICHELSON-MORLEY EXPERIMENT. The actual motion of the instrument relative to the hypothetical ether is unknown. Hence the instrument cannot be oriented such that one of its arms lies in the direction of motion. The instrument was therefore slowly rotated during the experiment. Irrespective of the direction of motion of the instrument relative to ether, each arm would twice be collinear with the direction of motion when the instrument is rotated by 360° and twice be perpendicular to the direction if the axis about which the instrument is rotated is perpendicular to the body's velocity. Of course, we can imagine a situation in which the axis of rotation of the instrument coincides with the direction of its motion in ether. *In this case, there would be no change in the difference between the paths traversed by the rays.* However, the experiment can be carried out in such a way that the direction of the axis of rotation would be varied and a change in the path difference would be ensured. Interference fringes would be observed in the interferometer in this case. If the path difference between the rays were varied as the instrument rotated, the position of the interference fringes would change in the field of view. The change in the path difference with time could then be calculated by measuring the displacement of the interference fringes, and the velocity of the instrument relative to ether could be determined.

Such an experiment was carried out in 1881 by Michelson and later, in 1887, by Michelson and Morley with a higher degree of accuracy. In order to increase the effective distance, Michelson and Morley used multiple reflections of a ray from the mirrors and increased the value of $l_1 + l_2$ to over 10 metres. The wavelengths of visible light range from 0.4×10^{-6} m to 0.75×10^{-6} m. The magnitude of delay, expressed by (12.15), assumes the following form when expressed in terms of wavelength:

!

In the Michelson-Morley experiment, the "arms" could not have the same length since this would require the measurement of a distance of a few metres with an accuracy of up to a millionth fraction of a metre. At that time, such precise measurements were impossible.
If the ballistic hypothesis were true, a variation of the brightness of a double star would be measured by observing its motion. Indeed, there are many variable stars, but the law of variation of their brightness is not in conformity with the ballistic hypothesis.

$$\Delta\lambda = \Delta t c = (l_1 + l_2)\frac{v^2}{c^2} \approx (l_1 + l_2) \times 10^{-8}. \tag{12.16}$$

Here, we have used the value $v^2/c^2 \simeq 10^{-8}$ for the velocity of the Earth around the Sun. The relative displacement of the interference fringes is

$$\frac{\Delta\lambda}{\lambda} = \frac{(l_1 + l_2) \times 10^{-8}}{\lambda}.$$

In Morley's experiments which were carried out in 1887, the effective distance $l_1 + l_2$ was 11 m. Hence the expected displacement was $\Delta\lambda/\lambda \simeq 1/5$ for $\lambda = 0.5 \times 10^{-6}$ m. This is much larger than the values which can be observed without difficulty. Actually, a displacement corresponding to a velocity of just 3 km/s of the instrument relative to the hypothetical ether could have been measured. However, no effect was observed. It turned out that the velocity of light in every direction is the same, and there is no ether wind. The experiment was repeated in 1905 with an even higher accuracy, but still yielded a negative result. The experiment was subsequently repeated by many researchers and has so far inevitably led to the conclusion that the velocity of light is the same in every direction, and there is no ether wind. The accuracy of the experiments was considerably increased with the advent of lasers. At present, it has been proved that the velocity of the ether wind, if one exists, would not exceed 10 m/s.

INTERPRETATION OF THE MICHELSON-MORLEY EXPERIMENT BASED ON THE CONCEPT OF ETHER. Two solutions were proposed to explain the situation described above by means of the ether concept:

1. It was suggested that in the vicinity of a massive object like the Earth the ether moves with the object, i. e. is entrained by its motion. Naturally, no "ether wind" would be observed in the vicinity of the object.

2. It was also suggested that the size of an object moving in the ether is not constant, but varies in such a way that the expected path difference (12.15) is not realized.

The suggestion about the entrainment of the ether has to be rejected because it is in contradiction with other observed facts. In particular, it is not in accord with the phenomenon of aberration of light. The second suggestion, which was put forth by Lorentz and Fitzgerald, successfully explains the absence of a delay. A comparison of (12.8) and (12.11) shows that for $l_1 = l_2 = l$, the time spent in traversing the path along the motion will be equal to the time taken by the ray to traverse the path perpendicular to the direction of motion if the length of the arm in the direction of motion shortens and becomes

$$l' = l\sqrt{1 - v^2/c^2}. \tag{12.17}$$

If we assume that bodies contract in the direction of motion in accordance with (12.17), it becomes obvious that the Michelson-Morley experiment would yield negative results.

However, this explanation is not convincing since it gives a logically unsatisfactory idea about the velocity of light. The velocity of light is assumed to be constant relative to ether and variable relative to bodies that are in motion relative to ether. However, measurements of this velocity relative to bodies always give the same result. In short, the velocity of light relative to bodies is variable, but the results of its measurements are always the same. Obviously, it is meaningless to state that the velocity of light is variable in this case and so it must be rejected. Instead, we have to agree that the velocity of light is constant. In this case, the results of the Michelson-Morley experiment can be explained in a natural way.

However, the following remark should be made. Strictly speaking, the Michelson-Morley experiment and the analogous experiments that were carried out subsequently do not lead to the conclusion that the velocity of light is constant. They simply indicate that the average velocity of light in opposite directions is identical in a given inertial coordinate system, and it is impossible to conclude that the velocity of light is constant in all directions.

However, if the homogeneity and isotropy of space and the homogeneity of time in inertial reference frames are assumed to have been proved, there can be no doubt that the velocity of light is constant in all directions.

BALLISTIC HYPOTHESIS. There is yet another way of explaining the result of the Michelson-Morley experiment: we can reject the ether concept entirely and treat light as a flow of material particles. In other words, we revert to Newton's point of view. It is natural to assume that the velocity of the particles relative to the source is constant and is added to the velocity of the source according to the parallelogram law.

Since the velocity of light relative to the source is assumed to be the same in all directions in the ballistic hypothesis, one would not expect any path difference in the Michelson-Morley experiment. Hence the ballistic hypothesis explains the result of this experiment in a natural way, thus avoiding the assumption that the velocity of light is constant, which is incomprehensible to a Galilean transformation. In spite of this, the ballistic hypothesis was found to be groundless.

?

How did ancient thinkers arrive at the idea that the velocity of light is very high? Were their arguments correct in the light of the concepts of modern physics?
Describe an experimental set-up that could be used to measure the velocity of sound by Roemer's method.
What is the effect of taking into account the velocity of Jupiter's motion in the calculation of the velocity of light by Roemer's method?

FLAW IN THE BALLISTIC HYPOTHESIS. De Sitter mentioned in 1913 that the ballistic hypothesis could be verified by observing double stars. A double star is a system of two close stars moving around a common centre of mass. If one of the stars is much larger than the other, then the smaller star can be assumed to move around the larger star which is at rest. A large number of double stars are known. The velocity of the stars can be measured from the Doppler effect, whence the orbital elements can be determined. It turns out that the components of double stars revolve in elliptical orbits in

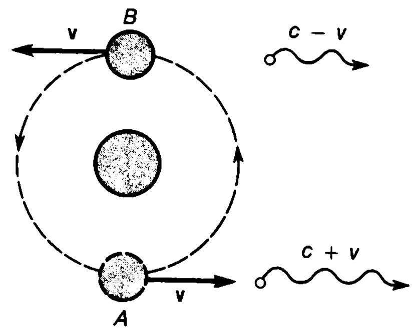


Fig. 33. Proving the flaw in the ballistic hypothesis.

If the ballistic hypothesis were true, it would be possible to observe a star both in the upper and the lower position simultaneously at a considerably large distance. This, however, is not the case in actual practice.

accordance with Kepler's laws. In other words, gravitational forces exist between them, decreasing in inverse proportion to the square of the distance between the components.

No peculiarity has been observed in the motion of the components of double stars. However, the motion of the stars would have been quite strange, had the ballistic hypothesis been valid.

Suppose that a double star is being viewed from a large distance s. For the sake of simplicity, we shall assume that the lighter star moves in a circular orbit at a velocity v around the heavier star which is assumed to be stationary (Fig. 33). The orbital period is taken as T. A ray of light, emitted when the star is at point B (upper point in Fig. 33) and moves away from the observer, will propagate in the direction of the observer at a velocity $c - v$. If t_1 is the instant of time at which the ray is emitted from the star, the instant at which it reaches the eye of the observer will be

$$T_1 = t_1 + \frac{s}{c - v}, \tag{12.18}$$

where s is the distance between the star and the observer. After half an orbital period $T/2$, a ray of light, emitted when the star will move from point A (lower point in Fig. 33), will propagate towards the observer at a velocity $c + v$. Consequently, the ray emitted by the star at A will reach the observer's eye at the instant

$$T_2 = t_1 + \frac{T}{2} + \frac{s}{c + v}. \tag{12.19}$$

If the distance s is large, this ray may overtake the ray emitted at B in view of its higher velocity. This will happen at a distance s for which $T_2 = T_1$. This distance can easily be calculated from (12.18) and (12.19). For very large distances s, the ray emitted at A may overtake the ray emitted at B in the previous revolution, and so on. In this case, an observer watching the star from a large distance will see the star at several points simultaneously on its orbit.

Thus, if the ballistic hypothesis were true, astronomers watching double stars would see a very intricate picture. In actual practice, this does not happen.

What is actually observed can be explained by assuming that the double stars move according to Kepler's laws, and that the velocity of light is constant and cannot be added to the velocity of the source as required by the ballistic hypothesis. Thus the ballistic hypothesis is refuted.

In view of the flaw in the ballistic hypothesis, it should be admitted that the velocity of light must be independent of the velocity of the source. It follows from the results of the Michelson-Morley experiment that it is also independent of the velocity of the observer. Hence we can conclude that the velocity of light is a constant quantity which does not depend on the velocity of the source or the observer.

INCOMPATIBILITY BETWEEN THE CONSTANCY OF THE VELOCITY OF LIGHT AND CONVENTIONAL CONCEPTS. The constancy of the velocity of light is in sharp contradiction with the accepted notions based on everyday experience, as well as with formulas (11.10) for velocity summation, which are derived from Galilean transformations. Thus, it can be stated that the Galilean transformations (11.2) contradict the experimentally observed fact that the velocity of light is a constant quantity. However, this contradiction becomes noticeable only at very high velocities.

Imagine a train moving at a velocity of 100 km/h relative to the railway bed. If a person walks towards the engine along a carriage at a velocity of 5 km/h relative to the train, his velocity relative to the railway bed is 105 km/h. This result is quite obvious and is in complete accord with the accepted notions about space and time which are expressed in the present case by the formula for velocity summation in classical mechanics. This formula has experimentally been verified many times.

Let us now consider a rocket travelling at a velocity of 100,000 km/s relative to the Earth. Suppose that an object in the rocket is moving in the same direction as the rocket at a velocity of 100,000 km/s relative to the rocket. What will be the velocity of the object relative to the Earth? If we measured this velocity, it would come to 164,000 km/s. Although such an experiment has actually not been carried out, many other similar experiments have shown that formulas (11.10) for velocity summation are not correct. For velocities much smaller than the velocity of light, this error is not noticeable since the deviations from (11.10) are extremely small. The

?

How is the aberration of light from a star observed if the true direction towards the star is not known?
Why would it be expedient to call the velocity relative to ether "absolute" if absolute ether existed?
Why cannot the results of the Michelson-Morley experiment be interpreted with the help of the Fitzgerald-Lorentz contraction?

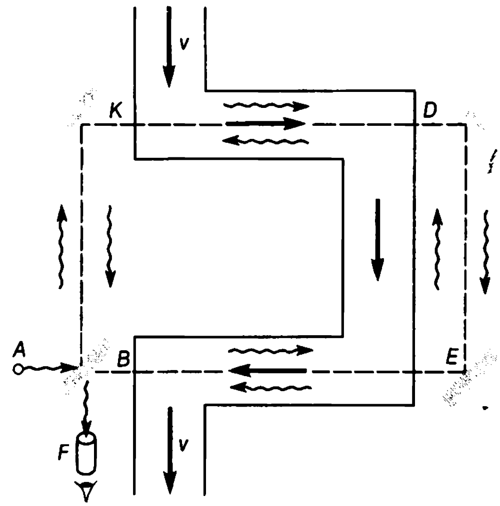


Fig. 34. Diagram of Fizeau's experiment.

inaccuracy of (11.10) was first experimentally revealed in the middle of the last century, although the scientists of the time did not realize a significance of this result.

THE IDEA BEHIND FIZEAU'S EXPERIMENT. Long before the concept emerged that the velocity of light is constant and it was realized that the Galilean transformations are approximate, an experiment was known in physics which indicated a strange rule for velocity summation comparable with the velocity of light. This experiment was first carried out by Fizeau in 1851.

Fizeau's experiment involved the measurement of the velocity of light in a moving medium, say, water. Let $u' = c/n$ be the velocity of light in the medium with a refractive index n. If the medium in which light is propagating is itself moving at a velocity v, the velocity of light relative to an observer at rest must be $u' \pm v$ depending on whether the velocities of light and the medium are in the same or opposite directions. In his experiment Fizeau compared the velocity of light in the direction of motion of the medium and against the motion.

A diagram of Fizeau's experiment is shown in Fig. 34. A monochromatic ray from the source A is incident on the translucent plate B and is then divided into two coherent rays. The ray which is reflected from the plate traverses the path $BKDEB$ (K, D and E are mirrors), while the transmitted ray traverses the path $BEDKB$, i.e. moves in a direction opposite to that of the first ray. Returning to the plate B, the first ray is partially reflected and enters the interferometer F. The second ray is partially transmitted upon returning to the plate B and also falls on the interferometer F. Both rays traverse the same

?

What is the astronomical evidence against the validity of the ballistic hypothesis?
How was the result of Fizeau's experiment interpreted at the time when it was carried out?
Why is the statement that the velocity of light is constant still a postulate in spite of the fact that so many experimental confirmations of the statement exist?

path and pass through the liquid flowing through the tube over the segments BF and KD. If the liquid is at rest, the paths of both rays are identical and will traverse this distance in the same time in both directions.

If, however, the liquid is in motion, the paths of the rays are not equivalent: one of them moves along the flow in the indicated segments, while the other ray moves against the flow in the corresponding segments. This results in a path difference, and one of the rays is delayed relative to the other. The path difference can be determined from the interference pattern, and thus the velocity of light can be calculated in the segments with the liquid since the velocity of light on the remaining segments and the pathlength of all the segments are known.

CALCULATION OF THE PATH DIFFERENCE BETWEEN RAYS. We introduce the following notation: l is the total length of the segments in which light passes through the liquid, t_0 is the time during which light passes through the entire path except the segments containing the liquid, $u^{(+)}$ is the velocity of light in the flow direction, and $u^{(-)}$ is the velocity of light against the flow direction. These velocities can be represented as

$$u^{(+)} = u' + kv, \qquad u^{(-)} = u' - kv, \tag{12.20}$$

where k is a coefficient which must be experimentally determined. If $k = 1$, the classical formulas (11.10) for velocity summation are valid. If, however, $k \neq 1$, there must be deviations from the formulas. It should be noted that we are dealing with very high velocities in this experiment since the refractive index of water is about 1.3 for visible light, and hence the velocity of light in water is about 230,000 km/s.

The time during which the first and second rays traverse the entire path is

$$t_1 = t_0 + \frac{l}{u' + kv}, \qquad t_2 = t_0 + \frac{l}{u' - kv}. \tag{12.21}$$

Hence the path difference in terms of time can be written in the form

$$\Delta t = t_2 - t_1 = \frac{2lkv}{u'^2 - k^2v^2}. \tag{12.22}$$

Measuring the path difference from the displacement of interference fringes and knowing the values of l, v and u', we can determine k from this formula.

RESULT OF FIZEAU'S EXPERIMENT. The following value was obtained for the coefficient k in Fizeau's experiment:

$$k = 1 - \frac{1}{n^2}, \tag{12.23}$$

where n is the refractive index of the liquid. Thus, the velocity of the liquid and the velocity of light in the liquid are not added in accordance with the classical formulas for velocity summation. From a layman's point of view, this result is as astonishing as the statement that the velocity of light in vacuum is constant. However, Fizeau's result did not cause any surprise at the time since Fresnel had shown much earlier that when matter moves in ether, it only partially entrains ether. The amount of entrainment exactly corresponds to the result of Fizeau's experiment.

It became clear only after the creation of the theory of relativity that Fizeau's experiment provided the first experimental evidence that the classical law for velocity summation and the Galilean transformations was incorrect.

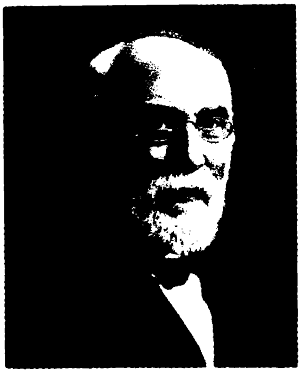
CONSTANCY OF THE VELOCITY OF LIGHT AS A POSTULATE. The statement that the velocity of light is constant in vacuum, i.e. the independence of the velocity of light of the velocity of the source or the observer, is the natural consequence of many experimental facts. We have only described such experiments and ideas which were chronologically the first. Later on, this statement was subjected to many experimental verifications. However, the main confirmation that the statement is true arises from the agreement of all experimental results with all the corollaries of the statement. There are many such confirmations since the whole of the modern physics of high velocities and high energies is based on the postulate that the velocity of light is constant.

In spite of all this, the statement that the velocity of light is constant remains a postulate or an assumption beyond the reach of direct experimental verification. This is due to a finite accuracy of experimental verifications, which is caused by the postulative nature of the relativity principle.

Sec. 13. LORENTZ TRANSFORMATIONS

Lorentz transformations are derived from the relativity principle and the postulate that the velocity of light is constant.

POSTULATES. Since the results obtained from Galilean transformations contradict experimental evidence at high velocities, and the constancy of the velocity of light does not follow from them, they do not correctly reflect the relation between the coordinates and time in two inertial coordinate systems moving relative to each other. We must find other transformations which correctly explain the experimental results and, among other things, lead to the concept that the velocity of light is constant. These transformations are called the Lorentz transformations. They can be derived from the two principles which were substantiated in previous sections:



Hendrik Antoon Lorentz (1853-1928)
Dutch physicist, the founder of the classical electron theory. He worked out the equations of electrodynamics for moving media and derived coordinate transformations which proved to be a significant step towards the creation of the theory of relativity.

(1) the relativity principle;
(2) the principle of the constancy of the velocity of light.

Although both these principles have been verified in many experiments, they are of postulative nature and are hence called sometimes the relativity postulate and the postulate that the velocity of light is constant.

The theory governing space-time relations, which is based on the Lorentz transformations, is called the (special) theory of relativity. It was created mainly on the basis of the works of Einstein (1879-1955), Poincaré (1854-1912) and Lorentz (1853-1928). A significant contribution was made by Minkowski (1864-1909) in developing a geometrical interpretation of the theory.

LINEARITY OF COORDINATE TRANSFORMATIONS. Using purely geometrical transformations to get a spatial rotation and displacement of the origin of coordinates within each reference body, we can always orient moving coordinate systems to the position shown in Fig. 29. Since the velocities are not added in accordance with the classical formulas (11.10), it can be expected that time in one coordinate system is not only expressed in terms of time in the other coordinate system, but also depends on coordinates. Hence the transformations have the following general form:

$$\begin{aligned} x' &= \Phi_1(x, y, z, t), \quad y' = \Phi_2(x, y, z, t), \\ z' &= \Phi_3(x, y, z, t), \quad t' = \Phi_4(x, y, z, t), \end{aligned} \tag{13.1}$$

where the right-hand sides contain certain functions Φ_i whose form remains to be found.

The general form of these functions is determined by the properties of space and time. While considering the geometrical relations in a chosen coordinate system and while carrying out measurements in this system, we assumed that each point is identical to any other point. This means that the origin of the coordinate system can be moved to any point, and all geometrical relations between any geometrical objects remain identical to those obtained with the origin at any other point.

This property is called the homogeneity of space, i.e. the property of invariance of the characteristics of space upon a transition from one point to another.

Similarly, we can align the axes of the coordinate system arbitrarily at all points in space. The relations between geometrical objects also remain unchanged in this case.

This means that the properties of space are the same in all directions. This property is called the isotropy of space.

The homogeneity and isotropy of space are its principal properties in inertial coordinate systems.

Time also has the important property of homogeneity. Physically, this can be explained as follows. Suppose that a certain physical situation arises at a certain instant of time. The evolution of this event will occur at subsequent instants of time. Suppose that the same physical situation arises at another instant of time. If this event evolves relative to this instant of time in exactly the same way as it did in the previous situation relative to the initial instant of time, it can be stated that time is homogeneous. In other words,

the homogeneity of time is the identity of evolution and variation of a given physical situation irrespective of the instant at which it arose.

It follows from the homogeneity of space and time that the transformations (13.1) must be linear. In order to prove this, let us consider an infinitesimal variation dx', i.e. the difference between the x'-coordinates of two infinitely close points. In a K coordinate system, this difference will correspond to the infinitesimal differences dx, dy, dz in coordinates and dt in time. The total variation dx' associated with the variations of the quantities x, y, z and t can be calculated from (13.1) via a total differential, whose expression is known from mathematics:

$$dx' = \frac{\partial \Phi_1}{\partial x} dx + \frac{\partial \Phi_1}{\partial y} dy + \frac{\partial \Phi_1}{\partial z} dz + \frac{\partial \Phi_1}{\partial t} dt. \tag{13.2}$$

In view of the homogeneity of space and time, these relations must be the same for all points in space at all instants of time. This means that the quantities $\partial \Phi_1/\partial x$, $\partial \Phi_1/\partial y$, $\partial \Phi_1/\partial z$ and $\partial \Phi_1/\partial t$ must be independent of coordinates and time, i.e. they must be constant. Hence the function Φ_1 has the following form:

$$\Phi_1(x, y, z, t) = A_1 x + A_2 y + A_3 z + A_4 t + A_5, \tag{13.3}$$

where A_1, A_2, A_3, A_4 and A_5 are constants. Hence the function $\Phi_1(x, y, z, t)$ is a linear function of its arguments. Similarly, we can prove that the other functions Φ_2, Φ_3 and Φ_4 in the transformations (13.1) will also be linear functions of their arguments due to the homogeneity of space and time.

The linearity of transformations also follows from the fact that if the acceleration of a point mass is zero in one inertial coordinate system, it will be zero in any other inertial coordinate system. This is a corollary of the definition of inertial systems.

TRANSFORMATIONS FOR y AND z. The origin of each system is defined by the equalities $x = y = z = 0$ and $x' = y' = z' = 0$. We shall assume that the origins coincide at the instant $t = 0$. Then the free term A_5 in the linear transformations like (13.3) must be zero, and the transformations for y

and z can be written in the form

$$\begin{aligned} y' &= a_1 x + a_2 y + a_3 z + a_4 t, \\ z' &= b_1 x + b_2 y + b_3 z + b_4 t. \end{aligned} \tag{13.4}$$

The alignment of the coordinate axes is shown in Fig. 29. The Y'-axis is parallel to the Y-axis, and the Z'-axis to the Z-axis. Since the X'-axis always coincides with the X-axis, the condition $y = 0$ always leads to the equality $y' = 0$, while the condition $z = 0$ leads to the equality $z' = 0$. In other words, we must have

$$\begin{aligned} 0 &= a_1 x + a_3 z + a_4 t, \\ 0 &= b_1 x + b_2 y + b_4 t \end{aligned} \tag{13.5}$$

for all values of x, y, z and t. This is only possible under the condition

$$a_1 = a_3 = a_4 = 0, \quad b_1 = b_2 = b_4 = 0. \tag{13.6}$$

Hence the transformations for y and z assume the following simple form:

$$y' = ay, \quad z' = az, \tag{13.7}$$

where the coefficients in the transformations must be identical: $y_3 = b_3 = a$ because the Y- and Z-axes are equivalent relative to the direction of motion. The coefficient a in (13.7) indicates the factor by which the length of a scale in the K' coordinate system is larger than that in the K coordinate system.

We can write (13.7) in an alternative form:

$$y = \frac{1}{a} y', \quad z = \frac{1}{a} z'. \tag{13.8}$$

The quantity $1/a$ indicates the factor by which the length of a scale in the K system is larger than that in the K' system. According to the relativity principle, both coordinate systems are equivalent, and hence the scale length must vary when going from one system to another in the same way as it does when going in the reverse direction. Hence the equality $1/a = a$ must be obeyed in (13.7) and (13.8), which means that $a = 1$ (the other mathematically possible solution $a = -1$ is unacceptable since we chose the orientations of the axes so that the positive Y- and Z-directions coincide with the Y'- and Z'-directions). As a result, the transformations for y- and z-coordinates assume the form

$$\boxed{y' = y, \quad z' = z.} \tag{13.9}$$

TRANSFORMATIONS FOR x AND t. Since the y- and z-variables are separately transformed, the x- and t-variables can only be connected through a linear transformation with each other. The origin of a moving coordinate system has the coordinate $x = vt$ in the stationary coordinate system and the coordinate $x' = 0$ in the moving system. Hence in view of linear transformations, we must have

$$x' = \alpha(x - vt), \tag{13.10}$$

where α is the proportionality factor which has to be determined.

Similar arguments can be used if we take the moving coordinate system to be at rest. In this case, the origin of the K system has the coordinate $x' = -vt'$ in the K' system since the K system moves in the K' system in the negative X-directions. The origin of the K system is characterized by the equality $x = 0$ in the K' system. Hence, if we assume the K' system to be at rest, we arrive at the following transformation instead of (13.10):

$$x = \alpha'(x' + vt'), \tag{13.11}$$

where α' is the proportionality factor. We shall prove that in accordance with the relativity principle, $\alpha = \alpha'$.

Suppose that a rod is at rest in a K' system and has a length l in the system. This means that the coordinates of the ends of the rod in the system differ by l:

$$x'_2 - x'_1 = l. \tag{13.12}$$

The rod moves at a velocity v in another K system. Its length will then be the distance between two points of the stationary system with which the ends of the moving rod coincide at the same instant of time. Let us mark the ends at the instant t_0. In accordance with (13.10), we obtain the following expressions for the x'_1- and x'_2-coordinates of the ends at this instant of time:

$$x'_1 = \alpha(x_1 - vt_0), \quad x'_2 = \alpha(x_2 - vt_0). \tag{13.13}$$

Consequently, the length of the moving rod in the stationary K system is

$$x_2 - x_1 = \frac{x'_2 - x'_1}{\alpha} = \frac{l}{\alpha}. \tag{13.14}$$

Let us now assume that the same rod is at rest in the K' system and has a length l in it. Consequently, the coordinates of the ends of the rod differ in the system by l, i.e.

$$x_2 - x_1 = l. \tag{13.15}$$

In the K' system, taken as the stationary system, the rod moves at a velocity v. In order to measure its length relative to the K' system, we must mark the ends of the rod at a certain instant t'_0 in the system. In accordance with (13.11), we have

$$x_1 = \alpha'(x'_1 + vt'_0), \quad x_2 = \alpha'(x'_2 + vt'_0). \tag{13.16}$$

Consequently, the length of the moving rod in the K' system which assumed to be stationary is

$$x'_2 - x'_1 = \frac{x_2 - x_1}{\alpha'} = \frac{l}{\alpha'}. \tag{13.17}$$

According to the relativity principle, both systems are equivalent, and the length of the same rod moving in these systems at the same velocity must be the same. Hence we must have $l/\alpha = l/\alpha'$ in (13.14) and (13.17), i.e. $\alpha = \alpha'$, Q. E. D.

Let us now make use of the postulate that the velocity of light is constant. Suppose that when the origins of the two systems coincide, at which time clocks at the two origins show the time $t = t' = 0$, a light signal is emitted from the origins. The propagation of light in the K' and K coordinate systems is described by the equalities

$$x' = ct', \quad x = ct, \tag{13.18}$$

where we have taken into account that the velocity of light in both systems has the same value c. These equalities describe the position of the light signal propagating in the X- and X'-directions at any instant of time in each coordinate system. Substituting (13.18) into (13.10) and (13.11) and considering that $\alpha = \alpha'$, we obtain

$$ct' = \alpha t(c - v), \quad ct = \alpha t'(c + v). \tag{13.19}$$

Multiplying both sides of these equalities by each other and cancelling out $t't$, we obtain

$$\boxed{\alpha = \frac{1}{\sqrt{1 - v^2/c^2}}.} \tag{13.20}$$

We obtain from (13.11) and (13.10) that

$$vt' = \frac{x}{\alpha} - x' = \frac{x}{\alpha} - \alpha(x - vt) = \alpha vt + x\left(\frac{1}{\alpha} - \alpha\right), \tag{13.21}$$

whence, from (13.20), we get

$$\boxed{t' = \alpha\left[t + \frac{x}{v}\left(\frac{1}{\alpha^2} - 1\right)\right] = \frac{t - (v/c^2)x}{\sqrt{1 - v^2/c^2}}.} \tag{13.22}$$

LORENTZ TRANSFORMATIONS. The transformations (13.9), (13.10) and (13.22) connect together the coordinates of two systems moving relative to each other at a velocity v. They are called Lorentz transformations. Let us give these transformations all together:

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \quad y' = y, \quad z' = z, \quad t' = \frac{t - (v/c^2)x}{\sqrt{1 - v^2/c^2}}. \tag{13.23}$$

In accordance with the relativity principle, the inverse transformations have the same form, the only exception being that the sign of the velocity must be reversed:

$$x = \frac{x' + vt'}{\sqrt{1 - v^2/c^2}}, \quad y = y', \quad z = z', \quad t = \frac{t' + (v/c^2)x'}{\sqrt{1 - v^2/c^2}}. \tag{13.24}$$

It is possible to get (13.24) from (13.23) without using the relativity principle. For this purpose, we must consider (13.23) as a system of equations in unprimed quantities and then solve the system. As a result, we obtain the expressions (13.24). We leave it for the reader to derive them by way of an exercise.

GALILEAN TRANSFORMATIONS AS LIMITING CASE OF LORENTZ TRANSFORMATIONS. At limiting velocities that are much lower than the velocity of light, quantities of the order of $v/c \ll 1$ in Lorentz transformations can be neglected in comparison with the unity. In other words, all quantities v/c in Lorentz transformations can be put equal to zero. These transformations then assume the form of Galilean transformations (11.1). At lower velocities, the difference between Galilean and Lorentz transformations is insignificant, and hence the error in Galilean transformations remained unnoticed for a long time.

!

It is insufficient to obtain Lorentz transformations only by means of the principles of relativity and the constancy of the velocity of light. The homogeneity and isotropy of space and the homogeneity of time must also be taken into consideration.

Lorentz transformations can also be achieved on the basis of other requirements, say, the requirement of the invariance of Maxwell's equations relative to linear transformations of spatial coordinates and time.

FOUR-DIMENSIONAL VECTORS. Four-dimensional vectors are defined in terms of the Lorentz transformations (13.23) which can be written in a form analogous to (6.20) with the help of the notation: $x_1 = x$, $x_2 = y$, $x_3 = z$ and $x_4 = ict$:

$$\begin{aligned}
x'_1 &= \frac{1}{\sqrt{1 - v^2/c^2}} x_1 + 0 \cdot x_2 + 0 \cdot x_3 + \frac{iv/c}{\sqrt{1 - v^2/c^2}} x_4, \\
x'_2 &= 0 \cdot x_1 + 1 \cdot x_2 + 0 \cdot x_3 + 0 \cdot x_4, \\
x'_3 &= 0 \cdot x_1 + 0 \cdot x_2 + 1 \cdot x_3 + 0 \cdot x_4, \\
x'_4 &= \frac{-iv/c}{\sqrt{1 - v^2/c^2}} x_1 + 0 \cdot x_2 + 0 \cdot x_3 + \frac{1}{\sqrt{1 - v^2/c^2}} x_4.
\end{aligned} \tag{13.25}$$

Consequently, the projections of the four-dimensional vector $\{A_1, A_2, A_3, A_4\}$ must be transformed in accordance with the formula

$$A'_\alpha = \sum_{\gamma=1}^{4} a_{\alpha\gamma} A_\gamma, \tag{13.26}$$

where the $a_{\alpha\gamma}$ coefficients are defined by the transformation matrices (13.25):

$$a_{\alpha\gamma} = \begin{vmatrix} \dfrac{1}{\sqrt{1 - v^2/c^2}} & 0 & 0 & \dfrac{iv/c}{\sqrt{1 - v^2/c^2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \dfrac{-iv/c}{\sqrt{1 - v^2/c^2}} & 0 & 0 & \dfrac{1}{\sqrt{1 - v^2/c^2}} \end{vmatrix}. \tag{13.27}$$

It can easily be proved through direct substitution that the four-dimensional velocity is characterized by the following projections:

$$\begin{aligned} u_1 &= \frac{u_x}{\sqrt{1 - u^2/c^2}}, & u_2 &= \frac{u_y}{\sqrt{1 - u^2/c^2}}, \\ u_3 &= \frac{u_z}{\sqrt{1 - u^2/c^2}}, & u_4 &= \frac{ic}{\sqrt{1 - u^2/c^2}}, \end{aligned} \tag{13.28}$$

where u_x, u_y and u_z are the projections of the three-dimensional velocity, and $u^2 = u_x^2 + u_y^2 + u_z^2$. In this book, we shall not use the four-dimensional notation to present the special theory of relativity. This example has been included only to show how the definition of a physical vector given in Sec. 6 can be used to analyze the vector nature of physical quantities in four-dimensional space-time.

If a set of four quantities forms a four-dimensional vector, they can be transformed from one coordinate system to another in accordance with (13.26), where the $a_{\alpha\gamma}$ coefficients are defined by the matrix (13.27). It can easily be verified by direct inspection that the $a_{\alpha\gamma}$ coefficients satisfy the following important relations:

$$\begin{aligned} &\sum_{\alpha=1}^{4} a_{\alpha\gamma} a_{\alpha\beta} = \delta_{\alpha\beta} = \begin{cases} 0 & \text{for } \gamma \neq \beta, \\ 1 & \text{for } \gamma = \beta; \end{cases} \\ &\sum_{\alpha=1}^{4} a_{\gamma\alpha} a_{\beta\alpha} = \delta_{\gamma\beta} = \begin{cases} 0 & \text{for } \gamma \neq \beta, \\ 1 & \text{for } \gamma = \beta. \end{cases} \end{aligned} \tag{13.29}$$

Coordinate transformations with $a_{\alpha\gamma}$ coefficients satisfying the relations (13.29) are called orthogonal transformations. The most important property of such transformations is that they do not change the value of the square of the four-dimensional vector which is defined as the sum of the squares of the projections of the vector:

$$\sum_{\alpha=1}^{4} A_\alpha^2. \tag{13.30}$$

In order to prove the statement, we express the square of the four-dimensional vector in a K' system in terms of its projections in another K system:

$$\sum_{\alpha=1}^{4} A_\alpha'^2 = \sum_{\alpha,\gamma,\mu} a_{\alpha\gamma} A_\gamma a_{\alpha\mu} A_\mu$$
$$= \sum_{\gamma,\mu} A_\gamma A_\mu \sum_{\alpha} a_{\alpha\gamma} a_{\alpha\mu} = \sum_{\gamma,\mu} A_\gamma A_\mu \delta_{\gamma\mu} = \sum_{\gamma} A_\gamma^2. \tag{13.31}$$

This means that the square of the four-dimensional vector is an invariant of the transformations (13.26):

$$\sum_{\alpha} A_\alpha^2 = \text{inv}. \tag{13.32}$$

Let us consider an example in which four-dimensional vectors can be used to solve some problems. In mechanics, considerable importance is attached to the energy-momentum vector of a particle whose projections are obtained as follows:

$$(p_1, p_2, p_3, p_4) = \left(p_x, p_y, p_z, \frac{iE}{c}\right), \tag{13.33}$$

where p_x, p_y, p_z are the projections of the three-dimensional momenta, $i = \sqrt{-1}$ is the imaginary unit, and E is the total energy of the particle.

Using (13.33), we obtain the following formulas from (13.26):

$$p_x' = \frac{p_x - vE/c^2}{\sqrt{1 - v^2/c^2}}, \quad p_y' = p_y, \quad p_z' = p_z,$$
$$E' = \frac{E - vp_x}{\sqrt{1 - v^2/c^2}}. \tag{13.34}$$

For the four-dimensional energy-momentum vector (13.33), the relation (13.32) assumes the form

$$p^2 - \frac{E^2}{c^2} = \text{inv}. \tag{13.35}$$

The value of this invariant can conveniently be calculated for a particle at rest, when $p = 0$, and $E = E_0$ is the energy of the particle with zero momentum. It will be shown later that $E_0 = m_0 c^2$, where m_0 is the rest mass of the particle. From (13.32) we obtain

$$\text{inv} = -\frac{E_0^2}{c^2} = -m_0^2 c^2. \tag{13.36}$$

Equation (13.35) then assumes the form $p^2 - E^2/c^2 = -m_0^2 c^2$. This leads to the following important formula connecting the total energy of the particle with its momentum:

$$E = c\sqrt{p^2 + m_0^2 c^2}. \tag{13.37}$$

Thus, equations (13.34) and (13.37) are corollaries of the statement that the set of quantities (13.33) forms a four-dimensional vector without any additional assumptions being made about these quantities.

PROBLEMS

3.1. An aeroplane flies north a distance l for a time t_1, and flies for a time t_2 on its return because of a north-easterly wind. Find the velocity of the aeroplane when there is no wind, and the velocity of the wind.

3.2. An aeroplane flying at a velocity v has a range l of flight when there is no wind. What will its range be if it flies in a head wind of velocity u directed at an angle φ to the plane of the trajectory relative to the Earth?

ANSWERS

3.1. $l(1/t_1^2 + 1/t_2^2)^{1/2}/\sqrt{2}$, $l(1/t_2 - 1/t_1)/\sqrt{2}$. **3.2.** $l(v^2 - u^2)/[v(v^2 - u^2 \sin^2\varphi)^{1/2}]$.

Chapter **4**

Corollaries of Lorentz Transformations

Basic idea:
The experimental verification of the results of Lorentz transformations serves as the verification of the basic principles of the special theory of relativity.

Sec. 14. RELATIVITY OF SIMULTANEITY

The relativity of simultaneity and the invariance of an interval are discussed.

DEFINITION. Two events occurring at different points x_1 and x_2 of a coordinate system are said to be simultaneous if they occur at the same instants of time *according to the clock belonging to the coordinate system. The instants at which these events occur at each point are registered by two clocks located at each point.* We shall assume that two events occur simultaneously at points x_1 and x_2 in a stationary coordinate system at the instant t_0.

The same events occur at points x'_1 and x'_2 in a moving coordinate system at the instants t'_1 and t'_2, where t'_1 and t'_2 are the readings on two clocks in the moving coordinate system located at points x'_1 and x'_2. The primed and unprimed quantities are related via the Lorentz transformations (13.23):

$$x'_1 = \frac{x_1 - vt_0}{\sqrt{1 - v^2/c^2}}, \quad x'_2 = \frac{x_2 - vt_0}{\sqrt{1 - v^2/c^2}},$$
$$t'_1 = \frac{t_0 - (v/c^2)x_1}{\sqrt{1 - v^2/c^2}}, \quad t'_2 = \frac{t_0 - (v/c^2)x_2}{\sqrt{1 - v^2/c^2}}. \tag{14.1}$$

Since the events occur at points along the X-axis, the y- and z-coordinates are zero in both systems. It can be seen from (14.1) that these events do not occur simultaneously (i.e. $t'_2 \neq t'_1$) in the moving coordinate system and are separated by

Albert Einstein (1879-1955)
One of the founders of modern physics, Einstein was born in Germany. He lived in Switzerland from 1893, in Germany from 1914, and migrated to the USA in 1933. He was one of the founders of the special theory of relativity and also created the general theory of relativity. He wrote a number of fundamental articles on the quantum theory of light (introduced the concept of photon, the photoelectric effect, predicted induced radiation). He developed the molecular-statistical theory of Brownian movement and quantum statistics.

a time interval given by

$$\Delta t' = t'_2 - t'_1 = \frac{(v/c^2)(x_1 - x_2)}{\sqrt{1 - v^2/c^2}}. \tag{14.2}$$

Thus, events that occur simultaneously in one coordinate system are not simultaneous in another system.

The concept of simultaneity is therefore not absolute in the sense that it depends on the coordinate system. To impart meaning to the statement that two or more events are simultaneous, we must indicate the coordinate system to which the statement refers.

The problem concerning the relativity of simultaneity and the physical meaning of the Lorentz transformations was solved by Einstein.

The relativity of simultaneity can also be demonstrated as follows. The readings on two clocks located at different points along the X-axis in a stationary coordinate system (Fig. 35) are compared with the readings on two clocks located at different points along the X'-axis in a coordinate system moving at a velocity v. Figure 35 shows the time at different points in the moving coordinate system corresponding to the instant $t = 0$ in the stationary coordinate system.

RELATIVITY OF SIMULTANEITY AND CAUSALITY. It can be seen from (14.2) that for $x_1 > x_2$, the inequality $t'_2 > t'_1$ is satisfied in the coordinate system moving in the positive X-direction ($v > 0$), while $t'_2 < t'_1$ in the coordinate system moving in the opposite direction ($v < 0$). Thus the consequence in which events occur in different coordinate systems is different. One might ask whether it is possible for a cause to precede its effect in one system, while in another system, on the contrary, for an effect to precede its cause. It is obvious that such a situation could not be admissible in a theory which recognizes the objective role of the cause-and-effect sequence in the world: a cause and its effect cannot be interchanged simply by looking at the events from a different point of view.

In order to make the cause-and-effect sequence objective and independent of the coordinate system in which it is considered, it is essential that no action which establishes a physical connection between two events occurring at different points in space can take place at velocities exceeding that of light.

In order to prove this, let us consider two events taking place in a stationary coordinate system. Suppose that an event occurring at point x_1 at an instant t_1 is responsible for an event occurring at point $x_2 > x_1$ at an instant $t_2 > t_1$. The velocity at which the "influence" of the action at point x_1 is perceived at point x_2 is denoted by v_{inf}. Obviously, we have

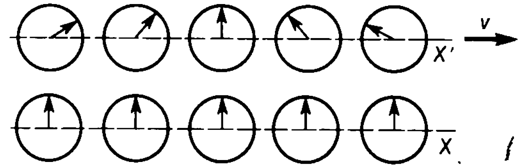


Fig. 35. Relativity of simultaneity.
In the stationary coordinate system, the clocks located at different points where some events are simultaneous show that the events are completed at the same time. In the moving coordinate system, the clocks show that some events are completed at different times, i.e. the events are not simultaneous there.

from the definition of velocity

$$\frac{x_2 - x_1}{v_{\text{inf}}} = t_2 - t_1. \tag{14.3}$$

These events occur at two points x'_1 and x'_2 in a moving coordinate system at instants t'_1 and t'_2. According to (13.22), we have

$$t'_2 - t'_1 = \frac{t_2 - t_1 - (v/c^2)(x_2 - x_1)}{\sqrt{1 - v^2/c^2}} = \frac{t_2 - t_1}{\sqrt{1 - v^2/c^2}}\left(1 - \frac{v}{c^2}v_{\text{inf}}\right), \tag{14.4}$$

where the difference $x_2 - x_1$ has been eliminated from the last equation using (14.3). Equation (14.4) shows that if

$$1 - \frac{v}{c^2}v_{\text{inf}} < 0, \tag{14.5}$$

then an effect would precede its cause in the moving coordinate system, which is impossible. Hence we must always have $1 - (vv_{\text{inf}}/c^2) > 0$, or

$$v_{\text{inf}} < \frac{c}{v}c. \tag{14.6}$$

Since the Lorentz transformations allow values of v which can approach c but do not exceed it (otherwise the transformations are no longer real), (14.6) can be written as:

$$\boxed{v_{\text{inf}} \leqslant c.} \tag{14.7}$$

Thus, a physical effect due to an event at one point cannot propagate to another point at a velocity exceeding that of light. Under this condition, the causality of events is absolute in nature: there is no coordinate system in which a cause and its effect can be interchanged.

INTERVAL INVARIANCE. The significance of transformation invariants for a theory was emphasized in Sec. 11. The invariants of Galilean transformations are the linear dimensions of a body and the interval of time between events. This is why the concepts of length and time interval play such an important role in classical physics.

However, neither length nor the time interval between events are invariants of Lorentz transformations. This means that they depend on the coordinate system. We shall consider this question in greater detail in later sections. For the present, we shall simply mention this fact in order to analyze an important invariant of a Lorentz transformation, viz. the space-time interval, or just interval.

Suppose that two events occur at points (x_1, y_1, z_1) and (x_2, y_2, z_2) at instants t_1 and t_2 respectively. The interval between these events or, in brief, the interval between points (x_1, y_1, z_1, t_1) and (x_2, y_2, z_2, t_2) is the quantity s whose square is given by the formula

$$s^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2. \tag{14.8}$$

This quantity has the same value in all coordinate systems and is thus an invariant of Lorentz transformations. In order to verify this, let us transform (14.8) in the K' coordinate system using (13.24). This gives

$$x_2 - x_1 = \frac{x'_2 - x'_1 + v(t'_2 - t'_1)}{\sqrt{1 - v^2/c^2}},$$

$$y_2 - y_1 = y'_2 - y'_1,$$

$$z_2 - z_1 = z'_2 - z'_1,$$

$$t_2 - t_1 = \frac{t'_2 - t'_1 + (v/c^2)(x'_2 - x'_1)}{\sqrt{1 - v^2/c^2}}.$$

Substituting these expressions into (14.8), we obtain

$$\begin{aligned} s^2 &= (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 - c^2(t_2 - t_1)^2 \\ &= (x'_2 - x'_1)^2 + (y'_2 - y'_1)^2 + (z'_2 - z'_1)^2 - c^2(t'_2 - t'_1)^2 = s'^2. \end{aligned} \tag{14.9}$$

This proves that the square of the interval is an invariant: $s^2 = s'^2 = \text{inv}$.

If the points under consideration are infinitesimally close to each other, Eq. (14.9) proves the invariance of the square of the differential of the interval:

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 - c^2\,\mathrm{d}t^2 = \text{inv}. \tag{14.10}$$

!

The theory of relativity does not prove the causality principle. The theory is based on the assumption that the causality principle holds and must be valid in all coordinate systems. This imposes a restriction on the velocity at which physical action can be transmitted. If the interval between events is spatially similar, we can choose a coordinate system in which the events occur simultaneously at different points in space.
If the interval between events is time similar, we can choose a coordinate system in which the events occur successively at the same point in space.

SPATIALLY SIMILAR AND TIME-SIMILAR INTERVALS. Let us denote the spatial distance between two events by l, and the interval of time between them by t. The square $s^2 = l^2 - c^2t^2$ of the interval between these events is an invariant.

Suppose that two events cannot causally be related in a coordinate system. For such events, $l > ct$ and hence $s^2 > 0$. It follows from the invariance of the interval that these events cannot causally be related in any other coordinate system. The converse is also true: if two events are causally related in a coordinate system ($l < ct$ and $s^2 < 0$), they are causally related in all other coordinate systems.

The interval for which

$$s^2 > 0 \tag{14.11}$$

is called a spatially similar interval, while the interval for which

$$s^2 < 0 \tag{14.12}$$

is called a time-similar interval.

The interval for which

$$s^2 = 0 \tag{14.13}$$

is called the zero interval. *Such an interval occurs between events that are connected via a signal that propagates at the velocity of light.*

If the interval between two events is spatially similar, we can choose a coordinate system in which the events occur at different points in space at the same instant of time ($s^2 = l^2 > 0$ and $t = 0$), and there is no coordinate system in which these events could occur at the same point (in this case, we would have $l = 0$, i.e. $s^2 = -c^2t^2 < 0$, which contradicts the condition $s^2 > 0$).

If the interval between two events is time similar, we can choose a coordinate system in which the events can occur in succession at the same point in space ($l = 0$ and $s^2 = -c^2t^2 < 0$), and there is no coordinate system in which these events could occur simultaneously (in this case, we would have $t = 0$, i.e. $s^2 = l^2 > 0$, which contradicts the condition $s^2 < 0$).

Thus, for events that are causally related, we can always choose a coordinate system in which these events occur at the same point in space at successive instants of time.

Spatial similarity, time similarity or the fact that a zero interval occurs between two events are independent of the coordinate system. This is just the property of invariance of the events themselves.

Example 14.1. Find a coordinate system in which two events separated by a spatially similar interval take place simultaneously at different points, while two events separated by a time-similar interval occur successively at the same point in space.

If the events occur at point $x_1 = 0$ at $t_1 = 0$ and at point x_2 at $t = t_2$, and if the interval between them is spatially similar ($s^2 = x_2^2 - c^2t^2 > 0$), the spatial separation between the events in a coordinate system in which the events occur simultaneously will be given by $x_2' = \sqrt{s^2}$. The velocity of this coordinate system is obtained from the Lorentz transformations. Assuming that the origins coincide at the instant of occurrence of the first event (i.e. assuming that $x_1' = 0$ and $t_1' = 0$), we can write the following expression for the second event:

$$t_2' = 0 = \frac{t_2 - x_2 v/c^2}{\sqrt{1 - v^2/c^2}}. \tag{14.14}$$

This gives $v = t_2 c^2/x_2$.

If the two events are separated by a time-similar interval ($s^2 = x_2^2 - c^2t^2 < 0$), the time interval between the events in a coordinate system in which they occur successively at the same point in space is $t_2' = \sqrt{-s^2}/c$. In order to determine the coordinate system, we can write

$$x_2' = 0 = \frac{x_2 - vt_2}{\sqrt{1 - v^2/c^2}}, \tag{14.15}$$

which gives $v = x_2/t_2$.

Sec. 15. LENGTH OF A MOVING BODY

The relativistic contraction of the length of a moving body and its reality are discussed.

DEFINITION OF THE LENGTH OF A MOVING BODY. The length of a moving rod is the distance between the points in a stationary coordinate system, which coincide with the ends of the moving rod at a certain instant of time indicated by the clock located in the stationary coordinate system. Thus, the ends of the moving rod are recorded at the same instant of time in the stationary coordinate system. Two clocks in the moving coordinate system, which are located at the ends of the rod when the ends are recorded, will show different times, as can be seen from Fig. 35. In other words, the ends of the rod are not simultaneously recorded in the moving coordinate system. *This means that the length of the rod is not an invariant of the Lorentz transformations and has different values in different coordinate systems.*

FORMULA FOR THE REDUCTION IN THE LENGTH OF A MOVING BODY. Let a rod of length l be at rest along the X'-axis in the K' coordinate system. It should be noted that when we speak of a body of a certain length, we always mean the length of the body at rest. We denote the coordinates of the ends of the rod by x'_1 and x'_2. Then, by definition, $x'_2 - x'_1 = l$. The quantity l is unprimed in this relation because it corresponds to the length of the rod in the coordinate system in which it is at rest. In other words, l corresponds to the length of the stationary rod.

Let us mark the position of the ends of the rod moving at a velocity v in a K coordinate system at the instant t_0. Using the Lorentz transformation formulas, we can write

$$x'_1 = \frac{x_1 - vt_0}{\sqrt{1 - v^2/c^2}}, \quad x'_2 = \frac{x_2 - vt_0}{\sqrt{1 - v^2/c^2}}, \tag{15.1}$$

whence

$$l = x'_2 - x'_1 = \frac{x_2 - x_1}{\sqrt{1 - v^2/c^2}} = \frac{l'}{\sqrt{1 - v^2/c^2}}, \tag{15.2}$$

where $l' = (x_2 - x_1)$ is the length of the moving rod. Writing (15.2) in the form

$$\boxed{l' = l\sqrt{1 - v^2/c^2},} \tag{15.3}$$

we note that *the length of the moving rod in the direction of motion of the coordinate system is smaller than the length of the stationary rod.* Naturally, if this argument is repeated by assuming that the K' coordinate system is the stationary system, we obtain (15.3) again for the reduction in the length of the moving rod in accordance with the requirements of the relativity principle.

If the rod is perpendicular to the direction of motion, i. e. if it lies along the Y'- or Z'-axis, equations (13.23) show that the length of the rod remains unchanged. *Thus, the size of a body does not change in a direction perpendicular to the direction of motion.*

CHANGE IN THE SHAPE OF A MOVING BODY. Since the length of a body is reduced in the direction of motion in accordance with (15.3), although its dimensions in the perpendicular direction remain unchanged, the shape of a moving body changes, and the body is "flattened" in the direction of motion.

Let us clarify the physical meaning of this change in the shape of a moving body. If the moving body is observed with an ordinary optical instrument, in, say, visible light, it will not appear to be flattened as predicted by the Lorentz transformations.

The change in the shape of a moving body has the following physical meaning. The coordinates of all the points on the surface of the moving body are marked in a stationary coordinate system at a certain instant of time. Thus, an instantaneous "mould" of the moving body is obtained at a certain instant of time. The shape of the "mould" which is at rest in the stationary coordinate system is taken as the shape of the moving object. The "mould" would not have the same shape as the body if the latter were at rest. In fact, the "mould" is found to be flatter than its original at rest.

In this sense, the effect of flattening of moving bodies is a real effect.

The situation is quite different if the shape of the moving body is visually observed. Two factors are responsible for this changed situation: firstly, rays from different parts of the body arrive at the observer at different instants of time; secondly, aberration takes place changing the apparent direction from which the rays arrive at the observer. It has been shown by calculation that as a result of these circumstances, the shape of the body being visually observed does not coincide with the shape obtained as a result of the Lorentz transformations.

ESTIMATION OF THE MAGNITUDE OF CONTRACTION. The velocity of any body is usually much smaller than the velocity of light, i. e. $v/c \ll 1$. Hence (15.3) can be represented in alternative form with an error not exceeding the first-order terms in v^2/c^2:

$$l' \approx l\left(1 - \frac{1}{2}\frac{v^2}{c^2}\right). \tag{15.4}$$

Consequently, the relative magnitude of the reduced length is

$$\frac{\Delta l}{l} = \frac{l' - l}{l} = -\frac{1}{2}\frac{v^2}{c^2}. \tag{15.5}$$

For velocities of the order of tens of kilometres per second, $v^2/c^2 \simeq 10^{-8}$, and hence the relative decrease in length is less than 10^{-8}. Such a small change is not easily observed. For example, a body whose length is 1 m will contract by less than 10^{-6} cm. The Earth's diameter is slightly more than 12,000 km. The Earth revolves around the Sun at a velocity $v = 30$ km/s, and this reduced its diameter by just about 6 cm in the reference frame fixed to the Sun. On the other hand, this reduction is significant at higher velocities. For example, if a body has a velocity of about $0.85c$, its length is reduced by half. If a body has a velocity close to the velocity of light, its length becomes very small.

!

The contraction of moving bodies and the deformation of their shape are real phenomena since they lead to observable physical consequences.
While considering the accelerated motion of bodies in the theory of relativity, we cannot use the concept of their perfect rigidity.

ON THE REALITY OF CONTRACTION OF A MOVING BODY. The following question may be asked in connection

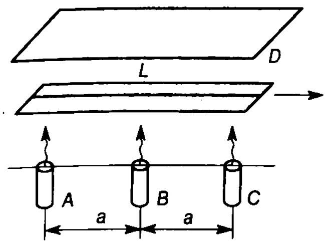


Fig. 36. Diagram of the experiment demonstrating how scales are contracted.

with what has been stated above about the shape of a moving body and the shape observed in light rays: Is the contraction of the body real, and if so, what does it mean? One can answer as follows: The change in the shape of the moving body is real since its physical consequences can actually be observed. In order to verify this, let us consider the following physical situation.

Three stationary sources of light (for example, lasers) A, B and C are situated on the same straight line at a distance a apart (Fig. 36). These sources can simultaneously emit short pulses of light in a direction perpendicular to the straight line on which they are situated. These pulses are recorded on a photographic plate D. A body (say, a ruler) can move between the light sources and the plane of the photographic plate in a direction parallel to AC. If the ruler intercepts the light pulse from a source to the photographic plate, the light ray will not be recorded on the plate.

Case 1. Suppose that a stationary ruler has a length $L < 2a$ ($L > a$). In this case, it can obstruct either one of the three sources A, B or C, or two of them, viz. A and B or B and C, depending on its position between the light sources and the photographic plate. It may so happen that it may screen the source B, while the sources A and C remain unobstructed. If all the three sources of light flash simultaneously for different positions of the ruler, the following combinations of spots will be observed on the photographic plate: (a) all the three spots are recorded (the ruler does not lie inside the segment AC); (b) any two of the three spots are recorded (the ruler obstructs one of the sources), this also includes the case when the source B is covered by the screen and hence only A and C are recorded; (c) either of the spots from A or C is observed, while the other two are obstructed by the ruler. The situation in which A and C are covered, while B is recorded, is not possible in practice.

Case 2. Let us consider a longer ruler for which $L > 2a$. Moving this ruler between the light sources and the photographic plate and recording the spots on the plate when all the three sources flash at the same time, we may observe the following combinations of spots: (a) all the three spots are recorded (in this case, the ruler does not lie inside the segment AC); (b) two spots (from A, B or B, C) are observed (the ruler screens one of the extreme sources). The situation in which A and C are observed, while B is covered, is not possible, i. e. it is impossible to obtain a photograph in which spots from A and C are observed, but there is no spot from B; note that such a situation would be possible if the ruler were smaller (as in Case 1); (c) only one spot from A or C is recorded, while the remaining two are covered by the ruler. The situation in which

A and C are covered and B leaves a spot on the photographic plate is impossible; (d) no ray is recorded on the photographic plate. In this case, all the three sources are covered by the ruler. Note that such a situation would be impossible if the ruler were smaller (as in Case 1).

Case 3. Let us now move the same ruler ($L > 2a$ long) between the light sources and the photographic plate parallel to AC at such a velocity v that $L\sqrt{1 - v^2/c^2} < 2a$ and record the flashes from the sources on the photographic plate. In this case, we obtain all the possible combinations of spots as in Case 1 when a smaller ruler was used. Among other combinations, we also encounter the situation in which two spots from A and C are recorded, while there is no spot from B, a situation which was impossible in Case 2. On the other hand, we shall never observe a photograph in which one of the three spots is not recorded, a situation which was possible in Case 2. Hence we conclude that this case is analogous to Case 1, i.e. the length of the moving ruler is smaller than $2a$. This means that there is as much meaning in the statement that the length of the moving ruler has been reduced to below $2a$ as in the statement that the length of the stationary ruler in the first case is smaller than $2a$. Consequently, the contraction of moving bodies is real and not a virtual phenomenon. It leads to tangible results whose reality is beyond doubt, as can be seen in the examples considered above.

Let us now analyze the same phenomena in a reference frame associated with the moving ruler (case $L > 2a$). In such a reference frame, the light sources and the photographic plate move in the negative direction at a velocity $-v$. In view of contraction, the distance between the sources A and C will be $2a\sqrt{1 - v^2/c^2}$, i.e. the sources will be separated by a distance much smaller than the length L of the stationary ruler. In spite of this, pulses from the sources A and C can by-pass the ruler and leave spots on the photographic plate. This is due to the relativity of simultaneity. Flashes from sources which are simultaneous in a reference frame in which they are at rest will not be simultaneous in a reference frame in which they are moving. Hence the flashes from the sources A, B and C do not occur simultaneously in the reference frame associated with the ruler. In the situation under consideration, the flash from the source C occurs earlier by an interval $\Delta t' = (2av/c^2)/\sqrt{1 - v^2/c^2}$. During this time, the light sources cover a distance $\Delta t'v$, and hence the second spot will be recorded at a distance

$$\frac{2av^2}{c^2\sqrt{1 - v^2/c^2}} + 2a\sqrt{1 - v^2/c^2} = \frac{2a}{\sqrt{1 - v^2/c^2}} \tag{15.6}$$

?

Are the definitions of length of moving bodies in classical mechanics and in the theory of relativity different?

Owing to what factors is the shape of moving bodies observed in light rays not so "flattened" as it follows directly from Lorentz transformations?

What is the physical meaning of the statement that the contraction of moving bodies is real?

from the first spot. Here, we have considered that the distance between the moving sources is $2a\sqrt{1 - v^2/c^2}$. At an appropriate velocity, when

$$\frac{2a}{\sqrt{1 - v^2/c^2}} > L, \qquad (15.7)$$

the pulse of light from the source A will also by-pass the ruler L and leave a spot on the photographic plate. Hence, in order to explain the observed phenomenon in the reference frame associated with the moving ruler, we must take into account both the contraction of the moving body and the relativity of simultaneity.

ON CONTRACTION AND PERFECT RIGIDITY OF A BODY. Let us consider two stationary isolated point masses x_1 and x_2 ($x_2 > x_1$) which are not connected to each other. The distance between the point masses is $l = x_2 - x_1$. Let us assume that the point masses are accelerated for a certain period according to the same law in the positive x-direction. This means that the velocities of the two point masses will be equal at each instant of time, and hence the paths covered by them from the starting point will also be the same. This leads to the conclusion that the distance between the moving point masses will remain equal to l during the course of their acceleration. However, the situation will be different from the point of view of observers moving with an acceleration and stationary relative to the point masses. If they could somehow measure the distance (say, using a light signal), they would observe that the distance between the point masses had increased as a result of the acceleration. Let us suppose that the acceleration is discontinued when the point masses attain a velocity v in a stationary reference frame, and that the point masses move uniformly at this constant velocity v. In this case, if the observers stationary relative to the point masses measured the distance between them, they would conclude that the distance between them had increased to $l/\sqrt{1 - v^2/c^2}$. However, this result can easily be explained. For a stationary observer, both point masses are synchronously accelerated and have the same velocity and acceleration at each instant of time. For the observers stationary relative to the point masses, however, the motion will not be synchronous due to the relativity of simultaneity. In the present case, they will find that the acceleration of the point mass x_2 precedes that of the point mass x_1, and hence the point mass x_2 moves farther than the point mass x_1.

Let us now suppose that the point masses are connected through a weightless spring and are identically accelerated in a stationary reference frame. It is clear from what has been stated above that to the observers stationary relative to the point masses x_1 and x_2 the spring will be stretched during the acceleration, and the potential energy of deformation will be accumulated in the spring. This energy is supplied by the engines which power the acceleration of the point masses. If the point masses are connected by a perfectly rigid rod, an infinite amount of energy is required to create an infinitesimal deformation. Obviously, it is impossible to describe the acceleration of both point masses through the same law. This leads to the conclusion that *when the acceleration of bodies is being considered, we cannot assume that they are perfectly rigid.*

Example 15.1. A rod of length l lies at an angle α' to the X'-axis in a K' coordinate system moving at a velocity v. Find the length l' of the rod and the angle α which it forms with the X-axis in a stationary coordinate system.

Denoting the coordinates of the ends of the rod in the K' system by (x'_1, y'_1) and (x'_2, y'_2), we can write

$$\begin{aligned} l_x &= x'_2 - x'_1 = l\cos\alpha', \\ l_y &= y'_2 - y'_1 = l\sin\alpha'. \end{aligned} \tag{15.8}$$

The coordinates (x_1, y_1) and (x_2, y_2) corresponding to the points where the two ends of the moving rod are at the same instant in the K system satisfy the following relations in accordance with Lorentz transformations:

$$\begin{aligned} x'_2 - x'_1 &= \frac{x_2 - x_1}{\sqrt{1 - v^2/c^2}}, \\ y'_2 - y'_1 &= y_2 - y_1. \end{aligned} \tag{15.9}$$

By definition, the length l' of the moving rod is $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. As a result, we can write the following expressions for l' in the stationary coordinate system:

$$\begin{aligned} l' &= \sqrt{l^2\cos^2\alpha'(1 - \beta^2) + l^2\sin^2\alpha'}, \\ \tan\alpha &= \frac{y_2 - y_1}{x_2 - x_1} = \frac{\tan\alpha'}{\sqrt{1 - \beta^2}}, \quad \beta = \frac{v}{c}. \end{aligned} \tag{15.10}$$

Sec. 16. PACE OF MOVING CLOCKS. INTRINSIC TIME

The slowing down of the pace of moving clocks and the twin paradox are discussed.

SLOWING DOWN OF THE PACE OF MOVING CLOCKS. Suppose that two events occur successively at point x'_0 in a moving K' coordinate system at the instants t'_1 and t'_2. The events occur at different points in a stationary K coordinate system at the instants t_1 and t_2. The time interval between the events in the moving coordinate system is $\Delta t' = t'_2 - t'_1$, while the time interval in the stationary coordinate system is $\Delta t = t_2 - t_1$.

On the basis of the Lorentz transformations, we can write

$$t_1 = \frac{t'_1 + (v/c^2)x'_0}{\sqrt{1 - v^2/c^2}}, \quad t_2 = \frac{t'_2 + (v/c^2)x'_0}{\sqrt{1 - v^2/c^2}}. \tag{16.1}$$

Consequently, we obtain

$$\Delta t = t_2 - t_1 = \frac{t'_2 - t'_1}{\sqrt{1 - v^2/c^2}} = \frac{\Delta t'}{\sqrt{1 - v^2/c^2}}. \tag{16.2}$$

Thus, the time interval

$$\boxed{\Delta t' = \Delta t\sqrt{1 - v^2/c^2}} \tag{16.3}$$

between the events measured by the moving clocks will be less than the time interval Δt between the same events measured by the stationary clocks. *This means that the pace of moving clocks is slower than that of stationary clocks.*

It may appear that this is in contradiction with the relativity principle since the moving clocks can be assumed to be stationary. However, any contradiction is only apparent. The time recorded at the same moving point is compared in (16.3) with the times at different stationary points. Hence, in order to apply the relativity principle, we must compare the times at the different points of a moving coordinate system with the time at the same point in a stationary coordinate system.

Let us make this comparison. Suppose that two events occurred successively at a certain point in a K coordinate system, say, point x_0 along the X-axis, at the instants t_1 and t_2. The time interval between these two events is $\Delta t = t_2 - t_1$. In a K' coordinate system, which is assumed to be stationary, these events occurred at different points at the instants t'_1 and t'_2. In accordance with (13.23), we have

$$t'_1 = \frac{t_1 - (v/c^2)x_0}{\sqrt{1 - v^2/c^2}}, \quad t'_2 = \frac{t_2 - (v/c^2)x_0}{\sqrt{1 - v^2/c^2}}. \tag{16.4}$$

Consequently,

$$\Delta t' = t'_2 - t'_1 = \frac{t_2 - t_1}{\sqrt{1 - v^2/c^2}} = \frac{\Delta t}{\sqrt{1 - v^2/c^2}}. \tag{16.5}$$

However, $\Delta t'$ is now the time interval between the events in the stationary coordinate system, while Δt is the time interval between the same events in the moving coordinate system. Thus, formula (16.5) has the same meaning as formula (16.2), and there is no contradiction with the relativity principle.

INTRINSIC TIME. Time measured with a clock associated with a moving point is called the intrinsic time of the point. For an infinitesimal time interval in (16.3), we can write

$$d\tau = dt\sqrt{1 - v^2/c^2}, \tag{16.6}$$

where $d\tau$ is the differential of the intrinsic time of the moving point, and dt is the differential of time in the inertial coordinate system in which the point has a velocity v at that instant. *It should be noted that* $d\tau$ *is the change in the reading of the same clock associated with the moving point, while* dt *is the difference between the readings of different clocks situated at adjacent points in the stationary coordinate system.*

It was shown in Sec. 14 (formula (14.10)) that the differential of the interval is an invariant. Since $dx^2 + dy^2 + dz^2 = d\mathrm{r}^2$ is the square of the differential of the distance between two adjacent points in space, we can transform (14.10) for the square of the differential of the interval as follows:

$$\frac{ds}{i} = c\,dt\sqrt{1 - \frac{1}{c^2}\left(\frac{d\mathrm{r}}{dt}\right)^2} = c\,dt\sqrt{1 - \frac{v^2}{c^2}}. \tag{16.7}$$

For the events between which the interval is calculated, we have taken two consecutive positions of the moving point and have considered that $(d\mathrm{r}/dt)^2 = v^2$ is the square of its velocity. The imaginary unit $i = \sqrt{-1}$ in this formula arises because $ds^2 = d\mathrm{r}^2 - c^2\,dt^2 = (-1)(c^2\,dt^2 - d\mathrm{r}^2)$. A comparison of (16.6) and (16.7) shows that the differential $d\tau$ of the intrinsic time can be expressed in terms of the differential of the interval as follows:

$$d\tau = \frac{ds}{ic}. \tag{16.8}$$

It can be seen from (14.10) that the differential of the interval is an invariant. Since the velocity of light is constant, we can conclude from (16.8) that *the intrinsic time is an invariant of Lorentz transformations.*

This is quite natural since intrinsic time is determined from the reading of the clock associated with the moving point, and the coordinate system in which the reading is taken is immaterial.

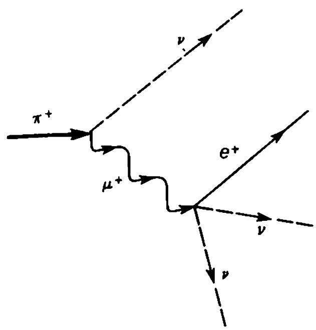


Fig. 37. Diagram of the creation and decay of a muon.

Experiments on the decay of this particle confirmed the time-dilatation formula obtained from the theory of relativity.

EXPERIMENTAL CONFIRMATION OF TIME DILATATION. Many experimental confirmations of time dilatation are known at present. One of the first confirmations was obtained during studies of muon decay. Most of the known elementary particles have very short lifetimes (10^{-6} s or less), after which an elementary particle decays and is transformed into some other particles.

The pi-meson (π-meson) is an elementary particle. There are three kinds of pi-mesons, viz. positive (π^+-mesons), negative (π^--mesons) and neutral (π^0-mesons). Positive pi-mesons decay into positive muons and neutrinos:

$$\pi^+ \to \mu^+ + \nu. \tag{16.9}$$

The neutrino is a neutral particle which weakly interacts with matter. After a certain interval of time, muons decay into positrons and two neutrinos:

$$\mu^+ \to e^+ + 2\nu. \tag{16.10}$$

The positron is a particle with a mass equal to that of an electron but with a positive charge. The creation and decay of a muon are shown in Fig. 37.

There are different ways of recording charged particles and of determining their velocities. This enables us to find the mean lifetime of a particle. If the time dilatation effect takes place, the mean lifetime of a muon must be the longer, the higher its velocity v:

$$\tau_{\mu+} = \frac{\tau_{\mu+}^{(0)}}{\sqrt{1 - v^2/c^2}}, \tag{16.11}$$

where $\tau_{\mu+}^{(0)}$ is the mean lifetime determined by a clock associated with the muon, i.e. the intrinsic lifetime, and $\tau_{\mu+}$ is the mean lifetime shown by a clock in the laboratory reference frame. In the absence of time dilatation, the dependence of the mean pathlength l on velocity is

$$l = \tau_{\mu+}^{(0)} v, \tag{16.12}$$

i.e. the pathlength is a linear function of velocity. In the case of time dilatation, we have

$$l = \tau_{\mu+}^{(0)} \frac{v}{\sqrt{1 - v^2/c^2}}, \tag{16.13}$$

which means that the pathlength is larger than in the absence of time dilatation and is not a linear function of velocity.

The study involved atmospheric muons created by cosmic rays. In this case, it was impossible to record the generation

!

Lorentz transformations are valid only for inertial reference frames. Therefore, the daily rate of a time pace cannot be analyzed in the reference frame fixed to the Earth's surface while flying around the Earth eastwards or westwards. Acceleration does not affect the daily rate of a time pace.

and decay of an individual muon, nor was it necessary to do so. The decay of a muon following the scheme in (16.10) is a random event, and the probability of the muon's decay over a path dx is dx/l. Consequently, the number density N of muons decreases over the path dx by $dN = -N\,dx/l$. Hence N decreases thus: $N(x) = N(0)\exp(-x/l)$, so that $l = x/\ln[N(0)/N(x)]$.

The problem therefore reduces to the measurement of the number density of muons at two points along their path. Of course, in this case, we must measure only muons with one particular momentum (velocity), but this can be done using methods developed in the physics of cosmic rays. As muons pass through the atmosphere, their number density decreases due to both their spontaneous decay and to absorption by atoms in the atmosphere. This muon absorption rate has been studied thoroughly and can be taken into account. Hence measuring the decrease in the number density of muons as they pass through the atmosphere, we can determine the mean path l traversed by a muon before it decays. Carrying out the measurements for muons with different momenta (velocities), we obtain the dependence of l on velocity v. This allows us to determine which of the two formulas (16.12) or (16.13) is confirmed experimentally. The studies showed that (16.13) holds with an intrinsic lifetime of $\tau_{\mu+}^{(0)} = 2\,\mu s$ for the muon.

Time dilatation plays a significant role in the operation of modern accelerators where it is frequently required to direct particles from their source to a distant target with which the particle is to interact. This would be impossible if there were no time dilatation since the time spent by the particles in covering such distances is tens and even hundreds of times longer than their intrinsic lifetimes. For example, the intrinsic lifetime of a π^+-meson is $\tau_{\pi+}^{(0)} \simeq 2.5 \times 10^{-8}$ s. After this, it decays into a muon and a neutrino. A π^+-meson can only cover a distance of $l \simeq 2.5 \times 10^{-8} \times 3 \times 10^8 \simeq 7.5$ m during its lifetime even at the velocity of light. But the π^+-meson targets are often situated some tens of metres away from the source, and still π^+-mesons safely reach such distant targets. For example, if a π^+-meson moves at a velocity differing from that of light only in the sixth decimal place (i.e. if $v/c \simeq 1\text{-}2 \times 10^{-6}$), its mean lifetime is $\tau_{\pi+} \simeq 2.5 \times 10^{-8}/\sqrt{1-(1\text{-}2 \times 10^{-6})^2} \simeq 1.25 \times 10^{-5}$ s. During this time, it can cover a distance of over 1 km and so reach the target much farther away than 7.5 m, the distance corresponding to its intrinsic lifetime.

An experiment was performed in 1972 on time dilatation by means of atomic clocks which can measure time with a very high precision. The basic idea of the experiment was to use

?

Why cannot experimental results on time dilatation with an atomic clock flying around the Earth be analyzed in a reference frame fixed to the Earth's surface?
Which factors of time dilatation should be taken into account in the experiment on orbiting of the Earth by an atomic clock?
What is the essence of the twin paradox and what is its solution?
What leads to the invariance of the differential of the intrinsic time?

three identical atomic clocks, leaving one unmoved and sending the other two by jet aeroplanes around the world, one westwards and the other eastwards. The readings of the clocks were compared after they had been returned to the starting point.

Let us consider the pace of all the clocks in a reference frame fixed to the Earth's centre. This reference frame is inertial, and we can use the formulas derived for time dilatation in this case. For the sake of simplicity, we shall assume that the round-the-world flight is accomplished at a single latitude. We denote the linear velocity of the points on the Earth's surface at the latitude by v, and the velocity of the aeroplane relative to the Earth's surface by u'. Since the Earth revolves from west to east, the velocity of the clock travelling westwards will be $v - u'$ relative to the stationary reference frame, while the velocity of the clock travelling eastwards will be $v + u'$ (Fig. 38). It should be noted that both clocks move in the same direction from east to west in the stationary reference frame since $v > u'$.

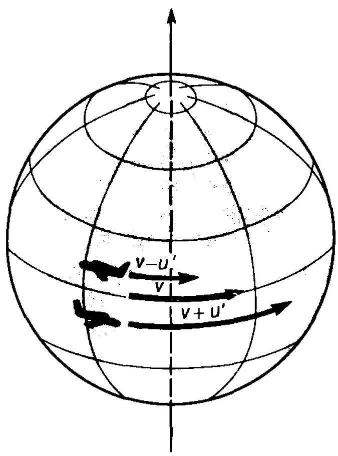


Fig. 38. Experiment for verifying time dilatation in flight around the Earth.

After a round-the-world flight, the clock mounted on an aeroplane flying eastwards is found to be slower than the clock on the Earth, while the clock on an aeroplane flying westwards is faster.

Let us denote the time shown by the clock in the stationary reference frame by t, the intrinsic time shown by the stationary clock on the Earth's surface by τ_0, the time shown by the clock travelling westwards by $\tau_{(+)}$ and the time shown by the clock travelling eastwards by $\tau_{(-)}$. Obviously, the clock which is stationary on the Earth's surface travels at a velocity v relative to the reference frame fixed to the Earth's centre and consequently is slower than the clock associated with the Earth's centre in accordance with (16.6):

$$d\tau_0 = dt\sqrt{1 - v^2/c^2}. \tag{16.14}$$

The clock travelling westwards has a lower velocity relative to the stationary reference frame and hence is slower to a smaller extent. Hence it is faster than the clock which is at rest on the Earth's surface. On the other hand, the clock travelling eastwards has a higher velocity relative to the stationary reference frame and hence is more slower than the clock which is at rest on the Earth's surface. Hence, after returning from its round-the-world flight, the clock which travelled westwards must be faster than the clock which remained stationary on the Earth's surface, while the clock which travelled eastwards must be slower than the stationary clock. By analogy with (16.14), we can write the following expressions for the clocks travelling westwards and eastwards respectively:

$$\begin{aligned} d\tau_{(+)} &= dt\sqrt{1 - (v - u')^2/c^2}, \\ d\tau_{(-)} &= dt\sqrt{1 - (v + u')^2/c^2}. \end{aligned} \tag{16.15}$$

Knowing the velocity of the aeroplane and other data, we can easily calculate the expected difference in the readings of the clocks from (16.14) and (16.15). However, this is not the only difference recorded by the clocks. As a matter of fact, it follows from the general theory of relativity that the gravitational field also affects the pace of the clocks: clocks are slowed down because of the gravitational field. A clock lifted above the Earth to a certain height by an aeroplane will slow down less than a clock left on the Earth's surface. Hence it runs faster than the clock on the Earth's surface. This difference in the times shown by the two clocks is of the same order of magnitude as the difference due to the velocities of the clocks and must be taken into consideration.

Calculations carried out for a set of experimental conditions show that the clock travelling westwards must advance by 275×10^{-9} s, and the clock travelling eastwards must slow down by 4×10^{-9} s relative to the clock left behind on the Earth's surface. The result of the experiment is found to be in good agreement with the predictions of the theory and confirms the effect of slowing down of clocks in motion.

Other experiments were also carried out with a view to compare the pace of clocks on the Earth and in an aeroplane without flying round the globe. One clock was carried in an aeroplane at a height of about 10 km for 15 hours, while an identical clock was allowed to remain on the Earth during this period. The beginning and the end of a measuring time interval were registered telemetrically with the help of laser pulses having a duration of about 10^{-10} s. The increase in the time difference registered by the clocks was recorded. The predictions of the theory were confirmed in these experiments with a very high degree of accuracy. For example, under typical experimental conditions (five independent experiments were carried out), the theory predicted a difference of $(+47.1 \pm 0.25) \times 10^{-9}$ s in the time intervals recorded by the clocks in the aeroplane and on the Earth, and the experimental value of this quantity was found to be $(+47 \pm 1.5) \times 10^{-9}$ s. It should be noted that the theoretical value $+47.1 \times 10^{-9}$ s is due to both the gravitational dilatation of time $(+52.8 \times 10^{-9}$ s) and the relative increase in the pace of the clocks $(-5.7 \times 10^{-9}$ s). It can be seen that the predictions of the theory are confirmed reliably and with a high degree of accuracy.

PACE OF CLOCKS MOVING WITH AN ACCELERATION. In the experiment involving the flight round the globe, it was assumed without reservations that formula (16.6) is applicable not only to clocks moving in a uniform straight line, but also to clocks being accelerated. However, the formula was derived only for motion in a straight line, and its generalization to accelerated motion requires additional substantiation. A theoretical proof of this generalization is hard to obtain since the special theory of relativity does not consider accelerated motion.

However, there are experimental indications that formula (16.6) is also applicable to accelerated motion. A charged particle is accelerated in a magnetic field by a Lorentz force. The magnitude of the particle's velocity remains unchanged in this case since the magnetic field exerts a force only in a direction perpendicular to the velocity and does not perform any work. Experiments on the measurement of lifetime of particles with a view to verify the time dilatation effect can be repeated in a magnetic field when the particles are accelerated, but the magnitude of their velocity remains unchanged. This enables us to verify the effect under conditions of acceleration. It can be concluded on the basis of the available experimental results that *formula* (16.6) *is also applicable to accelerated motion, at least to circular motion.*

In one experiment, μ-mesons having an energy of 1.27 BeV were made to move in a circle of diameter 5 m by a very strong magnetic field. The velocity of the μ-mesons was very close to the velocity of light, and their acceleration was 4×10^{16} m/s^2. Assuming that time dilatation is independent of acceleration, the lifetime of the mesons must increase by a factor of about 12 as compared to their intrinsic lifetime of 2.2×10^{-6} (i.e. in the reference frame where the μ-mesons are at rest). The measured value of the lifetime was $(26.37 \pm 0.05) \times 10^{-6}$ s, while the calculated value was 26.69×10^{-6} s. This shows that acceleration does not affect the pace of clocks.

TWIN PARADOX. Suppose that a rocket flies from a point in an inertial reference frame at the instant $t = 0$ and returns to the same point after completing its flight. The path s covered by the rocket is a known function of time $s = s(t)$, and the velocity at each instant of time is $v(t) = \mathrm{d}s/\mathrm{d}t$. At the instant marking the return of the rocket, the clock in the stationary reference frame shows the time t, while the clock associated with the rocket shows the time $\tau = \int_0^t \mathrm{d}t \sqrt{1 - v^2/c^2}$. Hence if one of the new-born twins were to travel on the rocket and the other were to remain in the inertial reference frame, the former would be younger after the rocket had returned to the Earth. There is nothing paradoxical about this. The apparent paradox arises because of an incorrect line of reasoning which goes as follows. Since the motion is relative, it might be stated that the second twin sets out on a journey, while the first one on the rocket does not move. In this case, the second twin should be younger than the first twin when they meet. Which of the twins is actually younger? This is the essence of the twin paradox. *The reasoning which leads do this paradox is incorrect because the reference frames associated with the twins are not equivalent since one of them is an inertial reference frame, while the other associated with the rocket is not an inertial reference frame.* Formula (16.6) is only valid for inertial reference frames. Hence the twin travelling on the rocket will be younger than the one left behind in the inertial reference frame. As was clearly demonstrated in the example on round-the-Earth flight, formula (16.6) is not applicable to noninertial reference frames. In the reference frame fixed to the Earth's surface, the passage of time in an aeroplane flying eastwards is slowed down, while in an aeroplane flying westwards it is speeded up. Hence there is no twin paradox, all the arguments and calculations must be carried out in an inertial reference frame. Let us consider a simple example. Suppose that a rocket is fired at the instant $t = 0$ at a velocity $v = \text{const}$ (instantaneous velocity: it does not affect the reading of the clock associated with the rocket in any way) in the positive X-direction of an inertial K reference frame (laboratory reference frame). After an interval of time τ_1, the velocity of the rocket is instantaneously reversed, and the rocket returns to the Earth. The calculations in the laboratory reference frame are obvious: the rocket's total flight time is $\tau = 2\tau_1$, and the clock associated with the rocket shows the time $\tau' = 2\tau_1\sqrt{1 - v^2/c^2}$ at the instant it returns.

The calculations in the reference frame associated with the rocket are much more complicated and cannot be carried out with the help of Lorentz transformations. Hence we shall not present them here.

Example 16.1. Find the height h above the Earth's surface at which muons generated at a height of 30 km and moving vertically at a velocity $v = (1\text{-}8 \times 10^{-4})c$ decay on the average. What is the distance $h_\mu^{(0)}$ between the Earth and a muon at the instant it is born in the reference frame associated with the muon?

The lifetime of the muon $\tau_\mu = \tau_\mu^{(0)}/\sqrt{1 - v^2/c^2} = 0.55 \times 10^{-4}$ s and the path traversed by it is $l \approx \tau_\mu v \approx \tau_\mu^{(0)} c/\sqrt{1 - v^2/c^2} = 3 \times 10^8 \times 0.55 \times 10^{-4}\,\text{m} = 1.65 \times 10^4\,\text{m} = 16.5$ km. Consequently, $h = 16.5$ km and $h_\mu^{(0)} = 1700$ m.

Sec. 17. COMPOSITION OF VELOCITIES AND TRANSFORMATION OF ACCELERATIONS

Formulas for the composition of velocities and transformation of accelerations are derived and their corollaries are discussed.

FORMULA FOR COMPOSITION OF VELOCITIES. Suppose that the motion of a point mass is given by the functions

$$x' = x'(t'),\quad y' = y'(t'),\quad z' = z'(t') \tag{17.1}$$

in a moving coordinate system and by the functions

$$x = x(t),\quad y = y(t),\quad z = z(t) \tag{17.2}$$

in a stationary coordinate system. The functions defined in (17.2) are derived from those in (17.1) by means of (13.24). We must establish the connection between the projections of the velocities of the point mass in the moving and stationary coordinate systems. These projections are respectively represented as

$$u'_x = \frac{\mathrm{d}x'}{\mathrm{d}t'},\quad u'_y = \frac{\mathrm{d}y'}{\mathrm{d}t'},\quad u'_z = \frac{\mathrm{d}z'}{\mathrm{d}t'}; \tag{17.3}$$

$$u_x = \frac{\mathrm{d}x}{\mathrm{d}t},\quad u_y = \frac{\mathrm{d}y}{\mathrm{d}t},\quad u_z = \frac{\mathrm{d}z}{\mathrm{d}t}. \tag{17.4}$$

From (13.24), we obtain

$$\begin{gathered} \mathrm{d}x = \frac{\mathrm{d}x' + v\,\mathrm{d}t'}{\sqrt{1 - v^2/c^2}},\quad \mathrm{d}y = \mathrm{d}y',\quad \mathrm{d}z = \mathrm{d}z', \\ \mathrm{d}t = \frac{\mathrm{d}t' + (v/c^2)\,\mathrm{d}x'}{\sqrt{1 - v^2/c^2}} = \mathrm{d}t'\frac{1 + vu'_x/c^2}{\sqrt{1 - v^2/c^2}}. \end{gathered} \tag{17.5}$$

Substituting the values of the differentials from (17.5) into (17.4) and comparing with (17.3), we obtain

$$\boxed{\begin{aligned} u_x &= \frac{u'_x + v}{1 + vu'_x/c^2}, \\ u_y &= \frac{\sqrt{1 - v^2/c^2}\,u'_y}{1 + vu'_x/c^2}, \\ u_z &= \frac{\sqrt{1 - v^2/c^2}\,u'_z}{1 + vu'_x/c^2}. \end{aligned}} \tag{17.6}$$

These are the formulas for the composition of velocities in the theory of relativity. In accordance with the relativity principle, the formulas for the inverse transformations are obtained, as usual, by interchanging the primed and unprimed quantities and by replacing v by $-v$.

It follows from (17.6) that the velocity of light is constant, and the composition of velocities never leads to velocities exceeding the velocity of light. Let us prove this statement. Suppose that $u'_y = u'_z = 0$ and $u'_x = c$. In this case, we get from (17.6)

$$u_x = \frac{c+v}{1+cv/c^2} = c\,, \qquad u_y = 0\,, \qquad u_z = 0\,. \tag{17.7}$$

Of course, this result is natural since the transformation relations themselves were obtained in their final form by requiring that the velocity of light be constant.

ABERRATION. Suppose that a ray of light propagates along the Y'-axis in a K' coordinate system, i.e.

$$u'_x = 0\,, \qquad u'_y = c\,, \qquad u'_z = 0\,. \tag{17.8}$$

In a stationary coordinate system, we obtain

$$u_x = v\,, \qquad u_y = \sqrt{1 - v^2/c^2}\,c, \qquad u_z = 0\,. \tag{17.9}$$

Consequently, the ray of light forms an angle β with the Y-axis in the stationary coordinate system. This angle is

$$\tan\beta = \frac{u_x}{u_y} = \frac{v}{c}\frac{1}{\sqrt{1 - v^2/c^2}}\,. \tag{17.10}$$

For $v/c \ll 1$, equation (17.10) becomes identical to formula (12.5) which comes from the classical theory:

$$\tan\beta = \frac{v_\perp}{c}\,. \tag{17.11}$$

However, this relation now has a different meaning. In the classical theory, it was necessary to distinguish between the situations involving a moving source and a stationary observer from those involving a moving observer and a stationary source. In the theory of relativity, there is just one case involving the relative motion of the source and the observer.

INTERPRETATION OF FIZEAU'S EXPERIMENT. The result (12.23) of Fizeau's experiment follows naturally from the formula for the composition of velocities in the theory of relativity.

The velocity of light in a stationary medium with refractive index n is c/n. If the medium moves along the x'-axis, we obtain the following expressions for the velocity of light in the moving coordinate system:

$$u'_x = \frac{c}{n}, \qquad u'_y = 0\,, \qquad u'_z = 0\,. \tag{17.12}$$

Hence in accordance with (17.6), we obtain the projections of the velocity of light in the coordinate system relative to which the medium moves at a velocity $\pm v$:

$$u_x = \frac{(c/n) \pm v}{1 \pm v/(cn)}, \qquad u_y = 0, \qquad u_z = 0. \tag{17.13}$$

The plus sign corresponds to the case when the direction in which light propagates in the medium and the direction along which the medium moves coincide, while the minus sign corresponds to the case when these directions are opposite.

Since the quantity $v/c \ll 1$ is small, we can transform (17.13) as follows:

$$u_x \approx \left(\frac{c}{n} \pm v\right)\left(1 \mp \frac{v}{cn}\right) \approx \frac{c}{n} \mp \frac{v}{n^2} \pm v = \frac{c}{n} \pm \left(1 - \frac{1}{n^2}\right)v, \tag{17.14}$$

where the terms of the first and higher orders in v/c have been neglected. This expression is identical to formula (12.23). Hence the result of Fizeau's experiment is an experimental confirmation of the formula for the composition of velocities in the theory of relativity.

TRANSFORMATION OF ACCELERATIONS. Suppose that a point mass undergoes in a primed coordinate system an acceleration whose projections are given by a'_x, a'_y and a'_z, while its velocity is zero at that instant of time. Thus, the velocity of the point mass in a K' coordinate system is described by the following formulas:

$$\frac{\mathrm{d}u'_x}{\mathrm{d}t'} = a'_x, \qquad \frac{\mathrm{d}u'_y}{\mathrm{d}t'} = a'_y, \qquad \frac{\mathrm{d}u'_z}{\mathrm{d}t'} = a'_z,$$
$$u'_x = u'_y = u'_z = 0. \tag{17.15}$$

Let us determine the velocity of the point mass in the unprimed coordinate system. The velocity is obtained from (17.6):

$$u_x = v, \qquad u_y = 0, \qquad u_z = 0. \tag{17.16}$$

The acceleration in the K coordinate system is

$$a_x = \frac{\mathrm{d}u_x}{\mathrm{d}t}, \qquad a_y = \frac{\mathrm{d}u_y}{\mathrm{d}t}, \qquad a_z = \frac{\mathrm{d}u_z}{\mathrm{d}t}. \tag{17.17}$$

The quantities $\mathrm{d}t$, $\mathrm{d}u_x$, $\mathrm{d}u_y$ and $\mathrm{d}u_z$ are determined using (17.5) and (17.6), while the quantities u'_x, u'_y and u'_z can be set equal to zero only after the differentials have been evaluated. For example, we have the following expression for $\mathrm{d}u_x$:

$$\begin{aligned} du_x &= \frac{du'_x}{1 + vu'_x/c^2} - \frac{(u'_x + v)(v/c^2)\,du'_x}{(1 + vu'_x/c^2)^2} \\ &= \frac{du'_x}{(1 + vu'_x/c^2)^2}\left(1 + \frac{vu'_x}{c^2} - \frac{vu'_x}{c^2} - \frac{v^2}{c^2}\right) \\ &= \frac{1 - v^2/c^2}{(1 + vu'_x/c^2)^2}\,du'_x. \end{aligned}$$

Given (17.15), we obtain

$$a_x = \frac{du_x}{dt} = \left(1 - \frac{v^2}{c^2}\right)^{3/2}\frac{du'_x}{dt'} = \left(1 - \frac{v^2}{c^2}\right)^{3/2} a'_x, \tag{17.18}$$

where in accordance with (17.15) we have set $u'_x = 0$.

The differentials du_y and du_z are calculated in a similar manner. Thus, we obtain the following transformation formulas for acceleration:

$$\begin{aligned} a_x &= \left(1 - \frac{v^2}{c^2}\right)^{3/2} a'_x, \\ a_y &= \left(1 - \frac{v^2}{c^2}\right) a'_y, \\ a_z &= \left(1 - \frac{v^2}{c^2}\right) a'_z. \end{aligned} \tag{17.19}$$

The point mass moves at a velocity v in the K system. Hence (17.19) has the following meaning. A moving point mass can be assigned an inertial coordinate system in which it is at rest at a given instant of time. Such a coordinate system is called the body axes system. If the point mass moves with an acceleration in this system, it will also move with an acceleration in any other system. However, this acceleration will be different but always smaller. The projection of the acceleration on the direction of velocity is reduced in proportion to the factor $(1 - v^2/c^2)^{3/2}$, where v is the velocity of the body axes coordinate system. The transverse component of acceleration perpendicular to the velocity of the particle varies to a smaller extent, its decrease being proportional to the factor $1 - v^2/c^2$.

?

How can the result of Fizeau's experiment be interpreted from the point of view of the theory of relativity?
What is the essential difference in the interpretation of aberration from the point of view of classical and relativistic concepts?
What physical factors explain the law of transformation of longitudinal and transverse projections of acceleration?

Example 17.1. Two protons move towards each other in a laboratory coordinate system at velocities v and $-v$. Find the relative velocity u' of the protons in the coordinate system associated with one of them and express the factor $\gamma' = 1/\sqrt{1 - u'^2/c^2}$ in terms of the factor $\gamma = 1/\sqrt{1 - v^2/c^2}$.

In the coordinate system associated with the proton moving in the positive X-direction of the laboratory system, we obtain the following relation from the velocity composition formula:

$$u' = \frac{u - v}{1 - uv/c^2} = \frac{-2v}{1 + v^2/c^2}, \tag{17.20}$$

where we have used the relation $u = -v$. Hence we obtain

$$\gamma' = 2\gamma^2 - 1. \tag{17.21}$$

Example 17.2. It was mentioned in Sec. 15 that the shape and geometrical size of a moving body apparent to a visual observation may differ from the shape and geometrical size determined from the Lorentz transformations. Let us analyze this topic for a thin rod moving along its length.

In order to ascribe a clear physical meaning to the idea of the visually observed length of the rod, we shall assume that a long stationary ruler is placed along the X-axis in a stationary coordinate system. A rod of length l at rest moves along the ruler at a velocity v in the positive X-direction. In order to measure the length of the rod in the stationary coordinate system, we must note at a certain instant of time the readings of the points on the stationary ruler, which coincide with the ends of the rod at this instant of time. By definition, the distance between these points is the length of the moving rod given in accordance with (15.3) by $l' = l\sqrt{1 - v^2/c^2}$.

Let us now observe the moving rod from a sufficiently large distance. The observation is carried out by a telescope through which the moving rod and the stationary ruler can simultaneously be seen. It is logical to define the observed length of the moving rod as the distance between its end points on the stationary ruler at the instant they are observed through the telescope. For the sake of definiteness, it can be assumed that a snapshot is taken in which the moving rod covers a part of the stationary ruler. The length of the covered part of the stationary ruler is the visually observed length of the moving rod, whose value has to be determined. Let us denote the angle between the positive X-direction and the direction of motion of photons from the moving rod to the telescope by α.

In order to solve this problem, we note that the photons which form an image of the rod on the photographic plate at a certain instant of time were not simultaneously emitted by different parts of the rod. The difference between the instants at which photons are emitted from the leading and trailing ends of the moving rod is $L' \cos\alpha/c$, where L' is the required visually observed length of the moving rod. Hence a photographic plate does not record the ends of the moving rod simultaneously, there is a time difference $t_2 - t_1 = L' \cos\alpha/c$.

If the coordinates of the ends of the rod in a coordinate system in which it is at rest are denoted by x'_1 and x'_2, then

$x'_2 - x'_1 = l$ will be the proper length of the rod. We denote the coordinates of the ends of the rod in a stationary coordinate system by x_1 and x_2, so that $x_2 - x_1 = L'$. From the Lorentz transformations

$$x'_1 = \frac{x_1 - vt_1}{\sqrt{1 - v^2/c^2}}, \qquad x'_2 = \frac{x_2 - vt_2}{\sqrt{1 - v^2/c^2}} \tag{17.22}$$

we obtain

$$x'_2 - x'_1 = \frac{(x_2 - x_1) - v(t_2 - t_1)}{\sqrt{1 - v^2/c^2}}. \tag{17.23}$$

Since $x'_2 - x'_1 = l$, $x_2 - x_1 = L'$ and $t_2 - t_1 = L' \cos\alpha/c$, we can write (17.23) in the form

$$l = \frac{L' - vL' \cos\alpha/c}{\sqrt{1 - v^2/c^2}}. \tag{17.24}$$

Hence

$$L' = \frac{l\sqrt{1 - v^2/c^2}}{1 - (v/c)\cos\alpha}. \tag{17.25}$$

PROBLEMS

4.1. Two point masses, which are at rest on the X-axis of a stationary coordinate system and a distance 100 m apart, are simultaneously accelerated in the same way in the positive X-direction. The acceleration is simultaneously discontinued when the points acquire a velocity $v = (1-2 \times 10^{-4})c$. What will be the distance between the points in the coordinate system in which they are at rest?

4.2. A proton covers a distance $l = 1.5 \times 10^8$ km between the Sun and the Earth at a velocity $v = 4c/5$. What will this distance appear to be in the reference system associated with the proton? How much time will be required to cover the distance in the reference system associated with the Earth and the proton?

4.3. Two projectors emit narrow beams of light in opposite directions. At what velocity must the projectors move in a direction perpendicular to the light rays so that the beams are at right angles to each other?

4.4. Two particles moving one after the other at a velocity $v = 3c/5$ strike a target with a time interval of 10^{-7} s. Find the distance between the particles in the laboratory reference frame and in the reference frame associated with the particles.

4.5. Two clocks A' and B' are located at points $(x'_1, 0, 0)$ and $(x'_2, 0, 0)$ in a K' coordinate system $(x'_2 - x'_1 = l)$, while two more clocks A and B are located at points $(x_1, 0, 0)$ and $(x_2, 0, 0)$ in a K coordinate system $(x_2 - x_1 = l)$. The K' system moves relative to the K system in the positive X-direction at a velocity v. When two points coincide spatially, the clocks B' and A show the same time (say, $t' = t = 0$).

What will be the time shown when the clocks A and A' and the clocks B and B' coincide spatially?

4.6. A rocket moves in a straight line with a uniform acceleration a measured by the passengers with the help of an accelerometer on board. How far will the rocket move in the laboratory reference frame before it attains a velocity v?

4.7. Two scales with identical proper length l move towards each other at the same velocities v along the X-axis. What will be the length of one scale in the coordinate system associated with the other scale?

4.8. A train is moving at a velocity v relative to the railway track. A bird is flying alongside at the same velocity v relative to the train. An aeroplane is also flying in the same direction at a velocity v relative to the bird. What will the velocity of the aeroplane be relative to the railway track?

ANSWERS

4.1. 5000 m. 4.2. $l = 0.9 \times 10^8$ km, $\tau_E = 625c$, $\tau_p = 375c$. 4.3. $c/\sqrt{2}$. 4.4. 18 m, 22.5 m. 4.5. $t_A = l\sqrt{1 - v^2/c^2}/v$, $t'_{A'} = l/v$, $t_B = l/v$, $t'_{B'} = l\sqrt{1 - v^2/c^2}/v$. 4.6. $c^2(1/\sqrt{1 - v^2/c^2} - 1)/a$. 4.7. $l(1 - v^2/c^2)/(1 + v^2/c^2)$. 4.8. $c\{1 - [(1 - v/c)/(1 + v/c)]^3\}/\{1 + [(1 - v/c)/(1 + v/c)]^3\}$.

Chapter 5

Dynamics of a Point Mass

Basic idea:
The main idea of Newtonian dynamics is that external agencies described by the force concept are responsible for the acceleration and not the velocity or the time derivative of the acceleration. Hence Newton's second law of motion is not a definition of force, although it is the only possible way of determining a force in many specific cases.

Ses. 18. FORCES

Different manifestations of force are described, which necessitate their introduction into classical mechanics independently of acceleration.

ORIGIN OF THE CONCEPT OF FORCE. For about 2000 years before Galileo, the task of a theory of motion was to explain why an object has a given velocity. In Aristotelian dynamics, it was assumed that each object is assigned a particular place: lighter objects occupy higher positions, while heavier objects occupy lower positions. All objects tend to move to their respective positions, and hence the upward motion of light objects and the downward motion of heavy objects do not require any explanation, as it is the natural motion of objects. The motion of stars and celestial bodies in the sky was also assumed to be natural. Motions that are not natural (for example, a billiard ball rolling along a table) are forced motions, and there must be a reason behind each occurrence.

What caused a ball to roll along a table was called a force by Aristotle. A force is imparted to the ball from the medium surrounding it. If Aristotle's law of motion were to be expressed in mathematical form, the force would be proportional to the velocity of the ball. Later, doubts began to be gradually raised as to whether the force is transferred to the ball from the surrounding medium. Kuzanskii (1401-1464) suggested that the force is imparted to the ball at the instant of impact, after which it remains in the ball and maintains its velocity. This eliminated the need for a surrounding medium. Afterwards, Galileo clearly showed that it is not the maintenance of the velocity that should be explained, but rather the

Aristotle (384-322 B. C.)
Greek scientist and philosopher, the founder of formal logic, the concept of motion, the basic principles of everyday life, ethics, the concept of man as a social animal and the concept of the state. He wrote a number of treatises, including *Physics* and *Metaphysics*.

variation in the velocity. He associated the concept of force with acceleration rather than with velocity. A body preserves its velocity due to the law of inertia according to which the body tends to maintain its uniform motion in a straight line. According to Galileo, this law should be considered a fundamental law which cannot be reduced to a simpler form.

In Newtonian dynamics, it is not the velocity but the change in velocity, i.e. the acceleration, which must be caused. The cause behind the change in velocity is called force. The problem is to quantitatively formulate the relation between force and acceleration. This can be done using Newton's laws of motion.

INTERACTIONS. Forces do not exist without objects. They are produced by objects. Hence one might say that objects act on one another through forces, i.e. they interact. The force is a quantitative vector measure of the interaction intensity.

MEASUREMENT OF FORCE. Forces not only change the velocity of an object, they also deform it. The simplest and the most visual example of a deformed object is a compressed or stretched spring which can conveniently be used as a standard measure of force. The standard unit of force can be defined in terms of a spring compressed or stretched to a certain extent. Two forces are said to have the same magnitude but opposite directions if they do not accelerate the object to which they are applied. This can be used to compare forces directed along the same straight line and leads to the conclusion that forces not only have numerical values, they also have directions. On the other hand, the standard can be used to construct a scale of forces. Figure 39 shows the composition of forces acting in different directions. Once again, the absence of acceleration in the object to which forces are applied indicates that the sum of

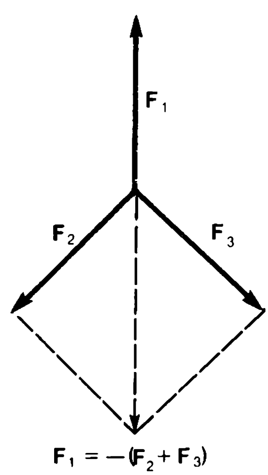


Fig. 39. Composition of forces according to the parallelogram law.

these forces is zero. It can be seen in Fig. 39 that the forces are composed according to the parallelogram law, i.e. as vectors. *This also proves that force is a vector, and a procedure for measuring forces is established independently of the measurement of acceleration.*

Sec. 19. NEWTON'S LAWS

The number of independent Newton's laws is discussed.

HOW MANY OF NEWTON'S LAWS OF MOTION ARE INDEPENDENT IN NATURE? It is well known that according to Newton's first law of motion, a body far removed from other bodies continues to be in the state of rest or uniform motion in a straight line, while the second law expresses the acceleration of a body subjected to the action of a force:

$$m\frac{d\mathbf{v}}{dt} = \mathbf{F}, \tag{19.1}$$

where m is the mass of the body, and $d\mathbf{v}/dt$ is the acceleration. It follows from (19.1) that in the absence of a force ($\mathbf{F} = 0$), $\mathbf{v} = \text{const}$. In other words, if no force acts on a body or if the resultant of the forces applied to a body is zero, the body will be either at rest or moving uniformly in a straight line. Hence it has been suggested that the first law is not independent and is just a corollary of the second law, while Eq. (19.1), which expresses this law, is just the definition of force, i.e. the convention needed to introduce the quantity $\mathbf{F}$ (called a force) into the theory using (19.1). This point of view is refuted in this book for the following reasons.

!

The physical meaning of Newton's second law of motion is that the external conditions are defined by the acceleration and not by the velocity or the time derivative of the acceleration. In classical mechanics, the external conditions are described using the concept of force.

We described at length in Secs. 5 and 7 that even a purely kinematic consideration of the motion of bodies in a reference frame is only possible if the scales and clocks behave in the manner prescribed in these sections. If, for example, we could not synchronize the clocks and so introduce a unified time base into the reference frame, as explained in Sec. 7, then not only the law of dynamics (19.1), but all the mathematical relations presented in Chapter 2 would be meaningless. This naturally gives rise to questions about how to verify the constructions described in Secs. 5 and 7 for a given reference frame. *Only after such a verification can we state that the physical quantities are clear and meaningful. This applies not only to dynamic quantities, but also to kinematic ones.* For example, even a simple concept of uniform motion becomes meaningless if we cannot synchronize clocks in the manner described in Sec. 7.

?

What is the subjective difference between action and reaction?
How are forces related to the deformations they produce?

In principle, we can verify the suitability of a reference frame by carefully studying the behaviour of the scales and clocks in the system. Such a verification must cover all space and must be quite precise. Only then can we write down Eq. (19.1) and derive the first law of motion from it. However, it is practically impossible to carry out such a verification, and in the absence of a verification, we cannot indicate the significance of (19.1).

In order to overcome this difficulty, we can choose a reference frame using Newton's first law: we must take a test object and place it quite far from every other object. If the observations of the motion of the test object reveal that it is moving uniformly in a straight line or is at rest, the reference frame is suitable for the kinematic and dynamic descriptions of the motion in accordance with the rules applied. This verification is equivalent to the one mentioned above. Once a verification has been carried out, we can write the law of motion in the form (19.1).

Hence Newton's first law of motion is an independent law which expresses a criterion for the suitability of a reference frame when considering motions in both dynamic and kinematic senses. The law is not only independent, it is also the first in the series because only with this law can we speak of the exact physical meaning and content of the second and third laws.

However, it is unacceptable to treat Eq. (19.1) as a definition of force not only because it expresses analytically a physical idea about the nature of the relation between the external conditions and the dynamic variables of the motion of a point mass, but also because the action of forces in mechanics is manifested in acceleration of bodies as well as, say, in their deformation when the acceleration is not considered at all. However, it must be emphasized that we are dealing with classical mechanics and not with physics in general. For example, it is meaningless to speak of forces in quantum mechanics since in this case we cannot describe the motion by using the concepts of velocities, accelerations, trajectories and point masses in the classical sense.

The unacceptability of the interpretation of Newton's second law of motion as definition of force also stems from an analysis of the connection between Newtonian mechanics and Aristotelian mechanics. Newtonian mechanics retains the concepts of natural and forced motion, but only the uniform motion in a straight line (the first law) is accepted as natural motion. Whereas Aristotle's law states that velocity is proportional to force, Newton's second law of motion states that acceleration is proportional to force. This means that external conditions determine the acceleration and not the velocity of an object. If the second law is just a definition of force, then why can we not describe force in such a way that the motion is defined in accordance with Aristotle's law or in such a way

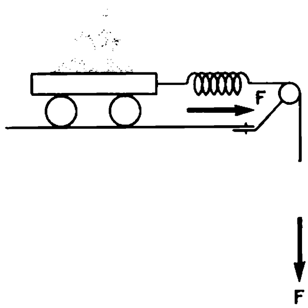


Fig. 40. Showing the dependence of acceleration on force.

The force is measured in terms of the extension of the spring.

that the force is proportional to the third time derivative of coordinates? The fact that this is not possible confirms that Newton's second law of motion cannot be treated as definition of force.

MASS. The simplest standard of force can be obtained from a spring graduated for different values of force in the manner described above. Thus, it is possible to apply different forces of known values to a body. The unit of force is an independent quantity which materializes in a spring stretched or compressed to a certain extent. The second measurable quantity is the acceleration of different bodies acted upon by a force. The experimental set-up for studying the dependence of acceleration on force is shown in Fig. 40. The results of such experiments show that the direction of acceleration coincides with that of the applied force. The same force imparts different accelerations to different bodies, and different forces impart different accelerations to the same body. However, the ratio of force to acceleration is always equal to the same quantity, i.e.

$$\frac{F}{a} = \text{const} = m. \tag{19.2}$$

The relation $F/a = \text{const}$ is valid only for quite small velocities. If the velocity of a body is increased, this ratio begins to vary, increasing with velocity. This means that the inertial properties of bodies are manifested more strongly with increasing velocity. This question will be considered in greater detail in the next section.

Writing Eq. (19.2) in vector form considering that the direction of acceleration coincides with that of the applied force, we arrive at (19.1), i.e. the expression for Newton's second law of motion. However, this equation can also be written conveniently in an alternative form:

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}, \tag{19.3a}$$

$$\mathbf{p} = m\mathbf{v}. \tag{19.3b}$$

The product of mass and velocity, $\mathbf{p} = m\mathbf{v}$, is called momentum. The force $\mathbf{F}$ on the right-hand side of (19.3a) is the sum of all forces acting on a body.

Although Eq. (19.3a) is obtained from (19.1) by a simple change of notation, it cannot be stated that it does not contain anything new. In the first place, Eq. (19.3a) contains a new physical quantity, viz. momentum, which is defined by (19.3b).

Sir Isaac Newton (1643-1727) English mathematician, mechanical engineer, physicist and astronomer. He founded classical physics and established the laws of classical mechanics. He discovered the law of universal gravitation and laid the foundations of celestial mechanics. He was the first to study the dispersion of light and chromatic aberration. He investigated the interference of light and developed the corpuscular theory of light. He worked out differential and integral calculi (independently of Leibnitz).

At low velocities, when the velocity and acceleration are directly measurable quantities, momentum is just an auxiliary quantity which can be determined from (19.3b). At very high velocities, however, momentum becomes the basic quantity which is measured in experiments, whereas a particle's velocity becomes practically constant and equal to the velocity of light. Hence at low velocities, Eqs. (19.1) and (19.3a) differ only in formal notation, while at high velocities, they differ in physical content. We shall show in more detail in the next section that Eq. (19.3a) and not (19.1) represents the extension of the equation of motion to relativistic velocities.

The property of a body which determines the value of the ratio of force to acceleration in (19.2) is called the inertia of the body. This quantity is called the inertial mass, or simply mass. In Newtonian mechanics, inertial mass has no meaning except as a property characterizing the inertia of a body.

At relativistic velocities, the inertial properties of bodies are conserved, but their manifestation becomes more complex. The direction of force generally does not coincide with that of acceleration, and relations like (19.2) become meaningless. However, such relations continue to be meaningful for the components of force and the components of acceleration in a particular direction. It characterizes the inertial properties of a body in this direction, but does not have a constant value and changes with the angle between this direction and the direction of motion. Hence we cannot extend the concept of mass as a scalar quantity to the relativistic case using Eq. (19.2). The relation (19.3b), in which mass appears as the proportionality factor between the momentum and velocity of a point mass, is more suitable for this purpose.

ON NEWTON'S SECOND LAW OF MOTION. Newton's second law of motion (19.1) can be treated as a law rather than a definition of force only if there exists an independent definition of force. One independent definition of force as the force of a deformed spring was considered in the preceding section. However, this is not enough for treating (19.1) as a law since such a definition of force is only valid for bodies at rest. Hence we must carry out experiments with moving deformed springs and with bodies accelerated by them to make sure that (19.1) is independent of velocity. This has been confirmed experimentally, and hence (19.1) represents a law rather than a definition of force. *The physical meaning of this law lies not in that the force has a specific expression, but in that the force defines the second time derivatives of coordinates* ($\mathrm{d}\mathbf{v}/\mathrm{d}t = \mathrm{d}^2\mathbf{r}/\mathrm{d}t^2$). *The invariance of acceleration relative to Galilean transformations leads to the invariance of force.*

In relativistic dynamics, however, (19.3a) is the equation of motion, which is invariant to the Lorentz transformations by definition, while the force F is not invariant. The situation becomes quite complicated, but it need not be analyzed in detail in order to say whether the relativistic equation is a law of motion or a definition of force. We might say, in view of the relativistic invariance of the equation and the relativity principle, that if this relation is an equation of motion and not a definition of force in the coordinate system in which the nonrelativistic approximation is valid, then it will be an equation of motion in all other coordinate systems as well.

!

Newton's second law of motion cannot be formulated until the first law and the concepts associated with an inertial reference frame have been established. Any attempt to do without these concepts will deprive the physical quantities appearing in the second law of a physical meaning. Forces for which Newton's third law of motion is valid do not satisfy the relativistic requirements and hence cannot exist in nature. In the nonrelativistic limit, however, Newton's third law of motion is obeyed with a high degree of accuracy. Hence the statement concerning the violation of Newton's third law of motion is of fundamental importance only for determining the limits of applicability of Newtonian mechanics.

ON NEWTON'S THIRD LAW OF MOTION. According to Newton's third law of motion, an interaction between two bodies involves the action of each body on the other by a force equal in magnitude and opposite in direction. Thus, different bodies are the sources of "action" and "reaction" forces. Similarly, the bodies to which these forces are applied are also different. Each of the interacting bodies is a source of "action" on the other body, and an object of "reaction" whose source is the other body. *Hence the difference between the forces of "action" and "reaction" is only subjective and depends on the point of view. The nature of "action" and "reaction" is essentially the same.*

The law of action and reaction is demonstrated in a visual form when the interaction between bodies takes place through other bodies, say, through a string or a spring. Figure 41 shows an experimental set-up which can be used to demonstrate the validity of Newton's third law of motion. In the case shown in Fig. 41*a*, the spring is compressed to a certain position under the action of external forces F_1 and F_2 applied to bodies m_1 and m_2 respectively. After the forces compressing the spring are withdrawn, the bodies m_1 and m_2 are set in an accelerated motion. Thus, each body is subjected to the action of a force which can be calculated from the acceleration acquired by the body. It is shown experimentally that the relation $m_1a_1 = m_2a_2$ is always satisfied, where a_1 anl a_2 are the accelerations of the bodies m_1 and m_2. This means that $F_1 = F_2$. In the case shown in Fig. 41*b*, one of the interacting bodies is connected to an electric motor. During the rotation of the motor, a string is wound on its shaft, and the other end of the string is rigidly attached to the second body. When the string is being wound on the shaft, the two bodies move towards each other with the accelerations a_1 and a_2, and the relation $m_1a_1 = m_2a_2$, i.e. $F_1 = F_2$, is always satisfied.

However, in such a simple form, the law of action and reaction is not always satisfied. Let us consider the interaction between two positive charges q_1 and q_2 moving at velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively and subjected to the action of forces $\mathbf{F}_1$ and

?

What are the prevailing views on the number of independent Newton's laws of motion? What do you think about this question?
What general relativistic concepts show that Newton's third law of motion cannot be obeyed in its simplest form?
What is the more profound view on Newton's third law of motion and how is it used to interpret the interaction between moving electron charges?
What is the role of field in an interaction?

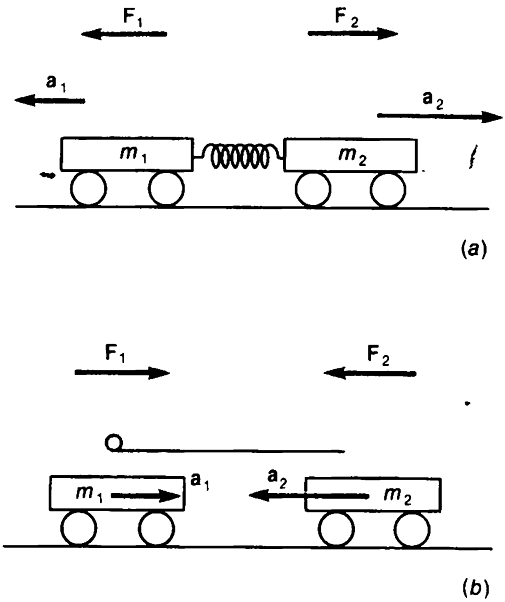


Fig. 41. When two bodies interact, one of them acts on the other with a force that is equal and opposite to the force exerted by the second body on it (Newton's third law).

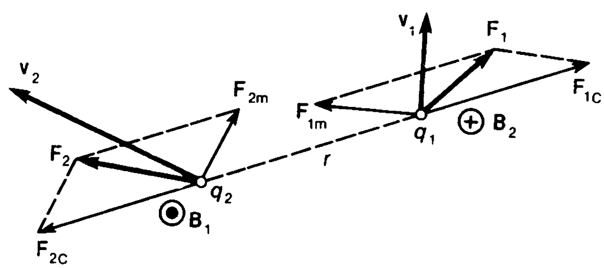


Fig. 42. Electrostatic interaction of moving charges.

In this case, Newton's third law is not satisfied since electric charges do not constitute an isolated system. Fields also participate in this interaction.

F_2 from the other charge (Fig. 42). Each of these forces can be represented as a sum of two components. The first component is the force of electrostatic interaction according to Coulomb's law. This force acts along the line joining the two charges, is equal to $q_1 q_2/(4\pi\varepsilon_0 r^2)$ and satisfies Newton's third law of motion, i. e. can be represented in the form $F_{1C} = -F_{2C}$. But besides the electrostatic interaction, a magnetic interaction also exists between the charges, as each moving charge produces a magnetic field with magnetic induction B at the point of location of the other charge. The field acts on a charge q moving at a velocity v with a Lorentz force given by

$$F_m = q\mathrm{v} \times \mathrm{B}. \tag{19.4}$$

The field B produced by a moving charge can be determined. For the present, there is no need to know the exact value of induction. It is important to note that in Fig. 42 the field $\mathbf{B}_1$ produced by the charge q_1 at the point of location of the charge q_2 is perpendicular to the plane of the figure and points towards the reader, while the field $\mathbf{B}_2$ produced by the charge q_2 at the point of location of the charge q_1 is perpendicular to the plane of the figure and points away from the reader. The Lorentz force (19.4) is perpendicular to the velocity $\mathbf{v}$ and to the magnetic field $\mathbf{B}$. It can be seen from Fig. 42 that the Lorentz forces $\mathbf{F}_{1m}$ and $\mathbf{F}_{2m}$ acting on the charges q_1 and q_2 are not collinear and hence cannot satisfy the law of action and reaction. The total force exerted by the first charge on the second ("action") is $\mathbf{F}_{2C} + \mathbf{F}_{2m} = \mathbf{F}_2$, while the total force exerted by the second charge on the first ("reaction") is $\mathbf{F}_{1C} + \mathbf{F}_{1m} = \mathbf{F}_1$. Obviously, these forces are not equal and are not directed opposite to each other:

$$\mathbf{F}_1 \neq -\mathbf{F}_2, \tag{19.5}$$

i.e. Newton's third law is not satisfied in this case.

It should be noted that for charged particles moving at velocities much lower than the velocity of light ($v/c \ll 1$), the forces of magnetic origin are much smaller than the electric forces. Since the departure from the law of action and reaction is due to magnetic forces, this departure is quite small at not too high velocities and can be neglected.

In order to analyze this question, we must bear in mind that Newton's third law of motion has a more profound meaning than a simple equality of the forces of action and reaction. Let us consider the interaction of carriages shown in Fig. 41 and write down the equation of motion (19.3b) for each of the interacting bodies:

$$\frac{\mathrm{d}p_1}{\mathrm{d}t} = F_1, \quad \frac{\mathrm{d}p_2}{\mathrm{d}t} = F_2, \tag{19.6}$$

where $p_1 = m_1 v_1$ and $p_2 = m_2 v_2$. The velocity v_1 or v_2 has a plus sign if its direction coincides with the positive x-direction. According to this convention, the velocity v_2 of the carriage m_2 in Fig. 41*a* will be positive, while the velocity v_1 of the carriage m_1 will be negative. In Fig. 41*b*, the signs of the velocities v_1 and v_2 will be reversed. The signs of the forces F_1 and F_2 in (19.6) are also determined by whether the vector of a given force coincides with the positive x-direction or has the opposite direction. In accordance with Newton's third law of motion, we must have $F_1 + F_2 = 0$. Hence, the termwise addition of (19.6) gives

$$\frac{\mathrm{d}p_1}{\mathrm{d}t} + \frac{\mathrm{d}p_2}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(p_1 + p_2) = F_1 + F_2 = 0, \tag{19.7}$$

whence

$$p_1 + p_2 = \text{const.} \tag{19.8}$$

Thus, when two bodies interact, the sum of their momenta is constant. Newton's third law of motion can also be formulated as the requirement for the sum of the momenta of the interacting bodies to be conserved in the absence of any external forces. This is where the third law of motion has a more profound physical meaning.

Let us now return to the interaction between moving charges (see Fig. 42). We have shown above that the forces with which the electrons act on one another are not equal and opposite. Hence, it follows from (19.7) that the sum of the momenta of the interacting electrons does not remain constant and keeps on changing, i.e. Newton's third law of motion is not satisfied in this case.

Let us, however, consider the interaction pattern in greater detail. The interaction involves not only charges q_1 and q_2, but also the electric field E and the magnetic field B. It can be asked whether these fields have a momentum. The answer to this question is in the affirmative. It is shown in electrodynamics that this momentum is distributed over entire space in which the electromagnetic field exists, and the momentum density (i.e. the momentum per unit volume) for the field in vacuum is $\mathbf{E} \times \mathbf{B}/(c^2\mu_0)$, where $\mu_0 = 4\pi \times 10^{-7}$ H/m is the magnetic constant. It is borne out by computations that when the sum of the momenta of the interacting electrons changes, the momentum of the electromagnetic field produced by the electrons also changes by the same amount simultaneously, but in the opposite direction. In other words, the total momentum of interacting moving electrons and of the electromagnetic field produced by them remains constant upon interaction. In such a formulation, the validity of Newton's third law of motion is restored for describing a given interaction.

It will be shown in Sec. 26 that the forces for which Newton's third law of motion is valid do not satisfy relativistic requirements and hence cannot exist in nature. Real forces satisfy Newton's third law of motion only in the nonrelativistic case, but this law was formulated just for this case. Hence the statement about the violation of Newton's third law of motion is significant only for determining the limits of applicability of Newtonian mechanics.

That is why the formulation of Newton's third law of motion as the requirement of conservation of the total momentum of interacting bodies and fields is physically more meaningful than the formulation which requires the equality of the forces of action and reaction.

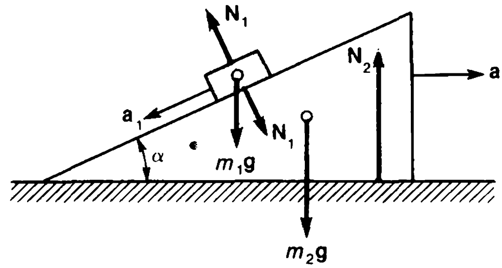


Fig. 43. Diagram of forces and accelerations in Example 19.1.

It can be shown that Newton's third law of motion cannot be satisfied in the relativistic case without considering the electromagnetic interactions. Suppose that two bodies interact in a coordinate system in such a way that the equality of action and reaction is observed, and the bodies are simultaneously set into motion upon being imparted equal and opposite accelerations. The two bodies will not simultaneously be set into motion in another coordinate system, and hence there will be an interval of time during which one of the bodies is being accelerated, while the other is at rest. It is clear that Newton's third law of motion in its simplest form will not be satisfied during this interval. Thus, the violation of Newton's third law in its simplest form is a consequence of the general relativistic properties of space and time.

Example 19.1. A body of mass m_1 can slide without friction along the inclined plane of a beam of mass m_2. The angle of inclination of the beam with the horizontal is α. The beam moves without friction along a horizontal plane (Fig. 43). Find the acceleration of the body and the beam.

Let us denote the acceleration of the body along the inclined plane relative to the beam by a_1, and the acceleration of the beam in the horizontal direction by a_2. The body is subjected to the supporting force N_1 and the weight $m_1 g$. The beam is subjected to the supporting force N_2 and the weight $m_2 g$.

Newton's equations of motion for the beam in projections on the horizontal and vertical directions are written in the form

$$\begin{aligned} m_2 a_2 &= N_1 \sin \alpha, \\ 0 &= m_2 g - N_2 + N_1 \cos \alpha. \end{aligned} \tag{19.9}$$

The corresponding equations for the body can be written as follows:

$$\begin{aligned} m_1 (a_1 \cos \alpha - a_2) &= N_1 \sin \alpha, \\ m_1 a_1 \sin \alpha &= m_1 g - N_1 \cos \alpha. \end{aligned} \tag{19.10}$$

There are four unknown quantities in (19.9) and (19.10). The

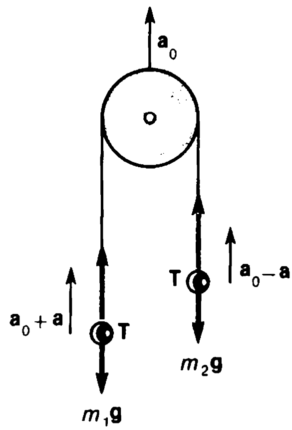


Fig. 44. Diagram of forces and accelerations in Example 19.2.

accelerations can be determined as follows:

$$a_1 = \frac{(m_1 + m_2) g \sin\alpha}{m_1 \sin^2\alpha + m_2}, \qquad a_2 = \frac{m_1 g \sin\alpha \cos\alpha}{m_1 \sin^2\alpha + m_2}. \tag{19.11}$$

Example 19.2. A weightless unstretchable string passes across a pulley, and loads of masses m_1 and m_2 are attached to its ends. There is no friction between the pulley and the string. The pulley moves vertically upwards with an acceleration a_0. Find the acceleration of the loads (Fig. 44).

Since the string is unstretchable, the accelerations of the loads of masses m_1 and m_2 relative to the pulley are equal in magnitude but opposite in sign. While the acceleration of the load of mass m_1 is $a_0 + a$, that of the load of mass m_2 is $a_0 - a$. The same tension T acts on the loads from the string. Hence the equations of motion for the loads have the form

$$m_1(a_0 + a) = T - m_1 g, \qquad m_2(a_0 - a) = T - m_2 g, \tag{19.12}$$

whence we get

$$a = \frac{(m_2 - m_1)(a_0 + g)}{m_1 + m_2}. \tag{19.13}$$

Example 19.3. A ring slides without friction along an undeformable rod rotating at a constant angular velocity ω about the vertical axis passing through the rod at right angles to its length. If r is the distance along the rod from the axis of rotation to the ring, find $r(t)$.

It is convenient to carry out the calculations in the polar coordinate system, the origin of the system coinciding with the point on the rod through which the axis of rotation passes. The plane of the polar coordinate system is assumed to coincide with the plane in which the rod moves. The radial acceleration in the polar coordinate system is $\ddot{r} - \omega^2 r$, where the dots indicate the time derivatives. There are no forces in the radial direction. Hence Newton's equation for the radial motion of the ring of mass m has the form

$$m(\ddot{r} - \omega^2 r) = 0, \tag{19.14}$$

whence

$$r = A_1 e^{\omega t} + A_2 e^{-\omega t}, \tag{19.15}$$

where A_1 and A_2 are constants determined by the initial conditions. For example, if $t = 0$, $r = 0$ and $\dot{r} = u$, then $A_1 = -A_2 = u/(2\omega)$, and hence

$$r = \frac{u}{2\omega}(e^{\omega t} - e^{-\omega t}) = \frac{u}{\omega}\sinh\omega t. \tag{19.16}$$

Sec. 20. RELATIVISTIC EQUATION OF MOTION

The relativistic equation of motion is derived and its corollaries are analyzed.

INERTIA IN THE DIRECTION OF VELOCITY AND PERPENDICULAR TO THE VELOCITY. If we continue the experiments with the carriages shown in Fig. 41 and keep on increasing the velocities of the carriages, it will be seen that the ratio F/a is not constant as indicated by (19.2), but varies with velocity. However, very high velocities are required for observing this phenomenon. It is easier to perform such experiments on charged elementary particles moving in electromagnetic fields (say, in accelerators). The force acting on a charged particle moving at a velocity v is calculated in accordance with the formula

$$\mathrm{F} = q(\mathrm{E} + \mathrm{v} \times \mathrm{B}). \tag{20.1}$$

Suppose that a charged particle, say, a proton, moves in a circular orbit in an alternating magnetic field B as in a cyclotron (Fig. 45). Over a certain region in the path of the proton, an electric field E is produced, whose magnitude is known and which varies in such a way that the proton gets accelerated upon passing through the region. Outside the accelerating region, the proton moves in a circle of known radius r under the action of the force $\mathrm{F}_n = e\mathrm{v} \times \mathrm{B}$. Specifying the magnetic induction B, determining the protons velocity from the time taken by it to traverse the circular trajectory in the accelerator, and taking into account the formula for centripetal acceleration $v^2/r = a_n$ during the circular motion of the proton, we can find the ratio $F_n/a_n = evBr/v^2$. Experiments give the dependence

$$\frac{F_n}{a_n} = \frac{\text{const}}{\sqrt{1 - v^2/c^2}}. \tag{20.2}$$

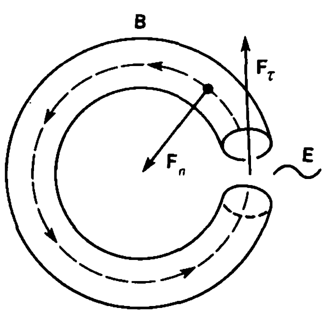


Fig. 45. Motion of a charged particle in a cyclotron.

When the proton passes through the accelerating region, its velocity increases under the action of the force $F_\tau = eE$. The change in the velocity during each cycle, i.e. the proton's acceleration a_τ, can be measured. Of course, it is not a simple task to make such measurements since the proton passes through the accelerating region under different phases of the accelerating field, i.e. for different values of E the radius of its orbit also keeps on changing. However, there is no need to discuss these problems at length here. It is obvious that these factors can be taken into account, and the electron's acceleration a_τ due to the force F_τ can be calculated. These results can be used to determine the ratio F_τ/a_τ. Experiments give the

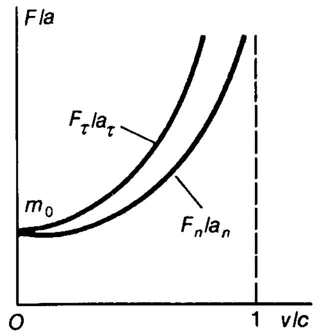


Fig. 46. Dependence of F/a on v/c for normal and tangential components of force.

dependence

$$\frac{F_\tau}{a_\tau} = \frac{\text{const}}{(1 - v^2/c^2)^{3/2}}. \tag{20.3}$$

For low velocities, i.e. for $v/c \ll 1$, (20.2) and (20.3) must be transformed into (19.2). Hence the constant quantity in these formulas is a particle's mass m_0, which is a measure of the inertia of the particle at rest and is called the rest mass. Equations (20.2) and (20.3) can be written in the following final form:

$$\frac{F_n}{a_n} = \frac{m_0}{\sqrt{1 - v^2/c^2}}, \quad \frac{F_\tau}{a_\tau} = \frac{m_0}{(1 - v^2/c^2)^{3/2}}. \tag{20.4}$$

These dependences are graphically shown in Fig. 46. The acceleration a_τ is the tangential acceleration, and the force F_τ is collinear with the tangent to the particle's trajectory. The acceleration a_n and the force F_n are normal to the velocity direction. *Equations* (20.4) *show that the inertia of a particle in the direction of its velocity is different from the inertia normal to the velocity.*

RELATIVISTIC EQUATION OF MOTION. Suppose that a particle moves along a trajectory. As in Sec. 8, we denote the unit vectors normal and tangential to the trajectory by n and τ (see Fig. 15). The total force F acting on the particle can be decomposed into normal and tangential components (Fig. 47):

$$\mathrm{F} = \mathrm{F}_n + \mathrm{F}_\tau. \tag{20.5}$$

Each component of the force produces an acceleration in the appropriate direction, determined by the inertia of a body in the direction. Since the normal acceleration is v^2/R (see (8.21), where R is the radius of curvature of the trajectory, and v is the particle's velocity) and the tangential acceleration is $\mathrm{d}v/\mathrm{d}t$, equations (20.4) for normal and tangential components of the force can be written as follows:

$$\mathrm{n}\frac{m_0}{\sqrt{1 - v^2/c^2}}\frac{v^2}{R} = \mathrm{F}_n, \quad \tau\frac{m_0}{(1 - v^2/c^2)^{3/2}}\frac{\mathrm{d}v}{\mathrm{d}t} = \mathrm{F}_\tau. \tag{20.6}$$

Adding these expressions termwise and taking into account (20.5), we obtain an equation of motion for a particle under the action of the resultant force F:

$$\mathrm{n}\frac{m_0}{\sqrt{1 - v^2/c^2}}\frac{v^2}{R} + \tau\frac{m_0}{(1 - v^2/c^2)^{3/2}}\frac{\mathrm{d}v}{\mathrm{d}t} = \mathrm{F}. \tag{20.7}$$

!

A relativistic mass as well as a nonrelativistic mass characterize the inertia of a point mass, the only difference being that in the relativistic case, the inertia of the point mass depends on velocity, while in the nonrelativistic case, this dependence can be neglected. In the relativistic case, the directions of acceleration and force do not coincide.

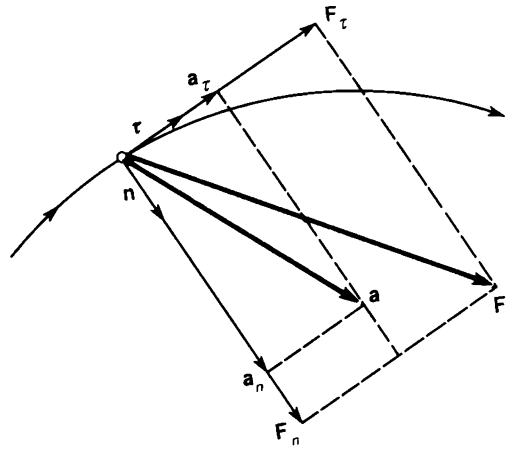


Fig. 47. In the relativistic case, the directions of force and acceleration generally do not coincide because of a difference in the inertia of a particle along its velocity and perpendicular to it.

The left-hand side of this equation can be simplified. Considering that $v(\mathrm{d}\boldsymbol{\tau}/\mathrm{d}s) = \mathrm{d}\boldsymbol{\tau}/\mathrm{d}t = (\mathrm{d}\boldsymbol{\tau}/\mathrm{d}s)(\mathrm{d}s/\mathrm{d}t)$ and representing (8.20) in the form

$$v\frac{\mathbf{n}}{R} = \frac{\mathrm{d}\boldsymbol{\tau}}{\mathrm{d}t}, \tag{20.8}$$

we can replace the quantity $\mathbf{n}v^2/R$ in (20.7) by $v\,\mathrm{d}\boldsymbol{\tau}/\mathrm{d}t$. In this case, the equation acquires the form

$$\frac{m_0}{\sqrt{1 - v^2/c^2}}\,v\frac{\mathrm{d}\boldsymbol{\tau}}{\mathrm{d}t} + \frac{m_0}{(1 - v^2/c^2)^{3/2}}\,\boldsymbol{\tau}\frac{\mathrm{d}v}{\mathrm{d}t} = \mathbf{F}. \tag{20.9}$$

By direct differentiation, we verify the equality

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{v}{\sqrt{1 - v^2/c^2}}\right) = \frac{1}{(1 - v^2/c^2)^{3/2}}\frac{\mathrm{d}v}{\mathrm{d}t},$$

which can be used to transform the left-hand side of (20.9) as follows:

$$\begin{aligned}
&\frac{m_0}{\sqrt{1 - v^2/c^2}}\,v\frac{\mathrm{d}\boldsymbol{\tau}}{\mathrm{d}t} + \frac{m_0}{(1 - v^2/c^2)^{3/2}}\,\boldsymbol{\tau}\frac{\mathrm{d}v}{\mathrm{d}t}\\
&= \frac{m_0 v}{\sqrt{1 - v^2/c^2}}\frac{\mathrm{d}\boldsymbol{\tau}}{\mathrm{d}t} + \boldsymbol{\tau}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_0 v}{\sqrt{1 - v^2/c^2}}\right)\\
&= \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_0 \mathbf{v}}{\sqrt{1 - v^2/c^2}}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_0 v\boldsymbol{\tau}}{\sqrt{1 - v^2/c^2}}\right),
\end{aligned}$$

where $v\boldsymbol{\tau} = \mathbf{v}$ is the particle's velocity. Thus, we arrive at the relativistic equation of motion of the particle

(Jules) Henri Poincaré (1854-1912)
French mathematician, physicist and philosopher, one of the founders of the special theory of relativity and relativistic dynamics. He wrote fundamental works on celestial mechanics and mathematical physics, the theory of differential equations, the theory of analytic functions and topology.

$$\frac{d}{dt}\left(\frac{m_0 \mathbf{v}}{\sqrt{1-v^2/c^2}}\right) = \mathbf{F}, \tag{20.10}$$

which is a generalization of Newton's equation of motion (19.1). It can be represented in a more convenient form analogous to (19.3a):

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}, \quad \mathbf{p} = m\mathbf{v}, \quad m = \frac{m_0}{\sqrt{1-v^2/c^2}}. \tag{20.11}$$

The quantity m is called the relativistic mass, or simply mass, m_0 is the rest mass, and p is the relativistic momentum, or simply momentum.

Usually, it is not necessary to specifically mention that the momentum is "relativistic", or that the mass is "relativistic", since when the velocities are very high, i. e. relativistic, we can use only relativistic expressions for momenta and masses. For low velocities, these expressions are automatically transformed into nonrelativistic ones.

Like nonrelativistic mass, relativistic mass characterizes the inertia of a point mass, the only difference being that in the relativistic case, the inertia of the point mass depends on velocity, while in the nonrelativistic case, this dependence can be neglected.

NONALIGNMENT OF FORCE AND ACCELERATION IN THE RELATIVISTIC CASE. Since the inertia of a body in the direction of motion is different from its value in the perpendicular direction, *the total force vector is not collinear with the total acceleration vector*, i.e. with the vector of the velocity change caused by the force as shown in Fig. 47. It can be seen from (20.11) that the momentum change vector coincides with the force vector in direction. This is why the difference between Newton's equations (19.1) and (19.3a) is formal in the nonrelativistic case and can be seen as a difference between notations.

In the relativistic case, however, the equation of motion for a point mass can be written only in the form (20.11). In principle, it can be written in the form (19.1) with mass $m = m_0/\sqrt{1-v^2/c^2}$, but then an additional term appears besides the force F on the right-hand side of the equation (see Eq. (47.11)).

?

What is the relativistic mass of a body and how do we write the relativistic equation of motion?
What factors specify the nonalignment of the direction of force and the acceleration caused by it?
Why is the rest mass an invariant?

Sec. 21. MOTION OF A SYSTEM OF POINT MASSES

The concepts and physical quantities characterizing the motion of a system of point masses are defined and the relevant equations are derived.

SYSTEM OF POINT MASSES. A system of point masses is an aggregate of a finite number of such point masses. Consequently, these point masses can be enumerated. An example of such a system is a gas contained in a certain volume if its molecules can be taken as point masses according to the conditions of the problem. The Sun and the planets constituting the solar system can be considered a system of point masses in all problems where the internal structure and the size of the Sun and the planets are not significant. The mutual arrangement of the points in such a system usually changes with time.

Each point in the system is under the action of two types of force: forces whose origin lies beyond the system, called the external forces, and forces exerted on the point by other points in the system, called the internal forces. Usually, internal forces are assumed to obey Newton's third law of motion. We shall enumerate the points using the subscripts, say, $i, j, \ldots$, each of which runs through the values 1, 2, 3, ..., n where n is the number of points in the system. We shall denote the physical quantities pertaining to the ith point by the same subscript as the point. For example, $\mathbf{r}_i$, $\mathbf{p}_i$ and $\mathbf{v}_i$ denote the radius vector, momentum and velocity of the ith point.

ANGULAR MOMENTUM OF A POINT MASS. Suppose that the position of a point mass is characterized by a radius vector $\mathbf{r}$ about point O taken as the origin. The angular momentum (moment of momentum) of the point mass about O is defined as the vector (Fig. 48)

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}. \tag{21.1}$$

This definition is valid both for relativistic and nonrelativistic momentum. In both cases, the direction of the momentum $\mathbf{p}$ coincides with that of the velocity of the point mass.

MOMENT OF FORCE ACTING ON A POINT MASS. The moment of force acting on a point mass about point O (see Fig. 48) is the vector

$$\mathbf{M} = \mathbf{r} \times \mathbf{F}. \tag{21.2}$$

As in the other cases, $\mathbf{F}$ in this formula is the resultant of all the forces acting on the point mass.

MOMENTAL EQUATION FOR A POINT MASS. Let us differentiate the moment of momentum (21.1) with respect to

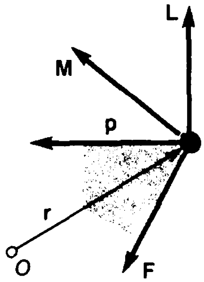


Fig. 48. Defining the angular momentum and the moment of force.

The angular momentum vector is perpendicular to the plane containing the radius vector and the momentum of the particle, while the moment of force vector is perpendicular to the plane containing the radius vector and the force. Point O is the origin of the radius vectors.

time:

$$\frac{d\mathbf{L}}{dt} = \frac{d\mathbf{r}}{dt} \times \mathbf{p} + \mathbf{r} \times \frac{d\mathbf{p}}{dt}. \tag{21.3}$$

Note that $d\mathbf{r}/dt = \mathbf{v}$ is the velocity whose direction coincides with the momentum $\mathbf{p}$, and the vector product of two parallel vectors is zero. Hence the first term on the right-hand side of (21.3) is zero, while the second term represents the moment of force (21.2) in view of the fact that $d\mathbf{p}/dt = \mathbf{F}$ according to (20.11). Consequently, Eq. (21.3) is transformed into the momental equation

$$\frac{d\mathbf{L}}{dt} = \mathbf{M}, \tag{21.4}$$

which is important for the analysis of the motion of point masses and bodies.

The momental equation for a point mass is not an independent law of motion and follows directly from Newton's laws of motion.

MOMENTUM OF A SYSTEM OF POINT MASSES. This concept is characteristic of the system as a whole. The momentum of a system of point masses is the sum of the momenta of the point masses constituting the system:

$$\mathbf{p} = \sum_{i=1}^{n} \mathbf{p}_i = \mathbf{p}_1 + \mathbf{p}_2 + \ldots + \mathbf{p}_n, \tag{21.5}$$

where $\mathbf{p}_i$ is the momentum of the point mass denoted by the subscript i, and n is the number of point masses in the system. Henceforth, we shall omit the indices on the summation symbol since the indices over which the summation is carried out are usually known even without a specific mention.

!

The momental equation for a point mass is not an independent law of motion, it follows directly from Newton's laws of motion.
The momental equation for a system of point masses is an independent law of motion and does not follow from Newton's laws of motion without additional assumptions.
In the relativistic case, the concept of the centre of mass is inapplicable since it is not an invariant of the Lorentz transformations. However, the concept of the centre-of-mass system has an exact meaning and proves to be very useful and important in mechanics.

ANGULAR MOMENTUM OF A SYSTEM OF POINT MASSES. The angular momentum of a system of point masses about point O taken as the origin is the sum of the angular momenta of the point masses constituting the system about point O:

$$\mathbf{L} = \sum_i \mathbf{L}_i = \sum_i \mathbf{r}_i \times \mathbf{p}_i, \tag{21.6}$$

where $\mathbf{L}_i = \mathbf{r}_i \times \mathbf{p}_i$ is the angular momentum of the point mass with the subscript i about the point O, defined by (21.1).

FORCE ACTING ON A SYSTEM OF POINT MASSES. This force is defined as the sum of all the forces acting on the point masses constituting the system, including the mutual forces of interaction of the point masses:

$$\mathbf{F} = \sum_i \mathbf{F}_i, \tag{21.7a}$$

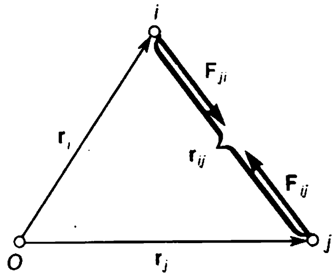


Fig. 49. Moment of the internal forces applied to points i and j is zero in accordance with Newton's third law.

where

$$F_i = F_i' + \sum_{j \neq i} F_{ji} \tag{21.7b}$$

is the force acting on the point mass with the subscript i. This force is composed of the external force F_i' acting on the point mass and the internal force $\sum_{j \neq i} F_{ji}$ acting on the point mass due to interactions with other point masses in the system. The index $j \neq i$ on the summation symbol indicates that the summation must be carried out over all values of j except the value $j = i$ since a point mass cannot act on itself. There is no need to make use of this notation if we observe that $F_{ii} = 0$.

Using Newton's third law of motion, we can considerably simplify the expression (21.7a) for the force acting on a system of point masses. Substituting (21.7b) into (21.7a), we obtain

$$F = \sum_i F_i' + \sum_i \sum_j F_{ji}.$$

The double sum in this formula can be represented in the following form:

$$\sum_i \sum_j F_{ji} = \frac{1}{2} \sum_i \sum_j (F_{ji} + F_{ij}) = 0. \tag{21.7c}$$

This is so because, in accordance with Newton's third law of motion, $F_{ji} + F_{ij} = 0$ (Fig. 49). Consequently, we obtain

$$F = \sum_i F_i = \sum_i F_i', \tag{21.7d}$$

i.e. *the force acting on a system of point masses is equal to the sum of the external forces acting on the point masses in the system.* Hence F_i in (21.7a) indicates the external forces only.

?

How are the momentum of the system of point masses and the force acting on it determined?
Is it possible to prove that the moment of internal forces in a system of point masses is zero only on the basis of Newton's laws of motion? What additional requirement should be taken to Newton's laws of motion to prove this?

MOMENT OF FORCE ACTING ON A SYSTEM OF POINT MASSES. This quantity is defined in the same manner as other quantities pertaining to the system. The moment of a force acting on a system of point masses about point O is the sum of the moments of forces applied to the point masses in the system about point O:

$$M = \sum_i M_i = \sum_i r_i \times F_i, \tag{21.8}$$

where M_i is defined by (21.2). The force F_i in (21.8) is the resultant of forces applied to the point i, including the internal forces. In other words, this force is defined by (21.7b).

Formula (21.8) can considerably be simplified under certain assumptions. Substituting (21.7b) into (21.8), we obtain

$$\mathbf{M} = \sum_i \mathbf{r}_i \times \mathbf{F}_i^e + \sum_i \sum_j \mathbf{r}_i \times \mathbf{F}_{ji}. \tag{21.9}$$

The first sum expresses the moment of external forces, while the double sum represents the moment of internal forces. The latter sum can be calculated as follows:

$$\sum_i \sum_j \mathbf{r}_i \times \mathbf{F}_{ji} = \frac{1}{2} \sum_i \sum_j (\mathbf{r}_i \times \mathbf{F}_{ji} + \mathbf{r}_j \times \mathbf{F}_{ij})$$
$$= \frac{1}{2} \sum_i \sum_j (\mathbf{r}_i - \mathbf{r}_j) \times \mathbf{F}_{ji}. \tag{21.10a}$$

Here, we have taken into account Newton's third law of motion $\mathbf{F}_{ij} + \mathbf{F}_{ji} = 0$. *Assuming that the moment of internal forces is zero*, i.e.

$$\frac{1}{2} \sum_i \sum_j (\mathbf{r}_i - \mathbf{r}_j) \times \mathbf{F}_{ji} = 0, \tag{21.10b}$$

we find that the expression (21.9) *for the moment of forces acting on a system contains only the moment of external forces.* This assumption is physically justified. In the case of central forces, for example, vectors $\mathbf{F}_{ji}$ are collinear with vectors $\mathbf{r}_i - \mathbf{r}_j$ so that each term in (21.10b) and hence the total sum of these terms are zero. Of course, the requirement that each term in the sum be zero is more rigorous than the requirement concerning the entire sum being zero. *In Newtonian mechanics, the requirement that the moment of internal forces in a system of point masses be zero is treated as a postulate supplementing Newton's laws of motion. Only under this condition can we assume that the moment of forces* (21.8) *can be reduced to the moment of external forces.*

It was mentioned in Sec. 19 that the interacting forces between two moving charges are noncentral. Charged particles constituting bodies are always in motion, and hence their interaction is noncentral. Therefore, strictly speaking, the momental equation of classical mechanics is not valid for a system of charged point masses in the basic sense of this term. In practice, however, the momental equation is found to be valid in most cases with a very high degree of accuracy. This is so because the magnitude of the noncentral part in each mutual interaction has a relativistic order of smallness relative to the magnitude of central interaction, and the sum (21.10b) of relativistically small terms vanishes on account of the statistical nature of these interactions.

In some cases, however, the moment of internal forces is clearly nonzero. As an example, we can consider a charged dielectric cylinder on whose surface coils of wires are tightly wound to form a solenoid. When an electric current varying in time is passed through the solenoid, the moment of internal forces arising in the charged dielectric cylinder-solenoid system tends to rotate the system about the cylinder's axis.

A rational interpretation of this phenomenon requires that the electromagnetic field should be taken into account, like in the case when Newton's third law of motion is not obeyed during the interaction of moving charges. Hence it can be concluded that the electromagnetic field has not only energy and momentum but also an angular momentum.

EQUATION OF MOTION FOR A SYSTEM OF POINT MASSES. Differentiating (21.5) with respect to time and considering from (20.11) that the equation of motion for the ith point has the form $\mathrm{d}\mathbf{p}_i/\mathrm{d}t = \mathbf{F}_i$, we obtain

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \sum \frac{\mathrm{d}\mathbf{p}_i}{\mathrm{d}t} = \sum \mathbf{F}_i, \quad \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{F}, \tag{21.11}$$

where

$$\mathbf{F} = \sum \mathbf{F}_i. \tag{21.12}$$

The quantity $\mathbf{F}$ is equal to the sum of the external forces since all the internal forces cancel each other in the sum (21.12) (see (21.7d)). Equation (21.11) has the same form as (20.11) for a point mass, but a different meaning since the physical carriers of the momentum $\mathbf{p}$ are distributed over the entire space occupied by the system of point masses. The points of application of the external forces constituting $\mathbf{F}$ are also distributed in a similar manner. An interpretation of (21.11) close to that of (20.11) is only possible in the nonrelativistic case.

CENTRE OF MASS. In the nonrelativistic case, i.e. for motion at low velocities, we can introduce the concept of centre of mass. To begin with, let us consider the expression for the momentum of a system of point masses in the nonrelativistic case:

$$\mathbf{p} = \sum m_{0i}\mathbf{v}_i = \sum m_{0i}\frac{\mathrm{d}\mathbf{r}_i}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\sum m_{0i}\mathbf{r}_i$$
$$= m\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{m}\sum m_{0i}\mathbf{r}_i\right), \tag{21.13}$$

where $m = \sum m_{0i}$ represents the mass of the system as the sum of the rest masses of the points constituting the system.

The radius vector

$$\mathbf{R} = \frac{1}{m}\sum m_{0i}\mathbf{r}_i \tag{21.14}$$

---

defines an imaginary point called the centre of mass of the system. The quantity $dR/dt = V$ is the velocity of motion of the imaginary point. If we take (21.14) into consideration, the momentum of the system (21.13) can be written in the form

$$p = m\frac{dR}{dt} = mV, \tag{21.15}$$

i.e. as the product of the mass of the system and the velocity of its centre of mass, exactly in the same way as the momentum of a point mass. The motion of the centre of mass can be observed in the same way as of a point mass.

Together with (21.14) and (21.15), equation (21.11) describing the motion of the system assumes the form

$$m\frac{dV}{dt} = F. \tag{21.16}$$

In this form, the equation of motion is equivalent to the equation of motion of a point mass whose mass is concentrated at the centre of mass, and all the external forces acting on the points of the system are applied to its centre of mass. The point representing the centre of mass (21.14) has a definite position relative to the point masses of the system. If the system is not a rigid body, the mutual arrangement of its points changes with the passage of time. Consequently, the position of the centre of mass also changes relative to the points of the system, but at each particular instant of time the centre of mass occupies a definite position. The expression "definite position" means that if we "glance" at the system of points from a different coordinate system at this instant, the position of the centre of mass relative to the points of the system remains unchanged. This can be proved as follows: it can be seen from the definition (21.14) of the centre of mass that if point O from which the radius vector R is measured is made to coincide with the centre of mass, R will obviously be zero. Hence if the radius vectors r_i of the individual points of the system are measured relative to the centre of mass, we obtain from (21.14)

$$\sum m_{0i} r_i = 0. \tag{21.17}$$

It should be recalled that the origin of the radius vectors r_i in Eq. (21.14) lies at an arbitrary point relative to which the position of the centre of mass of the system is given by the radius vector R.

Let us now imagine that the centre of mass has to be determined from (21.14) by using some other reference point

?

Why is the concept of the centre of mass inapplicable in the relativistic case and what does the concept of the centre-of-mass system mean?
What should be the state of motion of a point to fulfill the momental equation relative to which it is written?
Can you prove that the momental equation is valid relative to the centre of mass, although the motion of the latter is complex?

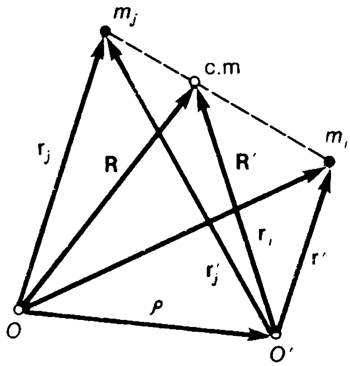


Fig. 50. Proving the invariance of the centre of mass of a system of point masses in the non relativistic case.

for radius vectors, i.e. some other coordinate system. One can ask if the same point serves as the centre of mass in this case. Let us determine the position of the centre of mass by taking the reference point at O', whose position relative to point O is characterized by the radius vector ρ (Fig. 50). The quantities corresponding to the reference point O' will be designated by primed letters. In order to determine the position of the centre of mass relative to point O', we can rewrite (21.14) as

$$\mathrm{R}' = \frac{1}{m}\sum m_{0i}\mathrm{r}_i'. \tag{21.18}$$

Considering that $\mathrm{r}_i' = \mathrm{r}_i - \rho$ and substituting this expression into (21.18), we obtain

$$\mathrm{R}' = \frac{1}{m}\sum m_{0i}\mathrm{r}_i - \frac{1}{m}\rho\sum m_{0i} = \mathrm{R} - \rho, \tag{21.19}$$

where $m = \sum m_{0i}$. Formula (21.19) shows that the radius vector R' drawn from O' terminates at the same point as the radius vector R originating from point O. This proves that the position of the centre of mass is independent of the coordinate system in which it is determined.

INAPPLICABILITY OF THE CONCEPT OF THE CENTRE OF MASS IN THE RELATIVISTIC CASE. The situation is quite different in the relativistic case. The expressions for momentum cannot be transformed in the manner indicated in (21.13) since the rest masses m_{0i} are now replaced by the relativistic masses which depend on time since the velocities vary with time. One could try to determine the centre of mass with the help of (21.14) by substituting the relativistic masses for the rest masses m_{0i} and assuming that m is the sum of the relativistic masses. This would naturally lead to a radius vector terminating at a certain point. This point could be called the centre of mass. However, there is no physical meaning of this point. If we tried to find the position of the centre of mass at a certain instant in another coordinate system, we would obtain a point having a different position relative to the points in the system. *Consequently, the concept of the centre of mass in the relativistic case is not an invariant concept independent of the choice of a coordinate system and is therefore inapplicable.* There is no sense in writing the equation of motion of this point and to observe its motion. However, the concept of the "centre-of-mass system" is generally accepted. Many relativistic concepts are considerably simplified in this system. *The centre-of-mass system is a coordinate system in which the sum of the momenta of the particles is zero.* It is always possible to find such a coordinate system. We shall do this later when considering the collisions. This system is characterized by its

?

The angular momentum and the moment of force are determined about a point. Is the state of motion of the point arbitrary?
What is the difference between the expressions for the angular momentum and the moment of force in the relativistic and nonrelativistic cases?
Under what conditions is the momental equation valid?
How do the angular momentum and the moment of force depend on the position of the point about which they are calculated?

velocity and not by the position of its origin. If the momenta of particles in this coordinate system are denoted by p_i, we must get

$$\sum p_i = 0. \tag{21.20}$$

This condition can only lead to the velocity of the coordinate system and not to its position. *Hence it can be stated that the centre-of-mass system exists in the relativistic case, although there is no centre of mass.*

MOMENTAL EQUATION FOR A SYSTEM OF POINT MASSES. Differentiating (21.6) with respect to time, we obtain the momental equation for a system of point masses:

$$\frac{dL}{dt} = \sum \frac{dr_i}{dt} \times p_i + \sum r_i \times \frac{dp_i}{dt}$$

$$= \sum v_i \times p_i + \sum r_i \times F_i = 0 + \sum M_i = M,$$

$$\boxed{\frac{dL}{dt} = M,} \tag{21.21}$$

where we have considered that the velocity vectors for a particle are parallel to the momentum vectors and have taken into account (21.8) for the moment of the force acting on the system. It should be recalled that M is the moment of the external forces applied to the system, as was explained in detail for the case of (21.8).

However, unlike Eq. (21.4) *for a point mass, Eq.* (21.21) *cannot be considered a simple corollary of Newton's laws of motion. Over and above Newton's laws, this equation also requires that internal forces should be central.* It should also be noted that (21.16) and (21.21) do not form a closed system of equations for a system of point masses. A system of N point masses has $3N$ degrees of freedom, while the number of equations given by (21.16) and (21.21) is just six. Hence $3N - 6$ more conditions or equations must be specified for a complete solution of the problem.

Example 21.1. Two point masses m_1 and m_2 are fastened on a horizontal table with the help of an absolutely unstretchable string OAB (Fig. 51) rigidly fixed at point O. A large force F is applied to the point mass m_2 for a very short time δt, imparting an impulse $\tilde{F} = F\,\delta t$ in the direction of the force F. Find the instantaneous velocities imparted to the points at the instant of impact. Frictional forces should be neglected.

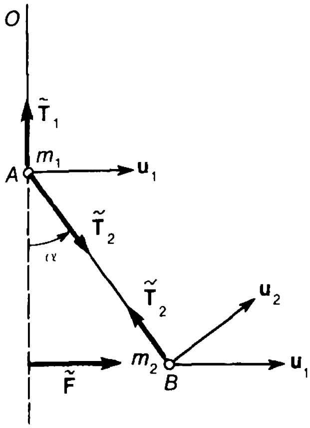


Fig. 51. Diagram of impulsive forces, tensions and velocities in Example 21.1.

The impact at the point mass m_2 produces impulsive tensions $\tilde{T}_1$ and $\tilde{T}_2$ in the absolutely unstretchable string. Obviously, the velocity u_1 of the point mass m_1 can be directed only at right angles to OA (see. Fig. 51). The velocity of the point mass m_2 is composed of the velocity u_1 and the velocity u_2 of the point mass m_2 relative to the point mass m_1. After the impact, the momentum conservation law for the closed system can be written in the form of the relations

$$\begin{aligned} &\tilde{T}_2 \sin\alpha = m_1 u_1, \quad \tilde{T}_1 - \tilde{T}_2 \cos\alpha = 0, \\ &\tilde{F} - \tilde{T}_2 \sin\alpha = m_2 (u_1 + u_2 \cos\alpha), \\ &\tilde{T}_2 \cos\alpha = m_2 u_2 \sin\alpha. \end{aligned} \tag{21.22}$$

There are four equations in four unknowns u_1, u_2, $\tilde{T}_1$ and $\tilde{T}_2$. Solving them, we get

$$\begin{aligned} u_1 &= \frac{\tilde{F} \sin^2\alpha}{m_1 + m_2 \sin^2\alpha}, \\ u_2 &= \frac{m_1 \tilde{F} \cos\alpha}{m_2 (m_1 + m_2 \sin^2\alpha)}. \end{aligned} \tag{21.23}$$

Example 21.2. In the relativistic case, a transition to the centre-of-mass system simplifies the solution of many problems of dynamics of a system of point masses. Suppose that the energy and momentum of each point of the system are specified. It is required to find the centre-of-mass system and to express in this system the energy of the system of point masses in terms of its value in the laboratory coordinate system.

As usual, we use the subscript i to enumerate the points of the system. Let us determine the momentum of the system of point masses:

$$\mathrm{p} = \sum_i \mathrm{p}_i, \tag{21.24}$$

and the energy of the system:

$$E = \sum_i E_i. \tag{21.25}$$

In the relativistic case, the concept of centre of mass does not exist, but the centre-of-mass system does exist and is characterized by the condition $\mathrm{p} = 0$.

The energy- and momentum transformations expressed by (13.34) are linear transformations. Hence their validity for the energy and momentum of a point mass leads to their applicability to the energy and momentum of a system of point masses described by the linear relations (21.24) and (21.25). This means that the energy and momentum in (13.34) signify the energy and momentum of both a point mass and a system of point masses. It is expedient to write these transformations in an incoordinate vector form. Let us denote the momentum component in the direction of the velocity by $\mathrm{p}_{\|}$, and the momentum component in a plane perpendicular to the direction of the velocity by $p_\perp$. The transformations (13.34) then assume the form

$$p'_\parallel = \frac{p_\parallel - vE/c^2}{\sqrt{1 - v^2/c^2}}, \quad p'_\perp = p_\perp, \quad E' = \frac{E - v \cdot p}{\sqrt{1 - v^2/c^2}}. \tag{21.26a}$$

The energy and momentum of a system of point masses are denoted by E and p in the laboratory coordinate system, and by E' and p′ in the centre-of-mass system. By definition of the centre-of-mass system, $p' = 0$. The first two equations in (21.26a) assume the form

$$p'_\parallel = \frac{p_\parallel - vE/c^2}{\sqrt{1 - v^2/c^2}} = 0, \quad p'_\perp = p_\perp = 0. \tag{21.26b}$$

This means that the velocity v is collinear with p and

$$v = \frac{pc^2}{E}. \tag{21.27}$$

Thus, the centre-of-mass system is a coordinate system which moves relative to the laboratory coordinate system at a velocity v determined by (21.27). Considering (see Chap. 6) that $p = \text{const}$ and $E = \text{const}$ for an isolated system of point masses, we conclude from (21.27) that the centre-of-mass system is an inertial system.

In order to determine the expression for energy in the centre-of-mass system, we make use of the invariant (13.35):

$$p'^2 - \frac{E'^2}{c^2} = p^2 - \frac{E^2}{c^2}. \tag{21.28}$$

Considering that $p' = 0$, we obtain from (21.28)

$$E'^2 = E^2 - p^2c^2. \tag{21.29}$$

It follows from (21.27) that $p^2 = E^2v^2/c^4$, and hence (21.29) assumes the form

$$E'^2 = E^2(1 - v^2/c^2). \tag{21.30}$$

Finally, we arrive at the expression

$$E' = E\sqrt{1 - v^2/c^2}. \tag{21.31}$$

PROBLEMS

5.1. A ring can slide without friction along an undeformable rod. The rod rotates at an angular velocity ω in the vertical plane about a horizontal axis passing through one of its ends at right angles to the rod. The distance between the ring and the axis of rotation is denoted by r. Assuming that the ring is in a state of rest at the point $r = 0$ at the instant $t = 0$ and that the rod points vertically downwards from the axis of rotation, find $r(t)$.

5.2. The flat base of a right circular cylinder of radius r_0 is rigidly fastened to a horizontal table. A thin weightless unstretchable thread of length $2l$ is rigidly fastened to the cylinder's surface at its base and is tightly wound on the cylinder near the base at a height l. (The thickness of the thread is neglected.) A mass m fixed to the other end of the thread can move in a horizontal plane so that the thread is always under tension. At the initial moment $t = 0$, a velocity u is imparted to the mass in a direction perpendicular to the part of the thread unwound on the cylinder so that the thread is unwound during the motion of the mass. In what time will the thread be completely unwound from the cylinder and what will be the time dependence of the tension of the thread?

5.3. A man puts a shot of mass m from shoulder height with his hand pointing vertically upwards. By applying a certain force, a man whose height up to the shoulder is 1.5 m and whose hand is 0.75 m long can throw the shot to a height of 4 m above the Earth's surface. To what height will the shot be thrown on the Moon's surface under the same conditions? The free fall acceleration on the Moon is about one-sixth that on the Earth.

5.4. A point mass begins to slide without friction from the top of a rigidly fastened sphere along its surface. The position of the sliding point can be characterized by the angle θ between the vertical and the radius vector joining the point to the centre of the sphere. For what value of θ will the point mass lose contact with the sphere's surface and begin to fall freely under the action of gravity?

5.5. Two point masses m_1 and m_2 are connected through a weightless unstretchable thread passing through a ring of mass m_3. The masses m_1 and m_2 are in close proximity on the surface of a table, while the ring is suspended by threads down the edge of the table. The threads are perpendicular to the edge, and there is no friction. Find the acceleration of the ring.

5.6. A perfectly rigid straight pipe OA of length l rotates in a horizontal plane about a vertical axis passing through point O at a constant angular velocity ω. A ball starts rolling from point O without friction along the pipe at an initial velocity u. Find the angle between the directions of the pipe and of the ball at the instant when the ball rolls off the pipe.

5.7. Two point masses m_1 and m_2 are connected through a perfectly rigid weightless rod and lie on an absolutely smooth horizontal surface. An impulsive force $\tilde{F}$ is applied to the point mass m_1 (see Example 21.1) at an angle α to the line joining the points where the masses m_1 and m_2 are located. Find the magnitudes of the velocities of the masses m_1 and m_2 immediately after the impact.

ANSWERS

5.1. $(g/(2\omega^2))(\cosh \omega t - \cos \omega t)$. 5.2. $3l^2/(2r_0u)$, $mu^2/\sqrt{l^2 + 2r_0ut}$. 5.3. 16.5 m. 5.4. $\arccos\ (2/3)$. 5.5. $m_3(m_1 + m_2)g/[4m_1m_2 + m_3(m_1 + m_2)]$. 5.6. $\arctan\ (l\omega/\sqrt{u^2 + l^2\omega^2})$. 5.7. $\{[\tilde{F}\cos\alpha/(m_1 + m_2)]^2 + [\tilde{F}\sin\alpha/m_1]^2\}^{1/2}$, $\tilde{F}\cos\alpha/(m_1 + m_2)$.

Chapter **6**

Conservation Laws

Basic idea:
Conservation laws hold for isolated systems and can mathematically be reduced to the first integrals of the equations of motion in mechanics. On the whole, the conservation laws for isolated systems express the fundamental properties of space and time, viz. the homogeneity and isotropy of space and the homogeneity of time.

Sec. 22. SIGNIFICANCE AND ESSENCE OF CONSERVATION LAWS

It is shown that the conservation laws in mechanics can be reduced to integrals of the equations of motion.

ESSENCE OF CONSERVATION LAWS. In principle, the laws of motion can provide answers to all questions concerning the motion of point masses and bodies. With enough skill and patience, one can determine the position of point masses at any instant of time, which means that the problem is completely solved. The possibilities of solving such equations have considerably been increased after the appearance of computers. For example, many problems associated with the motion of artificial satellites of the Earth and interplanetary flight of rockets could have been correctly formulated long ago in the form of equations, but it was only after the advent of computers that these equations could be solved with a view to obtain the necessary information. However, even at present, there are problems that can be formulated in terms of equations, but cannot be solved even with the help of computers. Hence the investigation of general properties of solutions of equations without obtaining the solution in concrete form continued to be important.

For example, suppose that we are interested in the motion of a body, but cannot solve the equation of its motion, and hence are unaware not only of its whereabouts at a particular instant of time, but also whether the body will remain near the Earth's surface during its motion or leave the Earth and set out on an interplanetary voyage. If, without solving the equations of motion, we can establish that the body will move near the

Earth and predict that it will not depart from the Earth's surface under any condition by more than, say, 10 km, it will mean a considerable progress. If on top of this we can establish that the velocity of the body will be zero at a height of 10 km and indicate the direction of the body's velocity on the Earth's surface at this height, then we essentially know everything about this motion for certain purposes, and there is no need to solve the equation of motion.

Conservation laws allow us to consider the general properties of motion without solving the equations of motion and without a detailed information about the evolution of the process. The analysis of the general properties of motion is carried out within the framework of the solutions of equations of motion and cannot contain more information than in the equations of motion. Hence the conservation laws do not contain more information than the equations of motion. However, the required information is contained in the equations of motion in such an implicit form that it is not easy to perceive it directly. Conservation laws allow this latent information to be presented in an observable form in which it can conveniently be used. An important feature of this information is its general nature: it can be applied to any specific motion irrespective of its detailed characteristics.

The general nature of conservation laws makes it possible to use them not only when the equations of motion are known and their solution is not known, but also when the equations of motion are unknown. This often helps reveal the same important features of motion without knowing the law according to which the forces act.

!

Conservation laws in mechanics are reduced to the first integrals of the equations of motion. However, their physical nature is based on the fundamental properties of space and time. Therefore the significance of the conservation laws is far beyond the scope of mechanics, they are fundamental laws of physics.

EQUATIONS OF MOTION AND CONSERVATION LAWS. Equations of motion are equations describing the change of physical quantities in space and time. We can imagine the motion as an infinite sequence of physical situations. Actually, we are not interested in any particular situation at any specific instant of time which is not associated with motion, but just in the sequence of situations through which the motion is accomplished. While considering the sequence of situations, we are interested not only in their contrasting features, but also in their common features and in quantities that are conserved in the processes. *Conservation laws provide an answer as to which quantities remain constant in a sequence of physical situations described by the equations of motion.* Clearly, the physical theory must formulate this constancy as the constancy of numerical values of the corresponding physical quantities or, as they are sometimes called, in the form of conservation laws.

?

Is it possible on the basis of the conservation laws to say how a motion will occur?
Is it possible on the basis of the conservation laws to judge whether a given motion of a point is, in principle, possible or impossible?

MATHEMATICAL MEANING OF THE MECHANICAL CONSERVATION LAWS. Let us consider as an example

Newton's one-dimensional equation of motion, which can be written in the form of two equations:

$$m_0 \frac{dv_x}{dt} = F_x, \quad (22.1a)$$

$$\frac{dx}{dt} = v_x. \quad (22.1b)$$

The problem is completely solved if the position of a moving point mass is known at any instant of time. Hence to solve this problem, we must first integrate (22.1a) and obtain v_x, after which we can consider v_x to be a known quantity and integrate (22.1b) to determine $x(t)$.

For a very wide range of forces, the first integration can be carried out in the general form, and the result can be represented as the constancy of the numerical value of a certain combination of physical quantities. This is nothing but the conservation law. *Thus, the conservation laws in mechanics can mathematically be reduced to the first integrals of the equations of motion.*

However, the significance of the quantities being conserved extends beyond the scope of mechanics and plays an important role in other fields. The physical quantities being conserved are fundamental, and their conservation laws are fundamental laws of physics and not just a result of mathematical manipulation of the equations of mechanical motion.

Sec. 23. MOMENTUM CONSERVATION LAW

The concept of an isolated system is introduced and the physical meaning of the momentum conservation law valid for such a system is discussed.

ISOLATED SYSTEM. A system of point masses or a point mass is called isolated if it is not subjected to any external forces. There can be no isolated systems in the Universe in the absolute sense of the term since all bodies are mutually connected, say, through gravitational forces. However, under certain conditions, we can treat bodies as isolated to a considerable extent. For example, a body in a certain region of space far from massive celestial bodies behaves like an isolated system. In other cases, the motion of a system in certain directions can be treated as the motion of an isolated system, although the system is certainly not isolated on the whole.

MOMENTUM CONSERVATION LAW FOR AN ISOLATED SYSTEM. There are no external forces in an isolated system. Hence we can put the force $\mathbf{F} = 0$ in the equation of motion (21.11), which then assumes the form

$$\frac{d\mathbf{p}}{dt} = 0. \quad (23.1)$$

Integrating this equation, we get

$$\boxed{\mathrm{p} = \text{const},} \tag{23.2}$$

$$\boxed{p_x = \text{const}, \quad p_y = \text{const}, \quad p_z = \text{const}.}$$

This equality expresses the momentum conservation law:

The momentum of an isolated system does not change during any process taking place in the system.

For a point mass, the momentum conservation law means that it moves in a straight line at a constant velocity in the absence of external forces. For a system of point masses, the momentum conservation law in the nonrelativistic case means that the centre of mass of the system moves uniformly in a straight line.

The momentum conservation law (23.2) is valid in both relativistic and nonrelativistic cases. In the relativistic case, however, it cannot be interpreted as a uniform motion of the centre of mass in a straight line since there is no centre of mass in this case (see Sec. 21). However, there is a centre-of-mass system in which the momentum conservation law reduces to the equality $\mathrm{p} = 0$ and indicates that the system remains a centre-of-mass system for any process taking place in it.

CONSERVATION LAWS FOR INDIVIDUAL PROJECTIONS OF MOMENTUM. It may so happen that a system of point masses or a point mass is not isolated, but the external forces act only in certain directions and are absent in the other directions. Then by an appropriate choice of the coordinate system, we can ensure that one or two projections of external forces vanish. Suppose, for example, that there is no force in the directions parallel to the XY-plane, i.e. $F_x = 0$, $F_y = 0$ and $F_z \neq 0$. Then the equation of motion (20.11) in terms of the components of quantities along the coordinate axes has the following form:

$$\frac{\mathrm{d}p_x}{\mathrm{d}t} = 0, \quad \frac{\mathrm{d}p_y}{\mathrm{d}t} = 0, \quad \frac{\mathrm{d}p_z}{\mathrm{d}t} = F_z. \tag{23.3}$$

Integrating the first two equations, we obtain

$$p_x = \text{const}, \quad p_y = \text{const}. \tag{23.4}$$

This means that the momentum of the system in the directions parallel to the XY-plane preserves its value, and the system behaves as an isolated system in these directions. For example, the gravitational forces near the Earth's surface act in the vertical direction, and there are no components along the horizontal. Hence a system of bodies can be treated as an

?

In what case can the momentum conservation law be applied to a nonisolated system?
A frictional force acts on a system of billiard balls moving along a table, and hence the system is not isolated from horizontal motions. Can the momentum conservation law be applied to the colliding balls? Why?
Can we consider a system of interacting electric charges to be an isolated system in general? What should we take into account in this case?

isolated system for horizontal motion as far as the gravitational forces are concerned.

APPLICATION OF THE MOMENTUM CONSERVATION LAW. Examples of the application of the momentum conservation law to solve specific problems will be considered in the next chapters. Here, we shall analyze an example of the ballistic pendulum which is a small body of mass m_1 suspended on a long thread (Fig. 52). The size of the body is such that when a bullet of mass m_2 travelling at a velocity v hits the body and is stuck in it, the ball is deflected. What will be the velocity u of the body with the bullet stuck in it? If we try to analyze the penetration of the bullet into the body, find the time dependence of the forces emerging in this case and then solve the equation of motion, it will require a lot of efforts. Even then the results would not be reliable since many assumptions will have to be made in order to obtain them and it is quite difficult to substantiate these assumptions rigorously.

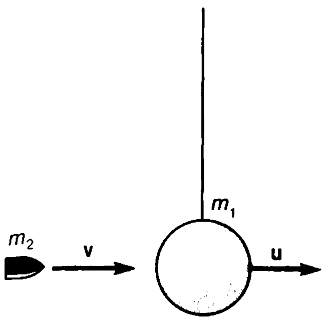


Fig. 52. Determining the bullet's velocity with the help of a ballistic pendulum.

If the body is quite small and can be replaced by a point mass, we can use the momentum conservation law for two interacting point masses without going into the details of the manner in which the bullet penetrates the body. In other words, we can write $m_2 v = (m_1 + m_2)u$ (see Fig. 52). The easiest way to determine the velocity u is to measure the height to which the body and the bullet in it rise and to substitute this value into the equation $(m_1 + m_2)u^2/2 = (m_1 + m_2)gh$ expressing the energy conservation law. This gives $v = \sqrt{2gh}(m_1 + m_2)/m_2$. However, such an approach is incorrect for a body of finite size, although it is widely used. The correct approach to this problem will be described in Sec. 34.

Sec. 24. ANGULAR MOMENTUM CONSERVATION LAW

The formulation and the conditions of applicability of the angular momentum conservation law are discussed.

FORMULATION OF THE LAW. Like the momentum conservation law, this law is also valid for isolated systems only. For such systems, the moment M of external forces is zero, and the momental equation (21.21) assumes the form

$$\frac{d\mathbf{L}}{dt} = 0. \tag{24.1}$$

Integrating this equation, we obtain

$$\boxed{\mathbf{L} = \text{const},} \tag{24.2}$$

$$\boxed{L_x = \text{const}, \quad L_y = \text{const}, \quad L_z = \text{const}.}$$

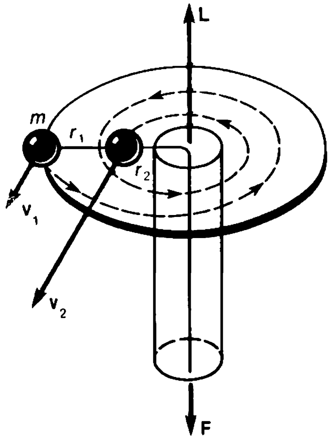


Fig. 53. An increase in the angular and linear velocities of a point mass in accordance with the angular momentum conservation law is due to a decrease in the distance of a rotating point mass from the axis of rotation under the action of a force applied to it.

This equation expresses the angular momentum conservation law:

The angular momentum of an isolated system does not change during any process taking place in the system.

CONSERVATION LAWS FOR INDIVIDUAL PROJECTIONS OF ANGULAR MOMENTUM. It may so happen that a system is not completely isolated, but the projection of the moment of force is zero, e. g. along the Z-direction. In this case, the momental equation (21.21) for the projections can be written in the following form:

$$\frac{dL_x}{dt} = M_x, \quad \frac{dL_y}{dt} = M_y, \quad \frac{dL_z}{dt} = 0. \tag{24.3}$$

Consequently, the system can be considered isolated only with respect to the z-projection of the angular momentum:

$$L_z = \text{const}. \tag{24.4}$$

Hence like the momentum conservation law, the angular momentum conservation law can be applied not only to completely isolated systems, but also to partially isolated ones.

Example 24.1. A thread is passed through a rigidly fastened tube. A body of mass m is suspended at the end of the thread and can rotate in a circle with the axis of rotation coinciding with the axis of the tube (Fig. 53). Suppose that at the initial instant of time the body rotates in a circle of radius r_1 at a velocity v_1. A force F is then applied to the thread, and as a result the body of mass m begins to move along a spiral with a decreasing radius and at a varying velocity. At the end of the process, the body moves in a circle of given radius r_2. Find the velocity v_2 of the body.

It is difficult to solve this problem using the equations of motion. As the body moves along the spiral, the force acting along the radius is directed at an angle to the velocity which consequently increases. Using the relevant data, we can calculate the change in the velocity and find the velocity v_2. However, it is much easier to solve the problem using the angular momentum conservation law. The force acting on the mass m is always directed along the radius, and hence its moment (21.2) is zero. Consequently, the angular momentum is conserved. In the present case, the angular momentum L is parallel to the axis of rotation and is $r_1 m v_1$ at the initial instant. It must have the same value, i. e. $r_1 m v_1 = r_2 m v_2$ at the final instant. This gives the required velocity to the body $v_2 = r_1 v_1 / r_2$.

?

In what case can the angular momentum conservation law be applied to a nonisolated system?

What is the property of space that specifies the validity of the angular momentum conservation law?

What physical circumstances specify whether the angular momentum conservation law can be applied to a nonisolated system?

Sec. 25. ENERGY CONSERVATION LAW

The energy conservation law is formulated and related concepts are discussed.

WORK DONE BY A FORCE. If the absolute value of velocity changes under the action of a force, work is said to be done by the force. The work of force is considered positive if the velocity increases, while the work of force is considered negative if the velocity decreases.

Let us find the relation between the work and change in velocity. We shall first consider the one-dimensional case when the force acts in the X-direction and the motion is in the same direction. For example, suppose that a point mass m_0 is displaced under the action of the force of compression or extension of a spring fastened at the origin of the coordinate system, viz. point O (Fig. 54). The equation of motion of the point has the form

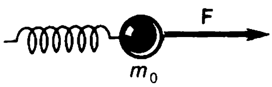


Fig. 54. Computing the work done by a force during one-dimensional motion.

$$m_0 \frac{dv_x}{dt} = F_x. \tag{25.1}$$

Multiplying both sides of this equation by v_x and considering that $v_x(dv_x/dt) = (1/2)(dv_x^2/dt)$, we get

$$\frac{d}{dt}\left(\frac{m_0 v_x^2}{2}\right) = F_x v_x. \tag{25.2}$$

We replace v_x on the right-hand side of this equation by dx/dt and multiply both sides by dt. This gives

$$d\left(\frac{m_0 v_x^2}{2}\right) = F_x dx. \tag{25.3}$$

In this form, the equation has a clear meaning: as the point is displaced by dx, the force performs the work $F_x dx$; consequently, the quantity $m_0 v_x^2/2$ characterizing the motion of the point and, in particular, the magnitude of its velocity, also changes. The quantity $m_0 v_x^2/2$ is called the kinetic energy of the point. If the point moves from position x_1 to x_2 and its velocity changes from v_{x1} to v_{x2}, the integration of (25.3) gives

$$\int_{v_x = v_{x1}}^{v_x = v_{x2}} d\left(\frac{m_0 v_x^2}{2}\right) = \int_{x_1}^{x_2} F_x dx. \tag{25.4}$$

Considering that

$$\int_{v_x = v_{x1}}^{v_x = v_{x2}} d\left(\frac{m_0 v_x^2}{2}\right) = \frac{m_0 v_{x2}^2}{2} - \frac{m_0 v_{x1}^2}{2},$$

we finally obtain

$$\boxed{\frac{m_0 v_{x2}^2}{2} - \frac{m_0 v_{x1}^2}{2} = \int_{x_1}^{x_2} F_x dx.} \tag{25.5}$$

The change in the kinetic energy of a point mass as a result of its motion from one position to another is equal to the work done by the force during this displacement.

The integral on the right-hand side of (25.5) is the limit of the sum of elementary works done during elementary displacements. The entire interval between points x_1 and x_2 is divided into subintervals Δx_i ($x_2 - x_1 = \sum \Delta x_i$) on each of which the force has a certain value F_{xi} (irrespective of the point on each of these subintervals where the value of the force F_{xi} is taken). The elementary work done over each of the subintervals Δx_i is $\Delta A_i = F_{xi}\Delta x_i$, and the total work done for the entire displacement from x_1 to x_2 will be

$$\sum_i F_{xi}\Delta x_i . \tag{25.6}$$

Making the length of each of the subintervals Δx_i tend to zero and their number to infinity, we obtain the work done by the force in displacing the point from x_1 to x_2:

$$A = \lim_{\Delta x_i \to 0} \sum_i F_{xi}\Delta x_i = \int_{x_1}^{x_2} F_x \,\mathrm{d}x . \tag{25.7}$$

It can be seen from (25.5) that the kinetic energy of a point mass changes if the forces are not zero. Thus, the kinetic energy is not conserved if a force is applied to the point mass. It remains constant only in the absence of a force since for $F_x = 0$ we obtain from (25.5)

$$\frac{m_0 v_{x2}^2}{2} = \frac{m_0 v_{x1}^2}{2} = \text{const} . \tag{25.8}$$

But in the absence of force this kinetic energy conservation law of a point mass is trivial since under these conditions the momentum conservation law itself establishes the constancy of velocity, and hence of its square as well.

If the displacement of a point mass does not coincide in direction with the force, work is done by the component of force along the displacement. Work is equal to the absolute value of force multiplied by the cosine of the angle between the force and the displacement. Since the elementary displacement of a point mass is $\mathrm{d}\mathbf{l}$ and the force $\mathbf{F}$ is also a vector, the elementary work can be represented in the form of the scalar product:

$$\mathrm{d}A = F\,\mathrm{d}l \cos(\widehat{\mathbf{F}, \mathrm{d}\mathbf{l}}) = \mathbf{F}\cdot \mathrm{d}\mathbf{l} . \tag{25.9}$$

Suppose that a point mass moves not in a straight line, as in (25.1), but along an arbitrary trajectory (Fig. 55). In this case, the work done by the force in displacing the point from position *1* to position *2* is also expressed as the limit of the sum

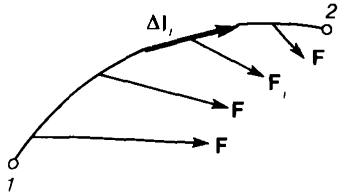


Fig. 55. Computing the work done by a force during motion along an arbitrary trajectory.

of elementary works (25.9) over the entire path. We divide the trajectory into subintervals $\Delta \mathbf{l}_i$ like the one shown in Fig. 55. The elementary work over each of these subintervals is $\Delta A_i = \mathbf{F}_i \cdot \Delta \mathbf{l}_i = F_i \, \Delta l_i \cos(\widehat{\mathbf{F}_i, \Delta \mathbf{l}_i})$.

The sum of all the elementary works is approximately equal to the work done in displacing the point from position *1* to position *2*. Making the length of each of the subintervals $\Delta \mathbf{l}_i$ tend to zero and their number to infinity, we obtain the work done by the force in moving a point mass along an arbitrary trajectory:

$$A = \lim_{\Delta \mathbf{l}_i \to 0} \sum_i \mathbf{F}_i \cdot d\mathbf{l}_i = \int\limits_{\substack{(1)\\L}}^{(2)} \mathbf{F} \cdot d\mathbf{l}. \tag{25.10}$$

The integral on the right-hand side of (25.10) is called the line integral along the curve L joining points *1* and *2*. In terms of the notation of the integration limits, the letter L indicates a specific line joining points *1* and *2*. This sign is usually omitted since we know the particular line along which the integration is carried out. The sequence of points (1) below and (2) above the symbol indicates the direction in which the point mass moves along this curve (in the present case, from point *1* to point *2*). Of course, we can move along the same curve from point *2* to point *1*, but in this case, the integration limits should be reversed in (25.10). If we reverse the direction of motion along the curve, only the sign of the integral changes. This is evident because the direction of all the elementary displacements $d\mathbf{l}_i$ is reversed, while the force remains the same at each point, and hence the signs of all the elementary works $\mathbf{F} \cdot d\mathbf{l}$ are reversed.

While considering the general case, we must replace (25.1) by a more general equation of motion:

$$m_0 \frac{d\mathbf{v}}{dt} = \mathbf{F}. \tag{25.11}$$

Subsequent computations are carried out in the same way as in the case of (25.1). Forming a scalar product of both sides of (25.11) and the velocity $\mathbf{v} = d\mathbf{r}/dt$, and considering that

$$\mathbf{v} \cdot \frac{d\mathbf{v}}{dt} = \frac{d}{dt}\left(\frac{\mathbf{v} \cdot \mathbf{v}}{2}\right) = \frac{d}{dt}\left(\frac{v^2}{2}\right), \tag{25.12}$$

we obtain an equation analogous to (25.3):

$$d\left(\frac{m_0 v^2}{2}\right) = \mathbf{F} \cdot d\mathbf{r}. \tag{25.13}$$

Vector $d\mathbf{r}$ has the same meaning as vector $d\mathbf{l}$. We used $d\mathbf{l}$ in (25.10) to emphasize that the integral is determined only by the

!

By definition, a line integral is the same as an integral of a variable. We should only divide the integration path into subintervals, calculate the value of the integrand for each subinterval and then the sum of the values for all the subintervals of the curve, and find the limit of the sum as the value of each subinterval tends to zero and their number to infinity.

line along which the integration is carried out and by the forces acting on the points of the line, and is independent of the position of the point relative to which the radius vector is measured.

Integrating both sides of (25.13) along the trajectory of a point mass between its positions *1* and *2*, we obtain

$$\frac{m_0 v_2^2}{2} - \frac{m_0 v_1^2}{2} = \int_{(1)}^{(2)} \mathbf{F} \cdot d\mathbf{l}. \tag{25.14}$$

The same remarks can be made concerning this equation as those made in connection with (25.8) if we consider that the trajectory of a moving point is a straight line in the absence of forces.

It can be stated that (25.14) *expresses the law of energy conservation if we take into consideration not just the mechanical forms of energy, but all other possible forms as well, i.e. if we extend the analysis of the problem beyond mechanics.* As a matter of fact, the right-hand side of this equation contains a quantity having the dimensions of energy. However, it may not be possible to find the physical meaning of this quantity within the framework of mechanics since it has a different, nonmechanical, origin. For example, if the force is frictional, the integral on the right-hand side of (25.14) expresses in certain units the degree of heating of the medium offering resistance to a body. A considerable amount of effort was required to find the form of energy called heat.

In many cases, however, the properties of forces are such that the right-hand side of (25.14) has a clear meaning within the framework of mechanics. Only such cases are of interest in mechanics and will be considered here.

POTENTIAL FORCES. Depending on their properties, forces can be divided into two categories. For forces of one type, the work of displacement between two points is independent of the path along which the displacement takes place, while for forces of the other type, the work depends on the path along which the displacement takes place.

By way of an example, let us consider the force of dry friction which acts against the velocity, but is independent of it within certain limits. Clearly, the work of the force is proportional to the length of the trajectory, and hence depends on the trajectory along which the body moves from one point to another.

Another well-known example is that of the work done in displacing a load from one point to another in the Earth's gravitational field. This work depends only on the difference

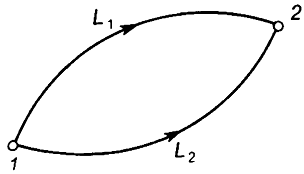


Fig. 56. Proving the equivalence of the path independence of work and the equality to zero of work in any closed path.

between the heights of the points and is independent of the specific shape of the trajectory, its length, etc.

A force whose work depends only on the initial and final points of the trajectory and which is independent of the shape of the trajectory is called a potential force. Gravitational forces belong to this category.

Instead of the expression "potential forces", the term "potential fields" is frequently employed. The field of a force is the region in space over which the force under consideration acts. The term "force" is often omitted from the expression "force field".

MATHEMATICAL CRITERION FOR THE POTENTIAL NATURE OF A FIELD. A field is called a potential field if the work done in this field, i.e. the integral

$$\int_{(1)}^{(2)} \mathbf{F}\cdot d\mathbf{l}, \tag{25.15}$$

depends only on the positions of points *1* and *2* and is independent of the path connecting the points. We can give a different mathematical expression for this definition: let us join points *1* and *2* by two different curves L_1 and L_2 (Fig. 56). According to the definition of the potential field, we can write

$$\underset{L_1}{\int_{(1)}^{(2)}} \mathbf{F}\cdot d\mathbf{l} = \underset{L_2}{\int_{(1)}^{(2)}} \mathbf{F}\cdot d\mathbf{l}. \tag{25.16}$$

Here, the integration paths between points *1* and *2* are different. If we move along the path L_2 from point *2* to point *1*, the sign of the integral will be reversed:

$$\underset{L_2}{\int_{(1)}^{(2)}} \mathbf{F}\cdot d\mathbf{l} = -\underset{L_2}{\int_{(2)}^{(1)}} \mathbf{F}\cdot d\mathbf{l}. \tag{25.17}$$

It should be noted that the direction of the path along which the integration is carried out has no relation whatsoever with the direction of motion of point masses. The evaluation of an integral is a purely mathematical operation. For example, the direction of the integration path coincides with the actual direction of motion of the point mass on the right-hand side of (25.14). However, nothing can stop us from putting a minus sign in front of the integral and evaluating it along the path in the reverse direction.

!

The equation of motion is always solved by means of two quadratures in the case of unidimensional motion when we know the force which depends only on coordinates. Any force which depends only on coordinates is a potential force in the unidimensional case.

Using (25.17), we can rewrite (25.16) in the form

$$\underset{L_1}{\int_{(1)}^{(2)}} \mathbf{F}\cdot d\mathbf{l} + \underset{L_2}{\int_{(2)}^{(1)}} \mathbf{F}\cdot d\mathbf{l} = 0. \tag{25.18}$$

The left-hand side is the sum of two integrals: the displacement in the first integral takes place from point *1* to point *2* along the path L_1, while the second integral indicates a return to the initial point along the path L_2. As a result, we obtain an integral along a closed contour, and (25.18) assumes the form

$$\oint \mathbf{F} \cdot d\mathbf{l} = 0. \tag{25.19}$$

The circle on the integral sign means that the integration is carried out over a closed contour. There is no need to specify this closed contour since it is obvious even otherwise. If it is necessary to distinguish between the contours, this can be done by using appropriate symbols under the integral. The initial definition of the potential field involved arbitrary paths connecting arbitrary points. Hence we can choose an arbitrary contour for the closed contour in (25.19).

The statement contained in Eq. (25.19) can be expressed in the following form.

(1) *A potential field is that in which the work done by the field forces over any closed contour is zero.*

Alternatively, we can present this statement in the form of a criterion.

(2) *The necessary and sufficient condition for a field to be a potential field is that the work done by the field forces over any closed contour be zero.*

WORK IN A POTENTIAL FIELD. We shall now use a mathematical theorem whose statement will be given without a proof: if F_x, F_y and F_z are the projections of a potential force, there is a function $E_p(x, y, z)$ such that these projections can be expressed through the formulas

$$F_x = -\frac{\partial E_p}{\partial x}, \quad F_y = -\frac{\partial E_p}{\partial y}, \quad F_z = -\frac{\partial E_p}{\partial z}. \tag{25.20}$$

The derivatives $\partial E_p/\partial x$, etc. are called partial derivatives and can be determined in the same way as normal derivatives of functions of single arguments assuming that all the remaining arguments of the functions are constants and have no connection with differentiation with respect to the given argument. For example, while evaluating $\partial E_p/\partial x$, we differentiate the function E_p with respect to x, assuming that y and z are constants.

Using the function E_p, we can calculate the work of the force on the right-hand side of (25.14). For this purpose, we first write the elementary work by considering that the projections of the displacement $d\mathbf{l}$ on the coordinate axes are dx, dy and dz

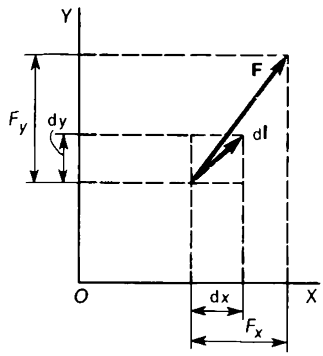


Fig. 57. Computing the work done by a force in a potential field.

Components of the quantities are shown for X- and Y-axes only.

(Fig. 57):

$$\mathbf{F}\cdot d\mathbf{l} = F_x dl_x + F_y dl_y + F_z dl_z = F_x dx + F_y dy + F_z dz. \tag{25.21}$$

Expressing the projections of the force in accordance with (25.20), we obtain

$$\mathbf{F}\cdot d\mathbf{l} = -\frac{\partial E_p}{\partial x}dx - \frac{\partial E_p}{\partial y}dy - \frac{\partial E_p}{\partial z}dz. \tag{25.22}$$

It is well known from the theory of functions of a single variable that $df = (\partial f/\partial x)dx$ is the differential of the function, which expresses the increment of the function upon a change of the argument x by dx. Hence, by analogy, we consider the quantity $(\partial E_p/\partial x)dx$ as the increment of E_p upon a change of the argument x by dx, assuming that the other arguments remain constant. For a displacement by dl, the total increment of E_p is obtained as the sum of the increments $(\partial E_p/\partial x)dx$, $(\partial E_p/\partial y)dy$ and $(\partial E_p/\partial z)dz$, caused by the respective displacements along the X-, Y- and Z-axes:

$$dE_p = \frac{\partial E_p}{\partial x}dx + \frac{\partial E_p}{\partial y}dy + \frac{\partial E_p}{\partial z}dz. \tag{25.23}$$

This increment is called the total differential, and the expression (25.22) for the elementary work assumes the form

$$\mathbf{F}\cdot d\mathbf{l} = -dE_p. \tag{25.24}$$

Integrating, we obtain the work done over the displacement from point *1* to point *2*:

$$\int_{(1)}^{(2)} \mathbf{F}\cdot d\mathbf{l} = -\int_{(1)}^{(2)} dE_p = -(E_{p2} - E_{p1}), \tag{25.25}$$

where E_{p1} and E_{p2} are the values of the function E_p at points *1* and *2*. Formula (25.25) clearly shows that the work in this case depends only on the initial and final points of the trajectory and is independent of the shape of the trajectory.

Taking (25.25) into consideration, we obtain instead of (25.14)

$$\frac{m_0 v_2^2}{2} - \frac{m_0 v_1^2}{2} = -(E_{p2} - E_{p1}). \tag{25.26}$$

Thus, the kinetic energy has changed between points *1* and *2* by the same amount as the quantity E_p (with opposite sign) as a result of a displacement between the same points. We can write (25.26) in a more convenient form:

$$\frac{m_0 v_2^2}{2} + E_{p2} = \frac{m_0 v_1^2}{2} + E_{p1}. \tag{25.27}$$

This means that the sum of the kinetic energy and the quantity E_p remains constant as a result of motion (any two points on the trajectory can be chosen as points *1* and *2* in (25.27)). Hence we can write

$$\frac{m_0 v^2}{2} + E_p = \text{const}. \tag{25.28}$$

The quantity E_p *is called the* potential energy of a point mass, *while Eq.* (25.28) *expresses the energy conservation law. It should be emphasized that this equation expresses not only the energy conservation law, but also the energy conversion law since it describes the mutual conversion of kinetic and potential energies.*

NORMALIZATION OF POTENTIAL ENERGY. So far, we have described the potential energy as a function whose partial derivatives with respect to coordinates with minus sign must be equal to the respective projections of the force as in (25.20). If instead of the potential energy E_p we take a different quantity $E'_p = E_p + A$, i.e. a quantity differing by a constant amount A throughout the space, the forces remain unchanged as a result of such a replacement. For example, we can write

$$F'_x = -\frac{\partial E'_p}{\partial x} = -\frac{\partial (E_p + A)}{\partial x} = -\frac{\partial E_p}{\partial x} = F_x, \tag{25.29}$$

where we have considered that the derivative of a constant is zero, i.e. $\partial A/\partial x = 0$. *Thus, the potential energy has been defined only to within an additive constant.*

Considering a point in space, we can state that the potential energy at the point is equal to any preassigned value. It follows from here that

a physical meaning cannot be assigned to the value of the potential energy itself; it can only be assigned to the difference between the potential energies at two points.

Using the arbitrariness in the choice of the potential energy, we can put it equal to any preassigned value at a point in space. Its value will then be uniquely determined at all other points in space. *This procedure of assigning a unique value to the potential energy is called* normalization.

As an example, let us consider the force of gravity near the Earth's surface. We direct the Z-axis along the vertical and take the origin at the Earth's surface. Then the projections of the force acting on a body of mass m will be $F_z = -mg$ and $F_x = F_y = 0$. Consequently, in accordance with (25.20), the potential energy is given by the expression $E_p(z) = mgz + A$,

?

What is the meaning of the line integral expressing the work done over the displacement between two points?
What does the integral depend on in the general case?
What are potential forces?
What criteria for the potential nature of forces do you know?
What is the connection between forces and a potential energy?
Is it possible to write the energy conservation law for forces which are not potential within the framework of mechanics?
What nonmechanical forms of energy do you know?

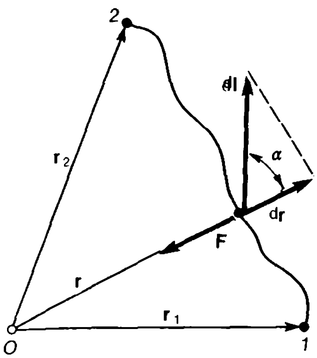


Fig. 58. Proving the potential nature of the force of gravity.

where A is a constant. If we assume that $E_p = 0$ at the Earth's surface ($z = 0$), then the constant A will be zero and $E_p(z) = mgz$. This is the expression for the potential energy whose value is normalized to zero at the Earth's surface. Likewise, it can be assumed that the potential energy at the Earth's surface is A_0. In this case, $A = A_0$ and $E_p(z) = mgz + A_0$. In this case, the potential energy is said to be normalized to the value A_0 at the Earth's surface.

INTERACTION ENERGY. The potential energy of a body is due to the interaction of the body with other bodies, in the present case with the Earth. If there is no interaction, there is no potential energy. Let us remove the body away from the Earth's surface. The force of gravity can be assumed to be constant only approximately when the distance of the body from the Earth's surface changes within certain small limits. If the body is removed from the Earth at large distances, we must take into account that the force of gravity decreases in inverse proportion to the square of the distance of the body from the Earth's centre. Let us make the origin of coordinates (point O) coincide with the Earth's centre. The force of gravity will be directed along the radius r. The components of the force perpendicular to the radius r will be zero, and the magnitude of the force will only depend on the distance from the Earth's centre. It can easily be verified that such a force is of potential nature. For this purpose, we calculate the elementary work done over a displacement by $d\mathbf{l}$ (Fig. 58). The force acting on a body of mass m will be

$$\mathbf{F} = -G\frac{Mm}{r^2}\frac{\mathbf{r}}{r}, \tag{25.30}$$

where M is the Earth's mass, G is the gravitational constant, and $\mathbf{r}/r$ is the unit vector along the radius from the Earth's centre. The minus sign indicates that the force is directed towards the Earth's centre. The elementary work done over a displacement by $d\mathbf{l}$ will be (see Fig. 58)

$$\mathbf{F}\cdot d\mathbf{l} = -G\frac{Mm}{r^2}\frac{\mathbf{r}}{r}\cdot d\mathbf{l} = -G\frac{Mm}{r^2}dl\cos\alpha = -G\frac{Mm}{r^2}dr, \tag{25.31}$$

where the vector $\mathbf{r}/r$ is a unit vector, and $dl\cos\alpha$ is the projection of the displacement on the direction of the radius, its magnitude being dr along the radius. Thus, the elementary work is defined only by the displacement along the radius and is independent of the displacement perpendicular to the radius. This means that there is no gravitational force in the plane perpendicular to the radius. The elementary work in (25.31) depends only on the variable r and its differential dr. Hence the

?

What does the normalization of the potential energy mean and what is it caused by? Which normalizations do you know?
What is the interaction energy?
What is a carrier of the potential energy?
What is the total energy of a body in the relativistic case?
What is the expression for the kinetic energy in the relativistic case?
What is the rest energy and how do we experimentally prove that it is an energy?
Why cannot we call the mass-energy relation the formula for converting mass into energy but call it the formula for relating these quantities?
What experimental proof of the mass-energy relation do you know?

work done over the displacement of the body from an arbitrary point at a distance r_1 to a point at a distance r_2 can be determined by integrating the function of one variable:

$$\int_{(1)}^{(2)} \mathbf{F} \cdot \mathrm{d}\mathbf{l} = -GMm \int_{r_1}^{r_2} \frac{\mathrm{d}r}{r^2} = -GMm\left(\frac{1}{r_1} - \frac{1}{r_2}\right). \tag{25.32}$$

It can be seen even at this stage that the gravitational force is of potential nature since the work done between points *1* and *2* depends only on the distances r_1 and r_2, and not on the path by which these points are connected. Obviously, the work done over a closed contour is also zero since if we return from point *2* to point *1* by a different path, the work will have the same value (25.32) but the opposite sign, and hence the total work over the closed path *1-2-1* is zero, as indeed it should be in the case of potential forces. This proves the potential nature of the gravitational force.

Comparing (25.32) with the general formula (25.25), we can find the potential energy E_p of a point mass m:

$$E_p(r) = -G\frac{Mm}{r} + A. \tag{25.33}$$

Let us consider the normalization of energy. It is desirable to choose the normalization condition so that it takes into account the physical characteristics of the interaction. In this case, the numerical value of the potential energy may acquire a more clear physical meaning and does not remain a purely formal number, as has so far been the case. Such a physical consideration does exist indeed. As a matter of fact, if a body is removed to infinity from the Earth's surface, there will be no interaction between the body and the Earth. This means that the existence of the body at infinity will not have any effect on the phenomena taking place at any finite distance from the Earth's surface. The same is true for the phenomena taking place at any finite distance from a body of mass m. Hence it can be logically concluded that in this case, the potential energy E_p associated with the interaction between the body and the Earth as the former is removed from the Earth's surface must also be zero. This leads to the normalization condition

$$E_p(\infty) = 0, \tag{25.34}$$

which is not an arbitrary requirement but takes into consideration the essence of physical processes taking place during interaction. It follows from the normalization condition (25.34) that the constant $A = 0$ in (25.33), and the potential energy of a point mass m in the Earth's gravitational field is

$$E_p(r) = -G\frac{Mm}{r}. \tag{25.35a}$$

It should be observed that under the normalization condition (25.34), the formula for the potential energy of a particle at a point B can be written in the form

$$E_p(B) = \int\limits_{(B)}^{\infty} \mathbf{F}(\mathbf{r})\cdot d\,\mathbf{r}, \tag{25.35b}$$

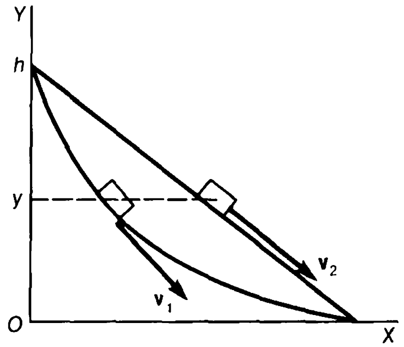
where the work is calculated along any path starting at point B and terminating at infinity where the force F vanishes and the interaction ceases.

APPLICATIONS. Many applications of the energy conservation law will be considered in the following chapters. For the present, it is sufficient to consider the effectiveness of using the energy conservation law in the well-known examples of sledges sliding down humps of complex shapes. If we are given a hump from whose top a sledge starts to slide down and it is required to determine the velocity of the sledge at any point on the hump (with or without friction), the solution of the problem with the help of equations of motion may turn out to be tiresome. The problem is considerably simplified if we make use of the energy conservation law (Fig. 59).

The energy conservation law can be used to carry out a rather simple analysis of the general peculiarities of motion without a detailed knowledge of the equations of motion if we know the law of change of potential, i.e. potential energy. Let us apply this method to the one-dimensional case. In this case, any force which depends only on the coordinates (and does not depend on velocity and time) is a potential force by definition. The determination of potential is reduced to the evaluation of the integral of a known force, which can always be carried out. Hence we can assume that the law of change of potential energy is known. Suppose that the potential energy changes as shown in Fig. 60.

Let us consider the motion of a particle whose total energy is E. The particle may be situated either in the region between points x_1 and x_2, or to the right of point x_3. Indeed, in accordance with the energy conservation law, the kinetic energy of a particle is equal to the difference between its total energy and the potential energy, i.e. to $E - E_p$, and can have only positive values. Hence only regions in which the total energy is higher than the potential energy are permissible regions for motion. For example, the motion in the region between x_2 and x_3 is not possible since the kinetic energy of the particle in the region would be negative.

Let us now consider the motion in the permissible region,



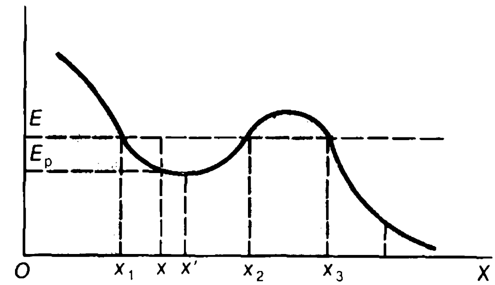


Fig. 59. Illustrating the energy conservation law: $v_1^2 = v_2^2 = 2g(h - y)$.

Fig. 60. A particle can move only in a region where its total energy is equal to or higher than the potential energy. This region is called a potential well.

say, $x_1 x_2$. Suppose that the particle is at point x. Its kinetic energy is then $E - E_p$, and it can move to the right or to the left. If it moves to the left, its potential energy increases and hence its kinetic energy decreases (since the total energy remains constant), i.e. the velocity of the particle decreases. This means that the particle at point x is subjected to the force directed to the right. This is also obvious from the formula expressing the force in terms of the potential energy:

$$F_x = -\frac{\partial E_p}{\partial x}. \tag{25.36}$$

At point x, the potential energy decreases with increasing x, and hence $\partial E_p/\partial x$ is negative while $F_x = -\partial E_p/\partial x$ is positive. In other words, the force is directed to the right, viz. in the positive X-direction. The particle will continue to move to the left until its velocity is zero, i.e. until its total energy is transformed into its potential energy. This will happen at point x_1. However, the particle cannot remain at rest at this point since it is subjected to the force directed to the right. Under the action of this force the particle will move to the right at an increasing velocity, attaining the maximum value at point x' where the potential energy of the particle is minimum. On the segment (x', x_2), the particle will be subjected to the force directed to the left, and the particle's velocity will be zero at point x_2. The particle will then start to move to the left, and so on. On the entire segment (x_1, x_2), there exists only one point where the particle may be at rest. This is the point x' where the potential energy is minimum, which is the condition for stable equilibrium.

The particle situated to the right of point x_3 can move from x_3 to infinity (if to the right of x_3 the potential energy does not rise above E at any point). The motion between points x_2 and x_3 is impossible. The region between x_1 and x_2 in which the particle is trapped is called a potential well, while the region between x_2 and x_3 which cannot be passed by the particle is called a potential barrier. In classical mechanics, a potential barrier is an unsurmountable obstacle for a moving particle. In quantum mechanics, the particle may pass through the potential barrier under certain conditions. This phenomenon is known as the tunnel effect and plays an important role in the microcosm. This effect is considered in greater detail in quantum mechanics.

TOTAL ENERGY AND REST ENERGY. All the arguments put forth in the previous section concerning the work of forces, the potential nature of forces and the potential energy also remain valid for motions at high velocities since in the considerations it is unimportant at which velocity a particle moves. The only difference is that the nonrelativistic equation of motion (25.11) must now be replaced by the relativistic equation of motion (20.10):

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_0 \mathbf{v}}{\sqrt{1 - v^2/c^2}}\right) = \mathbf{F}\,. \tag{25.37}$$

As in the nonrelativistic case (25.11), the multiplication of both sides of (25.37) by the velocity $\mathbf{v}$ gives

$$\mathbf{v}\cdot\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_0 \mathbf{v}}{\sqrt{1 - v^2/c^2}}\right) = \mathbf{F}\cdot\mathbf{v}\,. \tag{25.38}$$

Differentiating the left-hand side of this equation, we obtain

$$\begin{aligned}
&\mathbf{v}\cdot\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_0 \mathbf{v}}{\sqrt{1 - v^2/c^2}}\right)\\
&= m_0\mathbf{v}\cdot\left[\frac{1}{2}\frac{\mathbf{v}}{(1 - v^2/c^2)^{3/2}}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{v^2}{c^2}\right) + \frac{1}{\sqrt{1 - v^2/c^2}}\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\right]\\
&= \frac{m_0}{(1 - v^2/c^2)^{3/2}}\left[\frac{1}{2}v^2\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{v^2}{c^2}\right) + \left(1 - \frac{v^2}{c^2}\right)\left(\mathbf{v}\cdot\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}\right)\right].\\
&= \frac{1}{2}\frac{m_0}{(1 - v^2/c^2)^{3/2}}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{v^2}{c^2}\right)\left[v^2 + \left(1 - \frac{v^2}{c^2}\right)c^2\right]\\
&= \frac{1}{2}\frac{m_0 c^2}{(1 - v^2/c^2)^{3/2}}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{v^2}{c^2}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m_0 c^2}{\sqrt{1 - v^2/c^2}}\right).
\end{aligned}$$

Consequently, Eq. (25.38) assumes the form

$$\mathrm{d}\left(\frac{m_0 c^2}{\sqrt{1 - v^2/c^2}}\right) = \mathbf{F}\cdot\mathrm{d}\mathbf{r}, \tag{25.39}$$

where $\mathbf{v} = \mathrm{d}\mathbf{r}/\mathrm{d}t$ and both sides of the equation have been multiplied by $\mathrm{d}t$. Let us compare Eq. (25.39) with Eq. (25.13) of the nonrelativistic theory. It can be seen that as a work is done by a force, the quantity $m_0c^2/\sqrt{1 - v^2/c^2}$ changes instead of the kinetic energy.

Suppose that a particle moves in the potential force field so that the force acting on it is given by the relations (25.20). Then, proceeding from (25.39) and repeating all the steps in the computations from (25.20) to (25.28), we obtain the following relation instead of (25.28):

$$\frac{m_0c^2}{\sqrt{1 - v^2/c^2}} + E_p = \text{const}. \tag{25.40}$$

This formula expresses the energy conservation law in the relativistic case. The potential energy E_p has the same meaning as in the nonrelativistic case, while the quantity

$$E = \frac{m_0c^2}{\sqrt{1 - v^2/c^2}} \tag{25.41}$$

is called the total energy of a body. If the body is at rest, i.e. $v = 0$, then in accordance with (25.41), it has an energy

$$E_0 = m_0c^2, \tag{25.42}$$

which is called the rest energy of the body.

The expression "the total energy of a body" in the nonrelativistic case indicates the sum of its kinetic and potential energies, while in the relativistic case it is used not only for the quantity (25.41), but also for the sum of this quantity and the potential energy of the body. To avoid any confusion, one should be careful in order not to mix up two different meanings of the same term.

We should also mention that the self-energy of the body which creates the force field acting on the body under consideration is not taken into account in (25.40). It is assumed to be stationary and has only the rest energy.

KINETIC ENERGY. At low velocities, $v/c \ll 1$. Hence $1/\sqrt{1 - v^2/c^2} \simeq 1 + (1/2)v^2/c^2$, and Eq. (25.41) can be written in the form

$$E = m_0c^2 + \frac{1}{2}m_0v^2 + \dots\ . \tag{25.43}$$

Thus, as a body gathers velocity, its rest energy m_0c^2 is supplemented by the kinetic energy, their sum expressing the total energy of the moving body. *Hence the kinetic energy E_k of* *a body moving at an arbitrary velocity is given by the formula*

$$E_k = E - m_0c^2 = m_0c^2\left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1\right). \tag{25.44}$$

At low velocities, this relation can be transformed with the help of (25.43) into the classical expression $m_0v^2/2$ for the kinetic energy.

MASS-ENERGY RELATION. Taking into account (20.11) for the relativistic mass

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}, \tag{25.45}$$

we can represent (25.41) for the total energy in the following form:

$$E = mc^2. \tag{25.46}$$

A comparison of (25.46) and (25.42) shows that in both cases the energy is connected with the inertia of a body through the same formula. Hence two most important characteristics of the body, viz. energy and inertia (i.e. mass), are found to be mutually connected. The above derivation of the mass-energy relation shows that this relation is valid as a relation between the inertial mass of the body and its total energy, i.e. the sum of the kinetic and rest energies. But only experiment can show whether it is valid for other forms of energy, for example, the potential energy. The energy conservation law as expressed by Eq. (25.40) indicates that this relation is quite likely to be valid for the potential energy, i.e. that the potential energy has inertial properties. If Eq. (25.46) is found to be universal in nature, i.e. applicable for arbitrary forms of energy, it will be one of the most fundamental laws of physics. Experiment shows that this is indeed true. Equation (25.46) is called the mass-energy relation and was established by Einstein. Sometimes, this relation is referred to as the equivalence of mass and energy. It will be shown below that this expression is incorrect; we shall therefore not use this expression in this book.

EXPERIMENTAL VERIFICATION OF THE MASS-ENERGY RELATION. It follows from the experiments considered in the derivation of the relativistic equation of motion (20.10) that the inertia of a body depends on its velocity in the manner described by the formula for the relativistic mass appearing in the equation. It was shown in Sec. 20 that such a dependence of mass on velocity also follows from the relativity principle and the Lorentz transformations. Hence all experimental data which confirm the Lorentz transformations also confirm the relation (25.46).

Only one question remains unanswered in these experiments: Is the rest energy m_0c^2 indeed energy or just a quantity having the same dimensions as mass but not the same physical meaning? But what is the idea behind the question of whether the quantity m_0c^2 having the dimensions of energy is indeed energy? Isn't it tautology? No, it is not tautology, and this question has a clear physical meaning. Can the rest energy m_0c^2 be transformed into other forms of energy? If this is possible, the rest energy, like other forms of energy, is real; but if this is not possible, the rest energy is just an auxiliary quantity which does not have any physical meaning. Experiments show that the rest energy can indeed be transformed into other forms of energy and is therefore real.

One of the many experiments which confirms this statement is the annihilation of elementary particles. An electron and a positron can be treated as identical particles differing only in the sign of their electric charge and magnetic moment. They have the same mass which can be measured, for example, from their motion in a magnetic field, and the total energy can then be determined as the sum of the kinetic energy and the rest energy. Since the magnetic field does not do any work, the potential energy can be neglected. When an electron and a positron collide, they annihilate each other and cease to exist as particles having a rest mass. In their place a γ-quantum appears, i.e. a particle whose rest mass is zero and whose velocity is equal to the velocity of light. The energy of this quantum can be measured. It turns out that the energy of the γ-quantum is equal to the sum of the energies of the electron and the positron, including their rest energies. Thus, the rest energy can indeed be transformed into quite different forms of energy.

At the same time, these experiments explain the physical meaning of the mass-energy relation. It is stated sometimes that the relation (25.46) expresses the equivalence of mass and energy and the possibility of their mutual conversion. However, such statements are erroneous. We can speak of a conversion of mass into energy only in the case when mass disappears, and the disappearance of inertial properties results in the appearance of energy that has not existed before. Such processes are not known to exist.

In all processes, energy disappears in some form and appears in another form, its value remaining the same during such a conversion. Similarly, the form of existence of mass also changes, but its value is conserved. The relation (25.46) indicates that whatever the conversions of the forms of mass and energy in nature, this mass-energy relation is always satisfied.

INERTIAL NATURE OF POTENTIAL ENERGY. Let us apply the mass-energy relation to the potential energy. Since (25.40) proves the energy conservation law for the case of conversion of the total energy into the potential energy, the problem reduces to proving the inertial nature of the potential energy. It can be seen from (25.35a) that the potential energy is negative in the case of attraction in a gravitational field. This is not just a property of the gravitational forces: indeed, *any potential force of attraction is associated with a negative energy since a particle has to expend a part of its kinetic energy to overcome this attraction.* The sum of the kinetic and potential energies must remain constant, and at infinite separation the particle's velocity decreases, while its potential energy becomes zero. Consequently, the potential energy must be lower, i.e. negative, at finite distances.

If a particle moves in the gravitational force field at a finite distance from another, heavy, particle which can be assumed to be stationary, the sum of its total and potential energies $E + E_p$ must be less than its rest energy. Indeed, if $E + E_p > m_0c^2$, the energy conservation law allows the particle to move to infinity, when $E_p \to 0$. If, however, $E + E_p < m_0c^2$, the particle cannot be removed to infinity since in this case E would become smaller than m_0c^2, which is impossible since the energy of a particle cannot be lower than its rest energy. Hence the gravitational force confines the particle to a finite region under the condition

$$E + E_p < m_0c^2$$

or

$$(E - m_0c^2) + E_p = E_k + E_p < 0\,. \tag{25.47}$$

In other words, *the sum of the potential and kinetic energies must be negative.* This is the condition for the formation of bound states.

We assumed the body creating the gravitational field to be stationary. This is true only if its mass is much larger than the mass of the moving body. Otherwise, we must take into account the motion of the other body as well. Note that all the arguments remain valid in this case without any significant change.

If the motion of both particles is considered in the inertial coordinate system (and in no other system), the condition for the existence of bound states indicates that the sum of the kinetic and interaction energies of both particles must be negative. The interaction energy as the potential energy of one body in the field of another has to be taken into account only once. For example, the energy (25.35a) is the potential energy of a body of mass m in the gravitational field of another body of mass M. However, the same quantity can also be considered to be the potential energy of the body of mass M in the gravitational field of the body of mass m. This is the same quantity which is the interaction energy of the bodies of mass m and M, and need not be taken into account twice. Hence the condition for the existence of a bound state can be stated as: *the sum of the kinetic energy and the interaction energy of particles in the bound state must be negative.* The sum of the kinetic energy and the interaction energy is called the binding energy. *Hence it can be assumed that the binding energy is negative in the bound state.*

BINDING ENERGY. It is well known that the nuclei of atoms consist of neutrons and protons. We still do not know the exact law according to which nuclear forces act, but we do know that these are forces of attraction since they confine the neutrons and protons within the nucleus. Hence the binding energy in the nucleus is negative. Let us denote it in the form ΔE_{nuc}. The total energy of the nucleus is equal to the sum of the rest energies of protons, E_{0p}, and of neutrons, E_{0n}, minus the binding energy:

$$E_{\text{nuc}} = E_{0p} + E_{0n} - \Delta E_{\text{nuc}}. \tag{25.48}$$

If the mass-energy relation (25.46) is also applicable to the potential energy (its applicability to the rest energy and the kinetic energy has already been proved), the mass of the nucleus, M_{nuc}, must be less than the sum of the rest masses of protons, M_{0p}, and of neutrons, M_{0n}, since in this case it follows from (25.48) that

$$M_{\text{nuc}} = M_{0p} + M_{0n} - \Delta M_{\text{nuc}}, \quad \Delta M_{\text{nuc}} = \frac{\Delta E_{\text{nuc}}}{c^2}. \tag{25.49}$$

The quantity ΔM_{nuc} is called the mass defect of the nucleus. The rest masses of protons and neutrons can be measured in several ways and are precisely known. The mass of the nucleus can also be measured in experiments in which its inertial properties are manifested. The mass of the nucleus was indeed found to be smaller than the sum of the rest masses of the protons and neutrons constituting it. This means that the negative potential energy in the nucleus leads to a negative inertia in accordance with (25.46) or, in other words, *the mass-energy relation is applicable to the potential energy.*

The binding energy of nuclei has been studied extensively. The most convenient way to characterize it is as the binding

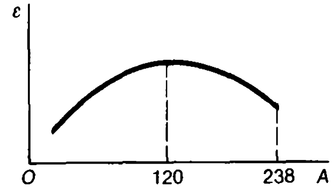


Fig. 61. Dependence of the binding energy on the mass number.

energy ε per nucleon (so far as nuclear forces are concerned, a proton and a neutron behave as identical particles):

$$\varepsilon = \frac{\Delta E_{\text{nuc}}}{A}, \tag{25.50}$$

where A is the total number of protons and neutrons in the nucleus and is called the mass number. The dependence of ε on A is shown in Fig. 61.

It can be seen that the nucleons (protons and neutrons) are weakly bound in elements appearing at the beginning of the Periodic Table. The binding becomes stronger as we move to heavier nuclei. For nuclei with mass number $A \simeq 120$, the binding energy attains its highest value of about 8.5 MeV. It should be recalled that an electron volt is the energy acquired by an electron or a proton upon passing a potential difference of 1 V ($1 \text{ eV} = 1.6 \times 10^{-19}$ J). Beyond this, the binding energy decreases. For nuclei of elements at the end of the Periodic Table, the binding energy is so small that nuclei with mass number above 238 are unstable. Such elements are obtained only artificially, exist for relatively short periods of time and spontaneously transform into lighter nuclei.

If the heavy nucleus of an element at the end of the Periodic Table is split into two nearly equal parts, we obtain two nuclei of elements lying closer to the middle of the Periodic Table. According to Fig. 61, the binding energy per nucleon for these nuclei is higher than for the parent nucleus, i.e. the nucleons are more strongly bound in these nuclei than in the parent nucleus. The sum of the rest masses of the nuclei obtained as a result of their fission is less than the rest mass of the parent nucleus. Hence the sum of the rest energies of the nuclei obtained as a result of their fission is less than the rest energy of the parent nucleus. The difference between the energies is liberated as the kinetic energy of the fission products and the radiation emitted during the fission. This is the atomic (nuclear) energy, which is used in atomic (nuclear) reactors and atomic bombs.

If two light nuclei of elements at the beginning of the Periodic Table are combined into one, their fusion results in a nucleus lying closer to the middle of the Periodic Table. According to Fig. 61, the nucleons in the resulting nucleus are more strongly bound than in the parent nuclei. Arguments similar to those put forth in the previous case lead to the conclusion that the fusion of light nuclei must result in the liberation of energy which can be used for making hydrogen bombs. The ways of liberating this energy in a controlled manner for peaceful purposes are not known yet and are intensively studied at present. Most scientists believe that the basic solution of this problem will be found by the end of the 20 century, and it should be possible to harness this energy in the 21 century.

The mass-energy relation was not only confirmed experimentally but also put to many important applications. At the same time, the phenomena mentioned above also prove the energy conservation law in the relativistic case.

ENERGY CONSERVATION LAW FOR A SYSTEM OF POINT MASSES. All that has been stated in this section about the energy conservation law refers to a point mass. The situation is considerably complicated for a system of point masses. Above all, it leads to a temptation to formulate the energy conservation law for a system of point masses in the same way as for a single point mass, i.e. to replace the equation of motion (25.11) for a point mass by Eq. (21.16) for the motion of the centre of mass of a system of point masses. Multiplying both sides of (21.16) by the velocity of the centre of mass and transforming the relation thus obtained by analogy with (25.12) and (25.13), we obtain the following relation instead of (25.14):

$$\frac{mV_2^2}{2} - \frac{mV_1^2}{2} = \int_{(1)}^{(2)} \mathbf{F} \cdot d\mathbf{R}, \tag{25.51}$$

where V_1 and V_2 are the velocities of the centre of mass of the system at the beginning and the end of the path, and the integral on the right-hand side is evaluated along the path of motion of the centre of mass. Equation (25.51) is identical to (25.14), but it cannot be stated that it expresses the energy conservation law in the same sense as (25.14). As a matter of fact, (25.51) is not an equation for the physically existing quantities. The centre of mass is an imaginary point, and the force $\mathbf{F}$ under the integral in (25.51) is applied to the centre of mass, while in actual practice it is composed of the forces applied to point masses forming the system. Hence the right-hand side of (25.51) does not represent the work of forces applied to a body, nor does the left-hand side represent a change in the kinetic energy of the system. The left-hand side simply takes into account the change in the kinetic energy associated with the motion of the centre of mass.

In order to obtain the energy conservation law for a system of point masses, we proceed from the equations of motion for each point of the system:

$$\frac{m_i d\mathbf{v}_i}{dt} = \mathbf{F}_i, \tag{25.52}$$

where the subscript i indicates the quantities corresponding to the ith point mass. The force $\mathbf{F}_i$ includes both the external and the internal forces acting on the ith point mass. Multiplying each equation in (25.52) by the displacement $d\mathbf{r}_i = \mathbf{v}_i\,dt$ of the corresponding point mass, summing the left- and right-hand sides of these equations over all the points and then integrating the sums thus obtained from the initial instant of time t_1 to the final instant t_2, we arrive at the equality

$$\sum_i \left(\frac{m_i v_{i2}^2}{2} - \frac{m_i v_{i1}^2}{2} \right) = \sum_i \int_{t_1}^{t_2} \mathbf{F}_i \cdot \mathbf{v}_i\,dt, \tag{25.53}$$

which expresses the energy conservation law for a system of point masses in the same sense as (25.14).

It is much more difficult to use the law (25.53) for specific cases of motion of a system of point masses than for the case of a single point mass even if only potential forces are involved. In particular, this is due to the fact that the work of internal forces is not necessarily zero and cannot be evaluated in a general form. The simplest motion is that of a rigid body in fairly uniform external potential fields. In this case, the kinetic energy on the left-hand side of (25.53) is the sum of the kinetic energy of the motion of the centre of mass and the rotational energy of the rigid body (see Sec. 33); the work of internal forces on the right-hand side is zero, while the work of external forces is expressed by the integral along the path of motion of the centre of mass. The consideration of the nonuniformity of the potential field considerably complicates the evaluation of the integral on the right-hand side of (25.53).

!

The homogeneity of space, the isotropy of space and the homogeneity of time form the basis of the momentum conservation law, the angular momentum conservation law and the energy conservation law respectively.

Example 25.1. Find the work done by the force $\mathbf{F} = y^2\mathbf{i}_x + x^2\mathbf{i}_y$ during displacement from point $x = 0$, $y = 0$ to point $x = 2$, $y = 1$ along a straight line L_1 given by $y = 2x$, and along a parabola L_2 given by $y = 2x^2$.

By definition,

$$\int_{L_1} \mathbf{F}\cdot d\mathbf{r} = \int_{L_1} (y^2\,dx + x^2\,dy) = \int_0^1 (4x^2\,dx + 2x^2\,dx) = 2, \tag{25.54}$$

$$\int_{L_2} \mathbf{F}\cdot d\mathbf{r} = \int_{L_2} (y^2\,dx + x^2\,dy) = \int_0^1 (4x^4\,dx + x^2 4x\,dx) = 1.8, \tag{25.55}$$

where the x-coordinate has been taken as the independent parameter for integration in both cases.

?

What is the physical meaning of the concepts of the homogeneity of space, the isotropy of space and the homogeneity of time?
What requirements are imposed by the homogeneity of space and the isotropy of space on the properties of forces?
What property of the potential energy of a system is due to its being isolated?

Example 25.2. Two identical particles, each having an energy E, move towards each other in the laboratory coordinate system. The rest energy of the particles is E_0. Find the energy E' of one of the particles in the coordinate system associated with the other particle.

Let us use the results of Example 17.1. Since $\gamma = E/E_0$ and $\gamma' = E'/E_0$, we obtain

$$E' = E_0\left[2\left(\frac{E}{E_0}\right)^2 - 1\right]. \tag{25.56}$$

For $\gamma \gg 1$, we can put $E' = 2E(E/E_0)$. If, for example, two protons move towards each other with an energy of about 30 GeV each, this is equivalent to the impingement of a proton with an energy of about 2000 GeV on a stationary proton.

Sec. 26. CONSERVATION LAWS AND THE SYMMETRY OF SPACE AND TIME

It is shown that the conservation laws are due to the homogeneity and isotropy of space and the homogeneity of time.

MOMENTUM CONSERVATION LAW AND THE HOMOGENEITY OF SPACE. Momentum can be conserved not only in an isolated system, but also in a nonisolated system if the resultant of all the external forces is zero. In this case, $\mathbf{F} = 0$ and hence $\mathbf{p} = \text{const}$ in Eq. (21.11), i.e. the momentum of the system is conserved even though the system is not isolated. For example, a drop of rain falls at a constant velocity in air. It is a nonisolated system of point masses subjected to external forces, viz. the force of gravity and the friction of air. The resultant of these two forces is zero, and the momentum of the drop remains constant. It can be stated that the momentum conservation of a nonisolated system of point masses is due to the properties of the external forces acting on the system. For other external forces, the momentum of such a system is not conserved. Hence we cannot speak of any general law of momentum conservation or nonconservation for a nonisolated system.

The situation is quite different in an isolated system. Momentum is always conserved in an isolated system. *The statement concerning the momentum conservation of an isolated system is universal in nature and is therefore called a law.* What is this law based on? While considering that in the absence of external forces we can put $\mathbf{F} = 0$ in (21.11) and bring it to the form (23.1), it was assumed that the condition (21.7c) expressing Newton's third law of motion is satisfied. Hence from a formally mathematical point of view it can be stated that the momentum conservation law for an isolated system of point masses is a corollary of Newton's third law of motion. It can be asked as to which basic factors are responsible for the validity of Newton's third law of motion. The answer to this question runs as follows: The validity of Newton's third law of motion and the momentum conservation law for an isolated system of point masses are based on the homogeneity of space. *By the homogeneity of space we mean the equivalence of all the points in space to one another. This means that if we have an isolated physical system, the course of events in it will not depend on the region of space in which the system is localized. When applied to an isolated system of point masses, this means that if all the points of the system are displaced by* $\delta \mathbf{r}$, *there will be no change in the state of the system or in its intrinsic motion. Hence it follows that the total work done by internal forces upon a displacement of the system by* $\delta \mathbf{r}$ *must be zero:*

$$\delta \mathbf{r} \cdot \sum_i \sum_j \mathbf{F}_{ij} = 0. \tag{26.1}$$

In view of the arbitrary nature of $\delta \mathbf{r}$, we obtain the equality

$$\sum_i \sum_j \mathbf{F}_{ij} = \frac{1}{2} \sum_i \sum_j (\mathbf{F}_{ij} + \mathbf{F}_{ji}) = 0, \tag{26.2}$$

which coincides with (21.7c). Since the interaction between each pair is independent of each other, from (26.2) we arrive at Newton's third law of motion:

$$\mathbf{F}_{ij} + \mathbf{F}_{ji} = 0. \tag{26.3}$$

This proves that

the momentum conservation law for an isolated system of point masses emerges from the fundamental property of space in inertial systems, viz. its homogeneity. Hence it can be concluded that the relativity principle is also connected with the homogeneity of space.

We can now return to the validity of Newton's third law of motion which was considered in Sec. 19. While deriving Eq. (26.3), it was assumed that point masses are the only carriers of momentum in a closed system, and that forces act at a distance without any material intermediary. Such forces must satisfy Eq. (26.3) if the space is homogeneous. Forces with a finite velocity of propagation do not satisfy these conditions, and Newton's third law of motion in its existing form cannot be applied to them. When such forces are present, the theorem on the momentum conservation of a closed system remains valid, but besides the point masses in the closed system, we must also consider the field in which the point masses interact and take into account the momentum of this field as well.

ANGULAR MOMENTUM CONSERVATION LAW AND THE ISOTROPY OF SPACE. The angular momentum conservation law for an isolated system follows from Eq. (24.1) whose derivation is based on the assumption that the moment of internal forces of interaction of point masses of the system satisfies Eq. (21.10b), i.e. the moment of these forces is zero. From a purely mathematical point of view, it is this circumstance that leads to the angular momentum conservation law for an isolated system of point masses. It was emphasized earlier that this relation does not follow from Newton's laws and is a requirement independent of the laws. Consequently, the momental equation (21.21) for a system of point masses cannot be reduced completely to Newton's laws. It can be asked as to which basic factors are responsible for the validity of Eq. (21.10b). The answer to this question runs as follows: (21.10b) is valid due to the isotropy of space. *By the isotropy of space we mean the equivalence of all the directions in space. This means that if we have an isolated physical system, the course of events in it will not depend on its orientation in space. When applied to an isolated system of point masses, this means that the angular displacement of the system by* $\delta\boldsymbol{\varphi}$ (see Sec. 9) *will not change its internal state and its intrinsic motion. Hence the total work done by internal forces upon an angular displacement must be zero. It can be seen from Eqs.* (9.3) *and* (9.5) that *the angular displacement by* $\delta\boldsymbol{\varphi}$ *transfers a point mass characterized by the radius vector* $\mathbf{r}_i$ *through* $\delta\mathbf{r}_i = \delta\boldsymbol{\varphi} \times \mathbf{r}_i$. *The fact that the total work done by internal forces upon an angular displacement of the system by* $\delta\boldsymbol{\varphi}$ *is zero can be expressed in the form*

$$\frac{1}{2}\sum_i\sum_j(\delta\mathbf{r}_i\cdot\mathbf{F}_{ji} + \delta\mathbf{r}_j\cdot\mathbf{F}_{ij}) = 0. \tag{26.4}$$

Consequently, we can write

$$\begin{aligned}\delta\mathbf{r}_i\cdot\mathbf{F}_{ji} + \delta\mathbf{r}_j\cdot\mathbf{F}_{ij} &= (\delta\boldsymbol{\varphi}\times\mathbf{r}_i)\cdot\mathbf{F}_{ji} + (\delta\boldsymbol{\varphi}\times\mathbf{r}_j)\cdot\mathbf{F}_{ij}\\ &= \delta\boldsymbol{\varphi}\cdot(\mathbf{r}_i\times\mathbf{F}_{ji}) + \delta\boldsymbol{\varphi}\cdot(\mathbf{r}_j\times\mathbf{F}_{ij}) = \delta\boldsymbol{\varphi}\cdot[(\mathbf{r}_i-\mathbf{r}_j)\times\mathbf{F}_{ji}],\end{aligned} \tag{26.5}$$

where we have taken into consideration both the well-known rule from vector analysis about the cyclic permutation of cofactors in a scalar triple product and Newton's third law of motion. Substituting (26.5) into (26.4), we obtain

$$\frac{1}{2}\sum_i\sum_j\delta\boldsymbol{\varphi}\cdot[(\mathbf{r}_i-\mathbf{r}_j)\times\mathbf{F}_{ji}] = 0. \tag{26.6}$$

Since the angular displacement $\delta\boldsymbol{\varphi}$ is arbitrary, we arrive at the following equality from (26.6):

$$\frac{1}{2}\sum_i\sum_j(\mathbf{r}_i-\mathbf{r}_j)\times\mathbf{F}_{ji} = 0, \tag{26.7}$$

which is identical to (21.10b). It can be stated that (21.10b) is a consequence of the isotropy of space.

This means that

the angular momentum conservation law for an isolated system of point masses is a consequence of the fundamental property of space in inertial systems, viz. its isotropy.

ENERGY CONSERVATION LAW AND THE HOMOGENEITY OF TIME. Let us denote the Cartesian coordinates of particles forming an isolated system by (x_i, y_i, z_i), the projections of their velocities on the coordinate axes by (v_{ix}, v_{iy}, v_{iz}), and their masses by m_i. The potential energy of the system is denoted by E_p, and the forces acting on the particles are given, in accordance with (25.20), by formulas

$$F_{ix} = -\partial E_p/\partial x, \quad F_{iy} = -\partial E_p/\partial y, \quad F_{iz} = -\partial E_p/\partial z.$$

Consequently, the equations of motion have the form

$$m_i \frac{dv_{ix}}{dt} = -\frac{\partial E_p}{\partial x_i}, \quad m_i \frac{dv_{iy}}{dt} = -\frac{\partial E_p}{\partial y_i}, \quad m_i \frac{dv_{iz}}{dt} = -\frac{\partial E_p}{\partial z_i}, \tag{26.8}$$

where $i = 1, 2, \ldots, n$, n being the total number of particles. Multiplying both sides of (26.8) by the projections dx_i, dy_i and dz_i of the displacement vector of a moving particle, we obtain

$$\begin{aligned} d\left(\frac{m_i v_{ix}^2}{2}\right) &= -\frac{\partial E_p}{\partial x_i} dx_i, \\ d\left(\frac{m_i v_{iy}^2}{2}\right) &= -\frac{\partial E_p}{\partial y_i} dy_i, \\ d\left(\frac{m_i v_{iz}^2}{2}\right) &= -\frac{\partial E_p}{\partial z_i} dz_i, \end{aligned} \tag{26.9}$$

where we have considered that $(dv_{ix}/dt)\,dx_i = dv_{ix}(dx_i/dt) = d(v_{ix}^2/2)$ and that m_i is constant. The summation of the left- and right-hand sides of (26.9) for all the particles gives

$$\begin{aligned} &\sum_i \left[d\left(\frac{m_i v_{ix}^2}{2}\right) + d\left(\frac{m_i v_{iy}^2}{2}\right) + d\left(\frac{m_i v_{iz}^2}{2}\right) \right] \\ &= -\sum_i \left(\frac{\partial E_p}{\partial x_i} dx_i + \frac{\partial E_p}{\partial y_i} dy_i + \frac{\partial E_p}{\partial z_i} dz_i \right). \end{aligned} \tag{26.10}$$

For further computations, we must take into consideration the homogeneity of time. *By the homogeneity of time we mean the equivalence of various instants of time. This means that a physical situation has the same evolution irrespective of the specific instant of time at which it is realized. Hence the* *homogeneity of time entails the absence of any explicit time dependence of the potential energy E_p on an isolated system* ($\partial E_p/\partial t = 0$). Consequently, the right-hand side of (26.10) is the total differential

$$\sum_i \left(\frac{\partial E_p}{\partial x_i} dx_i + \frac{\partial E_p}{\partial y_i} dy_i + \frac{\partial E_p}{\partial z_i} dz_i \right) = dE_p, \tag{26.11}$$

and (26.10) can be written in the form

$$d\left[\sum_i \frac{m_i v_i^2}{2} + E_p \right] = 0, \tag{26.12}$$

where we have considered that

$$\sum_i \left[d\left(\frac{m_i v_{ix}^2}{2} \right) + d\left(\frac{m_i v_{iy}^2}{2} \right) + d\left(\frac{m_i v_{iz}^2}{2} \right) \right]$$
$$= d \sum_i \frac{m_i v_i^2}{2}. \tag{26.13}$$

Equation (26.12) leads to the energy conservation law for an isolated system of point masses:

$$\boxed{\sum_i \frac{m_i v_i^2}{2} + E_p = \text{const.}} \tag{26.14}$$

It can be seen that the decisive step in the derivation of the energy conservation law was to establish the universal nature of (26.11) for an isolated system on the basis of the homogeneity of time.

Consequently,
the energy conservation law for an isolated system of point masses is due to the fundamental property of time in inertial systems, viz. its homogeneity.

In typical situations of interacting particles in the relativistic case, we must take into consideration the conversion of particles, radiation and other relativistic effects. Hence we cannot formulate the energy conservation law in a simple form analogous to (26.14).

UNIVERSALITY AND GENERAL NATURE OF CONSERVATION LAWS. The conservation laws emerge from the fundamental properties of space and time, and hence they are universal and general since space and time are the forms of existence of matter which cannot exist without space and time.

In the history of physics, situations arose when an experiment seemed to indicate a violation of conservation laws. For example, in the study of beta decay in which an atomic nucleus undergoes a radioactive transformation by emitting an electron, it was found that the energy and momentum conservation laws are not obeyed. This was indicated by very careful measurements of the energy and momentum of the electrons and nuclei. If the energy and momentum conservation laws were indeed violated in beta decay, this would have necessitated a serious reconsideration of our concepts of space and time, and of the very essence of the conservation laws which form the basis of physical concepts. Physicists would have been forced to take such a step only if quite a large number of physical phenomena indicated its necessity. Such an approach to the evolution of the physical theory was not adopted for explaining the seeming violation of the conservation laws in beta decay, and it was assumed that besides an electron, another unknown particle having the energy and momentum necessary for restoring the conservation laws is emitted by the nucleus. This particle was called a *neutrino*. It was many years later that the existence of this particle, which occupies a very important position in the family of elementary particles, was experimentally confirmed. Theoretically, this particle was discovered because the scientists were confident about the universal and general nature of the conservation laws.

PROBLEMS

6.1. Evaluate the integral $\oint_L xy^2\,dy$, where L is a circle $x^2 + y^2 = 1$ circumvented in the counterclockwise direction during integration.

6.2. Find the work done by the force $F = i_x y^2 + i_y x^2$ in the semicircle $y = \sqrt{1 - x^2}$ between points (0, −1) and (0, 1).

6.3. Find the length of the helical line given parametrically by: $x = \sin t$, $y = \cos t$, $z = t$ for $0 \leqslant t \leqslant 2\pi$.

6.4. A person of mass m_1 stands in a stationary lift of mass m_2, balanced by a counterweight of mass $m_1 + m_2$, and jumps in the lift with the same force which would have raised his centre of mass by h on the Earth's surface. What will be the displacement of his centre of mass relative to the lift's floor? The masses of the ropes and pulleys in the mechanism as well as the friction should be neglected.

6.5. Find an expression for the momentum of a particle in terms of its kinetic energy E_k and its rest mass.

6.6. Express the velocity of a particle through its momentum and rest mass.

6.7. Find the velocity of a particle having charge e and rest mass m_0 after passing through a potential difference U.

6.8. What is the binding energy of the lithium nucleus which consists of three protons and four neutrons if its rest mass is known to be 1.16445×10^{-26} kg? The rest masses of a proton and a neutron are 1.67265×10^{-27} kg and 1.67495×10^{-27} kg respectively.

6.9. Two particles of rest masses m_1 and m_2 move towards each other at relativistic velocities along a straight line in a laboratory coordinate system. The total energies of the particles are E_1 and E_2. Find the energy E'_1 of the first particle in the coordinate system in which the second particle is at rest ($E'_2 = m_2 c^2$).

6.10. A stationary particle of mass m_1 splits into two particles of rest masses m_2 and m_3. Find the energy of the particles and the magnitudes of their momenta which are directed opposite to each other.

ANSWERS

6.1. $\pi/4$. **6.2.** $4/3$. **6.3.** $2\pi\sqrt{2}$. **6.4.** $2h(m_1 + m_2)/(m_1 + 2m_2)$. **6.5.** $[E_k(E_k + 2m_0 c^2)]^{1/2}/c$. **6.6.** $cp/\sqrt{p^2 + m_0^2 c^2}$. **6.7.** $c[eU(eU + 2m_0 c^2)]^{1/2}/(eU + m_0 c^2)$. **6.8.** 0.0631×10^{-10} J $=$ 39.4 MeV. **6.9.** $\{[(E_1^2/c^2 - m_1^2 c^2)(E_2^2/c^2 - m_2^2 c^2)]^{1/2} + E_1 E_2/c^2\}/m_2$. **6.10.** $E_2 = (m_1^2 c^2 + m_2^2 c^2 - m_3^2 c^2)/(2m_1)$, $E_3 = (m_1^2 c^2 + m_3^2 c^2 - m_2^2 c^2)/(2m_1)$, $p_2 = p_3 = (E_2^2 - m_2^2 c^4)/c^2$.

Chapter 7

Noninertial Reference Frames

Basic idea:
Inertial forces exist in noninertial reference frames. They are real, have various manifestations and important applications. But the concept of inertial forces cannot be used in the analysis of motions in inertial reference frames since these forces do not exist in such frames.

Sec. 27. INERTIAL FORCES

Conditions for the emergence of inertial forces are analyzed and the question of their feasibility is discussed.

DEFINITION OF NONINERTIAL REFERENCE FRAMES. A noninertial reference frame is a reference frame moving with an acceleration relative to an inertial reference frame. The reference frame is associated with a reference body which, by definition, is assumed to be perfectly rigid. The accelerated motion of a rigid body includes the acceleration of both a translational and a rotational motion. Hence the simplest noninertial reference frames are the ones moving with an acceleration along a straight line as well as rotational systems.

TIME AND SPACE IN NONINERTIAL REFERENCE FRAMES. In order to describe the motion in a reference frame, it is necessary to explain the meaning of the statement that certain events occurred at certain instants of time at certain points. For this purpose, it is of prime importance that there should exist a unified time in a reference frame in the same sense as described in Sec. 7. *In a noninertial reference frame, a unified time in the sense mentioned in Sec. 7 does not exist.* Hence it is not clear how the duration of processes that begin at one point and end at another should be measured. The concept of duration of a process becomes meaningless for such processes since the clocks have different paces at different points. *The problem of measurement and comparison of lengths also becomes more complicated. For example, it is difficult to describe the concept of the length of a moving body if the concept of simultaneity at different points is not introduced.*

These difficulties can be partially overcome if we consider that an interval of proper time is independent of acceleration. Hence the analysis of space-time relations in an infinitely small space-time region of a noninertial reference frame can be carried out by using the space-time relations in an inertial reference frame which moves at the same velocity, but without any acceleration, as a corresponding infinitely small region of the noninertial reference frame. Such an inertial reference frame is called an accompanying frame. This enables us to establish the dependence between physical quantities if they are determined by space-time relations in an infinitely small region. These dependences can be then extended to finite regions. This, however, is a complicated approach and will not be used here.

We shall confine the analysis to the motion at a low velocity when these difficulties do not arise and we can make use of Galilean transformations by assuming that the space-time relations in a noninertial frame are the same as if it were an inertial frame.

INERTIAL FORCES. The only reason behind an accelerated motion of a body in inertial reference frames is the forces acting on it from other bodies. Force is always the result of interaction between bodies.

In noninertial reference frames, a body can be accelerated by simply varying the state of motion of the reference frame. For example, let us consider a noninertial reference frame fixed to a car. As the velocity of the car changes relative to the Earth's surface, all celestial bodies in this reference frame experience an acceleration. *Obviously, these accelerations are not the result of any forces acting on the celestial bodies on the part of other bodies. Thus, in noninertial reference frames there exist accelerations which are not associated with the forces of the same nature as in inertial reference frames.* Hence Newton's first law of motion becomes meaningless in such frames. So far as the interactions of bodies are concerned, Newton's third law of motion is generally satisfied in such systems. However, since accelerations of bodies in noninertial reference frames are caused not only by the "normal" forces of interaction between bodies, the manifestations of Newton's third law are distorted to such an extent that it becomes devoid of any clear physical meaning.

?

Indicate the manifestations of inertial forces in noninertial reference frames.
Describe how inertial navigation systems act.
The position of a rocket relative to the starting point can be determined by an inertial navigation system without observing any phenomena or objects outside the rocket. Doesn't this contradict the relativity principle? Why?

In principle, the theory of motion in noninertial reference frames could also be constructed by making radical changes in the concepts developed for inertial reference frames. To wit, it could be assumed that accelerations of bodies are caused not only by forces, but also by some other factors which have nothing in common with forces. However, a different approach was adopted historically. These other factors were considered to be forces which are related to accelerations in the same way as ordinary forces. *Moreover, it is assumed that as in inertial reference frames, accelerations in noninertial reference frames are caused only by forces, but in addition to the ordinary forces of interaction there also exist forces of a special kind, called inertial forces.* The formulation of Newton's second law of motion remains unchanged, but the inertial forces must also be considered in addition to the forces of interaction. The existence of inertial forces is due to the accelerated motion of a noninertial reference frame relative to an inertial reference frame. *Inertial forces are chosen in such a way as to ensure in the noninertial reference frame the accelerations that actually exist in the frame, but are only partially accounted for by the forces of interaction.* Hence Newton's second law of motion in a noninertial reference frame can be written in the form

$$m\mathbf{a}' = \mathbf{F} + \mathbf{F}_{in}, \tag{27.1}$$

where $\mathbf{a}'$ is the acceleration in the noninertial reference frame, $\mathbf{F}$ denote "ordinary" forces of interaction, and $\mathbf{F}_{in}$ are inertial forces.

ON THE REALITY OF EXISTENCE OF INERTIAL FORCES. Are inertial forces real forces? Inertial forces are real forces in the same sense in which accelerations are real in noninertial reference frames since the inertial forces are introduced only to account for these accelerations. They are also real in a more profound sense: while considering physical phenomena in noninertial frames, we can indicate specific physical consequences of the action of inertial forces. For example, inertial forces may jerk the passengers in a train out of their seats, which is quite a real and tangible result. *Hence inertial forces are as real as the fact that bodies move uniformly in a straight line in inertial reference frames if there are no "ordinary" forces of interaction, a formulation used for Newton's first law of motion.*

Inertial forces have an important practical application. For example, inertial navigation systems can be used to determine the position of an aeroplane or a rocket to a very high degree of precision with the help of instruments which measure inertial forces, without making any measurements of the position of the aeroplane or the rocket relative to the Earth.

DETERMINATION OF INERTIAL FORCES. In order to describe the motion of bodies in a noninertial reference frame with the help of Eq. (27.1), we must indicate the method of determining the inertial forces appearing on the right-hand side of this equation. Inertial forces characterize that part of the acceleration of a body which emerges due to the accelerated motion of the coordinate system relative to an inertial frame. Let us write the equation of motion of a body in noninertial and inertial reference frames:

$$m\mathbf{a}' = \mathbf{F} + \mathbf{F}_{in}, \tag{27.2}$$

$$m\mathbf{a} = \mathbf{F}, \tag{27.3}$$

where the "ordinary" forces of interaction $\mathbf{F}$ are identical in both reference frames, and $\mathbf{a}'$ and $\mathbf{a}$ are the accelerations in noninertial and inertial reference frames.

From (27.2) and (27.3), we obtain the following expression for the inertial forces:

$$\boxed{\mathbf{F}_{in} = m(\mathbf{a}' - \mathbf{a}).} \tag{27.4}$$

Usually, the following terminology is used in noninertial reference frames. The acceleration $\mathbf{a}$ relative to an inertial reference frame is called absolute, while the acceleration $\mathbf{a}'$ relative to a noninertial reference frame is called relative. *Formula* (27.4) *shows that the inertial forces cause a difference between the relative and absolute accelerations. This means that the inertial forces exist only in noninertial reference frames.* The introduction of these forces in the equations of motion, their application for explaining physical phenomena, etc. in noninertial reference frames are correct and logical. *However, the use of the concept of inertial forces in the analysis of motion in inertial reference frames is erroneous since such forces do not exist in these frames.*

Sec. 28. NONINERTIAL REFERENCE FRAMES OF TRANSLATIONAL MOTION IN A STRAIGHT LINE

Inertial forces in noninertial reference frames having a translational motion in a straight line are described and their manifestations are considered.

EXPRESSION FOR INERTIAL FORCES. Suppose that a noninertial system moves in a straight line along the X-axis of an inertial system (Fig. 62). Obviously, the coordinates of a point in these systems are connected through the relations

$$x = x_0 + x', \quad y = y', \quad z = z', \quad t = t'. \tag{28.1}$$

Consequently,

$$\frac{dx}{dt} = \frac{dx_0}{dt} + \frac{dx'}{dt}, \quad v = v_0 + v', \tag{28.2}$$

where $v = dx/dt$, $v_0 = dx_0/dt$ and $v' = dx'/dt$ are called the absolute, transport and relative velocities respectively.

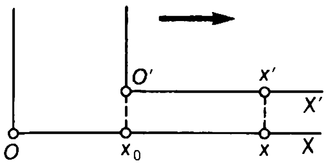


Fig. 62. A noninertial reference frame in a translational motion in a straight line.

Going over to accelerations in (28.2), we obtain

$$\frac{dv}{dt} = \frac{dv_0}{dt} + \frac{dv'}{dt}, \qquad a = a_0 + a', \tag{28.3}$$

where

$$a = \frac{dv}{dt}, \quad a_0 = \frac{dv_0}{dt}, \quad a' = \frac{dv'}{dt} \tag{28.4}$$

are called the absolute, transport and relative accelerations respectively. Consequently, in accordance with the definition (27.4), the expression for inertial forces in a noninertial reference frame moving in a straight line has the form

$$F_{in} = m(a' - a) = -ma_0, \tag{28.5}$$

or, in vector notation,

$$\boxed{\mathbf{F}_{in} = -m\mathbf{a}_0,} \tag{28.6}$$

i.e. *the inertial force is directed against the transport acceleration of a noninertial reference frame.*

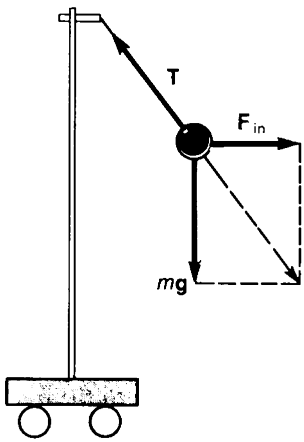


Fig. 63. Equilibrium of a pendulum in a noninertial reference frame.

PENDULUM ON A CART. Let us consider the equilibrium state of a pendulum in a noninertial reference frame moving in the horizontal direction with a translational acceleration a_0 (Fig. 63). The forces acting on the pendulum are shown in the figure. The equation of motion of the pendulum has the form

$$m\mathbf{a}' = \mathbf{T} + \mathbf{P} + \mathbf{F}_{in} = \mathbf{T} + \mathbf{P} - m\mathbf{a}_0 = 0, \tag{28.7}$$

i.e. $\mathbf{a}' = 0$. It is also clear that $\tan \alpha = a_0/g$, where α is the angle between the pendulum suspension and the vertical.

The acting forces and the equation of motion change in an inertial reference frame (Fig. 64). There is no inertial force in this case, and the only forces acting in the frame are the force $\mathbf{T}$ of the stretched string and the force of gravity $\mathbf{P} = m\mathbf{g}$. The equilibrium condition can be written in the form

$$m\mathbf{a} = \mathbf{T} + \mathbf{P} = m\mathbf{a}_0. \tag{28.8}$$

It is also clear that $\tan \alpha = a_0/g$.

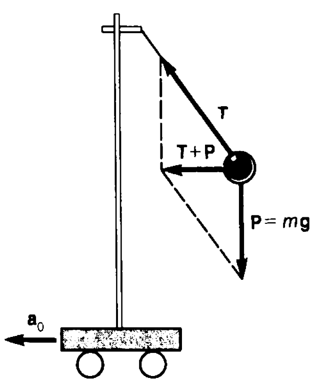


Fig. 64. Equilibrium of a pendulum under a uniform acceleration in an inertial reference frame.

A FALLING PENDULUM. A falling pendulum can be used for a very effective demonstration of phenomena in noninertial reference frames moving in a straight line. The pendulum is suspended from a massive frame which can fall freely by sliding down vertical ropes having a very low friction (Fig. 65*a*). When the frame is at rest, the pendulum performs natural oscillations. The frame can be brought into the state of free fall for any phase of the pendulum's oscillation. The motion of the pendulum during the free fall of the frame will depend on the phase in which the free fall began. If the deviation of the

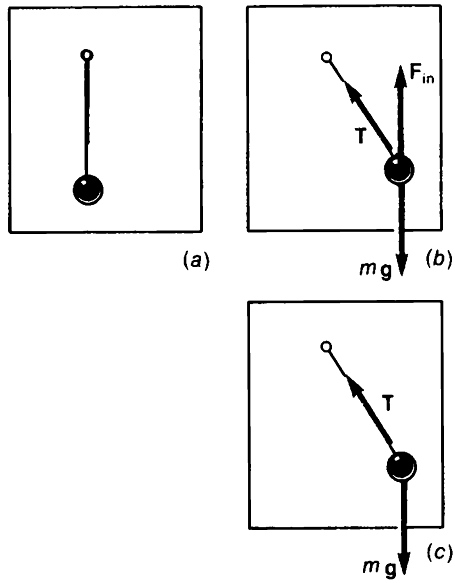


Fig. 65. Diagram of forces acting on a falling pendulum in (*b*) a noninertial reference frame fixed to the pendulum; (*c*) an inertial reference frame in which the pendulum falls with the free fall acceleration. The equilibrium position of the pendulum is shown in (*a*).

pendulum from the equilibrium position is maximum when the free fall begins, it will remain at the same point relative to the frame. If, however, its deviation at the start of its free fall is not maximum, it will have a certain velocity relative to the frame. During the fall of the frame, the magnitude of this velocity relative to the frame remains unchanged, and only its direction changes. As a result, the pendulum rotates uniformly about the point of suspension.

Let us consider this phenomenon in the noninertial reference frame fixed to the frame (Fig. 65*b*). The equation of motion has the form

$$m\mathbf{a}' = \mathbf{T} + \mathbf{P} + \mathbf{F}_{in} = \mathbf{T} + m\mathbf{g} - m\mathbf{g} = \mathbf{T}. \qquad (28.9)$$

This is the motion of a point mass rotating under the tension of the string in a circle whose centre lies at the point of suspension. The motion is circular with a linear velocity equal to the initial velocity. The tension of the string is the centripetal force which ensures the uniform motion of the pendulum in a circle and is mv'^2/l, where l is the length of the suspension, and v' is the velocity at which the pendulum moves relative to the frame.

There are no inertial forces in an inertial reference frame. The forces acting on the pendulum, which are shown in Fig. 65*c*, are just the tension of the string and the force of gravity. The equation of motion has the form

$$m\mathbf{a} = \mathbf{P} + \mathbf{T} = m\mathbf{g} + \mathbf{T}. \qquad (28.10)$$

In order to find the solution of this equation, let us represent the total acceleration of the pendulum as the sum of two accelerations: $\mathbf{a} = \mathbf{a}_1 + \mathbf{a}_2$. In this case, Eq. (28.10) can be decomposed into two equations:

$$m\mathbf{a}_1 = \mathbf{T}, \quad m\mathbf{a}_2 = m\mathbf{g}. \tag{28.11}$$

The first equation is identical to (28.9) and describes the rotation about the point of suspension, while the second equation has the solution $\mathbf{a}_2 = \mathbf{g}$, i.e. describes the free fall of the pendulum.

In the examples considered above, the analysis of motion was equally simple and illustrative both in noninertial and inertial reference frames. This is due to the fact that the examples were chosen specially with a view to illustrate the relations between noninertial and inertial reference frames. *However, quite often the solution of a problem in a noninertial reference frame is found to be much simpler than in an inertial reference frame.* For example, the analysis of a cylinder rolling down an inclined plane with a uniform acceleration in an arbitrary direction is much easier to obtain in the noninertial reference frame fixed to the inclined plane than in the inertial reference frame in which the plane moves with an acceleration.

Measurement of inertial forces allows us to determine the absolute acceleration of a reference frame relative to fixed stars. The instruments used for this purpose are called accelerometers.

Example 28.1. Solve the problem considered in Example 19.1 if the horizontal plane on which the beam is located moves upwards with an acceleration a_0.

To solve the problem in an inertial reference frame, we must consider in (19.9) and (19.10) the change in the reaction force of the supports on account of an additional acceleration of the masses in the vertical direction. However, it is easier to solve the problem in a noninertial reference frame moving upwards with a uniform acceleration a_0. In this case, an additional inertial force acts in the vertical direction, and the entire problem is reduced to a change in the force of gravity. The solutions for a_1 and a_2 have the same form as (19.11), except that g is now replaced by $g \to g + a_0$.

The problem can easily be solved for an arbitrary direction of the acceleration $\mathbf{a}_0$. In this case, we must take into account in (19.9) and (19.10) the action of inertial forces both in the vertical and horizontal directions.

!

Measuring inertial forces, we can determine the absolute acceleration of a reference frame relative to fixed stars. The corresponding instruments are called accelerometers. They are used in inertial navigation systems.

?

When and why must inertial forces be considered? What is the general method of determining inertial forces? What inertial forces exist in inertial reference frames in a translational motion?

Sec. 29. ZERO GRAVITY. THE EQUIVALENCE PRINCIPLE

Physical conditions behind the emergence of zero gravity are discussed and the equivalence principle is formulated.

ZERO GRAVITY. It was shown in the example concerning a falling pendulum that *in a freely falling noninertial reference frame, the inertial forces completely compensate for the action of the force of gravity, and the motion takes place as if there were neither inertial forces nor the forces of gravity. The state of zero gravity sets in.* This situation is widely used for simulating weightlessness under terrestrial conditions, for example, to train cosmonauts.

The state of zero gravity arises in a diving aeroplane if its acceleration towards the Earth is equal to the acceleration due to gravity at each instant of time. In order that the aeroplane may remain in the state of zero gravity for a long time, the entire operation is carried out through the "steep-climb" maneuver, which helps avoid large diving angles and a drastic increase in the velocity of the aeroplane. Cosmonauts can thus experience zero gravity for a long time and can work out the methods of moving inside the cabin and performing various operations.

GRAVITATIONAL AND INERTIAL MASSES. The emergence of the state of zero gravity during free fall is due to a very important physical factor, viz. the equality of the inertial and gravitational masses of a body. *The inertial mass characterizes the inertial properties of a body, while the gravitational mass characterizes the force with which the body is attracted according to the law of universal gravitation.* The gravitational mass has the same meaning as, say, an electric charge in an electrostatic interaction. Generally speaking, there are no indications that the inertial and gravitational masses of a body must be proportional or, which is the same, equal to each other (if two physical quantities are proportional to each other, they can be made equal to each other by a suitable choice of the units of these physical quantities). Let us prove that the inertial and gravitational masses of a body are proportional. The force exerted by the Earth whose gravitational mass is M_g on a body whose gravitational mass is m_g on the Earth's surface is

$$F = G\frac{m_g M_g}{R^2}, \tag{29.1}$$

where G is the gravitational constant, and R is the Earth's radius. If m is the inertial mass of the body, it will acquire an acceleration under the action of the force (29.1). This acceleration is given by

$$g = \frac{F}{m} = G\frac{M_g}{R^2}\frac{m_g}{m} = \text{const}\,\frac{m_g}{m}. \tag{29.2}$$

Since the acceleration g for all bodies is the same at the Earth's surface, the ratio of their inertial and gravitational masses must be the same, i.e. *the inertial and gravitational masses are proportional to each other. By an appropriate choice of the units of physical quantities, these two masses can be made equal to each other. We can then speak of mass in general without indicating which of the two masses is meant.* Owing to the circumstance that the inertial and gravitational masses are equal, the inertial and gravitational forces neutralize each other in the case of free fall and can be neglected.

In view of the fact that the equality of inertial and gravitational masses is of great importance, it was verified very thoroughly in various experiments. At present, the equality of these masses is assumed to have been proved to within 10^{-12} of their value, i.e. $|(m_g - m)/m_g| \leqslant 10^{-12}$.

The equality of inertial and gravitational masses has another corollary:

If a reference frame is in a uniformly accelerated motion in a straight line relative to an inertial reference frame (in which there is no gravitational field by definition), processes occur in it as if there were gravitational field in which the free fall acceleration is equal to the acceleration of the reference frame.

This is obvious for phenomena occurring in mechanics. The generalization of this statement to all physical phenomena is called the equivalence principle.

EQUIVALENCE PRINCIPLE. The equivalence principle states that *the acceleration of a reference frame is indistinguishable from the presence of a corresponding gravitational field.*

A specific gravitational field varies as we go from one point in space to another. Hence we cannot generally choose a reference frame which could move so that the action of its acceleration at each point in space is equivalent to the gravitational field at this point. However, if we have to consider a gravitational field in quite a small region in space, it can be assumed to be constant in this region to the first approximation. Hence the equivalence principle can always be applied in quite small regions of space, and certain conclusions can thus be drawn about the nature of processes occurring in this region. We shall illustrate this by considering red shift.

RED SHIFT. Gravitational field has a considerable influence on light in that it changes the frequency of light. *The inevitability of the variation of frequency of light in gravitational field is a consequence of the equivalence principle.*

Let us imagine the following experiment in the Earth's gravitational field. A ray of light of frequency ω is emitted from a point in space and propagates in the vertical direction (Fig. 66). What will be the frequency of this light at a height h?

•

!

Physically, zero gravity is due to the proportionality of inertial and gravitational masses which are physical quantities of different nature. At present, the proportionality of inertial and gravitational masses has been verified with a high degree of precision.
The equivalence principle can strictly be applied only to small regions of space.
The expression "red shift" has two meanings:
(1) this is the Doppler effect for a radiation source moving away (for example, the "red shift" in the spectrum of remote galaxies);
(2) when a change in frequency is due to gravitational field.

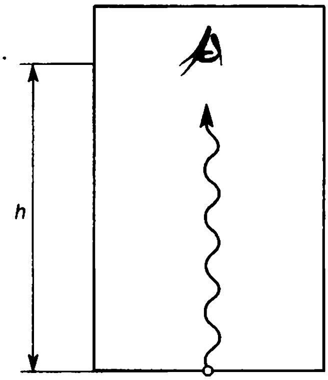


Fig. 66. Calculating red shift.

This question cannot be answered proceeding from general considerations since we are unaware of the action of gravity on frequency. We can make use of the equivalence principle to answer this question by considering that in the absence of gravitational field, the frequency of light does not change during its propagation.

We shall consider such an experiment in a reference frame falling freely in a uniform gravitational field. There are no forces in this reference frame, and all the processes occurring in it take place in the same way as in an inertial reference frame. Hence the frequency of light does not change upon propagation, and a stationary person at a point at height h in this reference frame must receive the same frequency that was emitted at point O in the same reference frame.

Let us now analyze the same experiment in the laboratory reference frame fixed to the Earth, in which a noninertial reference frame falls freely. We shall assume that at the instant of emission of the ray at point O, the velocity of this reference frame is zero (the acceleration, however, is not zero and is equal to the free fall acceleration). In time $\Delta t = h/c$ during which the ray of light propagates from point O to the point of observation at height h, the freely falling reference frame acquires a velocity $v = g\,\Delta t = gh/c$. Consequently, on account of the Doppler effect, an observer must receive in this reference frame a ray whose frequency is higher than that of the light emitted at point O by $\Delta\omega = \omega(v/c)$. However, an analysis of the phenomena in the noninertial reference frame showed that there was no change in frequency. Hence it can be concluded that during the propagation of light from point O to the point at height h, the frequency of the emitted light has decreased by $\Delta\omega = -\omega gh/c^2$. For visible light, this means a shift of the frequency towards the red colour of the spectrum. *Hence the effect of decrease in the frequency during the propagation of light against the force of gravity is called red shift.*

The magnitude of the red shift under terrestrial conditions is very small. For a difference of 10 m in height, we obtain the following estimate for the red shift:

$$\frac{\Delta\omega}{\omega} \simeq \frac{10 \times 10}{(3 \times 10^8)^2} \simeq 10^{-15}. \tag{29.3}$$

To note such a change in frequency is equivalent to observing a loss of one second in about 100 million years. However, the red shift, which is negligibly small under terrestrial conditions, was reliably detected in 1960 by using the Mössbauer effect according to which photons are emitted by a nucleus under

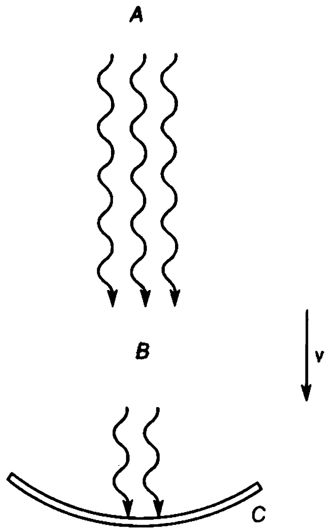


Fig. 67. Experimental set-up for detecting the red shift under terrestrial conditions.

certain conditions practically without recoil. The question of recoil during emission will be considered in Sec. 41. The condition for the recoilless radiation is that the recoil momentum upon a photon emission is taken not by an individual atom, but rather by the entire crystal lattice. The essence of the Mössbauer effect lies in the feasibility of such conditions. Because of the recoilless radiation, the emission line is found to be extremely narrow, i. e. the frequency spread of the emitted photons covers a very narrow region. On the other hand, the absorption of a photon also will take place only when its frequency is almost exactly equal to the recoilless emission frequency.

Suppose that a substance A (Fig. 67) emits recoilless photons of a certain frequency, and a similar substance B can absorb the photons of the same frequency under the same conditions. A certain number of photons pass through the substance B without getting absorbed and falls on a sensitive counter C which registers their number.

Suppose that the frequency of the photons changed for some reason during their propagation from A to B. In this case, they cannot be absorbed by the substance B, and the number of photons incident on the counter C will increase. Thus, the slightest variation in the frequency of the photons can be detected during their propagation from A to B. The same set-up can also be used to measure the variation in the emission frequency of the photons. For this purpose, it is necessary to displace the substance B along the line of propagation of a ray at a velocity v such that the frequency of the photons incident on it again becomes equal to the resonance absorption frequency owing to the Doppler effect. At this instant, the absorption again increases considerably, and the intensity of radiation registered by the counter C drops. The effect is very clearly manifested, and the velocity v can be determined with a very high precision. Consequently, we can measure the variation in the frequency of the photons during their propagation from A to B. In the experiments which were performed in 1960 and repeated several times afterwards, the height of the source A above the detector B was 15 m. The red shift was reliably measured, and the validity of formula (29.3) was confirmed.

Red shift becomes noticeable in the radiation emitted by stars since the stars have a larger mass than the Earth. For example, the formula for the red shift is confirmed by the data on the emission of radiation of Sirius.

The red shift caused by gravitational field should not be confused with the cosmological red shift due to the expansion of the Universe.

?

What physical factor is responsible for the emergence of zero gravity in the case of free fall?
What is gravitational mass?
What experiments demonstrate the proportionality of inertial and gravitational masses?
Formulate the equivalence principle.
What is the gravitational red shift? Can you evaluate its magnitude in the simplest case by means of the equivalence principle?
What experimental proofs of the gravitational red shift do you know?

The gravitational red shift is a direct consequence of the time dilatation in gravitational fields.

The passage of time at the Earth's surface is much slower than at a height. Consequently, one oscillation of the time standard at a height corresponds to more than one oscillation of the same time standard at the Earth's surface. This means that the frequency of light increases upon approaching the Earth's surface and decreases as we move away from the Earth.

Sec. 30. NONINERTIAL ROTATING REFERENCE FRAMES

It is shown that the transport acceleration leads to centrifugal inertial forces and a change in the transport acceleration from point to point is due to the emergence of Coriolis forces.

CORIOLIS ACCELERATION. The analysis of noninertial reference frames moving in a straight line showed that the relations between absolute, transport and relative velocities and the corresponding accelerations were identical (see (28.2) and (28.3)). *The situation is more complicated in the case of rotating reference frames. This difference is due to the fact that different points of a rotating reference frame have different transport velocities.* As before, the absolute velocity is the sum of the transport and relative velocities:

$$\mathrm{v} = \mathrm{v}_0 + \mathrm{v}', \tag{30.1}$$

while the absolute acceleration cannot be represented in this form.

As a point mass moves from one point of a rotating reference frame to another, its transport velocity changes. *Hence, even if the relative velocity of the point mass does not change during motion, it must experience an acceleration which is different from the transport acceleration.* As a result, the expression for the absolute acceleration in a rotating reference frame is not just the sum of the transport and relative accelerations, but contains an additional acceleration $\mathrm{a}_{\mathrm{Cor}}$ called the Coriolis acceleration:

$$\mathrm{a} = \mathrm{a}_0 + \mathrm{a}' + \mathrm{a}_{\mathrm{Cor}}. \tag{30.2}$$

EXPRESSION FOR CORIOLIS ACCELERATION. In order to find the physical nature of the Coriolis acceleration, let us consider the motion in the rotational plane. Primarily, we are interested in the motion of a point at a constant relative velocity along the radius (Fig. 68). Figure 68 shows the position of the point at two instants of time separated by an interval Δt during which the radius turns by an angle $\Delta\alpha = \omega\,\Delta t$. The velocity v_r along the radius varies during this time

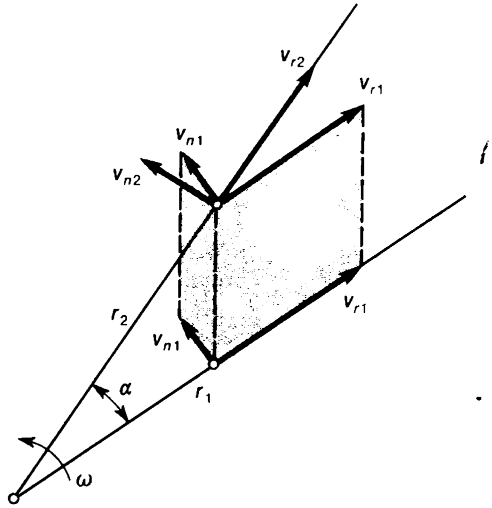


Fig. 68. The Coriolis acceleration emerging due to different values of the transport velocity at different points of a noninertial reference frame.

only in direction, while the velocity v_n, which is perpendicular to the radius, varies in both direction and magnitude. The magnitude of the total variation of the velocity perpendicular to the radius is

$$\begin{aligned}\Delta v_n &= v_{n2} - v_{n1}\cos\alpha + v_r\,\Delta\alpha\\ &= \omega r_2 - \omega r_1 \cos\alpha + v_r\,\Delta\alpha \approx \omega(r_2 - r_1) + v_r\omega\,\Delta t\\ &= \omega\,\Delta r + v_r\omega\,\Delta t,\end{aligned} \tag{30.3}$$

where we have considered that $\cos\alpha \simeq 1$.

Consequently, the magnitude of the Coriolis acceleration is

$$a_{\mathrm{Cor}} = \lim_{\Delta t\to 0}\frac{\Delta v_n}{\Delta t} = \omega\frac{\mathrm{d}r}{\mathrm{d}t} + v_r\omega = 2v_r\omega. \tag{30.4}$$

It can easily be seen from a consideration of the directions of various quantities in Fig. 68 that this expression can be represented in vector form as follows:

$$\mathbf{a}_{\mathrm{Cor}} = 2\boldsymbol{\omega}\times\mathbf{v}', \tag{30.5}$$

where $\mathbf{v}'$ is the relative velocity directed along the radius in the present case.

If a point moves in a direction perpendicular to the radius, i.e. in a circle, the relative velocity $v' = \omega r$, while the angular velocity of rotation of the point in a stationary reference frame will be $\omega + \omega'$, where ω is the angular velocity of the rotating reference frame. For the absolute acceleration, we obtain the following expression:

$$a = (\omega + \omega')^2 r = \omega^2 r + \omega'^2 r + 2\omega\omega' r. \tag{30.6}$$

The first term on the right-hand side is the transport acceleration, while the second term represents the relative acceleration. The last term $2\omega\omega' r = 2\omega v'$ is the Coriolis acceleration. All the accelerations in (30.6) are directed along the radius towards the centre of rotation. Taking the direction into account, the Coriolis acceleration in (30.6) can be written as

$$\boxed{\mathrm{a}_{\mathrm{Cor}} = 2\omega \times \mathrm{v}',} \tag{30.7}$$

where v' is the relative velocity directed perpendicular to the radius in the present case.

Any velocity can be represented as the sum of the components along the radius and perpendicular to it, both components obeying the same formula of the type (30.7). It follows hence that (30.7) is valid for the Coriolis acceleration irrespective of the direction of the relative velocity.

If the velocity is directed parallel to the axis of rotation, no Coriolis acceleration emerges since adjacent points on the trajectory have the same transport velocity in this case.

We can obtain an expression for the Coriolis acceleration by adopting a more formal approach, i.e. by directly computing the absolute acceleration. Representing the radius vector of a moving point in the form

$$\mathrm{r} = \mathrm{i}'_x x' + \mathrm{i}'_y y' + \mathrm{i}'_z z' \tag{30.8}$$

and differentiating with respect to t after taking into account the time dependence of i'_x, i'_y and i'_z, we arrive at the expression for the absolute velocity:

$$\mathrm{v} = \omega \times \mathrm{r} + \mathrm{v}' = \mathrm{v}_0 + \mathrm{v}', \tag{30.9}$$

where $\omega \times \mathrm{r} = \mathrm{v}_0$ is the transport velocity, and

$$\mathrm{v}' = v'_x \mathrm{i}'_x + v'_y \mathrm{i}'_y + v'_z \mathrm{i}'_z \tag{30.10}$$

is the relative velocity. Hence we arrive at the expression for the absolute acceleration:

$$\mathrm{a} = \frac{\mathrm{d}\mathrm{v}}{\mathrm{d}t} = \omega \times \frac{\mathrm{d}\mathrm{r}}{\mathrm{d}t} + \frac{\mathrm{d}\mathrm{v}'}{\mathrm{d}t} = \omega \times (\mathrm{v}_0 + \mathrm{v}') + \mathrm{a}' + \omega \times \mathrm{v}'. \tag{30.11}$$

The angular velocity of rotation is assumed to be constant, and the following relation has been taken into consideration:

$$\frac{\mathrm{d}\mathrm{v}'}{\mathrm{d}t} = \frac{\mathrm{d}v'_x}{\mathrm{d}t}\mathrm{i}'_x + \frac{\mathrm{d}v'_y}{\mathrm{d}t}\mathrm{i}'_y + \frac{\mathrm{d}v'_z}{\mathrm{d}t}\mathrm{i}'_z$$

$$+ v'_x\frac{\mathrm{d}\mathrm{i}'_x}{\mathrm{d}t} + v'_y\frac{\mathrm{d}\mathrm{i}'_y}{\mathrm{d}t} + v'_z\frac{\mathrm{d}\mathrm{i}'_z}{\mathrm{d}t} = \mathrm{a}' + \omega \times \mathrm{v}'. \tag{30.12}$$

Hence the absolute acceleration is

$$\mathrm{a} = \mathrm{a}_0 + \mathrm{a}' + \mathrm{a}_{\mathrm{Cor}}, \tag{30.13}$$

!

The Coriolis acceleration is due to a change in the transport velocity when passing from one point in a rotating reference frame to another.
The Coriolis force, as an inertial force, is directed opposite to the Coriolis acceleration and is applied to a body.
The possibility of decomposing angular velocity is due to the vector nature of angular velocity.

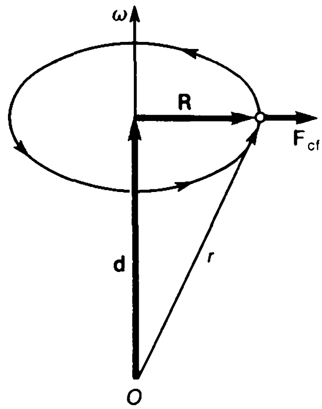


Fig. 69. The centrifugal inertial force.

where $a_0 = \omega \times v_0 = \omega \times (\omega \times r)$ is the transport acceleration, $a' = \frac{dv'_x}{dt} i'_x + \frac{dv'_y}{dt} i'_y + \frac{dv'_z}{dt} i'_z$ is the relative acceleration, and $a_{Cor} = 2\omega \times v'$ is the Coriolis acceleration. It is expedient to represent the transport acceleration in the form

$$a_0 = \omega \times (\omega \times r) = \omega(\omega \cdot r) - r\omega^2$$
$$= \omega^2 (d - r) = -\omega^2 R, \tag{30.14}$$

where $d = r - R$, and R is a vector perpendicular to the axis of rotation (Fig. 69). Thus, the transport acceleration is centripetal (it should be recalled that the angular velocity of rotation is assumed to be constant).

INERTIAL FORCES IN A ROTATING REFERENCE FRAME. The general formula (28.6) can be used to determine the inertial forces in a rotating reference frame by taking into account the expression (30.13) for the absolute acceleration. This gives

$$F_{in} = m(a' - a) = m(-a_0 - a_{Cor})$$
$$= m\omega^2 R - 2m\omega \times v' = F_{cf} + F_{Cor}. \tag{30.15}$$

The inertial force

$$\boxed{F_{cf} = m\omega^2 R} \tag{30.16}$$

connected with the transport acceleration is called the centrifugal inertial force. It is directed along the radius away from the axis of rotation. The inertial force

$$\boxed{F_{Cor} = -2m\omega \times v'} \tag{30.17}$$

connected with the Coriolis acceleration is called the Coriolis force. It is perpendicular to the plane containing the vectors of the angular and relative velocities. If these vectors have the same direction, the Coriolis acceleration is zero.

EQUILIBRIUM OF A PENDULUM ON A ROTATING DISC. As an example, let us consider the equilibrium position of a pendulum on a rotating disc (Fig. 70). In the noninertial reference frame, the pendulum is subjected to a centrifugal inertial force. There is no Coriolis force in the equilibrium position, and hence the relative velocity is zero ($v' = 0$). The equation of motion (Fig. 70*a*) has the form

$$ma' = T + mg + F_{cf} = 0. \tag{30.18}$$

In the inertial reference frame, the equation of motion for a pendulum in equilibrium (Fig. 70*b*) has the form

$$ma = T + mg. \tag{30.19}$$

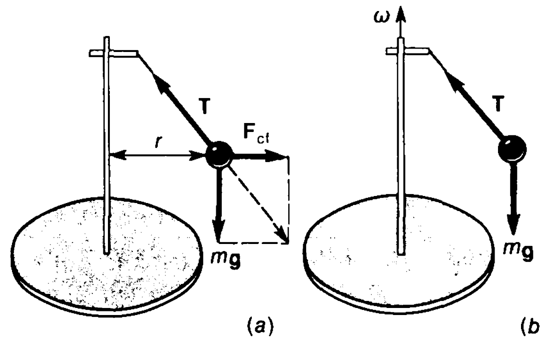


Fig. 70. Equilibrium of a pendulum in a rotating reference frame:
(*a*) noninertial; (*b*) inertial.

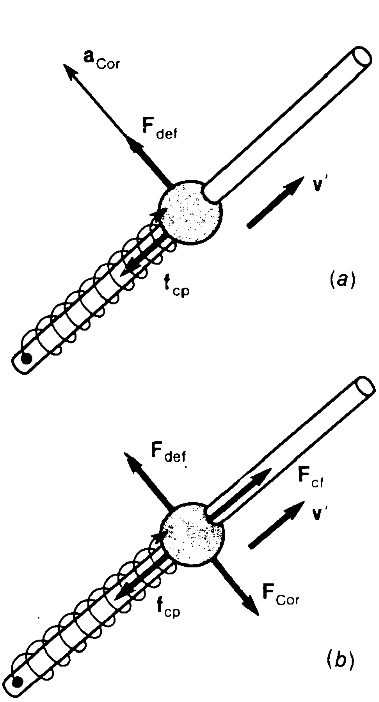


Fig. 71. The Coriolis force F_{Cor} applied to a body in a direction opposite to that of the Coriolis acceleration a_{Cor}.

It can be seen directly from Fig. 70 that $\tan\alpha = \omega^2 r/g$ and $a = \omega^2 r$ (α is the angle between the vertical and the pendulum suspension).

MOTION OF A BODY ALONG A ROTATING ROD. Suppose that a rigid rod rotates about an axis perpendicular to it and passing through one of its ends (Fig. 71). A body is fixed to the rod by a spring, and the force from the spring is proportional to the distance of the body from the axis of rotation ($F = -kr$). If $k = m\omega^2$, the centrifugal inertial force $F_{cf} = m\omega^2 r$ at any distance from the axis of rotation is balanced by the force of the spring. In this case, the body moves along the rod at a constant velocity v' (relative to the rod). The rod is slightly deformed (see Fig. 71). Let us consider the motion and forces in an inertial (stationary) reference frame and a noninertial reference frame fixed to the rod.

In the inertial reference frame, the body is subjected to the action of two forces (Fig. 71*a*): (1) the centripetal force $F_{cp} = m\omega^2 r$ from the spring directed towards the axis of rotation at each instant (this force ensures the motion of the body about the axis of rotation); (2) the force F_{def} exerted by the deformed rod (for a very rigid rod, this deformation may be indefinitely small, but the force has a finite value), which imparts the Coriolis acceleration a_{Cor} to the body (this is the usual force due to the deformation of the rod).

In the noninertial reference frame fixed to the rotating rod, there are four forces which balance one another, and hence the body in this system moves uniformly without an acceleration (Fig. 71*b*). These forces are: (1) the centrifugal inertial force $F_{cf} = m\omega^2 r$ directed along the rod away from the axis of rotation; (2) the centripetal force $F_{cp} = kr = m\omega^2 r$ from the spring directed along the rod towards the axis of rotation; (3) the Coriolis inertial force F_{Cor} applied to the body. It should be emphasized that this force is applied to the body and not to the rod. The rod is deformed due to the usual

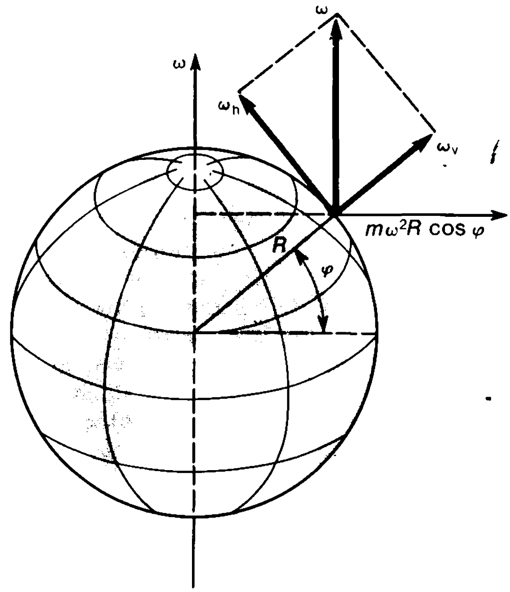


Fig. 72. A reference frame fixed to the Earth's surface.

interaction of deformed bodies and not because the Coriolis force is applied to it. The situation is analogous to that of a body lying on a table: the force of gravity acts on the body, while the table is subjected to the force exerted by the body due to its deformation but not the force of gravity. (4) The body is subjected to the force $\mathbf{F}_{def}$ exerted by the rod due to its deformation. This force is equal to the Coriolis force, but acts in the opposite direction.

NONINERTIAL REFERENCE FRAME FIXED TO THE EARTH'S SURFACE. Since the Earth rotates, the reference frame fixed to its surface is a noninertial rotating reference frame.

The angular velocity of rotation at any point on the surface can be conveniently decomposed into vertical and horizontal components (Fig. 72): $\omega = \omega_v + \omega_h$. At a latitude φ, these components are respectively equal to $\omega_v = \omega \sin\varphi$ and $\omega_h = \omega\cos\varphi$.

The centrifugal inertial force $m\omega^2 R\cos\varphi$, where R is the Earth's radius, lies in the meridional plane. In the northern hemisphere, it is inclined southwards from the vertical by an angle φ, while in the southern hemisphere, it is inclined northwards by the same angle. Thus, the vertical component of this force changes the force of gravity, while the horizontal component is directed tangentially to the Earth's surface along the meridian to the equator.

The Coriolis force depends on the relative velocity of the

?

What inertial forces arise in a rotating noninertial reference frame?
What are the factors responsible for the emergence of Coriolis forces?
Is any work done by Coriolis forces? By centrifugal forces?
Which two types of trajectory can be observed in the oscillations of the Foucault pendulum? How can this be realized?
Can you indicate how the inertial forces manifest themselves in the motion of bodies near the Earth's surface?

body. It is convenient to decompose this velocity into vertical and horizontal components: $\mathbf{v}' = \mathbf{v}'_v + \mathbf{v}'_h$. The Coriolis force can then be represented in the form

$$\begin{aligned} \mathbf{F}_{Cor} &= -2m(\boldsymbol{\omega}_v + \boldsymbol{\omega}_h) \times (\mathbf{v}'_v + \mathbf{v}'_h) \\ &= -2m\boldsymbol{\omega}_v \times \mathbf{v}'_h - 2m\boldsymbol{\omega}_h \times \mathbf{v}'_v - 2m\boldsymbol{\omega}_h \times \mathbf{v}'_h, \end{aligned} \tag{30.20}$$

where we have considered that $\boldsymbol{\omega}_v \times \mathbf{v}'_v = 0$.

The vertical component $\mathbf{v}'_v$ of the velocity is responsible for the component $-2m\boldsymbol{\omega}_h \times \mathbf{v}'_v$ of the Coriolis force in the horizontal plane perpendicular to the meridional plane. If the body moves upwards, the force is directed westwards, while if it moves downwards, the force is directed eastwards. Hence a body freely falling from a considerable height is deflected eastwards from the vertical pointing towards the Earth's centre. Obviously, this force deflecting the body from the vertical is $2m\omega \cos \varphi v'_v$.

The horizontal component $\mathbf{v}'_h$ of the velocity is responsible for two components of the Coriolis force. The component $-2m\boldsymbol{\omega}_h \times \mathbf{v}'_h$ depends on the horizontal component of the angular velocity of the Earth's rotation and is directed along the vertical. This force either presses the body against the Earth or, conversely, tends to detach it from the Earth's surface depending on the directions of vectors $\boldsymbol{\omega}_h$ and $\mathbf{v}'_h$. This must be taken into consideration while studying the motion of bodies at sufficiently large distances, for example, in the flight of ballistic missiles.

The second component of the Coriolis force associated with the horizontal component $\mathbf{v}'_h$ of the velocity is $-2m\boldsymbol{\omega}_v \times \mathbf{v}'_h$. This is a horizontal force perpendicular to the velocity. If we look in the direction of velocity, it always points to the right in the northern hemisphere. The Coriolis force is responsible, for example, for the unequal wear of the rails of a two-track railway if trains run along each track in the same direction all the time. In this case, the Coriolis force applied to the centre of mass of a carriage creates a moment relative to the right rail, which must be balanced by increasing the reaction force from the right rail to the wheels. Hence the pressure of the right wheels on the rails is higher than that of the left wheels, and this results in a certain increase in the wear of the right rails as compared to the left rails. An important manifestation of the Coriolis force is a change in the plane of oscillations of a pendulum relative to the Earth's surface.

FOUCAULT PENDULUM. Let us consider the oscillations of a pendulum by taking into account the action of the horizontal component of the Coriolis force on it. The projection of a point mass of the pendulum on the horizontal plane moves

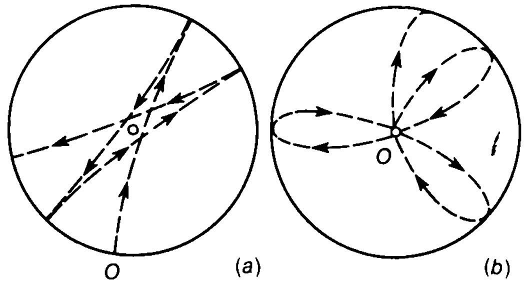


Fig. 73. Curves described by the tip of the Foucault pendulum:
(*a*) if it starts moving (from point *O*) from a nonequilibrium position without an initial velocity; (*b*) if it moves from the equilibrium position at a certain initial velocity.

along the curves shown in Fig. 73. The difference between the curves can be explained as follows.

If the pendulum is deflected from its equilibrium position and released at a zero initial velocity relative to the observer moving with the Earth, it begins to move towards the centre of equilibrium. However, the Coriolis force deflects it to the right, and it does not pass through the central point. Consequently, the projection of a point mass of the pendulum moves along the curve shown in Fig. 73*a*.

The pendulum can be set in motion in another way: a velocity can be imparted to it at the equilibrium position. In this case, the nature of its motion varies. As the pendulum moves away from the equilibrium position, the Coriolis force imparts to it an acceleration to the right. Hence when the pendulum is at the extreme position and its velocity along the radius from the centre is zero, it acquires the maximum velocity in the direction perpendicular to the radius. As a result, the trajectory of the pendulum touches a circle whose radius is equal to its maximum displacement from the equilibrium position. In this case, the projection of a point mass of the pendulum moves along the trajectory shown in Fig. 73*b*.

The deviation in the direction of the swing of the pendulum over one oscillation is extremely small. The entire process can be seen as a rotation of the plane of the swing of the pendulum about the vertical.

We can also analyze the oscillations of the Foucault pendulum in an inertial reference frame relative to fixed stars, the position of the plane of oscillations of the pendulum remaining unchanged relative to the stars. Due to the rotation of the Earth, the position of the plane in which the pendulum swings changes relative to its surface. This position is determined by the Foucault pendulum. It is easy to imagine this variation at the poles. The situation is more complicated for an arbitrary point on the Earth's surface, although the two cases are identical except that the angular velocity of rotation is ω_v in the latter case.

The angular velocity of rotation of the plane in which the pendulum swings is equal to $\omega_\perp$. Hence one revolution at the poles is completed in one day, while at a latitude φ this requires $1/\sin\varphi$ of a day. The plane of oscillations of the Foucault pendulum does not change at the equator.

CONSERVATION LAWS IN NONINERTIAL REFERENCE FRAMES. It was emphasized in Chap. 6 while considering the energy, momentum and angular momentum conservation laws that these laws in mechanics result from the equations of motion.

The energy, momentum and angular momentum of a system of point masses retain their values for closed systems, i.e. for the case when there are no external forces and momentum of external forces. The energy, momentum and angular momentum of a system change in the presence of external forces.

In noninertial reference frames, inertial forces appear in addition to the "ordinary" forces. These forces are always external forces with respect to the bodies under consideration. Consequently, there are no closed systems of bodies in these noninertial reference frames, and hence there are no energy, momentum and angular momentum conservation laws in the conventional sense.

However, there are no obstacles against the inclusion of inertial forces in the forces of a system, after which the system can be treated as a closed one. According to Eq. (27.2), the inertial forces must be considered exactly in the same way as ordinary forces. In particular, while calculating a change in energy, we must take into account the work of inertial forces, the moment of inertial forces in the momental equation, etc.

The nature of conservation laws in noninertial reference frames depends on the properties of inertial forces. In a noninertial reference frame rotating at a constant angular velocity, the inertial forces associated with the transport acceleration are central forces (to be more precise, axial forces directed in a straight line from the axis of rotation). It was shown above that the central forces are always potential forces. On the other hand, the Coriolis inertial force is perpendicular to the particle's velocity and hence does not perform any work. Consequently, the energy conservation law is valid in a noninertial reference frame rotating at a constant velocity if we consider, besides the usual potential energy, the potential energy associated with the inertial forces as well. It can easily be seen that the energy conservation law can also be formulated in a noninertial reference frame translating uniformly in a straight line if we take into account the work of inertial forces.

While considering a change in momentum and angular momentum, we must include the inertial forces and their moment in the equations of motion. To ensure the conservation of these quantities, it is necessary that the inertial forces satisfy the same requirements as those imposed on ordinary forces from the point of view of the conservation laws in inertial reference frames.

Example 30.1. Let us quantitatively consider the motion of a body near the Earth's surface in a noninertial reference frame fixed to this surface. We denote the acceleration due to gravity by g and neglect the air resistance.

It was assumed in Eqs. (30.8)-(30.14) that the origin of the radius vector r is at rest in the inertial reference frame. Let us take a point on the axis of rotation as such a point. Suppose that the origin of the noninertial reference frame on the Earth's surface is characterized by the radius vector $\mathbf{r}_0$, while the radius vector of the body near the Earth's surface is denoted by $\mathbf{r}'$ relative to this origin. We have

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{r}'. \tag{30.21}$$

The transport acceleration (30.14) is $\omega \times (\omega \times \mathbf{r}_0) + \omega \times (\omega \times \mathbf{r}')$ and Newton's equation of motion (27.2) assumes the following form if we take into account Eqs. (30.13) and (30.15):

$$m\ddot{\mathbf{r}}' = m\mathbf{g} - 2m\omega \times \mathbf{v}'$$
$$- m\omega \times (\omega \times \mathbf{r}_0) - m\omega \times (\omega \times \mathbf{r}'). \tag{30.22}$$

For the Earth, $\omega = 2\pi/86{,}164 = 7.29 \times 10^{-5}$ rad/s, and hence $|\omega \times (\omega \times \mathbf{r}_0)|/g \leqslant \omega^2 r_E/g = 3.45 \times 10^{-3}$, where $r_E = 6.37 \times 10^6$ m is the Earth's radius. The last term on the right-hand side of (30.22) is even smaller than the last but one since $r' \ll r_E$. Hence the last two terms in (30.22) can be neglected in comparison with the first, and the equation of motion can be written in the form

$$\ddot{\mathbf{r}}' = \mathbf{g} - 2\omega \times \mathbf{v}'. \tag{30.23}$$

For a small region near the Earth's surface in which the motion takes place, we can assume that $\mathbf{g} = \text{const}$ and is directed along the vertical. If necessary, the deviation from the vertical at different points on the Earth's surface and the variation of g with height and due to other factors can also be taken into account.

The consideration of the Coriolis force in (30.23) introduces only a small correction to the motion, which does not take this force into account. Hence we can write to the first approximation

$$\ddot{\mathbf{r}}' = \mathbf{g}, \quad \dot{\mathbf{r}}' = \mathbf{u} + \mathbf{g}t, \tag{30.24}$$

where u is the velocity at $t = 0$. It should be recalled that vector g points vertically downwards. Substituting $\dot{\mathbf{r}}' = \mathbf{v}'$ from (30.24) into (30.23), we obtain the equation of motion by taking into account the correction for the Coriolis force:

$$\ddot{\mathbf{r}}' = \mathbf{g} - 2\boldsymbol{\omega} \times (\mathbf{u} + \mathbf{g}t). \tag{30.25}$$

For the initial conditions $\mathbf{r}' = \mathbf{r}_0'$ and $\dot{\mathbf{r}}' = \mathbf{u}$ for $t = 0$, the solution of this equation is determined from two quadratures:

$$\dot{\mathbf{r}}' = \mathbf{u} + \mathbf{g}t - 2\boldsymbol{\omega} \times \left(\mathbf{u}t + \frac{\mathbf{g}t^2}{2}\right), \tag{30.26}$$

$$\mathbf{r}' = \mathbf{r}_0' + \mathbf{u}t + \frac{\mathbf{g}t^2}{2} - \boldsymbol{\omega} \times \left(\mathbf{u}t^2 + \frac{\mathbf{g}t^3}{3}\right). \tag{30.27}$$

The term $\boldsymbol{\omega} \times (\mathbf{u}t^2 + \mathbf{g}t^3/3)$ gives the correction to the coordinates of the body, induced by the Coriolis forces.

Let us calculate the deviation of a body from the vertical during a free fall. We direct the Z-axis along the vertical and the Y-axis eastwards. For $t = 0$, we have $x = 0$, $y = 0$, $z = h$, and (30.27) leads to the relations (see Fig. 65)

$$x = 0,\ y = \frac{1}{3}\omega g t^3 \cos\varphi,\ z = h - \frac{gt^2}{2}. \tag{30.28}$$

Consequently, at the point $z = 0$ where the body falls on the ground, the deviation of the body from the base of the vertical is given by

$$y = \frac{1}{3}\omega \cos\varphi \sqrt{\frac{8h^3}{g}}. \tag{30.29}$$

This deviation is very small. For example, $y = 0.022 \cos\varphi$ m for $h = 100$ m.

When a body moves at a high velocity along a nearly horizontal trajectory (for example, the flight of a bullet), we can neglect in (30.27) the term with $\mathbf{g}t^3/3$ in comparison with the term containing $\mathbf{u}t^2$. This gives

$$\mathbf{r}' = \mathbf{r}_0' + \mathbf{u}t + \frac{\mathbf{g}t^2}{2} + \omega t^2 \cos\varphi\, \mathbf{i}_x \times \mathbf{u} - \omega t^2 \sin\varphi\, \mathbf{i}_z \times \mathbf{u}. \tag{30.30}$$

The term $\omega t^2 \cos\varphi\, \mathbf{i}_x \times \mathbf{u}$ describes the deviation along the vertical caused by the Coriolis force, while the term $\omega t^2 \sin\varphi\, \mathbf{i}_z \times \mathbf{u}$ describes the deviation in the horizontal plane. In the northern hemisphere, $\sin\varphi > 0$ and the deviation occurs to the right of the velocity, while in the southern hemisphere, the deviation occurs to the left.

PROBLEMS

7.1. A bullet is fired at a target in the horizontal direction at a latitude $\varphi = 60°$. The initial velocity of the projectile is 1 km/s, and the distance from the target is 1 km. The target is situated precisely in the north-west direction. The air resistance is neglected. The correction for inertial forces was not taken into account in the calculations. By what distance will the bullet deviate from the target in the horizontal and vertical directions?

7.2. A river flows at a velocity v' in the northern hemisphere. What will be the angle of inclination α of the water surface to the horizontal at a latitude φ?

7.3. A weightless unstretchable string passes over a pulley. Loads of mass m_1 and m_2 are attached to the ends of the string. There is no friction between the pulley and the string. The pulley moves in the horizontal direction with an acceleration a. What will be the tension of the string?

7.4. Point masses m_1, m_2 and m_3 interacting in accordance with Newton's law are placed at the vertices of an equilateral triangle with side l. What should be the angular velocity of rotation of the system and how should the axis of rotation be arranged so that the position of the masses relative to one another remains unchanged?

ANSWERS

7.1. 5.8 cm and 2.4 cm. 7.2. $\arctan(2\omega \sin\varphi v'/g)$. 7.3. $2m_1m_2\sqrt{g^2 + a^2}/(m_1 + m_2)$. 7.4. $[G(m_1 + m_2 + m_3)/l^3]^{1/2}$, the axis of rotation passes through the centre of mass perpendicular to the plane of the triangle.

Chapter **8**

Dynamics of a Rigid Body

Basic idea:
The equations of motion of the centre of mass and momental equations of a system of point masses form a closed system of equations of motion of a rigid body.

Sec. 31. EQUATIONS OF MOTION

It is shown that the equations of motion of the centre of mass and momental equations of a system of point masses form a closed system of equations describing the motion of a rigid body.

SYSTEM OF EQUATIONS. In the sense in which a body was defined in Sec. 5, a rigid body can be considered a system of point masses with a constant distance between them. Hence all the statements and equations of Sec. 21 concerning a system of point masses are also applicable to a rigid body. Equations (21.11) and (21.21), which are reproduced below

$$\frac{d\mathbf{p}}{dt} = \mathbf{F}, \tag{31.1}$$

$$\frac{d\mathbf{L}}{dt} = \mathbf{M}, \tag{31.2}$$

do not generally form a closed system of equations since they represent only six scalar equations, while the number of degrees of freedom of a system of point masses is usually much larger. *However, these equations form a closed system for a rigid body, i.e. the motion of the rigid body in given external force fields can be completely defined with the help of these equations without any additional conditions and equations. Only the initial conditions of motion must be known.*

PROOF OF THE CLOSURE OF A SYSTEM OF EQUATIONS FOR A RIGID BODY. Let us recall the basic definitions presented in Sec. 9 for the kinematics of a rigid body. The orientation of a rigid body in space is completely defined by the direction of the axes of a rectangular Cartesian coordinate system rigidly fixed to the body, i.e. by the direction of the unit vectors i'_x, i'_y and i'_z in this coordinate system. The position of each point in this system is fixed and is specified either by the radius vector r' relative to the origin or by the Cartesian coordinates (x', y', z') of the point. Since the system of these coordinates is rigidly fixed to the body, the coordinates of each point of the body have a fixed value in this system. The orientation of this coordinate system relative to the inertial coordinate system, in which the motion of the body is considered and in which the equations of motion (31.1) and (31.2) hold, is completely defined by three Euler angles φ, ψ and θ (see Fig. 18). The position of the point with which the origin of the coordinate system $(\mathrm{i}'_x, \mathrm{i}'_y, \mathrm{i}'_z)$ is associated is given by the radius vector r_0 of this point relative to the inertial coordinate system, or by the Cartesian coordinates (x_0, y_0, z_0) of this point. Hence the position of a rigid body as a system with six degrees of freedom is defined by six quantities $(\varphi, \psi, \theta, x_0, y_0, z_0)$. The velocity of each point of the body is composed of both the translational motion at a velocity $\mathrm{v}_0 = \mathrm{dr}/\mathrm{d}t$ of the point of the rigid body coinciding with the origin of coordinates $(\mathrm{i}'_x, \mathrm{i}'_y, \mathrm{i}'_z)$ and the rotational motion at an instantaneous angular velocity ω about an axis passing through the origin. The combined velocity is thus given by (9.6):

$$\mathrm{v} = \mathrm{v}_0 + \omega \times \mathrm{r}'. \tag{31.3}$$

The angular velocity ω is expressed in terms of the derivatives of the Euler angles. Consequently, the velocity of all the points of a rigid body is completely defined by the coordinates $(\varphi, \psi, \theta, x_0, y_0, z_0)$ and their derivatives since the position of points relative to the coordinate system (X', Y', Z') is fixed. This means that both p and L in (31.1) and (31.2) can be expressed in terms of these coordinates and their derivatives. On the other hand, F and M in (31.1) and (31.2) can also be expressed in terms of these coordinates if the external forces are independent of the velocity, or in terms of both the coordinates and their derivatives if the forces depend on the velocity. Thus, we obtain six equations (31.1) and (31.2) for six unknown coordinates $(\varphi, \psi, \theta, x_0, y_0, z_0)$, i.e. the system is closed, and *these equations can rightfully be called the equations of motion of the rigid body. It must only be remembered that by forces and moments of forces on the right-hand side of these equations, we mean not only the ordinary forces and the moments of ordinary forces, but also the constraint forces imposed on the rigid body and their moments.*

!

The number of independent variables characterizing a system must be equal to the number of its degrees of freedom. Hence six independent variables are required to describe the motion of a perfectly rigid body. Then six independent equations of motion must be available to determine them.
In general, the inertial properties of a rigid body are characterized by six independent quantities, i.e. three axial and three centrifugal moments of inertia.
If the axes of a coordinate system are directed along the principal axes of inertia of a body, there are no centrifugal moments of inertia. The axial moments of inertia are called the principal moments of inertia.
Although strict mathematical rules exist for finding the principal axes, in many important cases, these axes can be determined from symmetry considerations, without resorting to mathematical calculations.

CHOICE OF A COORDINATE SYSTEM. The choice of point O' with which the coordinate system $(\mathrm{i}'_x, \mathrm{i}'_y, \mathrm{i}'_z)$ should be associated and the orientation of this system relative to a body are arbitrary and only a matter of convenience. A convenient choice can lead to a considerable simplification of these equations. One such choice for point O', viz. the centre of mass, was made in Sec. 21. In this case, (31.1) is transformed into (21.16):

$$m\left(\frac{\mathrm{d}\mathbf{v}_0}{\mathrm{d}t}\right) = \mathbf{F}, \tag{31.4}$$

called the equation of motion of the centre of mass. This is analogous to the equation of motion of a point mass.

It is assumed that F in (31.4) takes into account the constraint forces. For example, if a body is subjected to an external force, but is fixed at the centre of mass, then $\mathbf{F} = 0$. It should also be noted that the choice of point O' as the centre of mass is not always convenient. We shall prove this in the following section.

Sec. 32. MOMENTS OF INERTIA

The inertial properties of a rigid body are quantitatively analyzed with the help of the inertia tensor.

INERTIA TENSOR. In order to completely describe the motion of a rigid body, we must know not only the motion of one of its points, but also the motion of the body about this point. The most important concept in this case is that of the inertia tensor. In order to simplify the calculations, we consider the body to be an aggregate of point masses m_i (see Sec. 5).

Let us fix the body at point O (Fig. 74). The radius vector of point m_i drawn from O is denoted by $\mathbf{r}_i$. Let ω be the instantaneous angular velocity of the body. Then in accordance with (9.7), the velocity of the ith point of the body will be $\mathbf{v}_i = \omega \times \mathbf{r}_i$. Hence the angular momentum of the body as a whole relative to point O will be

$$\begin{aligned}\mathbf{L} &= \sum \mathbf{r}_i \times m_i\mathbf{v}_i = \sum m_i\mathbf{r}_i \times (\omega \times \mathbf{r}_i)\\ &= \omega\sum m_i r_i^2 - \sum m_i\mathbf{r}_i(\omega\cdot\mathbf{r}_i),\end{aligned} \tag{32.1}$$

where we have used the formula $\mathbf{A} \times (\mathbf{B} \times \mathbf{C}) = \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B})$ for the triple vector product.

The vector equality (32.1) can be written in the form of three projections on the coordinate axes:

$$\begin{aligned}L_x &= \omega_x\sum m_i r_i^2 - \sum m_i x_i(\mathbf{r}_i\cdot\omega),\\ L_y &= \omega_y\sum m_i r_i^2 - \sum m_i y_i(\mathbf{r}_i\cdot\omega),\\ L_z &= \omega_z\sum m_i r_i^2 - \sum m_i z_i(\mathbf{r}_i\cdot\omega).\end{aligned} \tag{32.2}$$

Considering that $(\mathbf{r}_i\cdot\omega) = x_i\omega_x + y_i\omega_y + z_i\omega_z$, we obtain in-

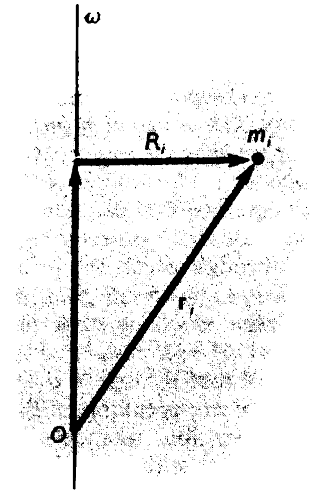


Fig. 74. To the concept of the inertia tensor characterizing the inertial properties of a rigid body.

stead of (32.2):

$$\begin{aligned} L_x &= J_{xx}\omega_x + J_{xy}\omega_y + J_{xz}\omega_z, \\ L_y &= J_{yx}\omega_x + J_{yy}\omega_y + J_{yz}\omega_z, \\ L_z &= J_{zx}\omega_x + J_{zy}\omega_y + J_{zz}\omega_z, \end{aligned} \tag{32.3a}$$

where

$$\begin{aligned} J_{xx} &= \sum m_i(r_i^2 - x_i^2), \\ J_{xy} &= -\sum m_i x_i y_i, \\ J_{xz} &= -\sum m_i x_i z_i. \end{aligned} \tag{32.3b}$$

The remaining quantities J_{yy}, J_{yx}, J_{yz}, etc. are also represented in a similar form. It can directly be seen from (32.3b) that $J_{xy} = J_{yx}$, $J_{xz} = J_{zx}$, and so on. Hence only six of the nine quantities J_{xx}, J_{xy}, etc. are different. The quantities J_{xx}, J_{yy} and J_{zz} are called the axial moments of inertia, while $J_{xy} = J_{yx}$, $J_{xz} = J_{zx}$ and $J_{yz} = J_{zy}$ are called the centrifugal moments of inertia.

Thus, the angular momentum of a body has quite a complex dependence on the mass distribution in the body, and its direction is generally not the same as that of the angular velocity of rotation of the body. The aggregate of the quantities

$$\begin{pmatrix} J_{xx} & J_{xy} & J_{xz} \\ J_{yx} & J_{yy} & J_{yz} \\ J_{zx} & J_{zy} & J_{zz} \end{pmatrix} \tag{32.4}$$

is called the inertia tensor. The quantities J_{xx}, J_{yy} and J_{zz} are called the diagonal elements of the tensor, while the other quantities are called the nondiagonal (off-diagonal) elements. In the present case, the quantities arranged symmetrically with respect to the diagonal are equal. Such a tensor is called a symmetric tensor.

PRINCIPAL AXES OF THE INERTIA TENSOR. Suppose that all the off-diagonal elements of a tensor are zero, and only the diagonal elements are not zero. Such a tensor has the form

$$\begin{pmatrix} J_x & & 0 \\ & J_y & \\ 0 & & J_z \end{pmatrix}. \tag{32.5}$$

In such cases, the axes of the body coinciding with the coordinate axes are called the principal axes of inertia, while the quantities $J_x = J_{xx}$, $J_y = J_{yy}$ and $J_z = J_{zz}$ are called the principal moments of inertia. *The tensor is said to have been diagonalized in this case.* Thus, if the axes of a coordinate system are directed along the principal axes of inertia of the body, the centrifugal moments of inertia will be zero.

?

What are the axial and centrifugal moments of inertia?
Define the principal axes of the inertia tensor. What is the form of the inertia tensor if the axes of a rectangular coordinate system coincide with the principal axes of the inertia tensor?
Are you able to determine the principal axes of the inertia tensor?
What are the central principal axes of the inertia tensor?

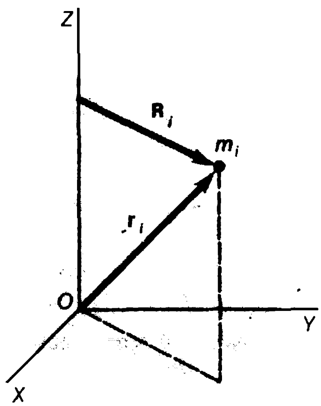


Fig. 75. The geometrical meaning of the quantities appearing in the definition of the axial moment of inertia.

The process of finding the principal axes is reduced to the mathematical procedure of diagonalization of the tensor. We shall not consider this procedure here but only formulate the result: *three mutually perpendicular principal axes can be passed through any point of a rigid body. The principal moments of inertia J_x, J_y and J_z will be different for different points of the body. If the principal axes pass through the centre of mass of the body, they are called the* central principal axes. Thus, there is no sense in speaking of the principal moments of inertia of a body without mentioning the point through which the principal axes are drawn. *In general, as we pass from one point to another, the principal axes change their direction, and the principal moments of inertia acquire different values.* For example, there is no sense in drawing an axis through a body and calling it a principal axis. When we are speaking of the central principal axes or the central principal moments of inertia, there is no need to indicate the point in the rigid body to which these quantities pertain since it is known by definition that this point is the centre of mass of the body.

Of special importance is the axial moment of inertia (Fig. 75) whose value is given by

$$J_{zz} = \sum m_i(r_i^2 - z_i^2) = \sum m_i R_i^2\,, \tag{32.6}$$

where R_i is the distance between point m_i and the axis. In many cases, the axial moment allows us to completely describe the dynamics of a rotating rigid body. It is also called the *moment of inertia of the body about an axis.*

DETERMINATION OF PRINCIPAL AXES. In many cases, the principal axes can be determined without any cumbersome mathematical manipulations like those required for diagonalization of the inertia tensor. For this purpose, it is sometimes sufficient to use the simple concepts of symmetry.

Suppose that we have a plane plate of a vanishingly small thickness. The point through which the principal axes pass lies on the plate. We direct the Z-axis perpendicular to the plate. Obviously, the z-coordinates of all the points on the plate are zero, i.e. $z_i = 0$. In this case, we obtain from (32.3b) $J_{zy} = 0$ and $J_{xz} = 0$. Consequently, any axis perpendicular to the plate will be a principal axis. The other two principal axes lie in the plane of the plate and are perpendicular to each other. Their direction depends on the shape of the plate.

Let us consider the case of a circular plate (Fig. 76) of finite thickness. Point O lying in the midplane of the plate is the point relative to which the principal axes are to be determined. Obviously, one of the principal axes is perpendicular to the plane of the plate. It can be stated that another principal axis lies in the midplane and passes through the given point and the centre of the disc. In Fig. 76, this axis is taken as the Y-axis. Let us verify this. We have

$$J_{yy} = \sum m_i(r_i^2 - y_i^2),$$
$$J_{yz} = -\sum m_i y_i z_i,$$
$$J_{yx} = -\sum m_i y_i x_i.$$

It can be seen that $J_{yx} = 0$ and $J_{yz} = 0$ in view of the symmetry of the plate relative to the planes $x = 0$ and $z = 0$. Thus, the axis chosen in the manner described above is indeed a principal axis. The third principal axis is uniquely defined by the two axes found above since it is perpendicular to both of them. Let us verify that the Z-axis is indeed a principal axis. We have

$$J_{zz} = \sum m_i(r_i^2 - z_i^2),$$
$$J_{zx} = -\sum m_i z_i x_i,$$
$$J_{zy} = -\sum m_i z_i y_i.$$

The equalities $J_{zx} = 0$ and $J_{zy} = 0$ emerge owing to the symmetry of the plate relative to the plane $z = 0$.

If the circular plate has a considerable thickness, it is called a circular cylinder. Naturally, all the arguments concerning the principal axes of a plate are also applicable to the case of a cylinder.

The principal axes relative to any point of a sphere can be found as follows. One of the principal axes passes through the centre of the sphere and the other two are oriented arbitrarily in a plane perpendicular to the first axis. The fact that these axes are the principal axes is proved by simple considerations of symmetry, from which it follows that the centrifugal moments J_{xy}, J_{xz} and others are zero in this case.

The central principal axes passing through the centre of mass are also determined with the help of similar considerations. For a plate of vanishingly small thickness, one of the central principal axes is perpendicular to the plane. The position of the other two in the plane of the plate depends on its shape. For a circular disc, any two mutually perpendicular axes can serve as the central principal axes. For a cylinder, the centre of mass lies in the middle of the height in the centre of the circular cross section. One of the central principal axes coincides with the cylinder's axis, while the other two are arbitrarily oriented in the middle of the circular plane of the cylinder and are perpendicular to each other. In the case of a sphere, any three mutually perpendicular axes passing through the centre of the sphere are its central principal axes.

COMPUTATION OF THE MOMENT OF INERTIA ABOUT AN AXIS. For this purpose, use is made of (32.6). However, it is

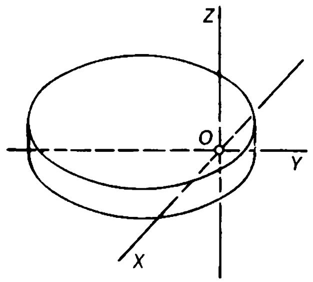


Fig. 76. The principal axes of a circular plate, which pass through a point on the midplane not coinciding with the centre.

?

Can the principal axes of the inertia tensor be determined by means of symmetry considerations? How can this be done?
What is the meaning of Huygens' theorem?
Suppose that we have a system of parallel axes passing through various points lying in a body and outside it. About which of these axes has the axial moment of inertia the minimum value?

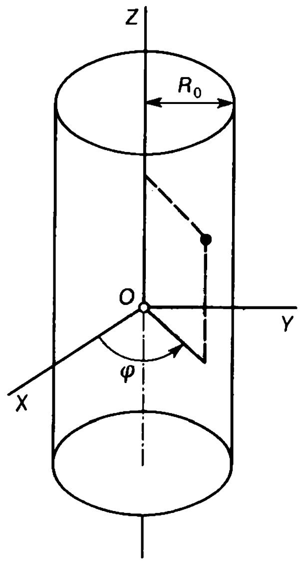


Fig. 77. Choice of the coordinate system for computing one of the principal moments of inertia of a cylinder.

more convenient to carry out integration by going over to the continuous distribution of masses. Suppose that the density of a body is given by $\rho(x, y, z)$. Then a mass $\rho\, \mathrm{d}V$ will be enclosed in a volume element $\mathrm{d}V = \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z$. If we calculate the moment of inertia of the body about the Z-axis, (32.6) assumes the form

$$J_{zz} = \int \rho(x, y, z)(y^2 + x^2)\, \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \tag{32.7}$$

and the integration can be extended to the entire volume of the body.

By way of an example, let us determine the moment of inertia of a uniform cylinder of radius R_0 and height h about an axis coinciding with the cylinder's axis. We direct the Z-axis of the coordinate system along the cylinder's axis and take the origin (point O) on the axis in the middle of the cylinder's height (Fig. 77). The cylinder has a constant density, i.e. $\rho = \rho_0 = \text{const}$. The integral (32.7) can be written in the following form:

$$J_{zz} = \rho_0 \int_{-h/2}^{h/2} \mathrm{d}z \int_S (y^2 + x^2)\, \mathrm{d}x\, \mathrm{d}y, \tag{32.8}$$

where S is the cross-sectional area of the cylinder. It is convenient to make the computations in the cylindrical coordinate system with the symmetry axis coinciding with the Z-axis. This gives

$$x = r\cos\varphi, \quad y = r\sin\varphi, \quad x^2 + y^2 = r^2, \quad \mathrm{d}x\, \mathrm{d}y = r\, \mathrm{d}r\, \mathrm{d}\varphi.$$

Hence, instead of (32.8), we can write

$$J_{zz} = \rho_0 \int_{-h/2}^{h/2} \mathrm{d}z \int_0^{R_0} r^3\, \mathrm{d}r \int_0^{2\pi} \mathrm{d}\varphi = \rho_0 h \frac{R_0^4}{4} 2\pi. \tag{32.9}$$

Considering that the volume of the cylinder is $\pi R_0^2 h$, and hence $m = \pi R_0^2 h \rho_0$ is the mass of the cylinder, we finally obtain

$$J_{zz} = \frac{1}{2} m R_0^2. \tag{32.10}$$

Other moments of inertia are determined in a similar manner. We leave it to the reader to calculate them. In particular, the moment of inertia of a homogeneous sphere about an axis passing through its centre is $(2/5)mR_0^2$, where m is the mass of the sphere, and R_0 is its radius. The moment of inertia of a thin disc about an axis passing through its centre and perpendicular to the plane of the disc is given by (32.10), while its moment of inertia about an axis passing through the centre of the disc and lying in its plane is $(1/4)mR_0^2$.

HUYGENS' THEOREM. In many cases, the computation of

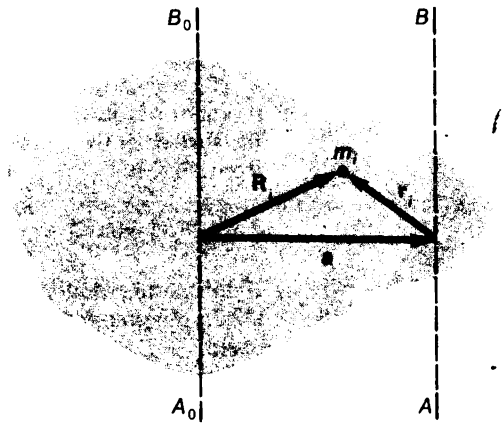


Fig. 78. The geometrical meaning of the vectors used to prove Huygens' theorem.

the moment of inertia about an axis is simplified by means of Huygens' theorem *connecting the moments of inertia about two parallel axes one of which passes through the centre of mass of a body* (Fig. 78). Let the axis A_0B_0 be the axis passing through the centre of mass. We denote the radius vector of a point mass m_i measured from this axis in a plane perpendicular to the axis by $\mathbf{R}_i$, and the radius vector of the same point mass measured from the axis AB which is parallel to the axis A_0B_0 but does not pass through the centre of mass by $\mathbf{r}_i$. Let us draw vector $\mathbf{a}$ from the axis A_0B_0 to the axis AB in this plane. Let J_0 be the moment of inertia about the axis passing through the centre of mass, and J be the moment of inertia about the axis AB which does not pass through the centre of mass. By the definition of the moments of inertia, we have

$$J_0 = \sum m_i R_i^2, \quad J = \sum m_i r_i^2. \tag{32.11}$$

It can directly be seen from Fig. 78 that $\mathbf{r}_i = -\mathbf{a} + \mathbf{R}_i$, and hence $r_i^2 = R_i^2 + a^2 - 2\mathbf{a}\cdot\mathbf{R}_i$. This gives

$$J = \sum m_i r_i^2 = \sum m_i R_i^2 + a^2 \sum m_i - 2\mathbf{a}\cdot\sum m_i \mathbf{R}_i. \tag{32.12}$$

Considering that $\sum m_i \mathbf{R}_i = 0$ by the definition of the axis passing through the centre of mass, and $\sum m_i = m$ is the mass of the body, we can rewrite (32.12) in the form

$$\boxed{J = J_0 + ma^2.} \tag{32.13}$$

This formula expresses Huygens' theorem. Knowing the moment of inertia of a body about an axis passing through its centre of mass, we can easily compute the moment of inertia about any other axis parallel to it.

For example, let us consider a cylinder whose moment of inertia about its axis is given by (32.10). The centre of mass of the cylinder lies on its axis, and hence (32.10) is the moment of inertia about the axis passing through its centre of mass. The moment of inertia of the cylinder about the axis AB lying on the surface of the cylinder parallel to its axis can be determined from (32.13):

$$J = \frac{1}{2} m R_0^2 + m R_0^2 = \frac{3}{2} m R_0^2. \tag{32.14}$$

If this moment of inertia were determined with the help of (32.7), the computations would have been a lot more complicated.

The moment of inertia of a sphere about the axis AB tangential to its surface is also easily found with the help of (32.13):

$$J = \frac{2}{5} m R_0^2 + m R_0^2 = \frac{7}{5} m R_0^2, \tag{32.15}$$

where we have considered that the moment of inertia of the sphere about the axis passing through its centre of mass is $(2/5) m R_0^2$.

Sec. 33. KINETIC ENERGY OF A ROTATING RIGID BODY

The formula for the kinetic energy of a rigid body rotating arbitrarily is derived.

EXPRESSING THE INERTIA TENSOR WITH THE HELP OF THE KRONECKER DELTA. It is convenient to perform calculations by introducing the Kronecker delta, which is defined as follows:

$$\delta_{\alpha\beta} = \begin{cases} 1 \text{ for } \alpha = \beta, \\ 0 \text{ for } \alpha \neq \beta. \end{cases} \tag{33.1}$$

The components J_{xx}, J_{xy}, etc. of the inertia tensor can be denoted by J_{11}, J_{12}, etc., i.e. in the form $J_{\alpha\beta}$. Formulas (32.3b) then acquire the form

$$J_{11} = \sum_{i,\gamma} m_i \left(x_{i\gamma} x_{i\gamma} - x_{i1}^2 \right),$$
$$J_{12} = -\sum_i m_i x_{i1} x_{i2}, \tag{33.2}$$
$$J_{13} = -\sum_i m_i x_{i1} x_{i3},$$

where $r_i^2 = x_{i1} x_{i1} + x_{i2} x_{i2} + x_{i3} x_{i3} = \sum_\gamma x_{i\gamma} x_{i\gamma}$, and γ is the dummy index. The other components are expressed in a similar manner.

Any projection A_α of the vector can be expressed in terms of other projections with the help of the symbol $\delta_{\alpha\gamma}$:

$$A_\alpha = \sum_{\gamma=1}^{3} \delta_{\alpha\gamma} A_\gamma. \tag{33.3}$$

This equality can be written in expanded form as follows:

$$A_\alpha = \delta_{\alpha 1} A_1 + \delta_{\alpha 2} A_2 + \delta_{\alpha 3} A_3. \tag{33.4}$$

From the symbols $\delta_{\alpha 1}$, $\delta_{\alpha 2}$ and $\delta_{\alpha 3}$, only the one in which α is equal to the other index will be different from zero. Suppose, for example, that $\alpha = 2$. In this case, (33.4) gives

$$A_2 = 0 \cdot A_1 + 1 \cdot A_2 + 0 \cdot A_3 = A_2.$$

Using the Kronecker delta, we can represent the expression (33.2) for the inertia tensor in the following form which is more convenient for calculations:

$$J_{\alpha\beta} = \sum_{i,\gamma} m_i (x_{i\gamma} x_{i\gamma} \delta_{\alpha\beta} - x_{i\alpha} x_{i\beta}). \tag{33.5}$$

KINETIC ENERGY OF ROTATION. If the transport velocity v_0 of a rigid body is zero, i.e. if the body rotates at an instantaneous velocity ω whose vector passes through a fixed point of the body, the velocity of its points will be

$$\mathbf{v}_i = \boldsymbol{\omega} \times \mathbf{r}'_i, \tag{33.6}$$

and hence its kinetic energy will be

$$E_k = \frac{1}{2} \sum m_i v_i^2 = \frac{1}{2} \sum m_i (\boldsymbol{\omega} \times \mathbf{r}'_i)^2. \tag{33.7}$$

Using the well-known formula from vector algebra, i.e.

$$(\mathbf{A} \times \mathbf{B}) \cdot (\mathbf{C} \times \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C})(\mathbf{B} \cdot \mathbf{D}) - (\mathbf{A} \cdot \mathbf{D})(\mathbf{B} \cdot \mathbf{C}),$$

we can write

$$(\boldsymbol{\omega} \times \mathbf{r}'_i)^2 = \omega^2 r'^2_i - (\mathbf{r}'_i \cdot \boldsymbol{\omega})^2. \tag{33.8}$$

This expression assumes the following form in terms of coordinates:

$$(\boldsymbol{\omega} \times \mathbf{r}'_i)^2 = \sum_{\alpha,\beta} \omega_\alpha \omega_\alpha x'_{i\beta} x'_{i\beta} - \sum_{\alpha,\beta} \omega_\alpha x'_{i\alpha} \omega_\beta x'_{i\beta}$$

$$J_{\alpha\beta} = \sum_{i,\gamma} m_i (x_{i\gamma} x_{i\gamma} \delta_{\alpha\beta} - x_{i\alpha} x_{i\beta}). \tag{33.5}$$

Substituting this formula into (33.7) and taking (33.5) into

!

The inertia tensor, in terms of which the kinetic energy of a rotating body is expressed, refers to the coordinate axes rigidly fixed to the body.
The kinetic energy of a rolling cylinder is the sum of the kinetic energies of the translation of the centre of mass and of rotation. Hence, while rolling down an inclined plane, the velocity of the centre of mass of the cylinder is lower than that of a cylinder slipping without rotation.

account, we obtain the kinetic energy of rotation:

$$E_k = \frac{1}{2}\sum_{\alpha,\beta} J_{\alpha\beta}\omega_\alpha\omega_\beta. \tag{33.10}$$

Here, $J_{\alpha\beta}$ is the inertia tensor referred to the coordinate axes rigidly fixed to the body and moving with it. The origin of the coordinate system is at rest, and ω_α are the projections of the instantaneous angular velocity of the body on the coordinate axes.

If the axes of a moving coordinate system are directed along the principal axes of inertia of a body, only the diagonal components are left in the inertia tensor, i.e.

$$J_{\alpha\beta} = J_\alpha\delta_{\alpha\beta}. \tag{33.11}$$

For such a choice of the coordinate axes which are rigidly fixed to the body, the expression (33.10) for the kinetic energy is simplified:

$$E_k = \frac{1}{2}(J_1\omega_1^2 + J_2\omega_2^2 + J_3\omega_3^2) = \frac{1}{2}\sum_\alpha J_\alpha\omega_\alpha^2. \tag{33.12}$$

If the instantaneous velocity of rotation coincides in direction with one of the principal axes, say, the X-axis of the moving system, then obviously $\omega_2 = \omega_3 = 0$, and (33.12) assumes a simpler form:

$$E_k = \frac{1}{2}J_1\omega_1^2. \tag{33.13}$$

Generally speaking, the angular velocity vector ω changes its direction during an arbitrary motion of a body and coincides in direction with one of the principal axes only over an instant of time. This formula is valid just for this particular instant of time. In the next instant, the angular velocity will no longer coincide with the principal axis, and (33.12) will again become an expression for the kinetic energy.

If, in addition to a rotation, a body also undergoes a translation at a velocity v_0, the velocity of its points is determined by (31.3). In this case, the expression for the kinetic energy becomes complicated. Substituting (31.3) into the expression for the kinetic energy, we obtain

$$E_k = \frac{1}{2}\sum m_i v_i^2 = \frac{1}{2}\sum m_i(\mathrm{v}_0 + \omega \times \mathrm{r}_i)^2$$
$$= \frac{1}{2}\left(\sum m_i\right)v_0^2 + \frac{1}{2}\sum m_i(\omega \times \mathrm{r}_i)^2 + \frac{1}{2}\sum m_i\cdot 2\mathrm{v}_0\cdot(\omega \times \mathrm{r}_i). \tag{33.14}$$

The first term in this formula expresses the kinetic energy of the translation of the body as a whole at a velocity v_0, the second term expresses the kinetic energy of rotation considered above, while the third takes into account the relation between the translational and rotational velocities. If the origin of a moving coordinate system is taken at the centre of mass of the body, then $\sum m_i r_i' = 0$, and hence the last term vanishes. For such a choice of the coordinate system, v_0 is the velocity of the centre of the body, and the formula for the kinetic energy becomes

$$E_k = \frac{1}{2} m v_0^2 + \sum_{\alpha, \beta} \frac{1}{2} J_{\alpha\beta} \omega_\alpha \omega_\beta. \tag{33.15}$$

All the remarks made about (33.10) are also valid for the corresponding quantity in (33.15). In particular, if the axes of a moving system are directed along the principal axes of the body, the kinetic energy is

$$E_k = \frac{1}{2} m v_0^2 + \frac{1}{2}(J_1 \omega_1^2 + J_2 \omega_2^2 + J_3 \omega_3^2). \tag{33.16}$$

Hence the kinetic energy, say of a cylinder rolling at a velocity $v_0 = \omega R_0$, will be

$$E_k = \frac{1}{2} m v_0^2 + \frac{1}{2} \frac{m R_0^2}{2} \left(\frac{v_0}{R_0}\right)^2 = \frac{3}{4} m v_0^2. \tag{33.17}$$

Sec. 34. PLANE MOTION. PENDULUMS

The most important cases of plane motion are analyzed.

PECULIARITIES OF THE DYNAMICS OF PLANE MOTION. It is known from the kinematics of plane motion described in Sec. 9 that in this case all points of a rigid body move in parallel planes. Hence it is sufficient to consider the motion of any cross section of the body in a plane. Formula (31.3) for the points of the body is considerably simplified since the angular velocity vector is always perpendicular to the plane, and hence always has the same direction. Therefore, if the Z-axis of the coordinate system fixed to the body is made perpendicular to the plane of motion, the angular velocity of rotation will always be directed along this axis, i.e. $\omega_z = \omega$ and $\omega_x = \omega_y = 0$. *In order to avoid the centrifugal moments of the inertia tensor, it is expedient to pass the axis of rotation through the centre of mass.* In this case, we just have to take into consideration the angular momentum

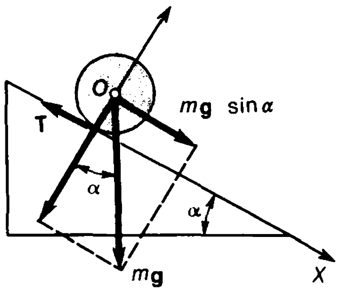


Fig. 79. Rolling of a cylinder down an inclined plane (without sliding).

about the axis of rotation:

$$L_z = L = J_{zz}\omega_z = J\omega,\ J_{zz} = J,\ \omega_z = \omega. \tag{34.1}$$

The subscript z on these quantities can be omitted since the Z-axis is the only axis of rotation. The forces acting on the body are parallel to the XY-plane, and the moments M_z of the forces are perpendicular to it. Thus, the equations of motion (31.1) and (31.2) for the plane motion acquire the form

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{F}, \tag{34.2}$$

$$J\frac{\mathrm{d}\omega}{\mathrm{d}t} = M, \tag{34.3}$$

where $M = M_z$, and $\mathbf{p}$ is the momentum.

Since the axis passes through the centre of mass of the body, (34.2) can be represented in the form (31.4) for the motion of the centre of mass in the plane of motion:

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \mathbf{F}. \tag{34.4}$$

For the x- and y-coordinates of the centre of mass, this equation assumes the form

$$m\ddot{x} = F_x, \quad m\ddot{y} = F_y. \tag{34.5}$$

In this case, the kinetic energy is expressed by (33.16):

$$E_k = \frac{1}{2}mv_0^2 + \frac{1}{2}J\omega^2. \tag{34.6}$$

ROLLING OF A CYLINDER DOWN AN INCLINED PLANE. We shall assume that the cylinder rolls without sliding. The forces acting on the cylinder are shown in Fig. 79. The force T is the frictional force which ensures the rolling of the cylinder without sliding. It is convenient to orient the X-axis along the inclined plane. Let us write the equation of motion for point O through which the central principal axis of inertia of the disc passes. Equations (34.4) and (34.3) have the form

$$m\frac{\mathrm{d}v_0}{\mathrm{d}t} = mg\sin\alpha - T, \quad J_0\frac{\mathrm{d}\omega}{\mathrm{d}t} = R_0T, \tag{34.7}$$

where $J_0 = (1/2)mR_0^2$, and the directions of rotation are chosen in such a way that ω is positive and increases as the cylinder rolls down.

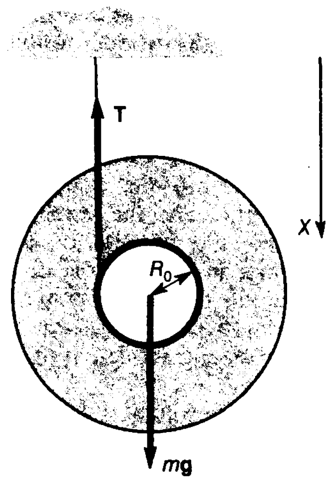


Fig. 80. Maxwell's pendulum.

Substituting T from the second equation of (34.7) into the first one and considering that $v_0 = \omega R_0$ (R_0 is the radius of the cylinder), we obtain

$$m\frac{dv_0}{dt} = mg\sin\alpha - \frac{J_0}{R_0}\frac{dv_0}{dt}, \tag{34.8}$$

or

$$\frac{3}{2}m\frac{dv_0}{dt} = mg\sin\alpha, \tag{34.9}$$

$$\frac{dv_0}{dt} = \frac{2}{3}g\sin\alpha. \tag{34.10}$$

Thus, the centre of the cydinder moves with a constant acceleration $(2/3)g\sin\alpha$.

MAXWELL'S PENDULUM. Maxwell's pendulum is an annular disc suspended on a string. The string is wound on the axle of the disc (Fig. 80). The equations of motion of the pendulum about the centre of mass have the form

$$m\frac{dv_0}{dt} = mg - T, \quad J_0\frac{d\omega}{dt} = R_0 T, \tag{34.11}$$

where T is the tension, J_0 is the moment of inertia of the entire system about the axle, and R_0 is the radius of the axle of the disc on which the string is wound.

So far as the forces and their moments are concerned, Maxwell's pendulum is analogous to a cylinder rolling down an inclined plane.

Thus, the equations of motion for Maxwell's pendulum have the same form as those for a cylinder rolling down an inclined plane and are solved in the same manner. We obtain

$$\frac{dv_0}{dt} = \frac{mg}{m + (J_0/R_0^2)}, \quad T = \frac{mg}{1 + (mR_0^2/J_0)}. \tag{34.12}$$

Let us analyze the dynamics of the pendulum. The acceleration of the disc is constant and always directed downwards. Its numerical value is the smaller, the larger the central moment of inertia J_0. For a sufficiently large moment of inertia J_0, the disc will have a very small acceleration. In the limit $J_0 \to \infty$, the acceleration of the disc $dv_0/dt \to 0$, and the tension $T \to mg$. Indeed, this must be so since the disc is simply hanging in the string without motion. As $J_0 \to 0$, $T \to 0$. In this case, the disc falls freely, and hence the string does not experience any tension.

Equations (34.11) and the solutions (34.12) do not describe the behaviour of the pendulum at the bottom dead centre where the string is thrown from one side of the cylinder to the other. The disc continues to rotate in the same direction, but the string is now wound on the cylinder instead of being unwound. Equations (34.11) and the solutions (34.12) are also valid for the winding of the string. During this process, the disc rises, its kinetic energy is transformed into the potential energy, and its velocity of ascent decreases. As the string is thrown over at the bottom dead centre, the direction of the velocity v_0 is reversed. Hence the centre of mass of the disc experiences a large acceleration during this time. According to Newton's third law of motion, this results in a considerable tension of the string. If the string is not strong enough, it may be snapped.

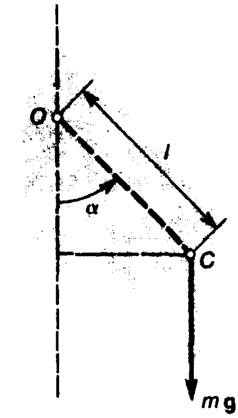


Fig. 81. The physical pendulum.

PHYSICAL PENDULUM. A physical pendulum is a rigid body suspended in the gravitational field from a horizontal axle (Fig. 81). The momental equation for a physical pendulum has the form

$$J\frac{d\omega}{dt} = -mgl\sin\alpha, \quad \omega = \frac{d\alpha}{dt}, \tag{34.13}$$

where the minus sign means that the moment of forces is directed opposite to the increasing angle α, and J is the moment of inertia about the axis passing through the point of suspension.

If the angle of deflection is small, it can be assumed to a high degree of accuracy that $\sin\alpha = \alpha$. We can then write (34.13) in the form

$$\frac{d^2\alpha}{dt^2} + \frac{mgl}{J}\alpha = 0. \tag{34.14}$$

The solutions of this equation are the functions $\sin(mgl/J)^{1/2}t$ or $\cos(mgl/J)^{1/2}t$. The pendulum performs small oscillations whose frequency and period are given by

$$\omega = \sqrt{\frac{mgl}{J}}, \quad T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{J}{mgl}}. \tag{34.15}$$

Such oscillations are called harmonic. Their properties will be described in Chap. 13. Here, we shall confine ourselves to just a few remarks.

Suppose that a physical pendulum consists of a point mass m suspended from a weightless rigid rod of length l and oscillating about point O. Such a pendulum is called a mathematical pendulum. Noting that $J = ml^2$ for this pendulum as for a rigid body, we obtain the period of oscillations of a mathematical pendulum from (34.15):

$$T = 2\pi\sqrt{\frac{ml^2}{mgl}} = 2\pi\sqrt{\frac{l}{g}}. \tag{34.16}$$

Let us denote the moment of inertia of a physical pendulum about an axis passing through its centre of mass by J_0. According to Huygens' theorem, $J = J_0 + ml^2$, and formula (34.15) for the period of oscillations of a physical pendulum becomes:

$$T = 2\pi\sqrt{\frac{J_0 + ml^2}{mgl}} = 2\pi\sqrt{\frac{J_0}{mgl} + \frac{l}{g}}. \tag{34.17}$$

A comparison of (34.16) and (34.17) shows that a mathematical pendulum whose length is equal to the distance between the point of suspension and the centre of mass of a physical pendulum has a smaller period than the physical pendulum. In order to make the period of oscillations of a mathematical pendulum equal to that of a physical pendulum, we must increase its length. *The length of a mathematical pendulum whose period of oscillations is equal to that of a physical pendulum is called the reduced length of the corresponding physical pendulum.* A comparison of (34.16) and (34.17) shows that the reduced length of a physical pendulum is $l_{red} = J/(ml)$. The point on a physical pendulum at a distance l_{red} from the point of suspension along a straight line passing through the centre of gravity is called the centre of oscillations. If a physical pendulum and a mathematical pendulum with reduced length oscillate about the same axis, the centre of oscillations of the physical pendulum and a point mass on the mathematical pendulum oscillate synchronously if in the beginning they are deflected to the same extent and released simultaneously.

The basic property of the centre of oscillations of a physical pendulum is that the period of oscillations does not change if the pendulum is suspended at an axis passing through this centre. Thus, when the point of suspension is transferred to the centre of oscillations, the previous point of suspension becomes the new centre of oscillations, i.e. the point of suspension and the centre of oscillations are interchangeable. The proof follows directly from Huygens' theorem and the formula for the period of oscillations of a pendulum.

If the amplitudes of oscillations of a physical pendulum are not very small, we cannot go over from (34.13) to (34.14). In this case, we must solve the nonlinear equation (34.13):

$$\frac{d\omega}{dt} = \ddot{\alpha} = -k\sin\alpha, \quad k = \frac{mgl}{J}. \tag{34.18}$$

!

In Maxwell's pendulum, the acceleration of a rotating disc is constant and always directed downwards. When the direction of velocity is reversed at the lower dead centre, a considerable increase in tension occurs.

?

Why is it expedient for plane motion to express the equation of motion and the momental equation relative to a point through which the central principal axis passes perpendicular to the plane of motion?

While integrating, it is convenient to measure the distance from the position of maximum deflection α_0, when the velocity of the pendulum is zero ($\dot{\alpha}_0 = 0$). We have

$$\int_{\alpha_0}^{\alpha} \ddot{\alpha}\,\mathrm{d}\alpha = -k\int_{\alpha_0}^{\alpha} \sin\alpha\,\mathrm{d}\alpha. \tag{34.19}$$

The integrands can be transformed as follows:

$$\ddot{\alpha}\,\mathrm{d}\alpha = \ddot{\alpha}\dot{\alpha}\,\mathrm{d}t = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{\alpha}^2}{2}\right)\mathrm{d}t = \mathrm{d}\left(\frac{\dot{\alpha}^2}{2}\right),$$

$$\sin\alpha\,\mathrm{d}\alpha = -\mathrm{d}\cos\alpha.$$

From (34.19) we obtain

$$\dot{\alpha}^2 = 2k(\cos\alpha - \cos\alpha_0). \tag{34.20}$$

This equality expresses the energy conservation law for a pendulum.

Writing (34.20) in the form

$$\frac{\mathrm{d}\alpha}{\sqrt{\cos\alpha - \cos\alpha_0}} = \sqrt{2k}\,\mathrm{d}t, \tag{34.21}$$

we can find the solution of the problem in implicit form by integration:

$$\int_0^{\alpha} \frac{\mathrm{d}\alpha}{\sqrt{\cos\alpha - \cos\alpha_0}} = \sqrt{2k}\,t.$$

Using the formula $\cos\alpha = 1 - 2\sin^2(\alpha/2)$, we obtain

$$\int_0^{\alpha} \frac{\mathrm{d}\alpha}{\sqrt{\sin^2(\alpha_0/2) - \sin^2(\alpha/2)}} = 2\sqrt{k}\,t. \tag{34.22}$$

We introduce a new integration variable θ with the help of the relation

$$\sin\theta = \frac{\sin(\alpha/2)}{\sin(\alpha_0/2)}. \tag{34.23}$$

In this case, (34.22) assumes the form

$$\int_0^{\beta} \frac{\mathrm{d}\theta}{\sqrt{1 - \sin^2(\alpha_0/2)\sin^2\theta}} = \sqrt{k}\,t. \tag{34.24}$$

The integral on the left-hand side is called an elliptic integral. This integral has thoroughly been studied, and tables of values of this integral can be used to analyze the oscillations of a pendulum with any angle of deflection. However, for angles which are not too large, it is expedient to present this integral by an approximate formula which is convenient for analysis. If $\sin^4(\alpha_0/2) \ll 1$, the integrand in (34.24) can be expanded into a series. Confining ourselves to the first two terms in the expansion, we can write

$$\int_0^\beta \frac{d\theta}{\sqrt{1 - \sin^2(\alpha_0/2)\sin^2\theta}}$$
$$= \int_0^\beta d\theta\left(1 + \frac{1}{2}\sin^2\frac{\alpha_0}{2}\sin^2\theta + \ldots\right)$$
$$= \beta + \frac{1}{4}\sin^2\frac{\alpha_0}{2}\left(\beta - \frac{\sin 2\beta}{2}\right) + \ldots. \tag{34.25}$$

Thus, the relation between the period of oscillations and the angle of deflection of the pendulum can be represented in the form

$$\beta + \frac{1}{4}\sin^2\frac{\alpha_0}{2}\left(\beta - \frac{\sin 2\beta}{2}\right) = \sqrt{k}\, t, \tag{34.26}$$

where $\sin\beta$ is defined by (34.23):

$$\sin\beta = \frac{\sin(\alpha/2)}{\sin(\alpha_0/2)}.$$

It can be seen from here that as the angle of deflection α varies from 0 to α_0, i.e. covers a quarter of the period of oscillations T, the quantity β varies from 0 to $\pi/2$. From (34.26), we obtain

$$\frac{\pi}{2} + \frac{1}{4}\sin^2\frac{\alpha_0}{2}\left(\frac{\pi}{2} - \frac{\sin 2\pi/2}{2}\right) = \sqrt{k}\,\frac{T}{4},$$

whence

$$T = \frac{2\pi}{\sqrt{k}}\left(1 + \frac{1}{4}\sin^2\frac{\alpha_0}{2}\right). \tag{34.27}$$

Comparing this equation with (34.15) for the period of linear oscillations and taking into account the expression for k in (34.18), we can write it in the form

$$T = T_0\left(1 + \frac{1}{4}\sin^2\frac{\alpha_0}{2}\right), \tag{34.28}$$

where $T_0 = 2\pi\sqrt{J/(mgl)}$ is the period of linear oscillations.

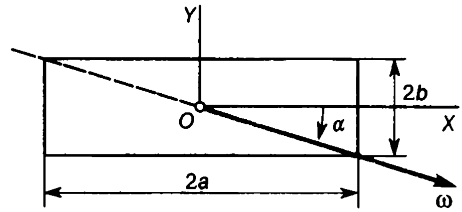


Fig. 82. The diagonal of a thin homogeneous plate is not the axis of its free rotation. Hence to keep the direction of the axis of rotation coinciding with the diagonal unchanged, a moment of force must be applied to the axis. This moment of force is produced by the constraints.

Suppose, for example, that the maximum deflection $\alpha_0 = 60°$. Since $\sin 30° = 1/2$, the period of nonlinear oscillations of the pendulum in this case will differ from the period of linear oscillations by about 6%. It can be concluded from here that the linear approximation gives a fairly accurate description of the motion of a physical pendulum not only for very small angles of deflection, but for quite large angles as well.

Let us now consider a ballistic pendulum, taking into account the finite size of the body into which the bullet is stuck (see Sec. 23, Fig. 52). When the bullet hits the pendulum, we should take into account not the momentum conservation law as in Sec. 23, but the angular momentum conservation law about the point of suspension of the pendulum, which can be written in the form $m_2 l v = J\omega$, where l is the distance from the point of suspension to the line of flight of the bullet passing through the centre of mass of the pendulum, J is the moment of inertia of the pendulum with a bullet stuck in it about the axis of oscillation of the pendulum, and ω is the angular velocity of motion of the centre of mass of the pendulum immediately after having been hit by the bullet. As the centre of mass is raised to a height h in the extreme position of the pendulum, the energy conservation law has the form $(m_1 + m_2)gh = J\omega^2/2$ since the kinetic energy immediately after the impact of the bullet cannot be reduced to the kinetic energy of the centre of mass of the pendulum since it also includes the kinetic energy of its rotation. Hence the velocity v of the bullet is connected with the height to which the centre of mass is raised through the relation $v = \sqrt{2gh}[J(m_1 + m_2)]^{1/2}/(m_2 l)$. Assuming the entire mass of the body to be concentrated at a point, we find that $J = (m_1 + m_2)l^2$, and hence $v = \sqrt{2gh}(m_1 + m_2)/m_2$, as was shown in Sec. 23.

Example 34.1. A very thin homogeneous rectangular plate with sides $2a$ and $2b$ rotates at a constant angular velocity ω about its diagonal (Fig. 82). The mass of the plate is m. Find the moment of the forces which must be applied to the axis of rotation to ensure that its direction remains unchanged in space.

The centre of mass of the plate lies at the intersection of the diagonals, two of its central principal axes are parallel to the sides of the plate, while the third is perpendicular to the plane of the plate. We introduce a coordinate system rigidly fixed to the plate, take the origin at the centre of mass and direct the X- and Y-axes parallel to the sides of length $2a$ and $2b$. The Z-axis is directed at right angles to the plane of the plate. The principal moments of inertia about these axes are

$$J_{xx} = \frac{1}{3}mb^2, \quad J_{yy} = \frac{1}{3}ma^2, \quad J_{zz} = \frac{1}{3}m(a^2 + b^2). \tag{34.29}$$

The angular velocity of rotation can be represented in the form

$$\omega = \omega(\mathrm{i}_x \cos\alpha - \mathrm{i}_y \sin\alpha), \tag{34.30}$$

where $\alpha = \arctan(b/a)$, and the angular momentum

$$\begin{aligned} \mathrm{L} &= J_{xx}\omega_x \mathrm{i}_x + J_{yy}\omega_y \mathrm{i}_y + J_{zz}\omega_z \mathrm{i}_z \\ &= \frac{1}{3}m\omega(\mathrm{i}_x b^2 \cos\alpha - \mathrm{i}_y a^2 \sin\alpha) \\ &= \frac{m\omega ab(b\mathrm{i}_x - a\mathrm{i}_y)}{3\sqrt{a^2 + b^2}}. \end{aligned} \tag{34.31}$$

The moment of the forces acting on the system is

$$\begin{aligned} \frac{d\mathrm{L}}{dt} &= \frac{\partial \mathrm{L}}{\partial t} + \omega \times \mathrm{L} = -\frac{1}{3}m\omega^2(a^2 - b^2)\sin\alpha\cos\alpha\,\mathrm{i}_z \\ &= -\frac{m\omega^2 ab(a^2 - b^2)}{3(a^2 + b^2)}\mathrm{i}_z. \end{aligned} \tag{34.32}$$

This moment of forces is exerted on the axis of rotation by the constraints which keep the direction of the axis of rotation unchanged in space.

Example 34.2. A homogeneous rod AB of mass m and length $2l$ is suspended from point O by two unstretchable weightless strings of the same length (Fig. 83). The strings form an angle α with the rod. At a certain instant of time, the string OB is snapped. Find the tension T of the string OA immediately after the snapping of the string.

At this moment of time, the acceleration of the centre of mass of the rod, viz. point C, is the sum of the acceleration of point A and the acceleration of point C relative to A. Since point A can move only in a circle, and its velocity at the moment of snapping of the string is zero, the radial acceleration is zero. The projection of the acceleration of point A along the tangent is denoted by a, and the magnitude of the angular acceleration of the rod is $\dot{\omega}$ at the same instant. Since

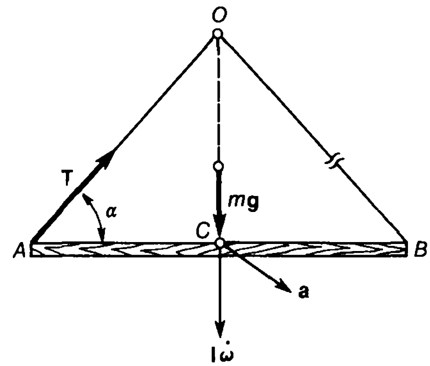


Fig. 83. At the instant of snapping of the string OB, the acceleration of the centre of mass of the rod, i.e. point C, is the sum of the acceleration of point A and the acceleration of point C relative to point A.

the angular velocity is zero at the initial instant of time, the component a of the acceleration of the centre of mass is perpendicular to OA, while the component $l\dot{\omega}$ is perpendicular to the rod. Consequently, the equations of motion have the form

$$mg\cos\alpha = m(a + l\dot{\omega}\cos\alpha),$$

$$mg\sin\alpha - T = ml\dot{\omega}\sin\alpha, \tag{34.33}$$

$$lT\sin\alpha = \frac{1}{3}ml^2\dot{\omega},$$

where we have considered that the moment of inertia of the rod about the axis passing through its centre of mass and perpendicular to it is $ml^2/3$. Hence we obtain

$$T = \frac{mg\sin\alpha}{1 + 3\sin^2\alpha}. \tag{34.34}$$

Example 34.3. Let us consider the problem concerning the sliding of a ladder, which was formulated in Example 9.1, to find that the result obtained there was meaningless from the physical point of view. For this purpose, we must consider the dynamics of the motion of the ladder, taking into account the force of gravity acting on it.

Let us denote the angle between the X-axis along the positive direction of which the lower end of the ladder slides and the ladder by θ $(\theta \geqslant \pi/2)$. The remaining notation is the same as in Example 9.1.

The ladder is subjected to the force of gravity applied to the centre of mass, the reaction of the walls and floor applied to the ends of the ladder and directed along the normals to the corresponding surfaces, and the force applied to the lower end of the ladder and ensuring its uniform motion. This force acts along the X-axis towards the origin and is equal in magnitude to the reaction of the vertical wall acting in the positive

X-direction. This follows from the fact that the lower end of the ladder and the centre of mass move at a constant velocity in the positive X-direction, and hence the X-projection of the total force acting on the ladder is zero. The momental equation about the lower end of the ladder taken as the origin of the inertial coordinate system has the form

$$\frac{J\mathrm{d}^2\theta}{\mathrm{d}t^2} = -\frac{mgl\cos\theta}{2} - Fl\sin\theta, \tag{34.35}$$

where $J = ml^2/3$ is the moment of inertia of the ladder about the lower end, m is the mass of the ladder, and F is the reaction of the vertical wall. According to the conditions of the problem, $u = -\mathrm{d}(l\cos\theta)/\mathrm{d}t$ and $\mathrm{d}u/\mathrm{d}t = 0$. Hence $u = l\sin\theta\,\mathrm{d}\theta/\mathrm{d}t$ and $\mathrm{d}^2\theta/\mathrm{d}t^2 = -\cot\theta u^2/(l^2\sin^2\theta)$. Hence we obtain from (34.35)

$$F = -\frac{mg\cot\theta}{2}\frac{1 - 2u^2}{3gl\sin^3\theta}. \tag{34.36}$$

The upper end of the ladder remains in contact with the wall when $F > 0$, i.e. $2u^2/(3gl\sin^3\theta) < 1$. At the initial instant $t = 0$, the angle $\theta = \pi/2$. Consequently, the problem will have a physical meaning only if the following condition is satisfied:

$$u < u_{\min} = \sqrt{\frac{3gl}{2}}. \tag{34.37}$$

If the velocity u is equal to or larger than $u_{\min}$, the upper end of the ladder loses contact with the vertical wall at the very beginning, and the problem about its sliding down the wall is devoid of any physical meaning. If, however, $u < u_{\min}$, the contact of the ladder with the vertical wall continues after the onset of motion until the critical angle $\theta_{cr} = \arcsin\{[2u^2/(3gl)]^{1/3}\}$, $\pi/2 < \theta < \pi$, is reached. In the interval of angles from $\theta = \pi/2$ to $\theta = \theta_{cr}$, the kinematic description of the sliding of the ladder considered in Example 9.1 is valid. It is not difficult to recalculate this interval of angles into the interval of time from the instant the ladder begins to slide. The centre of mass of the ladder describes in this time interval an arc of a circle, and the velocities of various points of the ladder can be calculated in the same way as in Example 9.1 and Problems 2.6 and 2.11. Starting from the critical angle, the motion is described by Eq. (34.35) with $F = 0$ under the initial conditions corresponding to the critical angle.

Sec. 35. MOTION OF A RIGID BODY FIXED AT A POINT. GYROSCOPES

The most important properties of the motion of a rigid body fixed at a point are described.

Leonhard Euler (1707-1783)
Mathematician, mechanical engineer, physicist and astronomer. Of Swiss origin, he lived in Russia from 1727 to 1741, and from 1766 until his death. He was a member of the Petersburg Academy of Sciences. He worked in Berlin from 1742 to 1766 and was also a member of the Berlin Academy of Sciences. Euler wrote over 800 papers on mathematics, physics, theory of music, etc., which considerably influenced the development of science.

CHOICE OF THE COORDINATE SYSTEM. The analysis of the plane motion is simplified by the fact that the angular velocity vector preserves in this case a constant direction in space, which is perpendicular to the plane of motion and does not change its orientation relative to the body. When a rigid body moves about a fixed point, all these simplifying factors disappear, and the angular velocity vector generally changes its direction in space as well as its orientation relative to the body, i.e. the instantaneous axis of rotation changes its orientation. It is convenient to consider this motion in a coordinate system rigidly fixed to the body. The origin of coordinates should naturally be located at the point where the body is fixed and is always at rest. The equations of motion obtained under these conditions are called Euler's equations.

EULER'S EQUATIONS. The equation of motion of the centre of mass of a body has the form

$$m\frac{d\mathbf{v}_0}{dt} = m\frac{d}{dt}(\omega \times \mathbf{r}_0) = \mathbf{F}, \tag{35.1}$$

where $\mathbf{r}_0$ is the radius vector of the centre of mass of the body, passing through the point where it is fixed. The constraints are included in F.

The axes of the coordinate system $(\mathbf{i}'_x, \mathbf{i}'_y, \mathbf{i}'_z)$ fixed to the body can conveniently be directed along the principal axes of inertia. In this case, the inertia tensor is reduced to its three principal values J_1, J_2 and J_3, while the angular momentum acquires a simple form $L_1 = J_1\omega_1$, $L_2 = J_2\omega_2$ and $L_3 = J_3\omega_3$, ω_1, ω_2 and ω_3 being the projections of the angular velocity onto the coordinate axes of the system moving along with the body. In the momental equation (31.2), the derivative $d\mathbf{L}/dt$ is calculated relative to the inertial coordinate system. This quantity has to be calculated relative to the moving coordinate system fixed to the body.

Suppose that a vector A is defined in terms of its components relative to the coordinate system $(\mathbf{i}'_x, \mathbf{i}'_y, \mathbf{i}'_z)$:

$$\mathbf{A} = \mathbf{i}'_x A'_x + \mathbf{i}'_y A'_y + \mathbf{i}'_z A'_z. \tag{35.2}$$

With the passage of time, the projections A'_x, A'_y and A'_z on the axes of the moving coordinate system change and so does the orientation of the coordinate axes relative to the inertial coordinate system. Hence we can write

$$\frac{d\mathbf{A}}{dt} = \mathbf{i}'_x\frac{dA'_x}{dt} + \mathbf{i}'_y\frac{dA'_y}{dt} + \mathbf{i}'_z\frac{dA'_z}{dt} + \frac{d\mathbf{i}'_x}{dt}A'_x + \frac{d\mathbf{i}'_y}{dt}A'_y + \frac{d\mathbf{i}'_z}{dt}A'_z. \tag{35.3}$$

The velocity of a point on the rotating body with radius vector r is given by $\mathrm{d}\mathbf{r}/\mathrm{d}t = \omega \times \mathbf{r}$. Similarly, by following the tip of the vector i'_x drawn from the point to the axes of rotation, we obtain $\mathrm{d}\mathrm{i}'_x/\mathrm{d}t = \omega \times \mathrm{i}'$. The derivatives of i'_y and i'_z have the same form. Thus,

$$\begin{aligned}
&\frac{\mathrm{d}\mathrm{i}'_x}{\mathrm{d}t}A'_x + \frac{\mathrm{d}\mathrm{i}'_y}{\mathrm{d}t}A'_y + \frac{\mathrm{d}\mathrm{i}'_z}{\mathrm{d}t}A'_z \\
&= \omega \times \mathrm{i}'_x A'_x + \omega \times \mathrm{i}'_y A'_y + \omega \times \mathrm{i}'_z A'_z \\
&= \omega \times (\mathrm{i}'_x A'_x + \mathrm{i}'_y A'_y + \mathrm{i}'_z A'_z) = \omega \times \mathbf{A}.
\end{aligned} \tag{35.4}$$

Hence (35.3) may be written in the form

$$\frac{\mathrm{d}\mathbf{A}}{\mathrm{d}t} = \frac{\partial \mathbf{A}}{\partial t} + \omega \times \mathbf{A},$$

where $\partial \mathbf{A}/\partial t = \mathrm{i}'_x(\mathrm{d}A'_x/\mathrm{d}t) + \mathrm{i}'_y(\mathrm{d}A'_y/\mathrm{d}t) + \mathrm{i}'_z(\mathrm{d}A'_z/\mathrm{d}t)$ is the derivative of A, obtained under the condition that the axes i'_x, i'_y and i'_z are stationary. This formula is valid for all vectors A. Applying it to the quantity L in (31.2), we can represent the momental equation in the following form:

$$\frac{\partial \mathbf{L}}{\partial t} + \omega \times \mathbf{L} = \mathbf{M}. \tag{35.5}$$

Considering that $L_x = J_x\omega_x$, $L_y = J_y\omega_y$ and $L_z = J_z\omega_z$, we can write (35.5) in terms of projections on the axes of the moving coordinate system:

$$\begin{aligned}
J_x\frac{\mathrm{d}\omega_x}{\mathrm{d}t} + (J_z - J_y)\,\omega_y\omega_z &= M_x, \\
J_y\frac{\mathrm{d}\omega_y}{\mathrm{d}t} + (J_x - J_z)\,\omega_z\omega_x &= M_y, \\
J_z\frac{\mathrm{d}\omega_z}{\mathrm{d}t} + (J_y - J_x)\,\omega_x\omega_y &= M_z.
\end{aligned} \tag{35.6a}$$

It should be emphasized once again that *all the quantities in this equation have been referred to the axes of the moving coordinate system rigidly fixed to the body. The primes have been omitted only in order to simplify the form of notation.*

These equations are called Euler's equations. In principle, they can always be used to determine the motion of a body fixed at one point, although the actual solution may be quite complicated and cumbersome.

FREE AXES. In order to completely define the motion with the help of (35.6a) without taking into account (35.1), we must take the origin of coordinates in which (35.6a) are written at the centre of mass of the body and consider the moment of constraints be zero in this case. Suppose that there are no forces acting on the body, and hence the moments of forces M_x, M_y and M_z are zero. We direct the axes of the coordinate system rigidly fixed to the body along the central principal axes. Consequently, J_x, J_y and J_z in (35.6a) are the central principal moments of inertia of the body. They are generally not equal to one another. Let us find out the possible free motion of the body.

It follows directly from (35.6a) that there can be no free rotation of the body during which the magnitude of the angular velocity and the orientation relative to the body are conserved, the angular velocity not coinciding in direction with any of the central principal axes with different moments of inertia. Let us suppose that such a situation is possible, i.e. $\omega_x = \text{const} \neq 0$, $\omega_y = \text{const} \neq 0$ and $\omega_z = \text{const} \neq 0$. It then follows from the equations that the following equations must be satisfied:

$$
\begin{aligned}
(J_z - J_y)\omega_y \omega_z &= 0,\\
(J_x - J_z)\omega_z \omega_x &= 0, \qquad (35.6b)\\
(J_y - J_x)\varphi_x \omega_y &= 0.
\end{aligned}
$$

These equations can simultaneously be satisfied only if two projections of the angular velocity simultaneously vanish. This means that the direction of the angular velocity coincides with one of the central principal axes. Suppose, for example, that $\omega_y = \omega_z = 0$. In this case, Eqs. (35.6b) will be satisfied. The angular velocity is directed along the X-axis, i.e. along a central principal axis.

Thus, a rigid body can freely rotate only about its central principal axes. These axes are called free axes. In general, the moments of inertia about these axes are different. It can be proved that *a rotation of a body is stable only about a central principal axis with the maximum or minimum moment of inertia. A rotation about a central principal axis with an intermediate moment of inertia is unstable.* Small random deviations of the axis of rotation from this direction lead to the emergence of forces which tend to increase the deviation. This can be demonstrated with the help of the following experiment. A body in the form of a rectangular parallelepiped has as its central principal axes three mutually perpendicular axes which are parallel to its sides and pass through the geometrical centre. The parallelepiped has the maximum and minimum moments of inertia about the axes parallel to its longest and

!

Only the central principal axes of the inertia tensor with the maximum and minimum values of the moment of inertia are the axes of the stable free rotation of a rigid body. A rotation about the central principal axis with an intermediate moment of inertia is unstable. No forces tending to change the direction of the axis of rotation or shift it parallel to itself in the body emerge upon rotation about its free axes.

The angular momentum of a rigid body fixed at a point does not coincide in direction with the angular velocity. They are related through the inertia tensor.

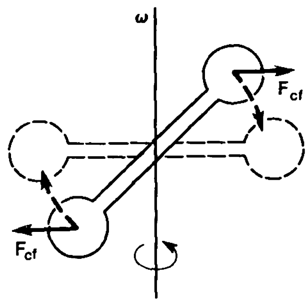


Fig. 84. The axis coinciding with the angular velocity vector is not free in the present case since the centrifugal forces of inertia in the reference frame fixed to the body tend to change the direction of this axis in space.

shortest sides. If the body is rotated about one of these axes and thrown, its motion will be stable and the direction of the axis of rotation will be conserved. If, however, the body is rotated about the axis parallel to its intermediate side, the motion will not be stable and the body will tumble at random.

In order to graphically demonstrate why the free axes must coincide with the central principal axes, let us consider a body in the shape of a dumb-bell. Let us choose the axis of rotation in a direction which does not coincide with any of the central principal axes. For example, we choose the axis shown in Fig. 84.

Obviously, the longitudinal axis of a rotating body tends to change its direction under the action of centrifugal forces in order to occupy the position shown by the dashed lines in Fig. 84. In this position, the rotation is stable, and the direction of ω coincides with the central principal axis about which the body has the maximum moment of inertia.

NUTATION. Let us consider a body which has an axial symmetry about a certain axis, i.e. which is a body of revolution (Fig. 85). Obviously, one of the central principal axes coincides with the axis of symmetry and the other two are perpendicular to this axis. We direct the X-axis along the axis of symmetry, and the Y- and Z-axes along the other two central principal axes. It follows from symmetry considerations that $J_x = J_1$ and $J_y = J_z = J_2$. In this case, Eqs. (35.6a) acquire the form

$$J_1 \frac{d\omega_x}{dt} = 0,$$

$$J_2 \frac{d\omega_y}{dt} + (J_1 - J_2)\,\omega_z\omega_x = 0, \tag{35.7}$$

$$J_2 \frac{d\omega_z}{dt} + (J_2 - J_1)\,\omega_x\omega_y = 0.$$

!

Precession is the motion of the axis of a gyroscope under the action of the moment of external forces applied to it.
Nutation is the motion of the axis of symmetry of a body about the total angular momentum vector which remains stationary in space.
The period of a gyroscopic pendulum characterizes the ability of its axis of rotation to retain the direction in space under the action of the moment of external forces.

In the first place, it can be seen from these equations that a motion is possible with $\omega_x = \omega_1 = \text{const}$ and $\omega_y = \omega_z = 0$, i.e. the body can rotate about the axis of symmetry at a constant angular velocity. However, this is not the only possible motion of the body. Let us write the second and third equations under the condition that $\omega_x = \omega_1 = \text{const}$ in the form

$$\frac{d\omega_y}{dt} + \gamma\omega_z = 0, \qquad \frac{d\omega_z}{dt} - \gamma\omega_y = 0, \tag{35.8}$$

where $\gamma = (J_1 - J_2)\,\omega_1/J_2$. These equations have the solution

$$\omega_y = A\cos\gamma t, \qquad \omega_z = A\sin\gamma t. \tag{35.9}$$

The angular velocity vector $\omega_\perp = \mathrm{i}_y\omega_y + \mathrm{i}_z\omega_z$ lying in the

Fig. 85. Nutation.
The axis of rotation, the angular velocity vector ω and the total angular momentum vector L lie in the same plane rotating about the latter with the velocity of nutation.

YZ-plane rotates about the origin at an angular velocity γ. The total angular velocity is given by

$$\omega = \mathrm{i}_x \omega_1 + \omega_\perp. \tag{35.10}$$

This resultant vector moves about the X-axis, forming the surface of a cone with angle α during its rotation ($\tan\alpha = \omega_\perp/\omega_1$), i.e. *the angular velocity of rotation of the body does not coincide in direction with the axis of symmetry of the body, viz. the X-axis.* In turn, the axis of symmetry does not remain stationary in space, and moves along the surface of a cone whose axis is stationary in space and coincides with the total angular momentum vector L. The angular velocity of the motion is also equal to γ. Consequently, the total motion can be described as follows: the plane containing the instantaneous velocity vector ω and the axis of symmetry rotates at an angular velocity γ about vector L, the relative arrangement of vector ω and the axis of symmetry remaining unchanged. This motion of the axis of symmetry of a body about the total angular momentum vector L which remains stationary in space is called nutation, and γ the velocity of nutation. In such a motion, vector ω rotates about the axis of symmetry at the same angular velocity γ as described above. ***The amplitude of nutation depends on the initial conditions responsible for nutation, but its frequency is determined only by the moments of inertia and the angular velocity of rotation about the axis of symmetry.*** A body can also rotate without nutation if its angular velocity exactly coincides with the axis of symmetry.

A sphere is also a body of revolution. In this case, $J_x = J_y = J_z$, and hence $\gamma = 0$. This means that in the absence of external forces, the axis of rotation of a sphere always maintains a fixed position relative to the body, and there can be no nutation. This is due to the fact that any axis passing through the centre of the sphere is a central principal axis of the inertia tensor. Nutation is possible in a nonhomogeneous sphere. In particular, nutation is observed for the axis of rotation of the Earth, which proves that the Earth cannot be considered a homogeneous sphere.

The moments of inertia of the Earth about the axes lying in the

?

What are the free axes of rotation? Which of them are stable?
What is nutation? What does the velocity of nutation depend on?
Why is nutation impossible for a homogeneous sphere?
Can you schematically draw a picture in which the total angular momentum, the instantaneous angular velocity and the axis of symmetry lie in the same plane rotating at the velocity of nutation about the total angular momentum vector?
What is gyroscopic precession?
What is the difference between precession and nutation?

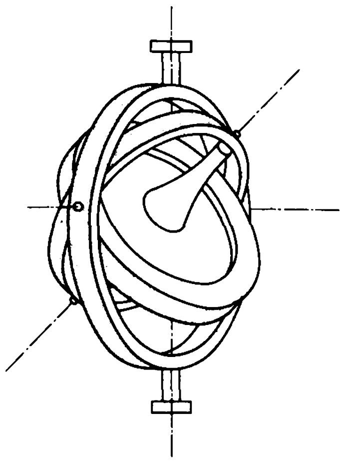


Fig. 86. The gimbal suspension allows an unimpeded variation of the orientation of a body and the suspension to which it is connected.

equatorial plane can be assumed to be equal. In (35.7) and (35.8), we can assume that the X-axis is directed along the axis of rotation of the Earth. When this factor is taken into account, the velocity of nutation γ is given, as in (35.8), by $\gamma = (J_1 - J_2)\omega_1/J_2$. Measurements of the moments of inertia of the Earth give the value $(J_1 - J_2)/J_2 \simeq 1/300$. This means that the period of nutation of the Earth's axis must be about 300 days, i.e. the axis of rotation of the Earth completes one revolution in about 300 days over the surface of a cone about the axis of symmetry. This axis is found from geodesic measurements, while the axis of rotation is determined by observing the motion of the stars. It passes through the centre of the circles described by the stars in twenty-four hours. However, the observed motion of the stars is much more complicated. In the first place, it is irregular and is seriously affected by earthquakes and seasonal changes taking place on the Earth's surface. Strictly speaking, it is this kind of reasons that are responsible for the nutation of the Earth since otherwise the axis of rotation of the Earth would coincide with the axis of symmetry to overcome the energy losses associated with viscosity, and no nutation would be observed. The actual period of nutation is about 440 days, due apparently to the fact that the Earth is not absolutely rigid. The maximum distance of a point on the Earth's surface through which the axis of rotation passes does not exceed 5 m (at the north pole) from the point lying on the axis of symmetry.

GYROSCOPES. An axially symmetric body set into a very fast rotation about its axis of symmetry is called a gyroscope. Examples of gyroscopes are a spinning top, or a disc rotating at a high speed about an axis passing through its centre and perpendicular to its surface. The body of revolution shown in Fig. 85 is also a gyroscope if its angular velocity ω_1 is quite high.

PRECESSION OF A GYROSCOPE. Let us assume that a gyroscope is fixed at its centre of mass, but its axis can freely rotate in any direction. Such a fastening is possible with the help of the gimbal (or Cardan) suspension shown in Fig. 86, which ensures a free variation of the orientation of the gyroscope's axis in three mutually perpendicular directions. There is no need to show the gimbal suspension in the figure (Fig. 87). Suppose that a moment of external forces is applied to a gyroscope. The gyroscope rotates about its axis at a very high angular velocity ω, and hence the possible nutation of its axis of rotation over the surface of a cone about the geometrical axis (see Fig. 85) is very small. The motion of the gyroscope's axis under the action of the moment of external forces can be neglected.

Hence we shall assume that the axis of rotation always coincides with the axis of symmetry of a gyroscope and that the angular momentum $L = J\omega$. The axis of rotation coincides

?

Under what conditions is it possible to assume that the angular momentum vector of a gyroscope, the instantaneous angular velocity of rotation and the axis of symmetry coincide?
What is the gimbal suspension?
What applications of gyroscopes do you know?
What does the velocity of precession depend on?
Can you explain the behaviour of an egg-shaped top? Why does its axis change the angle of inclination to the horizontal?

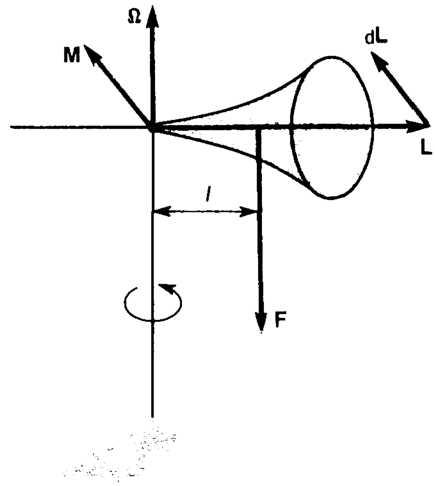


Fig. 87. Gyroscopic precession.
It is assumed that the axis of a rapidly rotating gyroscope coincides with the angular velocity vector ω and the angular momentum vector L.

with the central principal axis of the inertia tensor of the gyroscope and is chosen in such a way that it is stable during rotation.

Stable free rotation takes place about this axis. The direction of the axis of stable rotation remains fixed. For example, if we hold the base of a gimbal suspension and arbitrarily change its orientation, the joints will rotate in such a way that the direction of the axis of rotation remains unchanged in space. Hence if the gimbal is fastened to any object, say, a rocket, the orientation of the axis of rotation in space will remain constant relative to fixed stars. Knowing the position of the rocket relative to its axis of rotation, we can determine its orientation in space at any instant of time. This makes a gyroscope a very important navigational instrument in the flight of rockets. It is also the basic element of an autopilot, a device which automatically controls the flight of an aeroplane. Gyroscopes have many other important applications, some of which will be considered later in this book.

Let us suppose that the point of suspension of a gyroscope does not coincide exactly with its centre of mass. In this case, a moment of forces acts on the gyroscope's axis during the accelerated motion of a gimbal under the action of the inertial forces. If the gimbal is situated on the Earth, the force of gravity also gives rise to a moment of forces applied to the gyroscope's axis. *Under the action of the moment of forces, this axis begins to move and changes its direction in space. This motion caused by the moment of external forces is called the* precession of a gyroscope.

DIRECTION AND VELOCITY OF PRECESSION. The basic

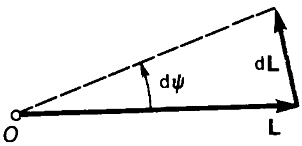


Fig. 88. Vector L changes only in direction, its absolute magnitude remaining unchanged.

property of a gyroscope which explains its behaviour under the action of forces is that the angular momentum vector L practically coincides with the angular velocity vector ω directed almost along the central principal axis of the gyroscope about which the rotation takes place. *Strictly speaking, the vectors do not have the same direction. However, the deviations from the common direction are very small and can be neglected.* Hence we shall assume that the vector $L = J\omega$ always concides with the central principal axis of the gyroscope. Such a coincidence is ensured by the gyroscopic forces whose origin will be discussed below. For the present, we simply note that these forces are Coriolis forces in nature.

It is convenient to apply the momental equation

$$\frac{dL}{dt} = M \tag{35.11}$$

to a gyroscope since the variation of L directly determines the motion of the gyroscope's axis. Knowing M, we can always find the direction of motion of the axis from the relation $dL = M\,dt$. In Fig. 87, the gyroscope's axis is horizontal, and the force F produces a moment $M = lF$ perpendicular to the plane of the figure. If the gyroscope were not in a state of rapid rotation, the force F would turn its axis to the right. However, the action of the force becomes quite different in the presence of rotation. Since $dL = M\,dt$, the tip of the axis begins to move in the horizontal plane. If F remains constant (for example, if it is the force due to a load suspended from the gyroscope at a certain distance from the point of support), the motion of the tip takes place at a constant angular velocity Ω. The gyroscope's axis rotates about the vertical axis passing through the point of support at the angular velocity of precession. As a result of precession, the total velocity of rotation $\omega + \Omega$ does not have the same direction as the gyroscope's axis. However, in view of the fact that $\omega \gg \Omega$, this disparity is extremely small, and in spite of the precession, it can be assumed as before that the angular velocity of rapid rotation coincides with the gyroscope's axis and the angular momentum L.

The angular velocity of rotation can easily be calculated. The precession of a gyroscope in the horizontal plane takes place as shown in Fig. 88. Point O represents the track of the axis of precession. Obviously, $dL = M\,dt = L\,d\psi$. Hence, by definition, we obtain the following expression for the angular velocity:

$$\Omega = \frac{d\psi}{dt} = \frac{M}{L} = \frac{M}{J\omega}. \tag{35.12}$$

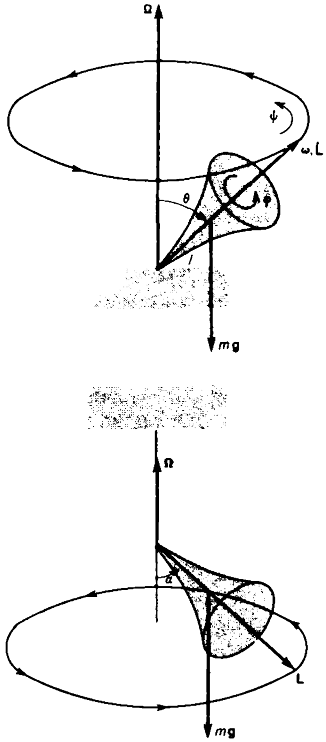


Fig. 89. The gyroscopic pendulum.

A peculiar feature of precession is that it has no "inertia", i.e. precession terminates as soon as the moment of external forces stops to act, as can be seen directly from (35.11). Hence precession does not behave like velocity, but like acceleration since the acceleration ceases to act as soon as the force is terminated.

GYROSCOPIC PENDULUM. Let us consider the case when the axis of the gyroscope is fastened at a point and suspended at the end on a thread (Fig. 89, cf. Fig. 87). Moreover, in this case, the axis is not horizontal, but is inclined at an angle θ to the vertical. It can clearly be seen that $M = mgl \sin\theta$, $dL = L \sin\theta \, d\psi = mgl \sin\psi \, dt$, and hence $\Omega = d\psi/dt = mgl/L$. Thus, the angular velocity does not depend on the angle at which the gyroscope's axis is inclined to the vertical. This is due to the fact that a change in the angle entails a simultaneous change in the moment of the force and the distance between the axis of rotation and the tip of vector **L** in the horizontal plane. The term "gyroscopic pendulum" reflects the independence of the velocity of precession of such a gyroscope of the angle of inclination of its axis. The period of revolution of this pendulum is $T = 2\pi/\Omega = 2\pi J\omega/(mgl)$; for quite large values of the moment of inertia J and the angular velocity of rotation ω and small l, this period can be very large and runs into several minutes or even hours. A mathematical pendulum with such a large period would have an enormous length. *The length of a mathematical pendulum with the period of oscillations equal to the period of precession of a gyroscopic pendulum is called the reduced length of the gyroscopic pendulum.* Since the period of a mathematical pendulum with length l_0 is $T = 2\pi\sqrt{l_0/g}$, the reduced length of the gyroscopic pendulum under consideration, $L_0 = g[J\omega/(mgl)]$, may indeed be very large for quite large $J\omega$ and small l.

Modern gyroscopes maintain a fixed orientation in space with a very high precision. The rate of their departure from the fixed orientation is less than 10^{-4} deg/h. It can be seen from Fig. 89 that the Euler angles φ, ψ and θ are rightfully called the angles of proper rotation, precession and nutation.

EGG-SHAPED TOP. If a top rests with its very sharp end on a platform, its axis precesses and moves along the surface of a cone as described above. This is a gyroscopic pendulum whose point of support lies below its centre of mass.

If, however, the top rests on a fairly broad end so that it cannot be assumed to be touching the surface at only one point on the axis of rotation, the effect becomes quite complicated. If the top is egg-shaped and rests with its sharper end on the surface during rotation, its axis tends to assume a vertical position. If it rotates with its broader end at the

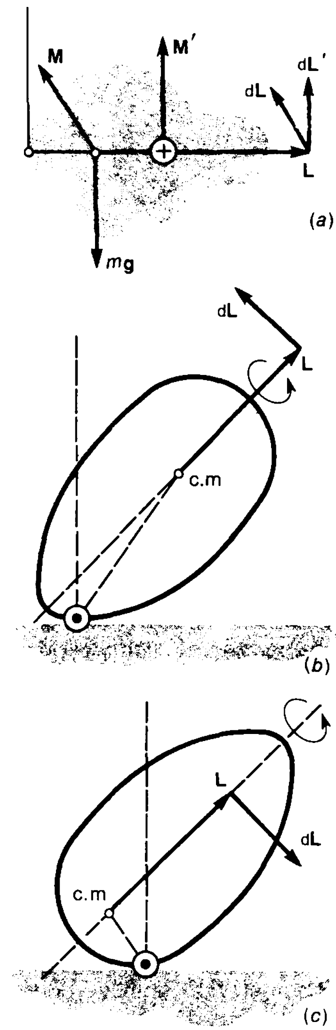


Fig. 90. The rise and fall of the axis of an egg-shaped top.

surface, the axis first assumes a horizontal position and then a vertical position, but in such a way that the top continues to rotate with its sharper end touching the surface.

Such a behaviour of the top is due to the frictional forces which produce a moment tending to orient the top's axis in the vertical plane. Suppose that a gyroscopic pendulum is acted upon by certain forces which tend to increase its velocity of precession, for example, a force F applied to the axis in the direction of its precession (Fig. 90*a*). It can easily be seen that this force produces a moment M about the point of support of the top. This moment is directed upwards and tends to raise the top's axis. It can be concluded from similar considerations that the factors responsible for a decrease in the velocity of precession tend to lower the axis.

Let us apply these arguments to the motion of egg-shaped tops. Such a top is shown in Figs. 90*b* and *c* with its sharp and broad ends downwards. If the top touches the table not along its axis of rotation, it begins to roll over the table owing to the frictional forces acting at the point of contact between the top and the table. It can be seen from Fig. 90*b* that the oscillations are responsible for an additional motion of the axis of rotation in the same direction in which it moves on account of precession. In this case, the velocity of precession increases, and hence the top's axis will be raised. For the situation shown in Fig. 90*c*, the pattern of motion of the egg-shaped top changes. In this case, the centre of mass lies on the other side of the vertical passing through the point of oscillation, while the direction of rotation of the top (i.e. the direction of L) remains the same. The precession now changes in the opposite direction. However, in this case, the oscillations lead to an additional motion of the axis against the direction of precession, and hence the top's axis will be lowered.

The same conclusions can also be drawn by directly considering the moments of frictional forces about the centre of mass of the top. In both cases shown in Figs. 90*b* and *c*, the frictional force acts perpendicular to the plane of the figure and is directed towards the reader. When the top rests with its sharp end touching the table, its centre of mass lies on the right of the vertical passing through the point of contact between the top and the table. Hence the moment of frictional force about the centre of mass is directed in such a way that it tends to turn vector L towards the vertical. Consequently, the top tends to assume a vertical position with its sharp end downwards. When the broad end of the top touches the table, its centre of mass lies on the left of the vertical drawn from the point of contact between the top and the table. In this case, the moment of frictional force about the centre of mass is directed

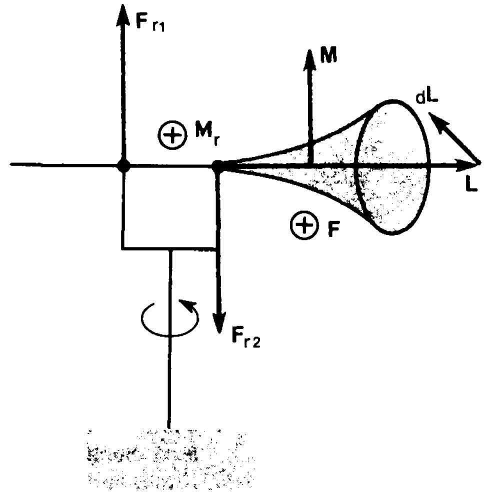


Fig. 91. Motion of a powered gyroscope.

in such a way that it tends to turn vector L towards the horizontal.

In practice, an egg-shaped top can have a stable motion only if its sharp end is in contact with the table. If the broad end is in contact with the table, the motion of the top is unstable and it quickly assumes a position in which its sharp end is in contact with the table. Expert demonstrators elegantly perform such experiments at their lectures.

POWERED GYROSCOPE. When the gyroscope's axis is fastened at one end only, it can move in any direction. Hence such a gyroscope is called a free gyroscope. If the gyroscope's axis is fixed at two points, its motion becomes restricted. Suppose, for example, that the axis is fastened in the way shown in Fig. 91: it can move freely in the horizontal plane but cannot move in the vertical plane. Such a gyroscope is called a powered gyroscope. *The motion of a powered gyroscope differs radically from that of a free gyroscope for the same moment of forces.* In order to analyze the motion of the axis of a powered gyroscope, we must take into account the moment of the reaction forces of a support at the points where the axis is fixed.

If a force F is horizontal (see Fig. 91), it creates a moment M directed upwards. If the gyroscope were free, this moment would raise the right end of the gyroscope. However, this motion is obstructed by the points at which the axis is fastened. Reaction forces F_{r1} and F_{r2} of these points produce a moment M_r perpendicular to the plane of the figure. Under the action of this moment, the right end of the gyroscope's axis moves in the horizontal plane in the direction of the initial force F. *Hence a powered gyroscope is compliant: its axis turns*

?

Why does a powered gyroscope become compliant?
What balances the moment of external forces during the precession of a gyroscope?
Can you explain why the precession of a gyroscope has no "inertia", i.e. the precession terminates as soon as the moment of external forces causing it stops to act?
Can you explain the emergence of gyroscopic forces?
What is the nature of gyroscopic forces?

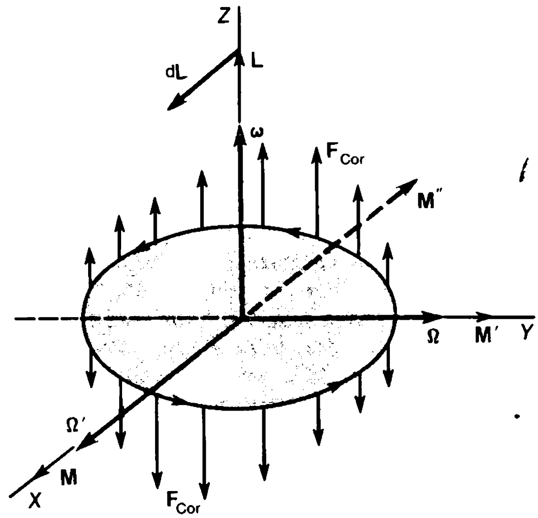


Fig. 92. Gyroscopic forces caused by Coriolis forces.

in the direction in which the external force tends to turn it. In a free gyroscope, however, the axis turns in a plane perpendicular to the applied force.

GYROSCOPIC FORCES. We have considered above the motion of gyroscopes. Let us now discuss the nature of gyroscopic forces. These forces are of Coriolis type.

Suppose that we have a rotating disc (Fig. 92) whose angular velocity of rotation is directed along the Z-axis. We shall assume that the disc is made of point masses m. Let us apply to the disc a moment of forces M in the positive X-direction. Under the action of this moment, the disc tends to start turning about the X-axis at an angular velocity Ω'. Consequently, Coriolis forces $\mathrm{F}_{\mathrm{Cor}} = -2m\Omega' \times \mathrm{v}'$ begin to act on the moving point masses of the disc. These forces produce a moment of forces along the Y-axis, causing the disc to rotate about this axis at an angular velocity Ω. As a result, the angular momentum vector L moves in the direction of vector M, i.e. the same precession takes place, which is performed by the gyroscope's axis under the action of the moment of external forces applied to it. Hence it can be stated that *gyroscopic forces are Coriolis forces.*

In order to follow the emergence of gyroscopic forces in greater detail, let us derive their value by proceeding directly from the computation of Coriolis forces. Figure 93 shows the distribution of the velocities of the points of a moving disc from the positive Z-direction. At different points of the disc above the Y-axis the Coriolis forces are directed perpendicular to the plane of the figure towards the reader, while at points

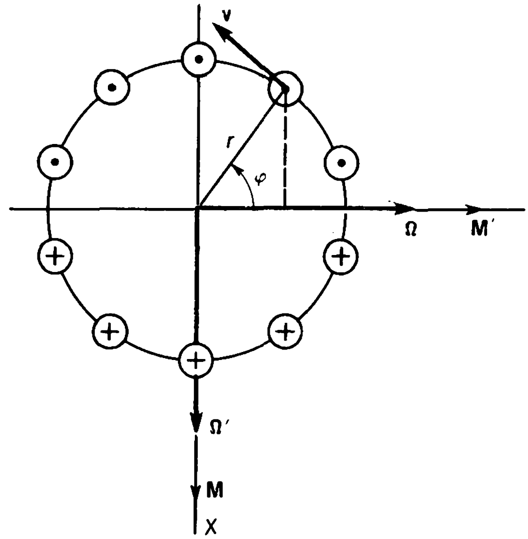


Fig. 93. Computing the moment of Coriolis forces.

below the Y-axis these forces are directed away from the reader. Further, considering that $F_{Cor} = -2m\Omega' \times v'$ and $v' = \omega r$, we can write the following expression for the Coriolis forces at point (r, φ):

$$F_{Cor} = 2m\Omega' v' \sin\varphi = 2m\Omega'\omega r \sin\varphi. \tag{35.13}$$

Hence we obtain the following relation for the moment of the Coriolis force about the Y-axis for the point under consideration:

$$M'_y = 2m\Omega'\omega r^2 \sin^2\varphi. \tag{35.14}$$

Since the average value $\langle \sin^2\varphi \rangle = 1/2$ over one cycle, we can write for $\langle M'_y \rangle$

$$\langle M'_y \rangle = mr^2\Omega'\omega = L\Omega', \tag{35.15}$$

where we have considered that $mr^2 = J$ is the moment of inertia of a point mass about the axis of rotation, and $L = J\omega$ is the angular momentum of the moving point about the same axis. Summing over all points of the disc, we find that the expression (35.15) does not change, but $\langle M'_y \rangle$ now represents the total moment of the Coriolis forces acting on the disc about the Y-axis. In this case, the quantity L indicates the angular momentum of the disc. It can be seen from Fig. 92 that the Coriolis forces also produce the moments of forces about the X-axis, but the sum of these moments is zero, and hence they need not be taken into account.

Under the action of the moment of forces $\langle M'_y \rangle$, the disc begins to rotate about the Y-axis. This rotation, like the one considered above, produces a moment of Coriolis forces about the X-axis, its direction being opposite to the initially applied moment of forces. The angular velocity of rotation increases until the moment of Coriolis forces arising about the X-axis compensates for the initially applied moment. For this, the following relation must be satisfied in accordance with (35.15):

$$M = L\Omega, \tag{35.16}$$

where M is the moment of external forces about the X-axis, and Ω is the angular velocity of rotation of the disc about the Y-axis. Thus, the moment of forces about the X-axis does not cause any rotation of the disc about this axis, but causes a rotation about the Y-axis. It can be seen from Fig. 92 that the tip of vector L moves in the direction of vector M. Considering that $\Omega = \mathrm{d}\theta/\mathrm{d}t$ and $\mathrm{d}L = L\,\mathrm{d}\theta$ (see Fig. 92), the relation (35.16) can be written in the form $M = \mathrm{d}L/\mathrm{d}t$ or, if we take into account the spatial directions of the vectors shown in Fig. 92, in vector form

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \mathbf{M}. \tag{35.17}$$

This is the momental equation which was used for a detailed analysis of the motion of a gyroscope.

Thus, it can be stated that the precession of the gyroscope's axis is due to Coriolis forces. During a steady precession, the angular velocity of the gyroscope's axis is responsible for the emergence of a moment of Coriolis forces which is equal to the moment of external forces acting on the gyroscope but has the opposite direction and the two moments neutralize each other.

Sec. 36. MOTION UNDER FRICTION

The physical pattern of the emergence of frictional forces is analyzed and their influence on the motion of bodies is considered.

DRY FRICTION. If two bodies are in contact so that their surfaces exert a pressure on each other, and if a small force is tangentially applied to these surfaces, there will be no sliding of one surface over the other (Fig. 94). In order to start sliding, we must apply a force greater than some minimum value. Consequently, when two bodies are in contact and a pressure exists between their surfaces, forces inhibiting their sliding appear and cause static friction. Sliding begins only when the external tangential force exceeds some value. *Thus, static friction F_{st} varies from zero to some maximum value F_{st}^{max} and is equal to the external force applied to the body.*

This force is directed against the external force and neutralizes it. As a result, the bodies remain stationary, and there is no sliding of one surface over the other.

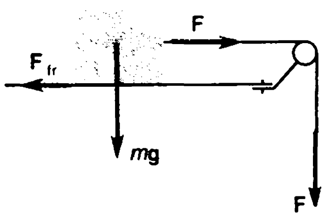


Fig. 94. Dry friction.
It is characterized by static friction.

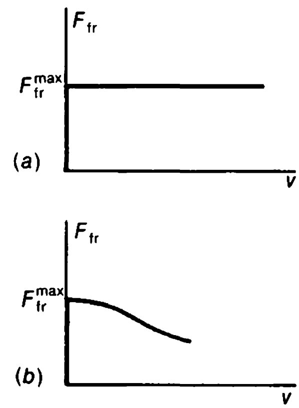


Fig. 95. Dependence of dry friction on velocity.

The frictional force depends on pressure, the material of bodies and the state of the contact surfaces. The static friction is higher for rough surfaces than for polished surfaces.

When the external tangential force exceeds the maximum static friction, one surface begins to slide over the other. *In this case, the frictional force is directed against the velocity.* For highly polished dry metallic surfaces at low velocities, the numerical value of the static friction is practically independent of velocity and is equal to the maximum static friction. Thus, the dependence of the frictional force on velocity has the form shown in Fig. 95*a*. For all the velocities $v \neq 0$, the frictional force has quite a definite magnitude and direction. For $v = 0$, its value is not unique and depends on the external force.

However, the frictional force is independent of velocity only at velocities that are not too high. It does not hold for all types of bodies and all types of finished surfaces. In some cases, the velocity dependence of the force of friction between solid surfaces has the form shown in Fig. 95*b*: as the velocity increases to a certain value, the frictional force first decreases in comparison with the static friction (for the sake of simplicity, the "maximum static friction" is sometimes called the "static friction") and then increases.

The most peculiar feature of the frictional force considered above is the existence of static friction: the frictional force does not vanish when the relative velocity of the contact surfaces becomes zero. Such a friction is called dry friction. For the case presented in Fig. 94, the frictional force is given by the formula $F_{fr} = k'mg$, where k' is called the coefficient of friction. This coefficient is determined experimentally.

The emergence of dry friction is due to the interaction of molecules, atoms and electrons near the contact surface, i.e. ultimately to the electromagnetic interaction.

FLUID FRICTION. If two metallic contact surfaces are well lubricated, they begin to slide under the action of very weak forces (practically equal to zero). This is due to the fact that instead of two solid metallic surfaces, thin fluid films of oil are now in contact formed on the surfaces as a result of lubrication. *Such frictional forces for which there is no static friction are called* fluid friction. For example, a metallic ball moves in a gas or a liquid even when a very weak force is applied to it. Frictional forces tending to inhibit the motion appear between the surface of the ball and the gas or the liquid. But as the velocity tends to zero, this frictional force also tends to zero. In other words, this is fluid friction.

The dependence of fluid friction on velocity is shown in Fig. 96. At very low velocities, this force is directly proportional to velocity: $\mathbf{F}_{fr} = -\beta\mathbf{v}$. The proportionality factor

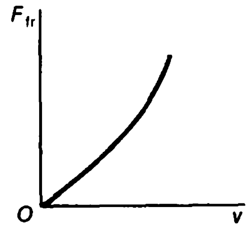


Fig. 96. Dependence of fluid friction on velocity.

The forces acting opposite to the velocity are laid off along the y-axis. A characteristic feature of this friction is that the frictional force vanishes at the zero velocity.

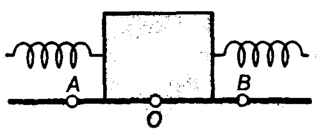


Fig. 97. Stagnation.

!

A characteristic feature of motion for fluid friction, which depends on velocity, is to attain a limiting velocity determined by the value of an applied force. There is no limiting velocity for dry friction.
Stagnation may drastically distort the results of measuring physical quantities by means of instruments if their pointers experience dry friction at their axes of rotation.
Skidding is due to sliding friction which is always directed against velocity, but which does not depend significantly on its magnitude.

β depends on the properties of a liquid or a gas, geometrical characteristics of a body and the quality of its surface.

When a rigid body moves in a liquid or a gas, it is subjected both to the frictional force directed along the tangent to the surface of the body at each point and to another type of force acting in a direction opposite to its velocity. This force is called drag, and its dependence on velocity may be rather complex indeed.

WORK OF FRICTIONAL FORCES. The work done by static friction is zero since there is no displacement. When a solid surface slides over another, the frictional force is directed against the displacement. The work done by this force is negative. Consequently, the kinetic energy of the bodies in contact is transformed into their internal energy; and the bodies rubbing against each other get heated. Fluid friction also performs a negative work; in this case also, the kinetic energy of the moving bodies is transformed into their internal energy, and the velocity of the bodies decreases.

Hence, in the case of motion under friction, the energy conservation law cannot be formulated as the constancy of the sum of the kinetic and potential energies. In the case of friction, this sum decreases, and the energy is transformed into the internal energy of the bodies in contact.

It should not be concluded that friction mainly plays a negative role in motion of bodies. If there were no friction, our lifestyle would have been different. It can be stated that the existing technology is based on the presence of friction. Without friction, there would be no motion of automobiles, people would not be able to walk on flat surfaces, nor would it be safe to sit in a chair, and so on.

STAGNATION. Let us imagine that a body moves with friction along a horizontal plane under the action of the forces shown in Fig. 97. In the equilibrium position of the body, the resultant of the forces exerted by the springs in the horizontal plane is zero. As the body is deviated from its equilibrium position, a force tending to return the body to its original position appears. However, if this force is smaller than the maximum static friction, it will not be able to move the body. Hence the equilibrium position of the body is not only its equilibrium position at point O, but all the other positions within the interval AB over which the body can deviate from the equilibrium position. Although the forces are exerted by the springs on the body in all these positions, the body remains at rest. If the body is deviated beyond the interval AB and then released, it will be set in motion by the forces exerted by the springs. Depending on the magnitude of the initial deviation, the body will either perform an oscillatory motion or simply move

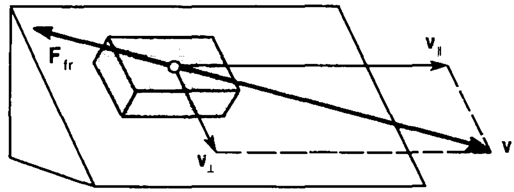


Fig. 98. Skidding.

in one direction, but come to rest after some time as a result of the energy losses due to friction. The body may come to rest in any position within the interval AB. Practically, the body will never stop in the equilibrium position. This phenomenon in which the body comes to rest and remains stationary at a position other than the equilibrium position while subjected to a nonzero force from the spring is called stagnation.

Obviously, if the friction between the body and the horizontal plane were fluid friction, there would be no stagnation since in this case even the smallest force exerted by the spring would set the body in motion. Hence the only equilibrium position in which the body can be at rest is the equilibrium position where the resultant of the forces exerted by the springs is zero.

Stagnation is of considerable importance in many cases. In measuring instruments, a measurable quantity or an effect produced by it is usually compared with the scale of the quantity or the effect, the result being shown as the reading of the pointer. If the pointer experiences dry friction at its axis of rotation, it will never exactly show the position corresponding to the equality of the measurable quantity to the scale. This will lead to a certain error of measurement which will be the larger, the larger the dry friction. Hence it is desirable to reduce the dry friction in measuring instruments to minimum and to create conditions close to those of fluid friction.

SKIDDING. Suppose that a body is at rest on an inclined plane (Fig. 98). This means that the maximum static friction is larger than the force $F = mg \sin \alpha$ which tends to cause the sliding of the body down the inclined plane (α is the angle of inclination to the horizontal). Let us now set the body in motion at a velocity v across the plane (see Fig. 98). The body immediately begins to slide down the inclined plane. This is due to the fact that as soon as the body starts moving in the direction of the velocity $v_\parallel$ across the plane, the friction between the plane and the body will be directed against the velocity. Hence there will be no force opposing the force $F = mg \sin \alpha$ which causes the sliding of the body. This results in the emergence of the velocity $v_\perp$ in the direction of sliding. The total velocity of the body along the inclined plane is $v = v_\parallel + v_\perp$. The frictional force F_{fr} is directed against the velocity. Only the component $F_{fr} \sin \beta$ ($\tan \beta = v_\perp / v_\parallel$) acts against the force $mg \sin \alpha$. If this component is equal to the force $mg \sin \alpha$, there will be no further increase in the velocity at which the body slides down the inclined plane, and the body will move at a constant velocity v at an angle β to $\mathrm{v}_\parallel$. In order to sustain such a motion across the inclined plane, the body must be subjected to a constant force $F \cos \beta$ equal to the component of the frictional force directed against the velocity $\mathrm{v}_\parallel$. The disappearance of static friction in the direction perpendicular to the velocity is called skidding. This term is borrowed from the phenomenon where it is manifested most clearly, viz. the skidding of automobiles.

Suppose that an automobile with a horizontal longitudinal axis rests on an inclined plane (see Fig. 98). Owing to the friction between the wheels and the plane, the automobile does not slide down the plane under the action of the force $F = mg \sin \alpha$. We can then set the automobile in motion across the inclined plane in the direction of the velocity $\mathrm{v}_\parallel$. If this is done very carefully and with a very small acceleration so that there is no sliding between the points on the wheels and the inclined plane, static friction will exist between them and will balance the force $F = mg \sin \alpha$. The automobile will safely move across the inclined plane without sliding. If, however, we try to move across the inclined plane with a large acceleration by revving up the motor, the driving wheels (usually the rear wheels) will begin to slide down the inclined plane, and the static friction balancing the component of the force of gravity along the inclined plane will vanish. As a result, the wheels begin to slide down the plane. If the rear wheels are the driving wheels, only these will move down the inclined plane, and the automobile will turn or, as is often said, "skid". It can easily be seen that skidding will also take place if brakes are applied suddenly when the decelerated wheels begin to slide down the plane.

It should not be concluded from here that a body begins to slide down an inclined plane only when there is a velocity across the plane. Let us consider the balance of the forces acting on a body after a force has been applied to it across the inclined plane (Fig. 99). Figure 99*a* shows a situation in which the force F is not very large. The resultant of the forces F and $mg \sin \alpha$ is balanced by the static friction F_{fr} which is less than the maximum static friction. All these forces lie on the inclined plane. Increasing the force F, we arrive at the critical situation shown in Fig. 99*b*. The resultant of F and $mg \sin \alpha$ attains a value equal to the maximum static friction. The body does not move in this case since all the forces balance one another. If the force F is increased further by a small amount, this equilibrium is violated (Fig. 99*c*). As before, the frictional force is directed

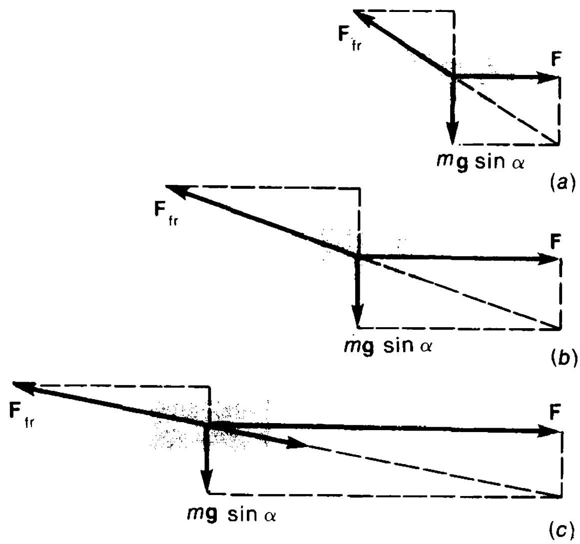


Fig. 99. Change in the balance of forces when approaching the skidding zone.

against the resultant of F and mg sin α. But since this force has already attained its maximum value, it is not equal to the resultant and cannot balance it. The most important point is that, primarily, the balancing of the force $mg \sin \alpha$, and not of the force F, is violated (see Fig. 99*c*). The component of F_{fr} directed against the force F balances the latter, while the component of F_{fr} directed against that of $mg \sin \alpha$ becomes smaller than this force. Hence the body begins to slide down the inclined plane, rather than to move across the plane in the direction of the applied force F as it might appear at first glance in the analysis of skidding. However, the body cannot just slide down the inclined plane since as soon as the sliding begins, the frictional force is reoriented in a direction against the sliding. Consequently, the force F is no longer balanced, and the body begins to move in the direction of this force. Thus, the body begins to slide and at the same time to move across the inclined plane. The process has been discussed as successive application of forces only to illustrate more clearly the essence of the physical phenomena. The body begins to move from the state of rest in the direction of the resultant of F and $mg \sin \alpha$, which attains the maximum value of static friction when the critical situation shown in Fig. 99*b* is reached.

?

What is the magnitude of dry friction for a body at rest and how is it directed?
What is the magnitude of fluid friction for a body at rest?
How does dry friction depend on velocity?
How does fluid friction depend on velocity?
How does stagnation affect measuring instruments?
What is skidding? When is it dangerous and when useful?

LIMITING VELOCITY. In the case of dry friction, bodies move with an acceleration when the external force exceeds the maximum frictional force. Under these conditions, a body may acquire an indefinitely high velocity when subjected to a constant external force. *The situation is quite different in the*

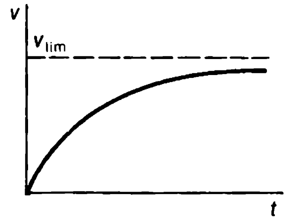


Fig. 100. Velocity approaches its limiting value in the presence of fluid friction.

case of fluid friction. In this case, a constant force applied to the body can accelerate it only to a certain velocity called the limiting velocity. When the limiting velocity is reached, the frictional force $F_{fr} = -\beta v$ balances the external force F, and the body moves at a constant velocity after this. Consequently, the limiting velocity is given by $v_{lim} = F/\beta$.

APPROACHING THE LIMITING VELOCITY. In the presence of fluid friction, the motion of a body in homogeneous space is described by the equation

$$m\frac{dv}{dt} = F_0 - \beta v. \tag{36.1}$$

We assume that the force F_0 is constant. Let $v = 0$ at the instant $t = 0$. Integrating (36.1), we obtain a solution of this equation:

$$\begin{gathered}\int\limits_0^v \frac{dv}{1-(\beta/F_0)v} = \frac{F_0}{m}\int\limits_0^t dt, \\ \frac{F_0}{\beta}\ln\left(1 - \frac{\beta}{F_0}v\right) = -\frac{F_0}{m}t,\end{gathered} \tag{36.2}$$

or, after integration,

$$v(t) = \frac{F_0}{\beta}\left\{1 - \exp\left[\left(-\frac{\beta}{m}\right)t\right]\right\}. \tag{36.3}$$

The plot of this function is presented in Fig. 100. The velocity $v(t)$ increases exponentially from 0 at $t = 0$ to its limiting value $v_{lim} = F_0/\beta$. The exponential curve depends very strongly on the value of the exponent. After the exponent has attained the value -1, the function rapidly vanishes. Hence it can be assumed that the velocity attains its limiting value over time τ during which the exponent in formula (36.3) becomes equal to -1. In other words, this value can be found from the condition $\beta\tau/m = 1$, i.e. $\tau = m/\beta$.

FREE FALL OF BODIES IN AIR. Bodies moving in air at very high velocities are subjected to fluid friction as well as to the aerodynamic drag proportional to the square of the velocity. For the free fall of a body in air, the limiting velocity is attained when the force of gravity acting on the body is equal to the drag. As an example, let us consider the jump of a parachutist from a balloon up to the instant of the opening of the parachute (we are considering the jump from a balloon which is stationary in air, and not from an aeroplane flying at a high speed). It has been established experimentally that the limiting velocity at which a person falls in air is about 50 m/s. We shall take this as the limiting velocity v_{lim}, although it depends to a certain extent on the height and mass of the jumper, the orientation of his body relative to the direction of motion, atmospheric conditions, etc. We direct the X-axis vertically upwards, and take the origin $x = 0$ at the surface of the Earth. Since at velocities under consideration the air drag is proportional to the square of the velocity, the equation of motion can be written in the form

$$m\dot{v} = m\ddot{x} = -mg + \varkappa v^2, \tag{36.4}$$

where $\varkappa$ is the coefficient of friction ($\varkappa > 0$). Assuming the limiting velocity v_{lim} to be known, let us present $\varkappa$ in terms of this velocity. For uniform motion with the limiting velocity, we have

$$m\ddot{x} = 0 = -mg + \varkappa v_{\text{lim}}^2, \qquad \varkappa = \frac{mg}{v_{\text{lim}}^2}.$$

Taking this expression into consideration, we can write (36.4) for $\varkappa$ in the form

$$\frac{dv}{dt} = -\frac{g}{v_{\text{lim}}^2}(v_{\text{lim}}^2 - v^2).$$

Integrating, we obtain

$$\int_0^v \frac{dv}{v_{\text{lim}}^2 - v^2} = -\frac{g}{v_{\text{lim}}^2}\int_0^t dt, \qquad \frac{1}{2v_{\text{lim}}}\ln\frac{v_{\text{lim}} + v}{v_{\text{lim}} - v} = -\frac{g}{v_{\text{lim}}^2}t.$$

Taking antilogarithms of this expression, we get

$$v = -v_{\text{lim}}\frac{1 - \exp(-2gt/v_{\text{lim}})}{1 + \exp(-2gt/v_{\text{lim}})}. \tag{36.5}$$

For the initial period of fall, when $2gt/v_{\text{lim}} \ll 1$, we can expand the exponents into series and confine ourselves to the term linear in t:

$$\exp\left(-\frac{2gt}{v_{\text{lim}}}\right) \approx 1 - \frac{2gt}{v_{\text{lim}}}. \tag{36.6}$$

In this case, (36.5) gives

$$v = -gt.$$

This means that a nearly free fall takes place at the initial stage, while the air drag does not play a significant role.

As the velocity increases, the role of air drag becomes more significant and is decisive at velocities close to limiting

?

Which factors determine the existence of limiting velocity for fluid friction?
Which factors determine rolling friction?
Let a cylinder rolling without friction stop as a result of the energy losses in overcoming rolling friction. Into what forms of energy and in what way is the kinetic energy of the rolling cylinder transformed?

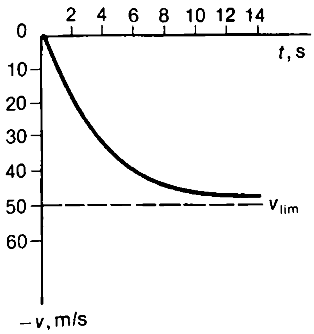


Fig. 101. Time dependence of the velocity of free fall of a parachutist.

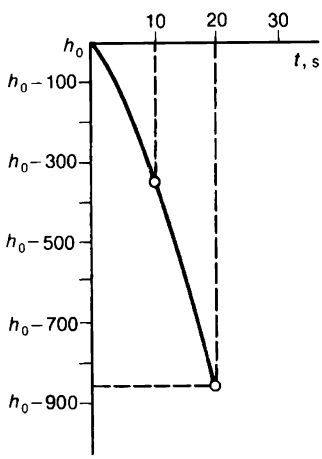


Fig. 102. Time dependence of the distance covered by a parachutist.

velocities. In this case, $2gt/v_{\text{lim}} \gg 1$, and hence

$$\frac{v_{\text{lim}} - v}{v_{\text{lim}}} = 1 - \frac{1 - \exp(-2gt/v_{\text{lim}})}{1 + \exp(-2gt/v_{\text{lim}})}$$

$$= \frac{2\exp(-2gt/v_{\text{lim}})}{1 + \exp(-2gt/v_{\text{lim}})} \approx 2\exp\left(-\frac{2gt}{v_{\text{lim}}}\right) \tag{36.7}$$

since $\exp(-2gt/v_{\text{lim}}) \ll 1$ can be neglected in this expression in comparison with unity in the denominator on the left-hand side of the last equality. Thus, for $t = 10$ s, the velocity differs from the limiting value by about $2e^{-4} \simeq 1/25$, i.e. by 2 m/s. Hence it can be assumed that the parachutist attains the limiting velocity in about 10 s after jumping. The time dependence of the velocity of the falling parachutist is shown in Fig. 101.

Integrating both sides of (36.5) with respect to time, we can find the distance covered by the parachutist during his fall:

$$\int_0^t v\,dt = -v_{\text{lim}} \int_0^t \frac{1 - \exp(-2gt/v_{\text{lim}})}{1 + \exp(-2gt/v_{\text{lim}})}\,dt$$

$$= -v_{\text{lim}} \int_0^t \left[1 - \frac{2\exp(-2gt/v_{\text{lim}})}{1 + \exp(-2gt/v_{\text{lim}})}\right] dt. \tag{36.8}$$

Considering that

$$-\frac{\exp(-2gt/v_{\text{lim}})}{1 + \exp(-2gt/v_{\text{lim}})}\,dt$$

$$= \frac{v_{\text{lim}}}{2g}\,d\ln\left[1 + \exp\left(-\frac{2gt}{v_{\text{lim}}}\right)\right],$$

$$v\,dt = dx,$$

we obtain from (36.8) the expression

$$h_0 - x = v_{\text{lim}}\left[t - \frac{v_{\text{lim}}}{g}\ln\frac{2}{1 + \exp(-2gt/v_{\text{lim}})}\right], \tag{36.9}$$

where h_0 is the height from which the parachutist begins to fall. This formula shows that the parachutist covers about 350 m during the first 10 s. The remaining distance until the opening of the parachute is covered almost uniformly at the limiting velocity. The time dependence of the distance is shown in Fig. 102.

The limiting velocity of descent of the parachutist with an open parachute is slightly less than 10 m/s. Hence the velocity of the parachutist drops from 50 m/s to about 10 m/s in a very short time after the opening of the parachute. This is due to the emergence of large accelerations, and hence large forces, which act on the parachutist. The action of these forces is termed dynamic shock.

When the parachutist jumps from an aeroplane flying at a velocity of several hundred metres per second, the pattern of his motion changes radically. After being thrown from the aeroplane, the parachutist's velocity drops within a few seconds from the velocity of the aeroplane to about 50 m/s. In this case, the acceleration of the parachutist is very large and so is the dynamic shock. Hence special measures are taken to ensure the safety of pilots being hurled out of aeroplanes flying at very high speeds, immediately after the instant at which they are catapulted.

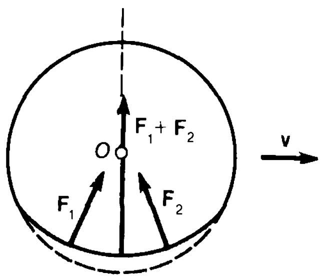
ROLLING FRICTION. Suppose that a cylinder rolls down an inclined plane without sliding. The dynamics of the motion of a cylinder subjected only to static friction was studied in Sec. 34. The assumption that the cylinder rolls without sliding means that the points of contact between the cylinder and the inclined plane do not slide relative to one another along the surface of contact. Hence the static friction acts between the cylinder and the plane. It is this friction that constitutes the tangential force T in Fig. 79. Together with the force $mg \sin \alpha$, this friction causes the rolling of the cylinder. Let us imagine that the surfaces of the cylinder and the inclined plane are absolutely undeformed. In this case, they must be in contact with each other along a geometrical line. There are no forces in this case except the static friction T. On the line of contact, the particles constituting the inclined plane and the cylinder do not experience any mutual displacement in the direction of the frictional force. Hence the work done by friction is zero and there are no losses due to friction. This means that

the rolling of a perfectly undeformed cylinder along a perfectly undeformed plane without sliding must not involve any losses due to friction, although static friction does exist and ensures the rolling of the cylinder.

In actual practice, however, there are losses of kinetic energy even in the case of rolling without sliding. For example, a cylinder rolling over a horizontal surface without sliding will ultimately come to a stop. If for a cylinder rolling down an inclined plane the kinetic energy at the end of the rolling is measured precisely, it will be found to be less than the potential energy which has been transformed into the kinetic energy. In other words, energy losses take place in this case. *These losses are due to rolling friction which cannot be reduced*

?

Why perfectly rigid bodies do not experience any rolling friction?
Why is rolling friction absent if deformations are perfectly elastic?
What is an approximate value of the limiting velocity of a person falling freely in air?
Can you show the difference in the dynamics of a parachutist when he jumps from a balloon and from an aeroplane flying at a very high speed?



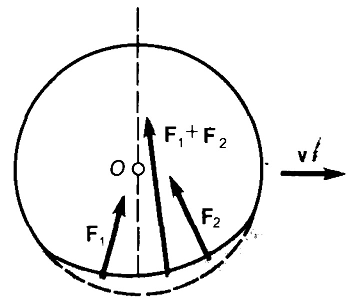


Fig. 103. During perfectly elastic deformation, the resultant $F_1 + F_2$ of forces passes through the axle of the wheel, and rolling friction does not set in.

Fig. 104. During inelastic deformation, the resultant $F_1 + F_2$ of forces does not pass through the axle of the wheel, and rolling friction sets in.

either to static friction or to sliding friction. The emergence of rolling friction is due to deformation. However, it can easily be seen that perfectly elastic deformations cannot give rise to any force impeding the motion (Fig. 103).

Naturally, it is assumed that there is no sliding of the cylinder. In case there is sliding, frictional forces appear in the contact zone even if the deformation is perfectly elastic. The deformation affects both a plane and a wheel. The wheel is slighty "flattened", as shown in Fig. 103 on a magnified scale. The dashed line shows the rim of the wheel in the absence of deformation. Forces F_1 and F_2 are the resultants of forces applied to the deformed wheel from the deformed parts of the surface in front of the vertical line and behind it. The total force acting on the wheel is $F_1 + F_2$, while the moment of forces about the axle of the wheel is equal to the sum of the moments of forces F_1 and F_2. The moment of force F_1 tends to increase the velocity of rotation of the wheel, while the moment of force F_2 tends to decrease the velocity. In the case of a perfectly elastic deformation, the entire pattern of forces is symmetric about the vertical line passing through the axle of the wheel. Consequently, the moments of forces F_1 and F_2 cancel out, and the resultant force $F_1 + F_2$ passes through the centre of the wheel. It has only a vertical component which balances the force of gravity and all that rests on the wheel. There is no horizontal force, and hence there is no rolling friction either.

The situation becomes different if the deformation is not perfectly elastic, as is indeed the case in actual practice. The forces acting on the wheel in this case are shown in Fig. 104. The forces F_1 and F_2 are different, and their resultant has a vertical component which balances the force of gravity as well as a horizontal component which is directed against the velocity and serves as the rolling friction. The moments of forces F_1 and F_2 are opposite but not equal. The moment of force F_2 which tends to retard the rotation is larger than the

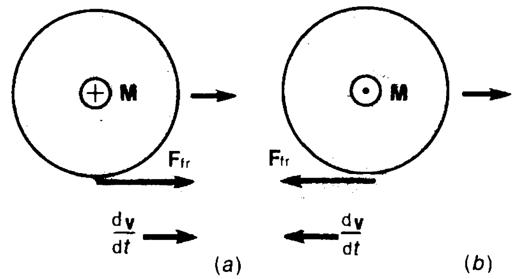


Fig. 105. Diagram of forces acting on the wheel of self-propelled means of transport.

moment of force ${}_1$ which tends to accelerate it. Hence the resultant moment of forces decelerates the rotating wheel. Under the action of the rolling friction, the kinetic energy is transformed into internal energy through inelastic deformation.

Thus, the rolling friction and the moment of forces retarding the rotation of the wheel emerge as a result of inelastic deformation of the wheel and the rolling surface in the vicinity of their contact. The effect of the rolling friction and the moment of forces can easily be taken into account. Only the determination of the rolling friction and the moment of forces is a difficult task. This is usually done experimentally, and their values are supplied in appropriate tabular form.

SELF-PROPELLED MEANS OF TRANSPORT. Two new questions arise in the analysis of motion of motorcars, locomotives and other self-propelled means of transport: How are they accelerated from rest and how are they brought to rest? It is sufficient to consider the motion of a wheel to study these problems. If the motion takes place without sliding of the wheels, there is no sliding friction. Rolling friction is always present and acts in the way described above. However, rolling friction does not play a significant role in starting up or stopping of carriages. Static friction plays a main role in this case.

When a vehicle is set into motion, the engine applies a moment of force M to the axle of the wheel (Fig. 105*a*). However, the static friction F_{fr} at the points of contact between the wheel and the road impedes the rotation of the wheel. This results in the static friction in the direction of motion.

The situation is reversed when brakes are applied. The moment of forces of brake blocks is directed in such a way (Fig. 105*b*) that the additional static friction resulting in this case is directed against the velocity of the vehicle. This additional static friction adds up to the static friction ensuring the rolling of the wheel without sliding when there are no internal moments of force acting on its axle.

If the total static friction during the interaction of the wheel and the road together with the additional static friction mentioned above exceeds the maximum static friction, the wheels slide. Hence the sliding of the wheels occurs when one wishes to start up a car too fast, or when one tries to apply brakes suddenly. In both cases, the skidding of the vehicle during a rapid take-off or braking may lead to sorrowful consequences. But even if nothing untoward takes place, a rapid take-off or braking will not be realized. As a matter of fact, as the relative velocity of sliding surfaces increases, the sliding friction decreases slightly in most cases in comparison with the maximum static friction. Hence when a wheel slides, the maximum take-off or braking force is smaller than in the absence of sliding. Hence the fastest take-off or braking is possible only when there is no sliding. An experienced driver always feels the grip of the wheels on the road and never allows the wheels to slide.

It was mentioned above that under certain conditions the frictional force begins to increase as a result of an increase in the sliding velocity. Obviously, in this case one can try to increase the tractive force of the wheels by increasing the velocity of skids. However, the lateral sliding force cannot be balanced even in this case.

ON THE NATURE OF FRICTIONAL FORCES. The emergence of frictional force is due to many processes occurring in the surface layers of the bodies in contact. The investigation of frictional forces is of considerable theoretical and practical significance and forms a separate branch of physics. These processes are reduced to intermolecular interactions in the regions where the bodies come in contact.

Dry friction emerges in the tangential plane of two bodies pressed against each other as a result of their relative motion. Static friction is due to very small (up to about 1 μm) relative displacement of the surfaces rubbing against each other. The region of displacements within which static friction is observed is called the region of preliminary displacements. When the force applied to a body exceeds a certain value, the preliminary displacement is transformed into sliding, and the frictional force slightly decreases in comparison with the maximum static friction.

The surfaces of real bodies are not perfectly smooth. They have waviness and roughness. Hence when two bodies touch each other, the contact does not take place over the entire surface, but only over certain regions called "spots" which are located at "protrusions" on the surfaces. At the "contact spots", forces of adhesion act between the surfaces. The total area of contact regions is about 2-3 orders of magnitude smaller than the apparent area of contact. As the surfaces slide past each other, the "contact spots" are destroyed and recreated, but their total area remains practically the same. The "spots" acquire an elongated form in the direction of motion. Stresses in the region of "spots" acquire very large values and are only a few times smaller than the theoretical values of the strength of a material.

Thus, it can be stated that friction appears as a result of a multitude of processes of interaction between contact surfaces of bodies. Heating of bodies and wear of surfaces are important practical consequences of friction.

PROBLEMS

8.1. A body of mass m is thrown from the origin $\mathbf{r} = 0$ of the coordinate system at an angle α to the horizontal plane at an initial velocity $\mathbf{u}$ at the instant $t = 0$. The force of friction of the body against air is $\mathbf{F}_{\mathrm{fr}} = -mk\mathbf{v}$. Find $\mathbf{r} = \mathbf{r}(t)$. The acceleration of free fall is $\mathbf{g}$. Find the equation of the trajectory if the X-axis is directed along the horizontal and the Y-axis along the vertical.

8.2. A stone is thrown horizontally from a high tower at an initial velocity u. The force of friction of the stone against air is $-mk\mathbf{v}$. Find the height of the tower if the stone lands at a distance l from the base.

8.3. The edges of a homogeneous rectangular parallelepiped of mass m are directed parallel to the axes of a rectangular Cartesian coordinate system whose origin coincides with one of the vertices of the parallelepiped. The length of the edges of the parallelepiped along the X-, Y- and Z-axes is a, b and c respectively. Find the moments of inertia J_{xx} and J_{yz}.

8.4. Two homogeneous right circular cylinders having the same radius and the same mass m are at rest on a horizontal surface, their axes being parallel and their surfaces being out of contact. A straight beam of mass M lies on top of the cylinders in a direction perpendicular to their length. The centres of mass of the cylinders and the beam lie in the same vertical plane. A force F acts on the beam parallel to its length along the line passing through its centre of mass. Find the acceleration of the beam if there is no sliding between the surfaces in contact.

8.5. A homogeneous rigid hemisphere of radius r rests on a horizontal surface in contact with a point on its convex side. Find the frequency of the small-amplitude oscillations of the hemisphere about the equilibrium position in the case when there is no friction and no sliding between the hemisphere and the horizontal surface at the point of their contact.

8.6. A homogeneous cube of mass m_1 and edge l, and a homogeneous circular cylinder of diameter d and length l both rest on a horizontal platform. The axis of the cylinder is horizontal, and its lateral surface is in contact with one of the faces of the cube all along its side. The platform is gradually inclined about an axis parallel to the line of contact between the cube and the cylinder towards the cube. The angle of inclination between the surface of the platform and the horizontal plane is α, and the coefficient of friction at each contact is k. For what value of α will the cube start sliding down the inclined plane?

8.7. A homogeneous sphere of mass m and radius r_1 rests in the state of unstable equilibrium on a rigidly fastened sphere of radius r_2 $(r_2 > r_1)$. The line joining the centres of the spheres is vertical. After the unstable equilibrium has been violated, the upper sphere of radius r_1 starts to roll without sliding over the surface of the lower sphere of radius r_2. If the angle between the vertical and the line joining the centres of the spheres is θ, find the value of θ for which the pressure at the point of contact between the spheres vanishes.

8.8. A body is thrown vertically upwards at an initial velocity u. The frictional force is $F_{fr} = -mkv$. After how much time will the body come to a halt?

8.9. A circular cylinder rolls without sliding down an inclined plane forming an angle α with the horizontal. What is the acceleration of the centre of mass of the cylinder along the inclined plane if the latter moves with an acceleration a (a) in the vertical direction, and (b) in the horizontal direction towards the rising plane?

8.10. A body falls freely in a medium, its velocity v being zero at the instant $t = 0$. The drag of the medium is $-m(av + bv^2)$, where a and b are positive constants. Find the velocity of the body at the instant t and the distance covered by it during this time.

8.11. A homogeneous disc of mass m is suspended in the vertical plane at points A and B on the circumference of the disc by two vertical strings. Points A and B are on the same horizontal level, and the arc AB of the circle forms an angle 2α at the centre of the disc. Find the tension of one of the strings immediately after the other string snaps.

8.12. A homogeneous rectangular parallelepiped with edges $2a$, $2b$ and $2c$ rotates at an angular velocity ω about an axis parallel to the principal diagonal but passing through a vertex not lying on this principal diagonal. Find the kinetic energy of the parallelepiped.

ANSWERS

8.1. $x = gt^2/2 + (u - g/k)(1 - e^{-kt})/k, y = (g/k^2)\ln[1 - kx/(u\cos\alpha)] + x[\tan\alpha + g/(ku\cos\alpha)]$. **8.2.** $(g/k)\ln[u/(u - kl)] - gl/u$. **8.3.** $m(b^2 + c^2)/3, \; -mbc/4$. **8.4.** $F/(M + 3m/4)$. **8.5.** $\sqrt{120g/(119r)}$, $\sqrt{15g/(26r)}$. **8.6.** $\arctan\{km_1/[m_1 + (1 - k)m_2]\}$. **8.7.** $\arccos(10/17)$. **8.8.** $(m/k)\ln[1 + ku/(mg)]$. **8.9.** $(2/3)(g + a)\sin\alpha$, $(2/3)(g\sin\alpha + a\cos\alpha)$. **8.10.** $(A/b)\tanh B - a/(2b)$, $(1/b)\ln(\cosh B/\cosh\alpha) - at/(2b)$, where $A = \sqrt{bg + a^2/4}$, $\tanh\alpha = a/(2A)$, $B = ct + \alpha$. **8.11.** $mg/(1 + 2\sin^2\alpha)$. **8.12.** $m\omega^2(b^2c^2 + 7a^2c^2 + 7a^2b^2)/[3(a^2 + b^2 + c^2)]$.

Chapter 9

Dynamics of Bodies of Variable Mass

Basic idea:

Reaction is directly proportional to the velocity of ejection of combustion products from the nozzle of a rocket and to their mass varying with time. Mass consumption can be reduced by increasing the velocity of ejection of combustion products, while an increase in reaction at a constant velocity of ejection of combustion products can be reached by increasing their mass consumption per second.

Sec. 37. NONRELATIVISTIC ROCKETS

The equation of motion for reaction propulsion is derived and the physical meaning of the basic quantities involved is discussed.

REACTION PROPULSION. The thrust of rocket engines is created by ejecting the fuel combustion products in a direction opposite to that of the force. The force appears in accordance with Newton's third law of motion and is therefore called reaction, while the rocket engines are called jet engines. However, it must be emphasized that every engine creating a thrust is essentially a jet engine. For example, the thrust of an ordinary propeller-type plane is the reaction produced as a result of the acceleration of a mass of air by the propeller in a direction opposite to that in which the plane flies. The thrust of a propeller-type plane is the force with which the masses of air thrown back by the propeller act on the plane. This force is applied to the propeller which is rigidly connected to the plane. A railway train starts moving under the action of the reactive thrust created by the acceleration of rails and the Earth in the opposite direction if the motion is considered in the inertial reference frame relative to fixed stars. Of course, the motion of the rails and the Earth cannot be practically observed in view of their enormous mass and a vanishingly small acceleration.

However, there is a significant difference between the reaction propulsion of rockets and other types of force. The thrust of a rocket engine is created by the ejection of combustion products which form a part of the mass of the rocket until the

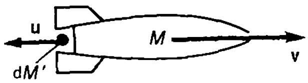


Fig. 106. Deriving the equation of motion of a rocket.

instant of their ejection. This is not so in other cases. For example, the air dispelled by the propeller of a plane never forms a part of the plane. Hence, when we speak of reaction propulsion, we mean the situation that is realized in a rocket engine. *This means that we are considering the motion of a body of variable mass, the thrust being created by ejecting a part of the mass of the body.*

EQUATION OF MOTION. Suppose that a rocket having a mass $M(t)$ at the instant t and moving at a velocity v throws away a mass dM' at a velocity u (Fig. 106). It should be borne in mind that M and dM' are relativistic masses, and the velocities v and u are measured relative to the inertial reference frame in which the motion is considered (and not relative to the rocket).

The mass conservation law has the form

$$dM + dM' = 0. \tag{37.1}$$

Obviously, $dM < 0$ since the mass of the rocket decreases. At the instant t, the total momentum of the system is $M\mathrm{v}$, while at the instant $t + dt$, it is given by the formula $(M + dM) \times (\mathrm{v} + d\mathrm{v}) + \mathrm{u}\,dM'$. The momentum conservation law for this isolated system can be written in the form

$$(M + dM)(\mathrm{v} + d\mathrm{v}) + \mathrm{u}\,dM' = M\mathrm{v}. \tag{37.2}$$

This leads to the equality

$$M\,d\mathrm{v} + \mathrm{v}\,dM + \mathrm{u}\,dM' = 0. \tag{37.3}$$

The term $d\mathrm{v}\,dM$ has been neglected since it is an infinitesimal of the second order of smallness. Taking (37.1) into consideration, we arrive at the equation of motion:

$$\boxed{\frac{d}{dt}(M\mathrm{v}) = \mathrm{u}\frac{dM}{dt},} \tag{37.4}$$

which is valid for both the relativistic and nonrelativistic cases.

If the velocities are small, they can be added by using the formula from classical mechanics, and hence u can be represented in the form

$$\mathrm{u} = \mathrm{u}' + \mathrm{v}, \tag{37.5}$$

where u' is the velocity of the ejected mass relative to the rocket. Substituting (37.5) into (37.4) and differentiating the left-hand side of (37.4) with respect to time, we obtain

$$\boxed{M\frac{d\mathrm{v}}{dt} = (\mathrm{u} - \mathrm{v})\frac{dM}{dt} = \mathrm{u}'\frac{dM}{dt}.} \tag{37.6}$$

Konstantin Eduardovich Tsiolkovsky (1857-1935)
Soviet scientist and inventor. The father of modern astronautics, he was the first to substantiate the possibility of using rockets for interplanetary flights and to indicate rational ways of developing astronautics and rocket building. He obtained a number of technical solutions in the design and construction of rockets and liquid-fuel rocket engines.

This is the equation describing the motion of rockets at nonrelativistic velocities in the absence of external forces.

If the rocket is subjected to the action of a force $\mathbf{F}$, (37.6) assumes the form

$$M\frac{d\mathbf{v}}{dt} = \mathbf{F} + \mathbf{u}'\frac{dM}{dt}. \tag{37.7}$$

Let us denote the consumption of fuel per second by μ. Obviously, $\mu = -dM/dt$. Hence (37.7) can also be written in the form

$$M\frac{d\mathbf{v}}{dt} = \mathbf{F} - \mu\mathbf{u}'. \tag{37.8}$$

The quantity $\mu\mathbf{u}'$ is called the reactive force. If the direction of $\mathbf{u}'$ is opposite to that of $\mathbf{v}$, the rocket is accelerated. If the two directions coincide, the rocket is retarded. For any other relation between $\mathbf{u}'$ and $\mathbf{v}$, the rocket's velocity changes not only in magnitude but also in direction.

TSIOLKOVSKY FORMULA. Let us consider the acceleration of a rocket moving in a straight line, assuming the velocity of the ejected gases to be constant relative to the rocket. In this case, (37.6) can be written in the form

$$M\frac{dv}{dt} = -u'\frac{dM}{dt}. \tag{37.9}$$

The minus sign on the right-hand side is due to the fact that in the case of acceleration, the velocity $\mathbf{u}'$ is directed against the velocity $\mathbf{v}$. Let us denote the velocity and mass of the rocket before acceleration by v_0 and M_0. In this case, writing (37.9) in the form

$$\frac{dM}{M} = -\frac{dv}{u'} \tag{37.10}$$

and integrating this equation, we obtain

$$\ln M - \ln M_0 = -\frac{v - v_0}{u'}. \tag{37.11}$$

This is the Tsiolkovsky formula which can be written in either of the following forms:

$$v - v_0 = u' \ln\frac{M_0}{M}, \tag{37.12a}$$

$$M = M_0 e^{-(v - v_0)/u'}. \tag{37.12b}$$

Formula (37.12a) expresses the change in the velocity of the rocket when its mass changes from M_0 to M, while formula (37.12b) expresses the mass of the rocket whose velocity changes from v_0 to v. When the rocket is accelerated from the state of rest, $v_0 = 0$.

The most important problem is to attain the maximum velocity with the minimum consumption of fuel, i.e. for the minimum difference between M_0 and M. It can be seen from (37.12a) that this can be done only by increasing the velocity u' of ejection of gases. However, the velocities of ejected gases are restricted. For example, velocities of about 4-5 km/s are attained for the gases ejected after the combustion of different types of fuel.

MULTISTAGE ROCKETS. All the load carried by a rocket is not useful right up to the end of the flight. For example, the fuel tanks are required only as long as they contain fuel. After this, when the fuel contained in them has been consumed, they are not only a useless, but also a harmful load since they complicate maneuvers and subsequent acceleration or deceleration of the rocket. Other parts and components of the rocket which become useless after fuel consumption also constitute a harmful load. Hence it is expedient to get rid of such a load as soon as possible. This can be done with the help of multistage rockets, first proposed by Tsiolkovsky (see Example 37.1).

If the rocket engines throw out mass in portions and not continuously, but at the same relative velocity, the efficiency of the rocket engines deteriorates, i.e. for fixed initial and final masses of the rocket, the final velocity decreases with increasing mass of the separately thrown-out installments, considering that each portion (installment) is ejected instantaneously.

CHARACTERISTIC VELOCITY. The concept of characteristic velocity is convenient for discussing various problems connected with space flights. Suppose that it is required to change the velocity of a rocket (acceleration, deceleration, change in direction of flight). In the reference frame, in which the rocket is at rest at the given instant of time, the problem can be reduced to imparting a certain velocity v to the rocket in a direction which would ensure the desired maneuver. The fuel consumption in this maneuver for a rocket flying outside the gravitational field can be taken into account with the help of formula (37.12b) with $v_0 = 0$, where M_0 is the mass of the rocket before the maneuver. The velocity v required to be imparted to the rocket is called the characteristic velocity of the maneuver. We can also introduce the concept of characteristic velocity for maneuvers carried out by applying external forces (gravitational force, air drag, etc.). In this case, the initial and final masses will be connected through a more complicated relation than (37.12b). However, it can be represented as before in the form (37.12b) (for $v_0 = 0$) by assuming this formula to be the definition of the characteristic velocity for the maneuver under consideration. As a crude approximation, the characteristic velocity can be taken equal to the characteristic velocity of the maneuver without any external force in order to obtain a rough estimate. It follows from the law of multiplication of exponents that the characteristic velocity of a complex maneuver consisting of several maneuvers is the sum of the characteristic velocities of the maneuvers comprising the complex maneuver. The concept of characteristic velocity makes it more convenient to describe some important features of interplanetary flights.

In order to raise a body above the Earth's gravity, it must be imparted a velocity of about 11.5 km/s (escape velocity). For a rocket, the velocity in formulas (37.12) should have such a value (for $v_0 = 0$) under the assumption that the fuel is consumed very rapidly and the velocity is acquired by the rocket in the immediate vicinity of the Earth's surface. Formulas (37.12) can be used to determine the part of a starting rocket that will ultimately fly into space. Assuming that the velocity of ejection of gases $u' \simeq 4$ km/s, we find that $M \approx M_0 e^{-3} \approx M_0/20$, i.e. about 5% of the total mass of the rocket will fly into space. In actual practice, the rocket gathers speed much more slowly than we assumed. This makes the situation even worse since it means a larger fuel consumption. In order to reduce the fuel consumption for accelerating a rocket in the gravitational field of the Earth, we must decrease the time of acceleration, i.e. we must increase the acceleration to the maximum possible extent. This involves considerable overloads. Hence it becomes necessary to select certain optimal conditions.

When a rocket returns from outer space, we can make use of aerodynamic braking, i.e. decrease the velocity in the Earth's atmosphere. However, the velocity can also be decreased by switching on the rocket engine. In this case, the velocity of 11.5 km/s must be decreased to zero for a soft landing. This is the characteristic velocity of return to the Earth. Hence the characteristic velocity of flight into space beyond the gravitational field of the Earth and of return without the use of aerodynamic braking is 23 km/s. One can ask as to what fraction of the initial mass will return to the Earth from such a flight. In accordance with formula (37.12b) we obtain the answer: $M \approx M_0 e^{-6} \approx M_0/400$.

The velocity required for overcoming the Moon's attraction

?

If a hole is drilled at the bottom of a water-filled bucket, a stream of water will flow through it. Can the force appearing as a result of flowing water be called reaction? And what if a hole is drilled in the wall of the bucket?

What factors determine the thrust of a rocket engine?

What is the characteristic velocity of a space flight?

is about 2.5 km/s. Hence the characteristic velocity for landing on the Moon and taking off from its surface is about 5 km/s, and the characteristic velocity for flight to the Moon and return to the Earth is estimated at about 28 km/s. This estimate, however, does not take into account the need for other maneuvers, and hence the above value should be increased appropriately. On the other hand, one can make use of the aerodynamic braking while returning to the Earth, and hence this quantity can be reduced to some extent. Consequently, it can be stated that the characteristic velocity for flight to the Moon does not differ significantly from 28 km/s. The characteristic velocity for flight to Mars and Venus is somewhat higher. Assuming that $u' \simeq 4$ km/s, the mass of the rocket returning to the Earth after a flight to the Moon will be about 1/1500 of the starting mass. Although this is a very rough estimate, it gives a fairly good idea of the potentialities of rockets with chemical fuel.

Example 37.1. **Find the optimal parameters for a two-stage rocket in order to impart a given velocity to a given payload. By way of an example, we can take the payload's mass $M = 500$ kg, the final velocity $v = 8$ km/s, and the velocity of ejected gas $u = 2$ km/s. Assume that according to the constructional requirements, the mass of each stage of the rocket without fuel is 10% of the mass of the fuel in this stage.**

We denote the masses of the fuel in the first and second stages of the rocket by M_1 and M_2. The total initial mass of the rocket is $m_0 = M + 1.1(M_1 + M_2)$. After the fuel in the first stage has been consumed, the remaining mass of the rocket is $m_1 = M + 1.1M_2 + 0.1M_1$, and the velocity of the rocket is given by

$$v_1 = u \ln \frac{m_0}{m_1}. \tag{37.13}$$

The second stage comes into operation after a load of mass $0.1M_1$ has been discarded, i.e. the acceleration starts when the mass of the rocket is $m_2 = M + 1.1M_2$ and terminates after the consumption of fuel in the second stage when the mass of the rocket is $m_3 = M + 0.1M_2$. As a result, the velocity attained by the rocket is given by

$$v_2 = v_1 + u \ln \frac{m_2}{m_3} = u \ln \frac{m_0 m_2}{m_1 m_3}. \tag{37.14}$$

It can be seen from this equation that $m_0 m_2/(m_1 m_3) = e^{v_2/u} = e^4$. Taking into account the equality $1.1m_1 = 0.1m_0 + m_2$, we can transform the above relation to the form

$$\frac{11 m_0 m_2}{(m_0 + 10 m_2) m_3} = e^4 .$$

This leads to the relation

$$m_0\left(\frac{11}{m_3} - \frac{e^4}{m_2}\right) = m_0\left(\frac{11}{M + 0.1M_2} - \frac{e^4}{M + 1.1M_2}\right) = 10e^4. \tag{37.15}$$

The mass m_0 has its minimum value when the expression within the parentheses has its maximum value as a function of M_2. The extremum condition for the expression within the parentheses gives the equality

$$\frac{M + 1.1M_2}{M + 0.1M_2} = e^2, \tag{37.16}$$

whence we get

$$M_2 = \frac{M(e^2 - 1)}{1.1 - 0.1e^2} = 8800 \text{ kg}. \tag{37.17}$$

Hence $m_0 = 209{,}400$ kg and $M_1 = 181{,}000$ kg.

Sec. 38. RELATIVISTIC ROCKETS

The equation of motion for a relativistic rocket is derived and the possibility of its practical realization is discussed.

EQUATION OF MOTION. While deriving Eq. (37.4), we noted that it is valid for low as well as high velocities. In the relativistic case, a mass M must be considered relativistic, i.e.

$$M = \frac{M'}{\sqrt{1 - v^2/c^2}}, \tag{38.1}$$

where M' is the variable rest mass of a rocket (we denote it by a primed letter to indicate that this is the mass in the moving reference frame fixed to the rocket). The rest mass of the rocket decreases during its motion. Taking this into account, we can write (37.4) for the relativistic case in the following form:

$$\frac{d}{dt}\left(\frac{M'\mathrm{v}}{\sqrt{1 - v^2/c^2}}\right) = \mathrm{u}\frac{d}{dt}\left(\frac{M'}{\sqrt{1 - v^2/c^2}}\right). \tag{38.2}$$

If necessary, we can easily take into account the external forces acting on the rocket. Let us transform this equation into the form of (37.6). For this purpose, we differentiate the left-hand side with respect to t and transfer the term proportional to v to the right-hand side. As a result, we arrive at the relation

$$\frac{M'}{\sqrt{1 - v^2/c^2}} \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = (\mathbf{u} - \mathbf{v}) \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{M'}{\sqrt{1 - v^2/c^2}}\right). \tag{38.3}$$

This relation is identical to (37.6) with relativistic mass $M = M'/\sqrt{1 - v^2/c^2}$. *However, the difference* $\mathbf{u} - \mathbf{v}$ *in* (38.3) *is not the velocity of ejection of gases relative to the rocket since we must use formula* (17.6) *for the velocity addition in the relativistic case.*

DEPENDENCE OF FINAL MASS ON VELOCITY. In order to obtain a formula similar to the Tsiolkovsky formula in the relativistic case, we must solve Eq. (38.3). We shall assume that the acceleration takes place in the positive X-direction. Then (38.3) assumes the form

$$\frac{M'}{\sqrt{1 - v^2/c^2}} \frac{\mathrm{d}v}{\mathrm{d}t} = (u_x - v) \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{M'}{\sqrt{1 - v^2/c^2}}\right). \tag{38.4}$$

From formula (17.6) for the velocity addition, we arrive at the following relation of the ejected gases relative to the rocket:

$$u'_x = \frac{u_x - v}{1 - vu_x/c^2}. \tag{38.5}$$

Further, we consider that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{M'}{\sqrt{1 - v^2/c^2}}\right) = \frac{1}{\sqrt{1 - v^2/c^2}} \frac{\mathrm{d}M'}{\mathrm{d}t}$$

$$+ \frac{M'}{c^2} \frac{v}{(1 - v^2/c^2)^{3/2}} \frac{\mathrm{d}v}{\mathrm{d}t}. \tag{38.6}$$

Substituting (38.6) into the right-hand side of (38.4), we obtain after simple transformations

$$\frac{M'}{1 - v^2/c^2}\left(1 - \frac{vu_x}{c^2}\right) \frac{\mathrm{d}v}{\mathrm{d}t} = (u_x - v) \frac{\mathrm{d}M'}{\mathrm{d}t}. \tag{38.7}$$

Expressing the quantity $u_x - v$ in terms of the velocity u'_x with the help of (38.5) and cancelling out the common factor $1 - vu_x/c^2$, we can represent the relativistic equation of motion in the following simple form:

$$M' \frac{\mathrm{d}v}{\mathrm{d}t} = \left(1 - \frac{v^2}{c^2}\right) u'_x \frac{\mathrm{d}M'}{\mathrm{d}t}. \tag{38.8}$$

Note that the velocity of ejection of gases must be directed against the motion of the rocket in order to accelerate it, i.e.

$u'_x = -u'$, where u' is the magnitude of this velocity. We can now represent (38.8) in a form similar to Eq. (37.10):

$$\frac{\mathrm{d}M'}{M'} = -\frac{1}{u'}\frac{\mathrm{d}v}{1 - v^2/c^2}. \tag{38.9}$$

Suppose that the mass of the rocket at the initial instant is M'_0 and its initial velocity is v_0. As in (37.10), we integrate both sides of this equation within appropriate limits. Since

$$\frac{1}{1 - v^2/c^2} = \frac{1}{2}\frac{1}{1 - v/c} + \frac{1}{2}\frac{1}{1 + v/c},$$

the integral of the right-hand side with respect to v is an elementary integral. As a result of integration, we obtain

$$\ln M' - \ln M'_0 = -\frac{c}{2u'}\left[\ln\left(1 + \frac{v}{c}\right) - \ln\left(1 - \frac{v}{c}\right)\right]_{v_0}^{v}$$

$$= -\frac{c}{2u'}\left[\ln\frac{1 + v/c}{1 - v/c} - \ln\frac{1 + v_0/c}{1 - v_0/c}\right].$$

Hence

$$\ln\frac{M'}{M'_0} = -\frac{c}{2u'}\ln\frac{(1 + v/c)(1 - v_0/c)}{(1 - v/c)(1 + v_0/c)},$$

or

$$\frac{M'}{M'_0} = \left[\frac{(1 + v/c)(1 - v_0/c)}{(1 - v/c)(1 + v_0/c)}\right]^{-c/(2u')}. \tag{38.10}$$

This formula for the relativistic case replaces formulas (37.12) for nonrelativistic rockets. It assumes an especially simple form for analysis if $v_0 = 0$, i.e. if the rocket is launched from the state of rest:

$$\boxed{M' = M'_0\left(\frac{1 - v/c}{1 + v/c}\right)^{c/(2u')}} \tag{38.11}$$

For low final velocities ($v \ll c$), this formula is transformed into (37.12b) for the nonrelativistic case (if $v_0 = 0$). Indeed, let us write the right-hand side of (38.11) in the following form for $v/c \ll 1$ and $u'/c \ll 1$:

$$\left(\frac{c + v}{c - v}\right)^{-c/(2u')} \approx \left[\left(1 + 2\frac{v}{c}\right)^{c/(2v)}\right]^{-v/u'} = \mathrm{e}^{-v/u'}, \tag{38.12}$$

where we have considered that

$$\frac{c + v}{c - v} = \frac{1 + v/c}{1 - v/c} \approx \left(1 + \frac{v}{c}\right)\left(1 + \frac{v}{c}\right) \approx 1 + 2\frac{v}{c},$$

$$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = \mathrm{e}.$$

Suppose that the rocket has to be accelerated to the velocity $c/2$ by means of chemical fuel, when $u' = 4$ km/s. Which part of the initial mass will be accelerated in this case? Considering that $c = 3 \times 10^5$ km/s, we obtain from (38.11) that

$$M' = M'_0 \left(\frac{1/2}{3/2}\right)^{3 \times 10^5/(2 \times 4)} \approx \frac{M'_0}{3^{(3/8) \times 10^5}} \approx \frac{M'_0}{10^{2 \times 10^4}}. \tag{38.13}$$

It is not possible to imagine the number $10^{20{,}000}$, hence there can be no question of accelerating rockets to relativistic velocities by means of chemical fuel.

However, the situation is not much better in the case of other fuels either. For nuclear rockets powered by fission energy, $u' \simeq 10^4$ km/s. In this case, we obtain instead of (38.13)

$$M' = \frac{M'_0}{3^{3 \times 10^5/(2 \times 10^4)}} \approx \frac{M'_0}{3^{15}} \approx \frac{M'_0}{10^6}, \tag{38.14}$$

i.e. only about a millionth part of the initial mass of the rocket attains the final velocity $c/2$.

Hence one can expect more or less promising results on the attainment of relativistic velocities only if u' is close to the velocity of light. This points towards the idea of creating rocket propulsion with the help of photon radiation. Such rockets, which belong only to the realm of fantasy at present, are called photon rockets.

PHOTON ROCKETS. For photon rockets, $u' = c$, and hence (38.11) assumes the form

$$M' = M'_0 \left(\frac{1 - v/c}{1 + v/c}\right)^{1/2} \tag{38.15}$$

It can be seen from this formula that a mass $M' = M'_0/\sqrt{3}$, i.e. more than half the initial mass, could be accelerated to the velocity $c/2$. Thus, these rockets could be highly effective. Suppose that v differs from the velocity of light by a very small quantity, say, by 10^{-4}, i.e. $v/c \simeq 1 - 10^{-4}$. In this case, we obtain from (38.15)

$$M' \approx M'_0 \frac{10^{-2}}{\sqrt{2}}, \tag{38.16}$$

which is quite a reasonable result. However, from the engineering point of view, photon rockets are just a fantasy. But this does not mean that the "reaction of radiation" is of no importance. On the contrary, it plays a very significant role in nature, e.g. in many astrophysical phenomena.

PROBLEMS

9.1. A raindrop begins to fall at the instant $t = 0$ in air containing water vapour at rest. As a result of condensation of water vapour on the drop, its mass increases during its fall according to the law $m = m_0 + \alpha t$. The moving drop is under the action of the frictional force $\mathbf{F}_{fr} = -mk\mathbf{v}$. Find the velocity $\mathbf{v}(t)$ of the drop.

9.2. Two passenger vehicles having a mass M each move on a horizontal surface without friction at an initial velocity v_0. Snow is falling vertically, and the mass of snow falling per second on a vehicle is a. In one of the vehicles, the snow is accumulated as it falls. In the other vehicle, the passengers immediately throw the snow out in a direction perpendicular to the motion of the vehicle. What distance will be covered by the vehicles in time t?

ANSWERS

9.1. $g/k - g[(\alpha - m_0 k)e^{-kt}]/[k^2(m_0 + \alpha t)]$.

9.2. $\dfrac{Mv_0}{a}\ln\left(\dfrac{M + at}{M}\right)$, $\dfrac{Mv_0}{a}[1 - e^{-(a/M)t}]$.

Chapter **10**

Collisions

Basic idea:
The knowledge of the result of a collision process and not of the process itself is of principal interest for discussing collisions. The theory aims at establishing the relation between the characteristics of the state of particles before and after a collision without answering the question as to how these relations occurred. Conservation laws do not govern collision processes, but are only observed in them.

Sec. 39. DESCRIPTION OF COLLISION PROCESSES

A method for analyzing collision processes by means of conservation laws is described.

DEFINITION OF COLLISION. The interaction of bodies is the most frequently encountered phenomenon in nature. When billiard balls approach one another, they interact at the moment of coming in contact. This results in a change in the velocities of the balls, their kinetic energies and, in general, their internal states as well (for example, their temperatures). Such interactions of billiard balls are referred to as collisions.

But the concept of collision is not confined just to interactions resulting from the contact of bodies. A comet flying from distant regions in the galaxy and passing in the vicinity of the Sun changes its velocity and again departs for distant regions. This is also a collision process, although there was no direct contact between the comet and the Sun, and the interaction took place through gravitational forces. A peculiar feature of this interaction, which enables us to consider it a collision, is that the region of space over which this interaction occurred is comparatively small. A significant change in the comet's velocity occurs in the region near the Sun. This is a large region when compared with terrestrial distances, but is quite small on the astronomical scale and, in particular, in comparison with the distances from the far-off regions from where the comet presumably came. Hence the process of collision of the comet with the Sun looks like this: for a long time, over which the comet has covered an enormous distance, it moved without interacting with the Sun; then, over a small range of the order of just a few hundred million kilometres, the comet interacts with the Sun, and as a result of this interaction, the velocity and some other characteristic parameters of the comet change; after this, the comet again takes off to distant regions, moving practically without any interaction with the Sun.

By way of another example, we can consider the collision of a proton with a nucleus. When the distance between the two is large, they move uniformly and in a straight line, practically without an interaction. At quite small distances, Coulomb's repulsive forces become quite large, and this leads to a change in the velocities of the proton and the nucleus. Quanta of electromagnetic radiation may be emitted in the process, and if their energy is quite high, the formation of other particles like mesons, or a decay of the nucleus, may be observed. Hence, as a result of this interaction, which also takes place over a comparatively small region of space, the proton and the nucleus will move in the simplest case at velocities and energies that are different from their values before the collision. Some quanta of electromagnetic radiation will be emitted and, in general, some other particles will be created.

On the basis of the above examples, we can give the following definition of collision: *Collision is the interaction of two or more bodies, particles, etc. which occurs over a comparatively small region of space in a comparatively small interval of time so that outside this region of space and beyond this interval of time we can speak of the initial states of the bodies, particles, etc. and of their final states after the collision as states in which these bodies, particles, etc. do not interact.*

!

We are not interested in what occurs in the region of colliding particles. It is important to know the relation between the characteristics of the colliding particles before and after a collision.

A collision of bodies is often called an impact. An impact is defined as a process in which the momenta of the colliding bodies change without any change in their coordinates. This is a special case of collision. Wherever appropriate, this term can be used to replace the term "collision".

?

What is the general definition of collision?
What is common between the collision of elementary particles, billiard balls and the passage of a comet near the Sun?
What is meant by the initial and final states?
What is an impact?
Give definitions of an elastic and an inelastic collision.

In mechanics, bodies and particles participating in a collision are characterized by their momenta, angular momenta and energies, while the process itself is reduced to a variation of these quantities. It can be stated that particles exchange energy and momentum. If new particles are formed as a result of collision and some of the particles that existed before the collision disappear, the carriers of energy and momentum change.

DIAGRAMMATIC REPRESENTATION OF COLLISION PROCESSES. At present, it is customary to represent collision processes in the form of diagrams. Particles or bodies participating in collisions are represented by their momentum vectors. The momentum vectors of particles before and after a collision are symbolically shown pointing towards the collision

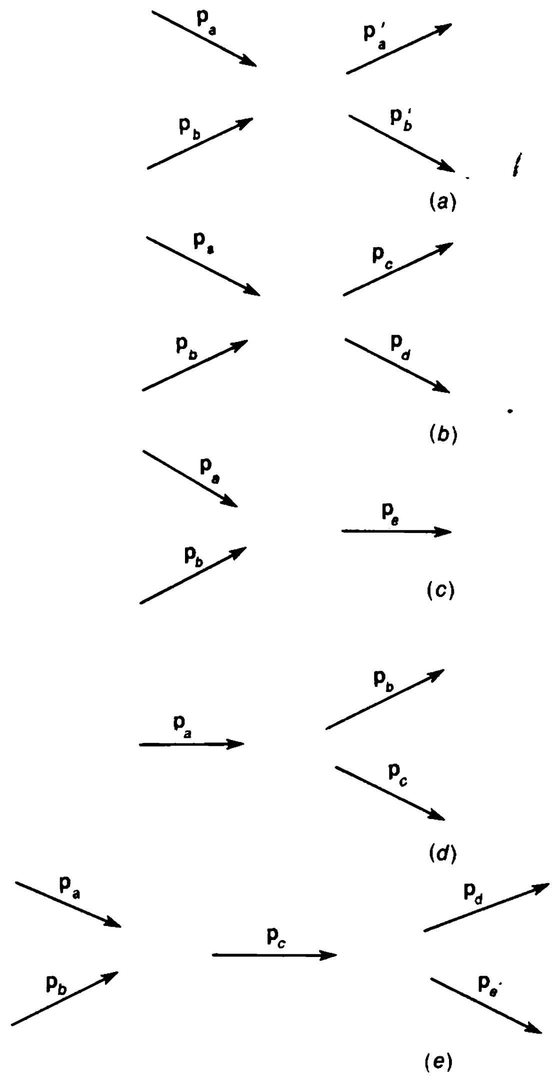


Fig. 107. Diagram of various collision processes.

region and away from it respectively. Obviously, there is a huge variety of collision processes. Some of the most characteristic collision processes are shown in Fig. 107. Figure 107*a* represents the collision of two particles *a* and *b* having momenta p_a and p_b respectively. The same particles exist after the collision, but their momenta have now changed to p'_a and p'_b. However, new particles *c* and *d* could be formed instead of the particles *a* and *b* as a result of the collision (Fig. 107*b*). Alternatively, the collision of the particles *a* and *b* could lead to the formation of a single particle e (Fig. 107c). It may so happen that one particle a disintegrates into two particles b and c as a result of some processes occurring inside it (Fig. 107d). There is no need to represent all imaginable diagrams of collisions. Let us just consider the possibility of a process which differs basically from all those considered above. In this process, an intermediate state is formed (Fig. 107e) and the collision process consists of two stages: first the particles a and b interact to form a particle c (called the intermediate particle) which then disintegrates into two particles d and e. In general, these particles could be the same as a and b, but they may as well be different. Thus, the final result of this process is equivalent to the collisions shown in Fig. 107a and b. The existence of an intermediate state, however, generally affects the process.

CONSERVATION LAWS AND COLLISIONS. Collisions are extremely complicated processes. Let us consider, for example, the simplest case of a collision of two billiard balls (see Fig. 107a). When the balls come in contact, they are deformed. As a result, a part of the kinetic energy is transformed into the potential energy of deformation (we speak of a part of the kinetic energy since the collision is not necessarily head-on). After this, the energy of elastic deformation is again transformed into the kinetic energy, but only partially since some of the energy is transformed into the internal energy, and the balls are heated as a result of the collision. Further, it must be noted that the surfaces of the balls are not absolutely smooth and frictional forces emerge between them. On the one hand, these forces also cause a part of the energy to be converted into the internal energy and, on the other hand, lead to a change in the rotation of the balls. Thus, even in the simplest case of collision, the picture is quite complicated.

However, the main interest in the study of collisions lies in the result of the collision and not in the knowledge of the process itself. The situation before the collision is called the initial state, and after the collision, the final state. Irrespective of the detailed nature of the interaction, certain relations are observed between the quantities characterizing the initial and final states.

The existence of these relations is due to the fact that the aggregate of particles participating in a collision forms an isolated system which obeys the energy, momentum and angular momentum conservation laws (see Chap. 6). Consequently, the relations between the quantities characterizing the initial and final states of a particle are expressed through the energy, momentum and angular momentum conservation laws upon a collision.

Conservation laws themselves do not provide an idea of what happens as a result of collision. However, if we know what happens during a collision, the conservation laws considerably simplify the analysis of the way in which the process occurs.

MOMENTUM CONSERVATION LAW. Let us denote the momenta of various particles before a collision by $\mathbf{p}_i$ ($i = 1, 2, \ldots, n$), and the momenta of particles after the collision by $\mathbf{p}'_j$ ($j = 1, 2, \ldots, k$). Since the momentum of a closed system is conserved, we can write

$$\sum_{i=1}^{n} \mathbf{p}_i = \sum_{j=1}^{k} \mathbf{p}'_j. \tag{39.1}$$

Obviously, the number and type of particles before and after a collision may be different. The momentum conservation law is valid both for the relativistic and nonrelativistic case.

ENERGY CONSERVATION LAW. The application of this law is more difficult than of the momentum conservation law. As a matter of fact, the energy conservation law was formulated (see Chap. 6) only for forms of energy encountered in mechanics. Hence in the nonrelativistic case, we have to take into consideration only the kinetic and potential energies, while in the relativistic case, we must consider the rest energy as well. However, there do exist other forms of energy as well which have to be taken into account. For example, when two billiard balls collide, they are slightly heated. Hence the sum of the kinetic energies of the balls before and after the collision is not the same, i.e. kinetic energy is not conserved during the collision. A part of this energy is transformed into the internal energy which is associated with the heat and is localized within the ball. There are other forms of internal energy as well. The mutual potential energy of the particles constituting the ball and their rest energy also belong to the internal energy. Hence, in order to apply the energy conservation law, we must take into account the internal energy of the particles or bodies participating in the collision. However, the potential energy of interaction between the colliding particles need not be taken into account since they are assumed to be noninteracting before and after the collision. Denoting the internal energy of the particles and the kinetic energy of translational motion of the body by E_{int} and E_{k} respectively, we can write the energy conservation law during the collision in the form

$$\sum_{i=1}^{n} (E_{\text{int}.\,i} + E_{\text{k}.\,i}) = \sum_{j=1}^{k} (E'_{\text{int}.\,j} + E'_{\text{k}.\,j}). \tag{39.2}$$

Note that it is convenient to treat the kinetic energy of rotational motion as the internal energy.

In the relativistic case, (39.2) assumes a much simpler form. As a matter of fact, the relativistic total energy of a body expressed in the form (26.10) includes the kinetic energy as well as the rest energy containing all forms of internal energy. For example, if a billiard ball is heated as a result of collision, this will lead to an increase in the rest mass and will be automatically taken care of by an appropriate change in its total energy. Hence in the relativistic case, (39.2) can be written in the form

$$\sum_{i=1}^{n} E_i = \sum_{j=1}^{k} E'_j, \tag{39.3a}$$

where

$$E_i = \frac{m_{0i}c^2}{\sqrt{1 - v_i^2/c^2}} \tag{39.3b}$$

is the total energy of the ith particle with a rest mass m_{0i}. Taking (39.3b) into account, we can represent (39.3a) in the following form:

$$\sum_{i=1}^{n} \frac{m_{0i}}{\sqrt{1 - v_i^2/c^2}} = \sum_{j=1}^{k} \frac{m'_{0j}}{\sqrt{1 - v_j'^2/c^2}}. \tag{39.4}$$

ANGULAR MOMENTUM CONSERVATION LAW. While applying the angular momentum conservation law, we must remember that bodies and particles may have intrinsic angular momentum. In bodies, this is associated with their rotation. Microparticles also have intrinsic angular momentum, called spin. For example, electrons, protons and many other elementary particles possess spin. It was described earlier that the existence of spin cannot be explained by the rotation of elementary particles. It should be considered the intrinsic angular momentum of the particle during collisions. Hence, if we denote the angular momenta of particles participating in a collision by L_i, and their intrinsic angular momenta by $\mathrm{L}_{\mathrm{int}.i}$, the angular momentum conservation law during the collision can be represented as follows:

$$\sum_{i=1}^{n} (\mathrm{L}_i + \mathrm{L}_{\mathrm{int}.i}) = \sum_{j=1}^{k} (\mathrm{L}'_j + \mathrm{L}'_{\mathrm{int}.j}). \tag{39.5}$$

ELASTIC AND INELASTIC COLLISIONS. Collision processes are divided into two types, viz. elastic and inelastic, depending on the nature in which the internal energy of particles changes as a result of their interaction. *If the internal energy of particles changes during an interaction, the collision is called inelastic; if it remains unchanged, the collision is said to be elastic.* For example, a collision of billiard balls during which the balls are heated to a certain extent is an inelastic collision since the internal energy has changed as a result of the collision. However, if the billiard ball is made of a suitable material (say, ivory), its heating is insignificant and the change in the rotational motion is negligibly small. Under this assumption, the impact of billiard balls can be treated as an elastic collision. *Sometimes, one speaks of a perfectly elastic collision in order to emphasize that the internal energy of the colliding particles is absolutely invariable. One also speaks of a perfectly inelastic collision if the entire energy is transformed into the internal energy in the final state.* For example, the head-on collision of two balls having the same mass and made of a soft material is considered to be a perfectly inelastic collision if the balls merge into a single body at rest after the collision.

CENTRE-OF-MASS SYSTEM. Introducing the centre-of-mass system, we can considerably simplify the analysis of collision phenomena (see Sec. 21). In this system, the energy and angular momentum conservation laws have the same form as (39.3) and (39.5), while the momentum conservation law having the form (39.1) can be represented in a simpler form since, by definition, the sum of the momenta of particles in a centre-of-mass system is zero:

$$\sum_{i=1}^{n} \mathbf{p}_i = \sum_{j=1}^{k} \mathbf{p}'_j = 0. \tag{39.6}$$

Sec. 40. ELASTIC COLLISIONS

Properties of elastic collisions are described and certain examples are considered.

COLLISION OF TWO PARTICLES IN A NONRELATIVISTIC CASE. Let us choose the coordinate system in such a way that one of the particles, say, the second, is at rest before collision, i.e. $\mathbf{p}_2 = 0$. In this case, the energy and momentum conservation laws can be written in the form

$$\frac{p_1^2}{2m_1} = \frac{p_1'^2}{2m_1} + \frac{p_2'^2}{2m_2}, \tag{40.1}$$

$$\mathbf{p}_1 = \mathbf{p}'_1 + \mathbf{p}'_2, \tag{40.2}$$

where the kinetic energy is expressed in terms of the momentum $[mv^2/2 = p^2/(2m)]$, and it is considered that the

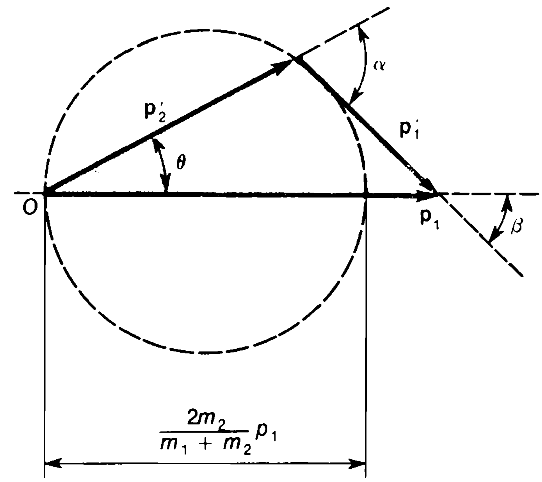


Fig. 108. Graphical solution of the collision problem for two particles ($m_1 > m_2$).

internal energy does not change in an elastic collision. Substituting the value of $\mathbf{p}_1' = \mathbf{p}_1 - \mathbf{p}_2'$ from (40.2) into (40.1), we obtain

$$\mathbf{p}_1 \cdot \mathbf{p}_2' = \mathbf{p}_2'^2\left(\frac{m_1 + m_2}{2m_2}\right). \tag{40.3}$$

Let us denote the angle between $\mathbf{p}_1$ and $\mathbf{p}_2'$ by θ. In this case, $\mathbf{p}_1 \cdot \mathbf{p}_2' = p_1 p_2 \cos\theta$, and (40.3) leads to the following expression for p_2', thus providing a complete solution of the problem under consideration:

$$p_2' = 2\left(\frac{m_2}{m_1 + m_2}\right)p_1 \cos\theta. \tag{40.4}$$

With the help of a simple geometrical construction, we can now describe the result of the collision. We draw vector $\mathbf{p}_1$ from a certain point O to represent the momentum of the incident particle (Fig. 108). After this, we draw a circle of radius $2[m_2/(m_1 + m_2)]p_1$ with its centre lying on the straight line coinciding with vector $\mathbf{p}_1$ in such a way that the circle passes through point O. Since the angle of a triangle inscribed in a circle and having its diameter as the base is $\pi/2$, all segments joining point O to points on the circle satisfy Eq. (40.4). Consequently, these segments describe the momentum after the collision of the particle that was at rest before the collision. It immediately follows from the momentum conservation law (40.2) that the momentum of the incident particle after the collision can be determined with the help of the drawing shown in Fig. 108. The angle between the momenta of the first and second particles after the collision is denoted by

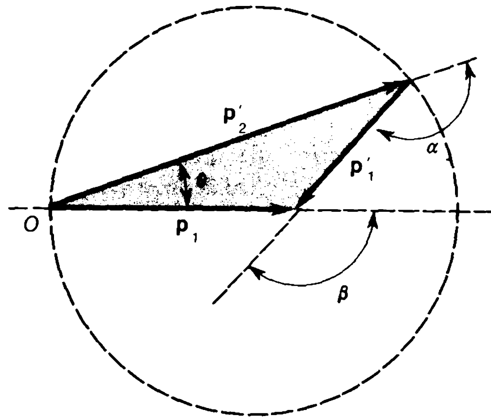


Fig. 109. Graphical solution of the collision problem for two particles ($m_1 < m_2$).

?

Under what condition does the angle of divergence between two particles lie between 0 and $\pi/2$ after an elastic collision?
When can a particle incident on a target be deflected at an arbitrary angle after an elastic collision? Under what conditions is scattering at an arbitrary angle impossible?
What are the factors determining the magnitude of the energy transferred by a moving particle to a target upon an elastic collision?
What is the condition for the most effective slowing down of fast particles as a result of elastic collisions?
What type of collisions leave the frequency of γ-quanta practically unchanged in the Compton effect?

α, while β is the angle by which the incident particle deflects from its initial direction before the collision. The value of p'_1 can also be determined geometrically without any difficulty. Thus, all quantities characterizing the collision have completely been determined. Figure 108 represents the case when $2m_2/(m_1 + m_2) < 1$, i.e. when the mass of the incident particle is larger than the rest mass ($m_1 > m_2$) of the particle called the target. It is clear from this figure that the angle α of divergence between the particles after the collision varies from $\pi/2$ to 0. The momentum p'_1 will have its maximum value when the target moves after the collision nearly at right angles to the velocity of the incident particle. Note that this particle cannot change its direction by an arbitrary angle, and there exists an angle β_{max} beyond which the direction of motion of the particle cannot change. In Fig. 108, this angle corresponds to the case when the line representing p'_1 is tangential to the circle.

Figure 109 represents geometrically the case of collision when the mass of a target is larger than the mass of an incident particle ($m_2 > m_1$). It can clearly be seen from the figure that the angle of divergence between the particles after the collision varies in the interval $\pi/2 < \alpha < \pi$. The angle β of deflection of the incident particle from its initial direction of motion varies from 0 to π, i.e. the particle may suffer an insignificant deflection or reverse its direction after the collision.

In each of the cases considered above, all the characteristics of the collision are determined by the angle θ. But what is the value of this angle in a particular collision? This question cannot be answered with the help of the conservation laws. Everything depends on the conditions under which the collision occurs and on the peculiarities of the interaction. Hence the conservation laws themselves do not provide a complete solution to the collision problem but can be used to analyze only its main features.

HEAD-ON COLLISION. It can be seen from Figs. 108 and 109 that the stationary particle receives the maximum momentum as a result of the collision if $\theta = 0$. In this case, the collision is called head-on, or central. Such a situation arises, for example, when two billiard balls approach each other along the line joining their centres (this line must not change its direction in space in an inertial coordinate system).

In this case, it follows from (40.4) that

$$\mathrm{p}'_2 = \left(\frac{2m_2}{m_1 + m_2}\right)\mathrm{p}_1. \tag{40.5}$$

The kinetic energy of the second particle after the collision, i.e. $E'_{k2} = p'^2_2/(2m_2)$, can be expressed in terms of the kinetic energy of the first particle before the collision, i.e. $E_{k1} = p_1^2/(2m_1)$, through the formula

$$E'_{k2} = \left[\frac{4m_1 m_2}{(m_1 + m_2)^2}\right] E_{k1}, \tag{40.6}$$

as can easily be seen from (40.5). It follows hence that the maximum transfer of energy takes place when the bodies have the same mass ($m_1 = m_2$). In this case,

$$E'_{k2} = E_{k1}, \tag{40.7}$$

i.e. the entire energy of the first particle is imparted to the second particle, and the first particle comes to a stop. This is obvious from the energy conservation law (40.7), as well as from Eq. (40.5) which assumes the form $\mathrm{p}'_2 = \mathrm{p}_1$, and together with the momentum conservation law (40.2) leads to the equality $\mathrm{p}'_1 = 0$.

When the masses of the colliding particles differ considerably, the transferred energy is very small. From (40.6), we obtain

$$E'_{k2} \approx 4\left(\frac{m_2}{m_1}\right) E_{k1} \quad \text{for} \quad m_1 \gg m_2, \tag{40.8a}$$

$$E'_{k2} \approx 4\left(\frac{m_1}{m_2}\right) E_{k1} \quad \text{for} \quad m_2 \gg m_1, \tag{40.8b}$$

i.e. in both cases $E'_{k2} \ll E_{k1}$. However, a considerable momentum transfer takes place. It can be seen from (40.5) that if the mass of the incident particle is much smaller than that of the stationary particle ($m_1 \gg m_2$), the momentum of the stationary particle after the collision, i.e. $\mathrm{p}'_2 \approx (2m_2/m_1)\mathrm{p}_2$, will be much smaller than that of the incident particle, but its velocity will not differ significantly from that of the incident particle. Considering that $p_2' = m_2 v_2'$ and $p_1 = m_1 v_1$, we obtain the following relation for the velocities:

$$v_2' = 2v_1 . \tag{40.9}$$

For $m_2 \gg m_1$, the transfer of momentum from the first particle to the second is quite significant, i.e. $p_2' \approx 2p_1$. However, although the momentum of the second particle is twice as large as that of the first particle, its velocity is very small in comparison with that of the first particle, i.e. $v_2' \approx (2m_1/m_2)v$. As a result of the collision, the velocity of the first particle is reversed in direction, but its magnitude remains practically unchanged.

MODERATION OF NEUTRONS. The peculiar features of elastic collisions find many practical applications. As an example, let us consider the moderation of neutrons. During the fission of a uranium nucleus, a considerable amount of energy is released in the form of kinetic energy of the fission fragments. Simultaneously, between two and three (2.3, on the average) neutrons are also produced as a result of fission. The fission of a uranium nucleus is caused by neutrons. When a neutron collides with a uranium nucleus, the collision is elastic in most cases, but sometimes it is captured by the nucleus, as a result of which the fission of the uranium nucleus takes place. The probability of capture is very low and increases with a decreasing neutron's energy. Hence, in order to ensure a quite intensive chain reaction, i.e. to ensure that the neutrons released as a result of fission of a uranium nucleus cause an intensive fission of other uranium nuclei, we must decrease the kinetic energy of neutrons. According to formula (40.8), each elastic head-on collision of a neutron with a uranium nucleus involves a transfer of only a small part (about 4/238) of the neutron's energy to the nucleus. This is a very small amount, and the neutrons are slowed down at a very low rate. In order to increase the rate of slow-down, a special material called moderator is introduced in the reactor's core where nuclear fission takes place. Obviously, the nuclei of the moderator must be very light. For example, graphite is used as a moderator. Carbon nuclei which constitute graphite are only 12 times heavier than a neutron. Hence each head-on collision of a neutron with graphite (a carbon nucleus) involves a transfer of about $4/12 = 1/3$ of the neutron's energy to graphite, and the moderation takes place quite rapidly.

COMPTON EFFECT. In the same way, let us consider a collision of two particles having relativistic velocities. Assuming that one of the particles is stationary before the collision, while the other moves at a relativistic velocity, the momentum conservation law will retain its form (39.1), but the energy conservation law (39.2) must be replaced by the total energy conservation law in the form

$$\frac{m_{01}c^2}{\sqrt{1 - v_1^2/c^2}} + m_{02}c^2 = \frac{m_{01}c^2}{\sqrt{1 - v_1'^2/c^2}} + \frac{m_{02}c^2}{\sqrt{1 - v_2'^2/c^2}}. \quad (40.10)$$

We shall not describe the properties of the solution of these equations in the general case because the analysis is very cumbersome. Instead, let us consider a particular case which played a significant role in physics. This is the Compton effect. All material particles have wave as well as corpuscular properties. This means that a particle behaves like a wave in some cases and like a corpuscle in some other cases. Light also possesses similar properties. The corpuscular properties of light are manifested in that the light radiation behaves like an aggregate of particles, viz. photons, under certain conditions. A photon has an energy E_{ph} and a momentum p, which are connected with the frequency ω and wavelength λ of light through the relations

$$\mathrm{p} = \hbar \mathrm{k}, \; E_{ph} = \hbar\omega, \quad (40.11)$$

where $|\mathrm{k}| = 2\pi/\lambda$, and $\hbar = 1.05 \times 10^{-34}$ J·s is Planck's constant. The smaller the wavelength, the more strongly are the corpuscular properties of light manifested. Photons corresponding to wavelengths of the order of 0.1 nm are called γ-quanta. The corpuscular properties of γ-quanta are manifested quite clearly. When γ-quanta collide with electrons, they behave like particles whose energy and momentum are given by (40.11).

Let us consider a collision between a stationary electron and a γ-quantum (Fig. 110). Before the collision, the incident quantum had a momentum $\mathrm{p}_1 = \hbar\mathrm{k}$ and an energy $E_{ph1} = \hbar\omega$. After the collision, the quantum moves at an angle β and its momentum and energy are given by $\mathrm{p}_1' = \hbar\mathrm{k}'$ and $E'_{ph2} = \hbar\omega'$ respectively. The energy and momentum of the electron after the collision are $E_2' = mc^2$ and $\mathrm{p}_2' = m\mathrm{v}$. Before the collision, its energy was equal to the rest energy $E_2 = m_0c^2$ and its momentum p_2 was zero. Let us write down the energy conservation law (40.10) and the momentum conservation law (39.1), taking into account (40.11):

$$m_0c^2 + \hbar\omega = mc^2 + \hbar\omega', \; \hbar\mathrm{k} = \hbar\mathrm{k}' + m\mathrm{v}. \quad (40.12)$$

These equalities can also be written in the form

$$mc^2 = \hbar(\omega - \omega') + m_0c^2, \quad m\mathrm{v} = \hbar(\mathrm{k} - \mathrm{k}').$$

Squaring both sides, we get

$$m^2c^4 = \hbar^2(\omega^2 + \omega'^2 - 2\omega\omega') + m_0^2c^4 + 2\hbar m_0c^2(\omega - \omega'),$$

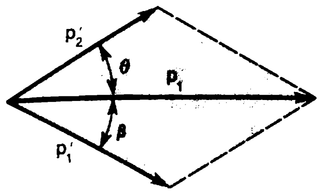


Fig. 110. To the Compton effect.

$$m^2v^2 = \hbar^2(k^2 + k'^2 - 2kk'\cos\beta).$$

Considering that $k = 2\pi/\lambda = \omega/c$, where λ is the wavelength, we multiply the second equality by c and subtract it termwise from the first. This gives

$$m^2c^4(1 - v^2/c^2)$$
$$= m_0^2c^4 - 2\hbar^2\omega\omega'(1 - \cos\beta) + 2\hbar m_0c^2(\omega - \omega'). \tag{40.13}$$

Considering that

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}, \quad 1 - \cos\beta = 2\sin^2\frac{\beta}{2},$$

we obtain from (40.13) the relation

$$\frac{c}{\omega'} - \frac{c}{\omega} = \frac{2\hbar}{m_0c}\sin^2\frac{\beta}{2}. \tag{40.14}$$

The wavelength and the frequency are connected through the relation $c/\omega = \lambda/(2\pi)$. Hence (40.14) can finally be represented in the form

$$\Delta\lambda = \lambda' - \lambda = 2\Lambda\sin^2\frac{\beta}{2}, \tag{40.15}$$

where $\Lambda = 2\pi\hbar/(m_0c) = 2.42 \times 10^{-10}$ cm is called the Compton wavelength of the electron.

Thus, we have found that if a γ-quantum collides with a free electron and is deflected by an angle β, its momentum will vary in accordance with the laws of an elastic collision, and this decrease in the momentum leads to an increase in the wavelength given by (40.15). The change in the wavelength of γ-quanta can be measured directly. Observations made by Compton fully confirmed the validity of formula (40.15). Thus the original assumptions on which the derivation of this formula was based (including formulas (40.11)) were also confirmed experimentally.

Of course, γ-quanta can collide not only with free electrons situated outside the atoms, but also with electrons in the atoms. The result of collision depends on how strongly an electron is bound to the atom. For outer electrons which are situated away from the nucleus and whose attraction by the nucleus is screened by the electric charge of electrons situated closer to the nucleus, this bonding force is very weak. Hence when a γ-quantum collides with an outer electron, the situation is the same as if the electron were not bound to the atom, i.e. as if it were free. As a result of the collision, the electron is detached from the atom, and the γ-quantum is scattered in accordance with (40.15). A different situation is observed when a γ-quantum collides with an inner electron, situated at a small distance from the nucleus and having a quite strong bond with it. The electron cannot be detached from the atom in this case, and the collision practically takes place not with an individual electron, but rather with the atom as a whole. Of course, the conservation laws (40.12) remain valid in this case; however, by m_0 and m we now mean not the mass of the electron, but the mass of the whole atom, which is many thousand times larger than the mass of the electron. For the change in the wavelength of the γ-quantum, we also obtain formula (40.15) in which m_0 now stands for the rest mass of the atom. It follows hence that $\Delta\lambda$ is practically zero, i.e. the γ-quantum does not change momentum as a result of the collision, as indeed should be the case when it collides with a particle of a very large mass.

Hence in Compton's experiments, we observe at any angle both γ-quanta of the same wavelength as that of the incident γ-quanta and γ-quanta whose wavelength increases in accordance with (40.15).

Sec. 41. INELASTIC COLLISIONS

Properties of inelastic collisions are described and some examples are considered.

GENERAL PROPERTIES OF INELASTIC COLLISIONS. The main feature of such collisions is that the internal energy of particles or bodies participating in the interaction undergoes a change. This means that inelastic collisions involve not only a transformation of kinetic energy into potential energy, or vice versa, but also a transformation of the internal energy of one particle into the internal energy of the other particle. The particle or body whose internal energy changes (and hence its internal state also changes) is either converted into a different particle or body, or retains the same identity but in a different energy state. Hence a mutual conversion of particles takes place as a result of inelastic collisions. If, for example, a light quantum is absorbed by an atom, not only does the quantum disappear, but the atom is also transferred into some other energy state. Many nuclear reactions are examples of inelastic collisions.

INELASTIC COLLISION OF TWO PARTICLES. In this case, a part of the kinetic energy of particles must be transformed into internal energy, or vice versa. Of course, the energy and momentum conservation laws are also valid in this case. But they do not provide any indication as to which part of the kinetic energy of the particles is transformed into internal energy, or vice versa. This depends on the specific nature of the collision. The collision may be nearly elastic when only a small part of energy is transformed into internal energy or nearly perfectly inelastic when practically all kinetic energy is transformed into internal energy. Let us imagine that we can change the elastic properties of a stationary body from a purely elastic state to a purely inelastic state when a body incident on it simply sticks to it. We can then follow the collisions for all degrees of "inelasticity". Let us consider a perfectly inelastic collision. In this case, both bodies merge into one as a result of the collision and move like a single entity. Assuming that the second body of mass m_2 was stationary before the collision, we can write the conservation laws in the following form:

$$E_{\text{int}1} + E_{\text{int}2} + E_{\text{k}1} = E'_{\text{int}(1+2)} + E'_{\text{k}(1+2)}, \tag{41.1}$$

$$\mathrm{p}_1 = \mathrm{p}'_{(1+2)}, \tag{41.2}$$

where $E_{\text{int}1}$ and $E_{\text{int}2}$ are the internal energies of the first and the second body before the collision, $E_{\text{k}1}$ is the kinetic energy of the moving body, p_1 is its momentum, while $E'_{\text{int}(1+2)}$, $E'_{\text{k}(1+2)}$ and $\mathrm{p}'_{(1+2)}$ represent the internal energy, the kinetic energy and the momentum of the body obtained as a result of merging of the two bodies after the collision.

If the mass-energy relation is not taken into consideration, (41.2) makes it possible to find the velocity of the body obtained as a result of merging:

$$m_1 \mathrm{v}_1 = (m_1 + m_2)\,\mathrm{v}_2, \tag{41.3}$$

whence

$$\mathrm{v}_2 = \frac{m_1}{m_1 + m_2}\mathrm{v}_1. \tag{41.4}$$

Using these formulas, we can also determine the kinetic energy ΔE_k transformed into internal energy due to the collision:

$$\Delta E_\text{k} = \frac{m_1 v_1^2}{2} - \frac{(m_1 + m_2)v_2^2}{2} = \frac{m_2}{m_1 + m_2}E_{\text{k}1}. \tag{41.5}$$

If the mass of the stationary body is very large ($m_2 \gg m_1$), then $\Delta E_\text{k} \approx E_{\text{k}1}$, i.e. nearly all the kinetic energy is transformed into internal energy. In this case, the body formed by merging of two colliding bodies is practically at rest. If the mass of a body at rest is very small ($m_2 \ll m_1$), then $\Delta E_\text{k} \simeq 0$, i.e. no substantial transformation of the kinetic energy into internal energy occurs.

The body formed by merging of two colliding bodies moves at nearly the same velocity as that of the first body before the collision.

ABSORPTION OF A PHOTON. The absorption of a photon by an atom is a typical example of an inelastic collision which can be described by using the diagram shown in Fig. 98*c*. Before the absorption, we have a photon and an atom, while after the absorption, we are left only with the atom. Assuming that the atom was stationary before absorbing the photon, we can apply the energy and momentum conservation laws to this process by taking into account the relations (40.11) for the photon:

$$M_0 c^2 + \hbar\omega = M' c^2, \tag{41.6a}$$

$$\frac{\hbar\omega}{c} = M' v'. \tag{41.6b}$$

From (41.6a), we obtain the mass of the atom after the absorption of the photon:

$$M' = M_0 + \frac{\hbar\omega}{c^2},$$

while (41.6b) gives the velocity of the atom if we take into account the last equality:

$$v' = \frac{c\hbar\omega}{M_0 c^2 + \hbar\omega}. \tag{41.7}$$

Assuming that the energy of the photon is much smaller than the rest energy of the atom ($\hbar\omega \ll M_0 c^2$), we can represent this formula in a more convenient form:

$$v' \approx c\frac{\hbar\omega}{M_0 c^2}\left(1 - \frac{\hbar\omega}{M_0 c^2}\right) \approx c\frac{\hbar\omega}{M_0 c^2}. \tag{41.8}$$

Thus, after absorbing the photon, the atom acquires a kinetic energy

$$\Delta E_k = \frac{M_0 v'^2}{2} = \frac{(\hbar\omega)^2}{2M_0 c^2}. \tag{41.9}$$

!

When a photon is absorbed by an atom at rest, not all its energy is transformed into the internal energy of the atom. A certain part of the photon's energy is transformed into the kinetic energy of the atom as well.
When an atom emits a photon, it does not transfer to the photon all the internal energy liberated during emission. A certain part of the internal energy is transformed into the kinetic energy of the atom.

This means that not all the photon's energy has been transformed into the internal energy of the atom, the amount of energy not transformed being given by (41.9). This energy ΔE_k was used for imparting a kinetic energy to the atom.

EMISSION OF A PHOTON. The emission of a photon by an atom is also a typical collision process, which is represented diagrammatically in Fig. 98*d*. Such a process is usually called a decay. When a photon is emitted, the internal energy of an atom changes; a part of the energy is transformed into the photon's energy, while another part is used up as the kinetic energy of the atom. This part of the energy is called the recoil energy. Consequently, the energy of the emitted photon is less than the change in the internal energy of the atom by an amount ΔE_k. This energy can be calculated with the help of the energy and momentum conservation laws which have the following form in this case:

$$M_0c^2 = M'c^2 + \hbar\omega', \tag{41.10a}$$

$$0 = \frac{\hbar\omega'}{c} + M'v'. \tag{41.10b}$$

Obviously, ΔE_k is the kinetic energy of the atom after the emission of the photon. From (41.10b), we obtain

$$\Delta E_k = \frac{M'v'^2}{2} = \frac{(\hbar\omega')^2}{2M'c^2}. \tag{41.11}$$

The quantity M' can be found from (41.10a). However, for $\hbar\omega' \ll M_0$, this quantity does not differ significantly from M_0 and there is no need to take into consideration its deviation from M_0. In other words, it can be assumed that (41.11) contains M_0 instead of M'.

Thus, an emitted photon does not carry all the internal energy of the atom emitting it, and not all the energy of a photon absorbed by an atom is transformed into the internal energy of the atom.

Sec. 42. REACTIONS OF SUBATOMIC PARTICLES

Basic concepts associated with the reactions of subatomic particles are described.

THRESHOLD ENERGY. It was mentioned above that inelastic collisions include all the processes of interconversion of particles. Some of these conversions involving photons were considered in the previous section. We shall now discuss some other concepts associated with these processes.

Suppose that particles a and b are transformed into c and d as a result of collision. It is customary to consider collisions in the centre-of-mass system. In this system, the momentum conservation law is reduced to the equality of the sum of the momenta of the particles to zero before and after the collision. For the present, we are not interested in this law. The energy conservation law has the form

$$E_a + E_b + E_{ka} + E_{kb} = E'_c + E'_d + E'_{kc} + E'_{kd}, \tag{42.1}$$

where E represents the internal energies of the particles with the appropriate subscripts, and E_k are their kinetic energies. The quantity

$$Q = E_a + E_b - E'_c - E'_d = E'_{kc} + E'_{kd} - E_{ka} - E_{kb} \tag{42.2}$$

is called the energy of reaction. *This quantity is equal to the total kinetic energy of the particles participating in a reaction, or the change in the internal energy with the opposite sign.* If the kinetic energy of the reaction products is higher than that of the original particles, then $Q>0$. For $Q<0$, the sum of the internal energies of the reaction products is more than the sum of the internal energies of the original particles. Thus, for $Q>0$, internal energy is transformed into kinetic energy, while for $Q<0$, on the contrary, kinetic energy is absorbed and transformed into internal energy.

Let $Q>0$. In this case, the reaction can take place for any kinetic energies of the particles, which may be quite low. For example, the reaction may take place for $Q = 0$ as well.

However, the situation is quite different for $Q<0$. In this case, the sum of the kinetic energies must exceed a minimum value at which the reaction can take place. If this minimum value is not attained, the reaction cannot proceed. Obviously, this minimum is equal to the absolute magnitude $|Q|$. It is called the threshold energy of the reaction. *Thus, the threshold energy of a reaction is the minimum kinetic energy of the reacting particles at which the reaction can proceed.*

ACTIVATION ENERGY. For $Q>0$, a reaction may occur spontaneously for any values of the kinetic energy. However, this does not mean that the reaction does indeed occur. For example, if two protons are brought quite close together, they will interact. This will result in the formation of a deuteron, a positron, a neutrino, and in the liberation of kinetic energy equal to 1.19 MeV. In this reaction, $Q>0$. However, this reaction cannot proceed unless the forces of Coulomb repulsion appearing when the two protons approach each other are overcome.

For this purpose, protons must have a certain minimum kinetic energy which is conserved after a reaction, but does not participate in the reaction and only ensures that it takes place. Hence this energy is called the activation energy.

!

The threshold energy of a reaction is the minimum kinetic energy of the reacting particles for which the reaction can still proceed.
The activation energy is the minimum kinetic energy of the reacting particles which is conserved after the reaction; it does not participate in the reaction, but simply ensures that the reaction does take place.

TRANSITION TO LABORATORY COORDINATE SYSTEM. The activation energy and the threshold energy are determined in the centre-of-mass system. But how can we find the threshold energy in the laboratory system if we know its value in the centre-of-mass system? Obviously, this requires a transition from the centre-of-mass system to the laboratory coordinate system.

Let us consider this transition by taking the example of a collision of two particles. Obviously, in the general case, we should make use of the relativistic formulas. We assign the superscript c.m to quantities pertaining to the centre-of-mass system, and l to those corresponding to the laboratory system. Suppose that in the laboratory system, particle *2* is stationary and particle *1* is incident on it. In the centre-of-mass system, the particles move towards each other. As a result of the collision of these particles, a reaction may occur in which the internal energy of the particles formed is $E_i'^{(c.m)}$ in the centre-of-mass system. The threshold energy of this reaction is Q, while the internal energies of the colliding particles in the centre-of-mass system are $E_1^{(c.m)}$ and $E_2^{(c.m)}$ respectively. Obviously, the condition under which a reaction can take place between these particles in the centre-of-mass system can be written on the basis of (42.2) in the following form:

$$E^{(c.m)} = E_1^{(c.m)} + E_2^{(c.m)} + Q \geqslant \sum_i E_i'^{(c.m)}. \tag{42.3}$$

Obviously, the aggregate of two particles having a threshold energy Q can be considered in the centre-of-mass system to be a single particle with internal energy $E^{(c.m)}$ defined by (42.3). Upon a transition to the laboratory system, this "particle" has a momentum p_1 equal to the momentum of the first particle, which is in motion in this system, and an intrinsic (internal) energy $E^{(c.m)}$. Consequently, as a result of a transition to the laboratory system, $E^{(c.m)}$ in (42.3) is transformed into the energy

$$E^{(l)} = \sqrt{c^2p_1^2 + (E^{(c.m)})^2}\,. \tag{42.4}$$

On the other hand, the total energy of these individual particles can be represented in the form

$$E^{(l)} = \sqrt{c^2p_1^2 + (E_1^{(c.m)})^2} + E_2^{(c.m)}. \tag{42.5}$$

It follows from (42.4) and (42.5) that

$$(E^{(c.m)})^2 = (E_1^{(c.m)})^2 + (E_2^{(c.m)})^2 + 2E_2^{(c.m)}\sqrt{c^2p_1^2 + (E_1^{(c.m)})^2}\,. \tag{42.6}$$

The kinetic energy of the first particle in the laboratory system is

$$E_{k1}^{(l)} = \sqrt{c^2p_1^2 + (E_1^{(c.m)})^2} - E_1^{(c.m)}. \tag{42.7}$$

Finding the value of $\sqrt{c^2p_1^2 + (E_1^{(c.m)})^2}$ from (42.6) and substituting this quantity into (42.7), we obtain

$$E_{k1}^{(l)} = \frac{(E^{(c.m)})^2 - (E_1^{(c.m)})^2 - (E_2^{(c.m)})^2}{2E_2^{(c.m)}} - E_1^{(c.m)}$$
$$= \frac{(E^{(c.m)})^2 - (E_1^{(c.m)} + E_2^{(c.m)})^2}{2E_2^{(c.m)}}. \tag{42.8}$$

With the help of this relation, we can represent (42.3) in the form

$$E_{k1}^{(l)} \geqslant \frac{(\sum E_i'^{(c.m)})^2 - (E_1^{(c.m)} + E_2^{(c.m)})^2}{2E_2^{(c.m)}}. \tag{42.9}$$

This is the required inequality for calculating the threshold energy in the laboratory coordinate system. Let us apply this for determining the threshold of the most familiar reactions involving two protons.

THRESHOLD OF GENERATION OF π^0-MESONS. When two protons collide, a π^0-meson may be created according to the reaction

$$p + p \to p' + p' + \pi^0, \tag{42.10}$$

where p' are the same protons, but with different energies and momenta. The intrinsic energy of the proton is $E_{p0} =$ 980 MeV, and of a π^0-meson is $E_{\pi 0} = 135$ MeV. Hence, using (42.9), we can determine the following value of the threshold energy of the reaction:

$$E_{k1}^{(l)} \geqslant \frac{(2E_{p0} + E_{\pi 0})^2 - (2E_{p0})^2}{2E_{p0}} = 280 \text{ MeV}. \tag{42.11}$$

THRESHOLD OF GENERATION OF PROTON-ANTIPROTON PAIRS. The collision of two protons may also result in the formation of a proton-antiproton pair according to the reaction

$$p + p \to p + p + p + \bar{p}, \tag{42.12}$$

where $\bar{p}$ is the symbol for the antiproton. It has the same intrinsic energy as the proton, and hence we obtain the threshold energy of this reaction with the help of (42.9):

$$E_{k1}^{(l)} \geqslant \frac{(4E_{p0})^2 - (2E_{p0})^2}{2E_{p0}} = 6E_{p0} \simeq 6 \text{ GeV}. \tag{42.13}$$

ROLE OF COLLISIONS IN PHYSICAL INVESTIGATIONS. Investigation of collisions is the principal method for analyzing properties, interactions and structure of atomic and subatomic particles.

When the interaction of macroparticles is considered, it is possible to study the evolution of the process. For example, when two billiard balls collide, it is possible to follow the evolution of the process, the deformation of the shape of the balls after the collision and the transformation of the kinetic energy of the balls into their potential energy of deformation. The interval of time over which this process takes place is very small on the ordinary time scale. However, if we take into consideration the time scales that can be measured by modern experimental techniques, this time interval is extremely large, and a detailed investigation of the process is quite possible. Hence the impact of billiard balls can be considered not only to be a collision, but also a process in which the geometrical and physical properties of the balls undergo a change. This makes it possible to carry out a continuous observation of the chain of events connecting the state of the billiard balls long before the impact with the state long after the impact. An investigation of this process provides information about the physical properties of the balls and their interaction. If this information is of no significance in a particular case (for example, in a billiards game), the impact can be treated as a collision.

A different situation prevails for the phenomena in the physics of atomic and subatomic particles, when the evolution of the process of their interaction cannot be studied experimentally in space and time, and only the result can be investigated. This means that in the physics of atomic and subatomic particles, a collision always has the meaning of the definition given at the beginning of Sec. 39. Studies of the collisions make it possible to verify the theoretical concepts about the collision process and is a principal method for investigating interactions, interconversions, structure and other important properties of microparticles and the processes occurring in the microcosm.

Let us consider certain examples involving the studies of the structure of microparticles. It was assumed at the beginning of this century that the positive charge of an atom and the main part of its mass (pertaining to the positive charge) are "smeared" over the entire volume of the atom whose linear dimensions are of the order of 10^{-8} cm. Electrons move in this positively charged cloud. The total charge of the electrons is equal in magnitude to the positive charge of the atom, and hence the total charge of the atom is zero. Investigations were undertaken to study the collision of alpha particles with atoms. The electric charge and the mass of alpha particles, as well as the forces of interaction between their charges, were already known at that time. According to the theoretical results on the collision of alpha particles with atoms, if alpha particles have a very high energy, collisions with atoms cannot reverse the direction of their velocity or even deflect them in the backward direction. However, such a situation was observed experimentally, and hence it became necessary to change the concept of the structure of the atom. The result was the planetary model of the atom, according to which all the positive charge of the atom and nearly all its mass are concentrated in the nucleus, the electrons revolving around it like planets around the Sun. The size of the atom was estimated at about 10^{-8} cm.

Collisions of electrons of very high energies with protons were studied at the end of the fifties. In principle, these experiments are similar to the ones that led to the planetary model of the atom. The distribution of the electric charge over the volume of the proton was determined with the help of these investigations.

Collisions of electrons of extremely high energies with protons were studied in the seventies. It was found that an electron collides not with a proton as a whole, but with the particles (quarks) constituting it. This led to the quark model for the structure of the proton, which had been predicted theoretically before these experiments.

These examples are but a small part of the fundamental discoveries made in the physics of microcosm with the help of the collision studies.

PROBLEMS

10.1. A stationary particle of mass m_0 decays into two particles of rest mass m_{10} and m_{20}. Find the kinetic energy of the decay products.

10.2. A particle of rest mass m_{01} and energy E impinges on a stationary particle of rest mass m_{02}. Find the velocity of the centre-of-mass system.

10.3. A particle of rest mass m_{01} impinges at a velocity v on a stationary particle of mass m_{02}. The collision is perfectly inelastic. Find the rest mass and the velocity of the particle formed as a result of the collision.

10.4. A particle of rest mass m_0 is scattered at an angle θ after undergoing an elastic collision with a stationary particle of the same mass. Find the kinetic energy of the particle after the scattering if it had an energy E_k before the scattering.

ANSWERS

10.1. $[(m_0 - m_{10})^2 - m_{02}^2]\, c^2/(2m_0)$, $\quad [(m_0 - m_{20})^2 - m_{01}^2]\, c^2/(2m_0)$.

10.2. $c\sqrt{E^2 - m_{01}^2 c^4}/(E + m_{02}c^2)$. 10.3. $(m_{01}^2 + m_{02}^2 + 2m_{01}m_{02}/\sqrt{1 - v^2/c^2})^{1/2}$, $\quad m_{01}v/(m_{01} + m_{02}\sqrt{1 - v^2/c^2})$.

10.4. $E_k \cos^2\theta/[1 + E_k \sin^2\theta/(2m_0c^2)]$.

Chapter **11**

Motion in a Gravitational Field

Basic idea:
Gravitational forces significantly affect the motion only when at least one of the interacting bodies has quite a large (astronomical) mass.

Sec. 43. GRAVITATIONAL FORCES

Gravitational forces and their manifestations are described.

NEWTON'S LAW OF GRAVITATION. This law defines the force F of attraction between point masses m_1 and m_2 separated by a distance r:

$$F = G\frac{m_1 m_2}{r^2}, \tag{43.1}$$

where $G = 6.7 \times 10^{-11}\ \mathrm{N\cdot m^2/kg^2}$ is the gravitational constant. Bodies having a spherically symmetric mass distribution over their volume interact as if their masses were concentrated at the centres of these spheres. The potential energy of the point mass m_2 in the gravitational field of the point mass m_1 can be written in accordance with (25.35a) as follows:

$$E_p = -G\frac{m_1 m_2}{r}. \tag{43.2}$$

But this quantity is the same as the potential energy of mass m_1 in the gravitational field of mass m_2. Hence E_p in (43.2) is the energy of interaction of the point masses m_1 and m_2.

GRAVITATIONAL FIELD NEAR THE EARTH'S SURFACE. Let us denote the radius of the Earth by R_0, and the distance between a point mass m and the Earth's surface by h, where $h \ll R_0$. The total distance between the material point and the centre of the Earth is $R_0 + h$. Hence, in accordance with (43.1), the force of gravity is given by

$$F = \frac{GMm}{(R_0 + h)^2}. \tag{43.3}$$

We assume that

$$\frac{1}{(R_0 + h)^2} = \frac{1}{R_0^2}\frac{1}{(1 + h/R_0)^2} \approx \frac{1}{R_0^2}\left(1 - 2\frac{h}{R_0} + \ldots\right), \tag{43.4}$$

where we have neglected the quadratic and higher-order terms in $(h/R_0)^2$ in view of the fact that the quantity h/R_0 is very small. For example, for distances of the order of 20 km, i.e. the altitudes attainable by aeroplanes, we have $h/R_0 \simeq 3 \times 10^{-3}$. The square of this quantity is a millionth part of unity. In most cases, there is no need to take into account insignificant variations in the force of gravity. For example, when a body falls from an altitude of up to 1 km, the variation in the force of gravity is less than $2(h/R_0) \simeq 3 \times 10^{-4}$. Within this accuracy, we can consider that the force of gravity is constant and independent of the altitude. According to (43.3) and (43.4), we can write this quantity as

$$F_0 = \frac{GMm}{R_0^2} = gm, \tag{43.5}$$

where $g = GM/R_0^2 = 9.8\ \mathrm{m/s^2}$ is the acceleration due to gravity near the Earth's surface. In this approximation, we can analyze the problems associated with the force of gravity near the Earth's surface.

GRAVITATIONAL ENERGY OF A SPHERICAL OBJECT. Suppose that we have a sphere of radius R and mass M. The energy of the gravitational field, or the gravitational energy, is associated with the mutual interaction of the particles constituting this sphere. This energy is numerically equal to the work that must be done in order to scatter the material of the sphere, treated as a continuous medium, over the entire infinite space. Naturally, in this case, we have to take into account the work done in overcoming the forces of gravitational attraction. Forces like the electromagnetic forces that confine atoms in molecules and molecules in solids, liquids, etc. can be neglected.

For the sake of simplicity, we shall assume that the mass of a body is distributed uniformly over the sphere with a density $\rho = 3M/(4\pi R^3)$. It is convenient to remove the substance from the sphere to infinity layer by layer, starting from the surface. The layers that have been removed cannot influence the removal of the subsequent layers in any way since the

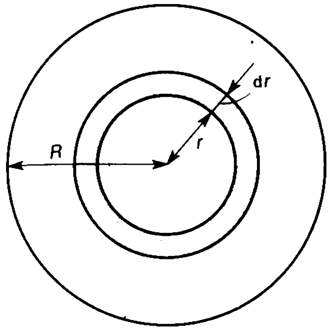


Fig. 111. Computing the gravitational energy of a sphere.

Since the potential energy of a spherical layer of a substance depends only on the internal layers, calculations should be started from the outer layers of a body and terminated at the centre.

subsequent layers are assumed to lie inside the spherical layers that have been removed.

Here, we accept without a proof the statement that a homogeneous spherical layer does not create any gravitational field in the spherical cavity formed by it.

A layer of thickness dr at a distance r from the centre of a sphere (Fig. 111) has a mass $\rho 4\pi r^2\,dr$. When this layer is being removed, it is subjected only to the action of the mass of the sphere of radius r, enclosed within the cavity formed by this layer. The work done in removing this spherical layer is equal to its potential energy in the gravitational field created by all the inner layers:

$$dF_{g.f} = -G\frac{(\rho 4\pi r^3/3)\,\rho 4\pi r^2\,dr}{r}. \tag{43.6a}$$

Integrating this expression over the entire volume of the sphere, i.e. from $r = 0$ to $r = R$, we obtain the total gravitational energy of the sphere:

$$E_{g.f} = -G\frac{16\pi^2}{3}\rho^2\int_0^R r^4\,dr = -G\frac{16\pi^2}{15}\rho^2 R^5. \tag{43.6b}$$

Considering that $\rho = 3M/(4\pi R^3)$, we get

$$E_{g.f} = -\frac{3}{5}G\frac{M^2}{R}. \tag{43.7}$$

This is the gravitational field energy associated with the gravitational attraction of the mass elements comprising the sphere.

GRAVITATIONAL RADIUS. The rest energy of a body of mass M is equal to Mc^2. One can ask if it is possible to visualize this energy as the gravitational field energy transformed into the rest energy when the substance constituting the body is contracted from a scattered state at infinity, where the particles do not interact with one another.

To compute the radius of the sphere to which the substance is contracted from infinity, we equate the gravitational energy to the rest energy (after neglecting the numerical factors):

$$\frac{GM^2}{r} \sim Mc^2.$$

Hence we obtain $r \sim GM/c^2$. The quantity

$$r_g = \frac{GM}{c^2} \tag{43.8}$$

is called the gravitational radius.

?

What is the gravitational energy of a spherically symmetric body?
What is a gravitational radius?
What are the gravitational radii of the Earth and the Sun?
What are "black holes"? What is the evidence in favour of their existence?

By way of an example, let us calculate the gravitational radius of the Earth whose mass is $M = 6 \times 10^{24}$ kg:

$$r_{g.E} = \frac{(6.7 \times 10^{-11})(6 \times 10^{24})\ \text{m}}{(3 \times 10^{8})^2} = 4 \times 10^{-3}\ \text{m} = 0.4\ \text{cm}.$$

This means that the entire mass of the Earth would have to be concentrated in a sphere of diameter of about 1 cm if its gravitational energy were equal to its rest energy. In actual practice, the Earth has a diameter of about 10^9 cm. This indicates that the gravitational energy plays an insignificant role in the overall energy balance of the Earth, including its rest energy. A similar situation prevails on the Sun which has a gravitational radius of about 1 km, while its actual radius is about 700,000 km.

SIZE OF THE UNIVERSE. This, however, is not always the case. For some astronomical objects, the gravitational energy is nearly equal to their rest energy and always plays a significant role. The Universe as a whole can be considered an object of this type.

The average density of distribution of matter over the Universe can be determined by observation, by estimating the mass of astronomical objects and their distance from the Earth. The accuracy of these estimates is not very high since, firstly, there are significant errors in the determination of their distances and, secondly, it is very difficult to take into account the mass of the interstellar gas and nonluminous objects that cannot be observed. At present, it is assumed that the average density ρ has an order of magnitude $\sim 10^{-25}$ kg/m^3. This means that a cubic metre of the interstellar space contains about 100 protons, i.e. the average distance between protons would be about 30 cm if the entire mass of the Universe were distributed uniformly over its volume in the form of protons.

This situation can be visualized as follows. It is well known that the electric charge of a proton is distributed over a volume having linear dimensions of about 10^{-13} m. Hence, if a proton were a pea of about 1 cm in diameter, the average separation between the protons corresponding to their average separation in the Universe would be about twenty times the distance between the Sun and the Earth.

Let us calculate the value of the radius R_0 of a sphere in the Universe for which the rest energy of the mass contained in the sphere is equal to the gravitational energy or, in other words, the radius of the sphere is equal to the gravitational radius of the mass enclosed within this sphere. Since the mass of the sphere is $M \sim \rho_0 R_0^3$, the anticipated condition can be written on the basis of (43.8) in the form

$$R_0 \approx \frac{G\rho_0 R_0^3}{c^2}. \tag{43.9}$$

Hence

$$R_0 \approx \frac{c}{\sqrt{G\rho_0}} \simeq \frac{3 \times 10^8}{\sqrt{6.7 \times 10^{-11} \times 10^{-25}}}\ \mathrm{m} \simeq 10^{26}\ \mathrm{m}.$$

Thus, the required gravitational radius has the same order of magnitude as the quantity which is currently accepted as the radius of the Universe. This means that on the scale of the Universe, gravitation plays a very significant, and at times decisive, role.

"BLACK HOLES". The most important physical meaning of the concept of gravitational radius lies in the idea that the region within a sphere of such a radius as if loses all contact with the region outside it, with the exception of the gravitational interaction. If the entire mass of the Earth could be enclosed within a sphere of diameter 1 cm, the inner regions of this sphere would have no contact with the outer regions except for the gravitational action. This means that light could not emerge from such an inner region and only the enormous gravitational forces would indicate the existence of such a region.

The particles and radiation quanta flying in the vicinity of such a sphere of gravitational radius would be drawn towards it and disappear there. Hence such a region is called a "black hole".

Do "black holes" actually exist in the Universe? It is borne out by theoretical calculations that if the mass of a star is more than twice the mass of the Sun, it will shrink uncontrollably under the action of gravitational forces. At a certain instant, the radius of the star will become equal to its gravitational radius and it will be converted into a "black hole".

"Black holes" have not been discovered so far, but at least one astronomical object that possibly contains a "black hole" can be observed at present.

It can also be assumed that if a rather small mass, say, of a few tons, were somehow confined to a small volume comparable to the value given by (43.8), it would then be transformed into a small "black hole". According to one hypothesis, a few such "black holes" were left from the initial superdense state of the Universe. These objects are called relic "black holes" and have also not been discovered in nature so far.

Apparently, "black holes" do not exist in nature any more since the quantum theory leads to the conclusion about the "evaporation" of matter from "black holes" having a mass of a few tons; consequently, it is believed that such "black holes" have already disappeared.

Sec. 44. MOTION OF PLANETS AND COMETS

Kepler's laws of planetary motion are derived and applied to the motion of planets and comets, as well as to the propagation of light in the gravitational field of the Sun and to interplanetary flights.

Johannes Kepler (1571-1630)
German astronomer who discovered the laws of planetary motion. These laws were used for compiling the tables of planets. He laid the foundations of the theory of eclipses, and invented a telescope in which the objective and the eyepiece are both double-convex lenses.

EQUATION OF MOTION. The mass of the Sun (2×10^{30} kg) is 332,000 times the mass of the Earth (6×10^{24} kg) and about 1000 times the mass of Jupiter, the heaviest planet in the solar system. Hence the Sun can be considered to be stationary to a fairly high degree of accuracy, and the planets can be assumed to be revolving around it. The distance between the Sun and the planets is much larger than the size of the Sun and the planets. For example, the distance between the Sun and the Earth is nearly equal to 150 million kilometres, while the diameters of the Sun and the Earth are about 1.4 million kilometres and 12,700 kilometres respectively.

Thus, while considering the motion of the Earth and other planets around the Sun, we can treat them as point masses with a high degree of accuracy.

We denote the mass of a planet and the Sun by m and M respectively and treat them as a point mass and the centre of force. We take the centre of the Sun as the origin of the coordinate system. The equation of motion for the planet can be written in the form

$$m\frac{d\mathbf{v}}{dt} = -G\frac{mM}{r^2}\frac{\mathbf{r}}{r}, \tag{44.1}$$

where $\mathbf{r}$ is the radius vector of the planet.

MOMENTAL EQUATION. The force acting on a point mass is directed along the radius vector. The moment of this force about the centre of force is zero, and the momental equation (21.4) has the form

$$\frac{d\mathbf{L}}{dt} = \mathbf{M} = \mathbf{r} \times \mathbf{F} = 0. \tag{44.2}$$

Hence the angular momentum of a point mass about the centre of force has a constant magnitude as well as direction:

$$\mathbf{L} = \mathbf{r} \times m\mathbf{v} = \text{const}. \tag{44.3}$$

PLANE OF MOTION. Equation (44.3) can be written in the form

$$m\mathbf{r} \times \dot{\mathbf{r}} = \text{const}. \tag{44.4}$$

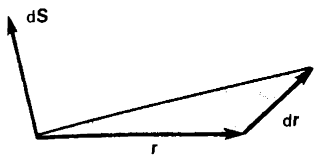


Fig. 112. Representation of an area by a vector perpendicular to the surface on which this area lies.

Hence the elementary displacement $d\mathbf{r} = \mathbf{v}\,dt$ and the radius vector $\mathbf{r}$ lie in a plane perpendicular to $\mathbf{L}$.

This means that the motion takes place in the same plane, i.e. can be called plane motion.

KEPLER'S SECOND LAW. This law states that

the segment joining the Sun to a planet sweeps equal areas in equal intervals of time.

This law follows directly from the angular momentum conservation law (44.3). Indeed, (44.3) can be written in the form

$$\mathbf{r} \times d\mathbf{r} = \frac{\mathbf{L}}{m} dt. \qquad (44.5)$$

Let us find the geometrical meaning of the left-hand side of this equation. It can directly be seen from Fig. 112 that the vector product $\mathbf{r} \times d\mathbf{r}$ is equal in magnitude to twice the area of the triangle formed by vectors $\mathbf{r}$ and $d\mathbf{r}$:

$$|\mathbf{r} \times d\mathbf{r}| = |\mathbf{r}|\,|d\mathbf{r}| \sin(\widehat{\mathbf{r}, d\mathbf{r}}) = r\,dr \sin\alpha = r\,dh = 2dS. \qquad (44.6)$$

Denoting the area of a surface element by a vector perpendicular to this surface (see Fig. 112), we can write $d\mathbf{S} = \mathbf{r} \times d\mathbf{r}/2$ and represent (44.5) in the form

$$d\mathbf{S} = \frac{\mathbf{L}}{2m} dt. \qquad (44.7)$$

Since $\mathbf{L} = \text{const}$, integrating both sides of this equation with respect to time, we obtain

$$\mathbf{S} - \mathbf{S}_0 = \frac{\mathbf{L}}{2m}(t - t_0), \qquad (44.8a)$$

or

$$\Delta S = \frac{L}{2m} \Delta t. \qquad (44.8b)$$

This is Kepler's second law which states that the radius vector of a planet sweeps equal areas in equal intervals of time.

KEPLER'S FIRST LAW. According to this law,

all planets move in elliptic orbits with the Sun as one of the foci.

In order to prove this law, we must find the orbit. It is easier to carry out the calculations in a polar coordinate system whose plane coincides with the plane of the orbit. First of all, we must write the energy and angular momentum conservation laws in polar coordinates. For this purpose, we decompose the elementary displacement $d\mathbf{r}$ into two components:

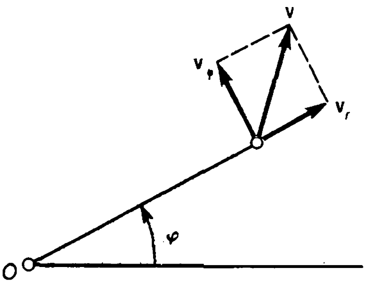


Fig. 113. Decomposition of the velocity vector into two components, one along the radius and the other perpendicular to it, in the polar coordinate system.

$(\mathrm{dr})_\varphi$ perpendicular to the radius r in the polar coordinate system, and $(\mathrm{dr})_r$ along this radius (Fig. 113). The first displacement is associated with the change in angle φ during the motion, while the second one is due to the change in the distance r of the planet from the origin of coordinates. We denote the unit vector perpendicular to r and directed towards increasing φ by e_φ and the unit vector directed towards the increasing radius by e_r. The displacement dr can be expressed through the formula

$$\mathrm{dr} = \mathrm{e}_\varphi(\mathrm{d}r)_\varphi + \mathrm{e}_r(\mathrm{d}r)_r. \tag{44.9}$$

Since $(\mathrm{d}r)_\varphi$ is an elementary arc of a circle of radius r, we can write $(\mathrm{d}r)_\varphi = r\,\mathrm{d}\varphi$; the quantity $(\mathrm{d}r)_r$ is the projection of dr on r i.e. $(\mathrm{d}r)_r = \mathrm{d}r$. Hence (44.9) assumes the form

$$\mathrm{dr} = \mathrm{e}_\varphi r\,\mathrm{d}\varphi + \mathrm{e}_r\,\mathrm{d}r. \tag{44.10}$$

Dividing both sides of (44.10) by the displacement time, we obtain (see Fig. 113)

$$\mathrm{v} = \frac{\mathrm{dr}}{\mathrm{d}t} = \mathrm{e}_\varphi r\frac{\mathrm{d}\varphi}{\mathrm{d}t} + \mathrm{e}_r\frac{\mathrm{d}r}{\mathrm{d}t} = \mathrm{e}_\varphi v_\varphi + \mathrm{e}_r v_r, \tag{44.11}$$

where $v_\varphi = r\,\mathrm{d}\varphi/\mathrm{d}t = r\dot{\varphi}$ and $v_r = \mathrm{d}r/\mathrm{d}t = \dot{r}$. Squaring both sides of this equation and considering that the scalar product of the vectors e_φ and e_r is zero in view of their orthogonality, i.e. $(\mathrm{e}_\varphi, \mathrm{e}_r) = 0$, we obtain the following expression for the square of the velocity:

$$v^2 = v_\varphi^2 + v_r^2 = r^2\dot{\varphi}^2 + \dot{r}^2. \tag{44.12}$$

Substituting the expression $\mathrm{r} = \mathrm{e}_r r$ for the radius vector r and the expression (44.11) for the velocity v into (44.3), we obtain from the law of vector product the expression

$$\mathrm{L} = mr^2\dot{\varphi}\mathrm{e}_r \times \mathrm{e}_\varphi = \mathrm{const}. \tag{44.13}$$

The vector $\mathrm{e}_r \times \mathrm{e}_\varphi$ is a unit vector perpendicular to the plane of motion. This vector determines the direction of vector L. According to (44.13), the conservation law for L has the form

$$\boxed{L = mr^2\dot{\varphi} = \mathrm{const}.} \tag{44.14}$$

The energy conservation law can be written with the help of (44.12) without any additional computations:

$$\frac{mv^2}{2} - G\frac{mM}{r} = \frac{m}{2}(\dot{r}^2 + r^2\dot{\varphi}^2) - G\frac{mM}{r} = \mathrm{const}. \tag{44.15}$$

Thus we have obtained two equations (44.14) and (44.15) in two unknown functions $r(t)$ and $\varphi(t)$, which are sufficient for describing the motion completely. However, we are not interested in finding the course of motion in time, but just in the shape of its trajectory. Hence we eliminate the time dependence from these equations. It follows from (44.14) that $\dot{\varphi} = L/(mr^2)$. Substituting this expression into (44.15), we can eliminate $\dot{\varphi}$. Next, we represent r as a composite function of time: $r(t) = r[\varphi(t)]$. For the sake of convenience, we introduce the function

$$\rho = \frac{1}{r} \tag{44.16}$$

instead of r. This gives

$$\frac{dr}{dt} = \frac{dr}{d\varphi}\frac{d\varphi}{dt} = \frac{d}{d\varphi}\left(\frac{1}{\rho}\right)\frac{d\varphi}{dt} = -\frac{1}{\rho^2}\frac{d\rho}{d\varphi}\frac{L}{mr^2}$$
$$= -\frac{L}{m}\frac{d\rho}{d\varphi}.$$

Substituting this expression for $\dot{r}$ and the expression $1/\rho$ for r in accordance with (44.16) into (44.15), we obtain the equation

$$\left(\frac{d\rho}{d\varphi}\right)^2 + \rho^2 - G\frac{2m^2M}{L^2}\rho = \text{const.} \tag{44.17}$$

Differentiating this equation again with respect to φ, we obtain

$$\frac{d^2\rho}{d\varphi^2} + \rho = C, \tag{44.18}$$

where $C = Gm^2M/L^2 > 0$. The general solution of Eq. (44.18) is well known:

$$\rho = C + A\cos\varphi + B\sin\varphi, \tag{44.19}$$

where A and B are arbitrary constants which must be determined from the initial conditions. The right-hand side of this equation can be transformed as follows:

$$\rho = C + A\cos\varphi + B\sin\varphi$$
$$= C + \sqrt{A^2+B^2}\left(\frac{A}{\sqrt{A^2+B^2}}\cos\varphi + \frac{B}{\sqrt{A^2+B^2}}\sin\varphi\right)$$
$$= C + \sqrt{A^2+B^2}\,(\cos\varphi_0\cos\varphi + \sin\varphi_0\sin\varphi)$$
$$= C + \sqrt{A^2+B^2}\cos(\varphi - \varphi_0)$$
$$= C\left[1 + \frac{\sqrt{A^2+B^2}}{C}\cos(\varphi-\varphi_0)\right]$$
$$= \frac{1}{p}[1 + e\cos(\varphi-\varphi_0)], \tag{44.20}$$

where the following notation has been used: $p = 1/C =$

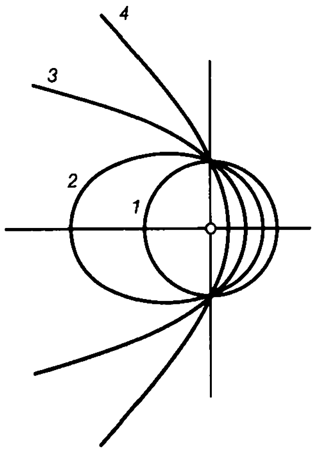


Fig. 114. Various possible trajectories of motion in the gravitational field of a point object: *1*–circle; *2*–ellipse; *3*–parabola; *4*–hyperbola.

$L^2/(Gm^2M)$ and $e = \sqrt{A^2 + B^2}/C$. The angle φ_0 is defined by the expressions

$$\cos \varphi_0 = \frac{A}{\sqrt{A^2 + B^2}}, \qquad \sin \varphi_0 = \frac{B}{\sqrt{A^2 + B^2}}.$$

Thus, the equation of the curve along which a body (planet) moves has the following form in polar coordinates:

$$\frac{1}{\rho} = r = \frac{p}{1 + e\cos(\varphi - \varphi_0)}. \tag{44.21a}$$

It is well known from analytic geometry that this is the equation of a conic section, i.e. the curve formed by the section of a cone by a plane. The quantity p is called the orbital parameter, while e is the eccentricity. The conic section may be an ellipse ($e > 1$), a circle ($e = 0$), a parabola ($e = 1$), or a hyperbola ($e < 1$).

It can be seen from (44.21a) that the distance r from the body assumes its minimum value r_{min} for $\varphi = \varphi_0$. Hence it is convenient to direct the axis of the polar coordinate system through the point closest to the centre of attraction. This point is called perihelion. The point opposite to perihelion on the orbit is called aphelion. When the axis of the polar coordinate system is chosen in this way, we must put $\varphi_0 = 0$ in (44.21a), which thus assumes an even simpler form:

$$r = \frac{p}{1 + e\cos\varphi}. \tag{44.21b}$$

The curves described by this equation are shown in Fig. 114.

Let us consider the motion along an ellipse in greater detail. For the smallest separation r_{min} from the centre of attraction, the body is at $\varphi = 0$, while for the largest separation, the angle φ subtended by the body is π. Hence we can write from (44.21b):

$$r_{min} = \frac{p}{1 + e}, \qquad r_{max} = \frac{p}{1 - e}. \tag{44.22}$$

At the instants of the smallest and largest separations of the body from the centre of attraction, the radial velocity $\dot{r} = 0$. Hence the energy conservation law (44.15) at these points can be written in the following form by taking (44.14) into account:

$$\frac{L^2}{2m}\left(\frac{1}{r_{min}}\right)^2 - G\frac{mM}{r_{min}} = \frac{L^2}{2m}\left(\frac{1}{r_{max}}\right)^2 - G\frac{mM}{r_{max}} = E_0, \tag{44.23}$$

where E_0 denotes the total energy of the body, i.e. the sum of

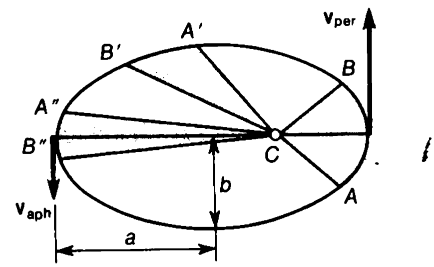


Fig. 115. Planetary motion in an elliptic trajectory.

the kinetic and potential energies. The value of the quantity $p = 1/C$ is obtained from (44.18). Substituting (44.22) into (44.23), we obtain the following relation between the eccentricity e and the energy E_0:

$$e = \left(1 + \frac{2E_0L^2}{G^2m^3M^2}\right)^{1/2}, \tag{44.24a}$$

$$E_0 = -\frac{Gm^3M^2}{2L^2}(1 - e^2). \tag{44.24b}$$

Formula (44.24b) confirms the statement proved earlier, according to which bound states are possible only for negative values of the binding energy, i.e. only when the sum of the kinetic and potential energies is negative. If the total energy is positive, motion in a finite region is impossible.

The particle moves along a hyperbola and goes to infinity. In the limiting case, when the total energy is zero, the particle again goes to infinity, but it follows a parabolic trajectory.

KEPLER'S THIRD LAW. According to this law,

the squares of the periods of revolution of various planets about the Sun are proportional to the cubes of the major semiaxes of their ellipses.

In order to prove this law, let us write down .Eq. (44.8b) connecting the period of revolution T with other characteristics of motion:

$$S = \frac{L}{2m}T, \tag{44.25}$$

where S is the area of the ellipse. It is well known from geometry that the area of an ellipse is $S = \pi ab$, where a and b are its semiaxes (Fig. 115). Formula (44.21b) directly leads to the following expression for the semiaxes in terms of the eccentricity e and the parameter p:

$$a = \frac{p}{1 - e^2}, \quad b = \frac{p}{\sqrt{1 - e^2}}. \tag{44.26}$$

It follows from (44.26) that

$$b^2 = \frac{p^2}{1 - e^2} = pa. \tag{44.27}$$

On the other hand, let us take into consideration the relation between L and p indicated in (44.20):

$$L = m\sqrt{GMp}. \tag{44.28}$$

Let us write down the value of T^2 from (44.25) and make use of (44.28) and (44.27):

$$T^2 = \frac{4m^2S^2}{L^2} = \frac{4m^2\pi^2a^2b^2}{m^2GMp} = \frac{4\pi^2a^3}{GM}. \tag{44.29}$$

This means that the square of the period of revolution of a planet depends only on the major semiaxis and is proportional to its cube. This proves Kepler's third law.

These laws were established by Kepler from an analysis of the planetary motion. This was a significant achievement of the scientific approach and opened the way towards the formulation of the law of gravitation.

MOTION OF COMETS. Comets are small celestial bodies which move near the Sun in strongly prolate elliptic orbits. The existence of some two dozen comets is reliably established and their periodic motion near the Sun has actually been observed. The motion of comets in elliptic orbits with the Sun as one of the foci is analogous to the motion of planets and has been investigated in detail recently.

The most exotic feature of comets is their "tail" which can be seen as a luminous plume repelled by the Sun. The "tail" is formed by the gas which reflects the rays of the Sun. The physical reason behind the "repulsion" of the "tail" by the Sun is the pressure of the electromagnetic radiation corresponding to the visible and invisible parts of the spectrum, as well as the effect of the particle flux (mainly protons) emitted by the Sun. The latter factor is mainly responsible for the "repulsion".

The curvature of a comet's trajectory near the Sun depends on its velocity. The higher the velocity, the smaller the curvature.

REPULSION OF LIGHT RAYS IN THE GRAVITATIONAL FIELD OF THE SUN. The curvature of the trajectories of bodies in a gravitational field raises the question of the effect of this field on light rays. If this action is the same as on a body, then light will not propagate in a straight line in a gravitational field. This hypothesis was put forth long ago, and the

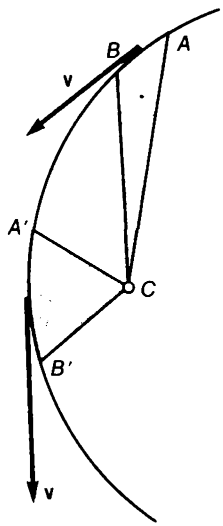


Fig. 116. Motion of a body in a hyperbolic trajectory.

As the body approaches the centre of attraction, its velocity increases.

curvature of a ray in the gravitational field of the Sun was calculated as far back as 1804.

For this purpose, Newton's concepts about the corpuscular nature of light were used. The corpuscles can be treated as point masses having an arbitrarily small mass m and moving at the velocity of light. The mass plays only an auxiliary role in calculations.

In view of the large velocity of corpuscles, their motion near the Sun takes place along a hyperbola with a very small curvature (Fig. 116). Consequently, the initial direction of motion changes by an angle $\Delta\varphi$ which can be calculated (Fig. 117).

In order to do so, we must calculate exactly the trajectory of a point mass, as shown above. We obtained formula (44.19) which can be applied to all types of motion in a gravitational field. Using this formula, we obtain

$$\frac{1}{r} = G\frac{m^2M}{L^2} + A\cos\varphi + B\sin\varphi. \tag{44.30}$$

Constants A and B are determined from the conditions of motion of light rays. In the polar coordinate system shown in Fig. 117, the angle φ decreases with time, varying from π to 0 and then to a negative value $-\Delta\varphi$ equal to the required angle of deflection of a light ray by the Sun.

At the initial instant of time, when the light ray is directed towards the Sun but is at a very large distance from it, we have $\varphi = \pi$, $1/r = 0$, $\cos\varphi = -1$ and $\sin\varphi = 0$. Substituting these values into (44.30), we obtain $A = Gm^2M/L^2$. Consequently, (44.30) assumes the form

$$\frac{1}{r} = G\frac{m^2M}{L^2}(1 + \cos\varphi) + B\sin\varphi. \tag{44.31}$$

In order to determine B, we divide this formula by $\sin\varphi$ and take into account that $1 + \cos\varphi = 2\cos^2(\varphi/2)$ and $\sin\varphi = 2\sin(\varphi/2)\cos(\varphi/2)$. This gives

$$\frac{1}{r\sin\varphi} = G\frac{m^2M}{L^2}\cot\left(\frac{\varphi}{2}\right) + B. \tag{44.32}$$

Let us now use the initial condition once again and make φ tend to π. The distance r_0 at which the light ray would have passed by the Sun in the absence of the force of attraction is called the impact parameter (see Fig. 117).

It can be seen from Fig. 117 that $r\sin\varphi = r_0$ as $\varphi \to \pi$, while $\cot(\varphi/2) = 0$. Hence for $\varphi \to \pi$, we obtain from (44.32) $B = 1/r_0$. The angular momentum L is conserved during the motion. Let us calculate its value at the initial instant of time

!

Bound states can exist only for negative values of the binding energy, i.e. for a negative sum of the kinetic and potential energies. If the binding energy is positive, motion in a finite region is not possible.
The shape of the trajectory of a body moving in a central force field is determined by the total energy of the body.

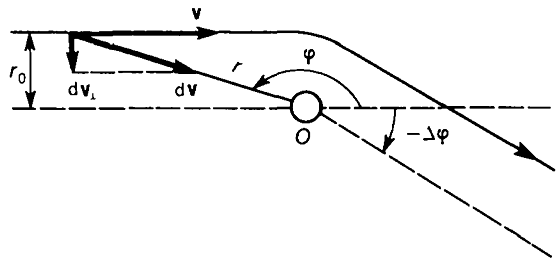


Fig. 117. Calculations of the deflection of light rays approaching the Sun, based on classical concepts.

for $\varphi = \pi$. Obviously, $L = mcr_0$, where c is the velocity of light. Substituting this quantity into (44.32), we finally obtain a formula describing the trajectory of a light ray:

$$\frac{1}{r \sin \varphi} = G \frac{M}{c^2 r_0^2} \cot\left(\frac{\varphi}{2}\right) + \frac{1}{r_0}. \tag{44.33}$$

This formula does not contain the mass m of the particle. This means that the trajectory of a particle in a given gravitational field is independent of its mass.

After deflection, the ray moves away from the Sun to a distance $r \to \infty$ in the direction of the angle $-\Delta\varphi$. Let us determine this angle from (44.33):

$$0 = G \frac{M}{c^2 r_0^2} \cot\left(-\frac{\Delta\varphi}{2}\right) + \frac{1}{r_0}. \tag{44.34}$$

Since the angle $\Delta\varphi$ is very small, we can assume that $\cot(-\Delta\varphi/2) \approx -2/\Delta\varphi$ and hence

$$\Delta\varphi = \frac{2GM}{c^2 r_0}. \tag{44.35}$$

Assuming that the light ray passes close to the Sun and putting r_0 in (44.35) equal to the radius of the Sun, we obtain $\Delta\varphi \simeq 0''.87$. It should be recalled that this value of the angle of deflection was obtained in 1804. However, this value could not be verified experimentally for a long time. Later, the effect of deflection of a ray was also discovered in the general theory of relativity. In this case, however, the deflection was found to be $1''.75$, i.e. twice as large as in the first case. Such a deviation from the predictions is conducive for an experimental verification of the theory. Hence the results of the first measurement of the effect during the solar eclipse in 1919 were awaited eagerly everywhere.

The idea behind the measurements can be described as follows. The position of the stars near the solar disc must be registered on a photographic plate during the solar eclipse. In view of the deflection of a light ray, the position of the stars on the photographic plate will correspond to the apparent separation between these stars and the solar disc. The true position of these stars is known precisely from everyday astronomical observations. Hence the apparent displacement of the stars, which is registered on the photographic plate, can be used to determine the angle of deflection $\Delta\varphi$. The result was found certainly to be closer to the predictions of the theory of relativity than to the predictions of the classical theory. However, the accuracy of measurements was not high enough to dispel all doubts. Most scientists treated these observations as a verification of the theory of relativity, although some scientists still had their reservations on this issue. Other measurements carried out subsequently yielded results still closer to the theory of relativity.

INTERPLANETARY FLIGHTS. A flight around the Earth at a low altitude is the first step towards space research. *The circular (or orbital) velocity v_1 is the velocity of flight in a circular orbit of radius r_E equal to the Earth's radius.* The centripetal acceleration v_1^2/r_E is equal to the acceleration due to gravity g at the surface of the Earth. Hence

$$v_1 = \sqrt{gr_E} = 7.9 \text{ km/s}, \tag{44.36}$$

where $r_E = 6371$ km.

The next step is to overcome the forces of the Earth's attraction. *The escape velocity v_2 is the velocity that a body must possess at the surface of the Earth in order to be able to go beyond the gravitational field of the Earth.* The air resistance is not taken into account. From the energy conservation law $mv_2^2/2 = GmM_E/r_E$, we obtain

$$v_2 = \sqrt{2gr_E} = 11.2 \text{ km/s}, \tag{44.37}$$

where $GM_E/r_E^2 = g$.

The solar escape velocity v_3 is the velocity which must be imparted to a body at a distance equal to the Earth's orbital radius R_E from the Sun so that the body can leave the solar system. This velocity can be determined from the energy conservation law:

$$\frac{mv_3^2}{2} = \frac{GmM_S}{R_E},$$

where M_S is the mass of the Sun. Hence

$$v_3 = \sqrt{\frac{2GM_S}{R_E}} = 42 \text{ km/s}. \tag{44.38}$$

According to another widely used definition, the solar

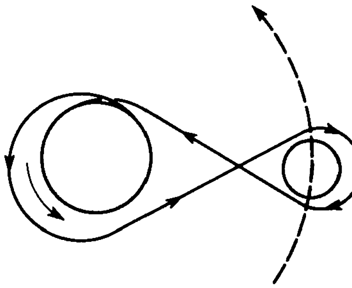


Fig. 118. A typical trajectory of a flight to the Moon and back to the Earth.

Using the linear velocity of the Earth's rotation, the launching and landing of a rocket are carried out in the eastern direction.

?

Prove that the motion in a central force field is a plane motion.
Which conservation law forms the basis of Kepler's second law?
What are the possible trajectories of a point mass in the gravitational field of a point body and under what conditions are they realized?
How do the predictions concerning the deflection of a light ray in the field of the Sun, which were made more than 150 years ago on the basis of the classical theory, differ from the predictions of the general theory of relativity?
What is the experimental state of affairs in this connection?

escape velocity is the velocity that must be possessed by a body near the Earth's surface relative to it in order to enable it to leave the solar system. This definition has a significant drawback in that the numerical value of this velocity is indeterminate. If a spaceship is launched in a direction opposite to the velocity of the Earth's motion around the Sun, the solar escape velocity will be equal to about 72 km/s. If, however, the launching direction coincides with the direction of the Earth's orbital motion, the solar escape velocity will be equal to about 16.5 km/s.

For a flight to any celestial body, we must overcome the force of the Earth's gravity first of all, i.e. we must reach a point where the gravitational fields of the Earth and the celestial body balance each other (we do not take into account the presence of other celestial bodies). The velocity required for this purpose is somewhat smaller than v_2, but this difference is quite small and is not taken into consideration in rough estimates.

A typical trajectory of a flight to the Moon and back to the Earth is shown in Fig. 118. The motion of the Moon during the flight of the rocket is neglected for the sake of clarity of illustration. The actual trajectory is displaced along the Moon's orbit in the course of the flight. It can be seen from the figure that the linear velocity of the Earth's rotation is completely utilized in such a flight: during launching it is added to the velocity of the rocket, while during landing the relative velocity of the spaceship and the Earth's surface is reduced on account of this motion.

For flights to remote planets like Jupiter and Uranus, we must overcome not only the Earth's attraction, but also the attraction of the Sun between the Earth's orbit and the orbit of the planet. In this case, it is expedient to launch the rocket in the direction of the Earth's orbital motion to make use of the kinetic energy. The simplest trajectory of such a flight is shown in Fig. 119.

The motion takes place in an elliptic orbit with the Sun at one of the foci. At the instant of launching, the Earth is at the perihelion of the ellipse, and hence the minimum distance $r_{\min}$ is equal to the radius R_E of the Earth's orbit, while the maximum distance $r_{\max}$ is equal to the radius R_{pl} of the orbit of the planet to which the flight is directed. The distances are determined from the formulas (see (44.22)) $R_E = p/(1+e)$ and $R_{pl} = p/(1-e)$.

Hence we obtain the ellipse's parameters:

$$e = \frac{R_{pl} - R_E}{R_{pl} + R_E}, \qquad p = \frac{2R_{pl}R_E}{R_{pl} + R_E}. \tag{44.39}$$

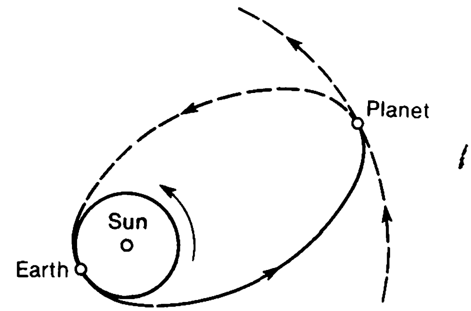


Fig. 119. Possible trajectory of flight to a remote planet of the solar system.

Knowing the orbital parameters, we can use Kepler's law to calculate the period of revolution of a body moving in this orbit around the Sun:

$$\frac{T^2}{T_E^2} = \left(\frac{R_{pl} + R_E}{2R_E}\right)^3, \quad T = T_E\left(\frac{R_{pl} + R_E}{2R_E}\right)^{3/2}, \tag{44.40}$$

where $T_E = 1$ year is the period of revolution of the Earth around the Sun. The time of flight to the planet is $\tau = T/2$. For example, $R_{pl} \simeq 19R_E$ for Uranus, and hence the time of flight will be about 16 years.

The trajectory of flight considered above is not the most suitable one from the point of view of the time of flight. It is possible to select a trajectory along which the time of flight is considerably reduced. The main idea is to use the forces of attraction of the planets whose orbits intersect during flight. This considerably increases the average velocity and reduces the duration of the voyage. For example, in one of the possible versions of the flight to Uranus, the trajectory passes near Jupiter, and thus the velocity is considerably increased. In this case, the flight to Uranus will take just 5 years instead of 16.

Sec. 45. MOTION OF THE EARTH'S ARTIFICIAL SATELLITES

The distinction between Kepler's laws and the laws of motion of the Earth's artificial satellites is analyzed. The consequences of this difference are discussed.

DISTINCTION BETWEEN THE LAWS OF MOTION OF THE EARTH'S ARTIFICIAL SATELLITES AND KEPLER'S LAWS OF PLANETARY MOTION. The motion of the Earth's artificial satellites cannot be described by Kepler's laws due to two reasons:

1. The Earth is not exactly spherical and the density distribution over its volume is not uniform. Hence its gravitational field is not equivalent to the gravitational field of a point mass situated at the geometrical centre of the Earth.

2. The atmosphere of the Earth slows down the motion of artificial satellites, as a result of which the shape and size of

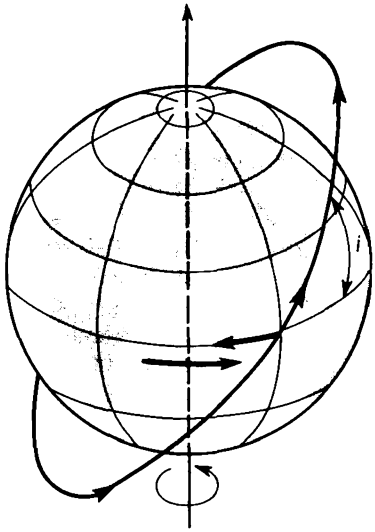


Fig. 120. Variation of a satellite's trajectory due to the Earth's rotation and the deviation of its shape from sphericity.

The arrow pointing to the right along the equator indicates the direction of motion of the points on the Earth's surface due to its revolution about its axis. The arrow pointing to the left (westwards) has a dual meaning: it indicates the westward displacement of the satellite's trajectory, as well as the direction of rotation of the plane of its orbit about the Earth's axis.

!

Since the rotational plane of a satellite remains practically unchanged relative to fixed stars even as the Earth is rotating, after one revolution the satellite will cross a certain fixed latitude westwards to the same extent as a point situated on the Earth's surface at this latitude will be displaced eastwards relative to fixed stars as a result of one revolution of the Earth about its axis.

their orbit change and the satellites eventually fall down on the Earth.

The deviation of the flight of satellites from the predictions of Kepler's laws can be used to draw conclusions about the shape of the Earth, the density distribution over its volume and the structure of the Earth's atmosphere. Hence the most comprehensive data were obtained only from an investigation of the motion of the Earth's artificial satellites. We shall briefly describe each investigation.

The problem of launching of artificial satellites is considered in the chapter on the motion of bodies of a variable mass.

TRAJECTORY OF A SATELLITE. If the Earth were a homogeneous sphere without any atmosphere, a satellite would move in an orbit whose plane preserves its orientation relative to the reference frame of fixed stars. The orbital parameters are determined in this case with the help of Kepler's laws. As the Earth rotates, the satellite moves over different points on the Earth's surface in successive turns. Knowing the trajectory of the satellite for any cycle, we can easily predict its position at all subsequent instants of time. For this purpose, we must consider that the Earth rotates from west to east at an angular velocity of about 15° per hour. Hence in its next turn, the satellite will cross the same latitude displaced to the west by the same number of degrees as the Earth has turned to the east during a period of revolution of the satellite (Fig. 120).

On account of the resistance offered by the Earth's atmosphere, satellites cannot have prolonged flights at altitudes below 160 km. The minimum period of revolution at such an altitude in a circular orbit is about 88 min, i.e. about 1.5 h. During this period, the Earth turns by about 22°30′. At a latitude of 50°, this angle corresponds to a distance of about 1400 km westwards as compared to the initial position.

However, such a method of calculations gives quite accurate results only for a few revolutions of a satellite. If long intervals of time are considered, we must take into account the difference between a sidereal day and 24 hours. Since the Earth completes one revolution around the Sun in 365 days, in one day it subtends an angle of about 1° (to be more precise, 0.99°) around the Sun in the direction of its rotation about its axis. Hence the Earth turns by 361° instead of 360° relative to fixed stars in one day, and hence one revolution is completed in 23 h 56 min instead of 24 h. Hence the trajectory of a satellite is displaced westwards by $(15 + 1/24)°$ instead of 15° in an hour. In the course of a few days, this correction runs into several degrees.

If the Earth were a homogeneous sphere devoid of an atmosphere, the above method of calculations would lead to quite accurate predictions concerning the position of a satellite over a considerably long time period in advance. However, the nonspherical shape of the Earth, the nonuniformity of its density and the existence of the atmosphere considerably change the nature of motion of satellites.

THE SHAPE OF THE EARTH. It has been known since long that the Earth is not exactly spherical. The first quantitative estimate of this deviation was provided by Newton who used the law of universal gravitation. Newton's calculations were based on a simple principle. Let us imagine a channel passing from a pole to the centre of the Earth and then to a point on the equator along a radius. Obviously, the pressure in each channel at the centre of the Earth must be the same. Due to the rotation of the Earth, the weight of a certain element of the liquid's column in a channel going towards the equator will be less than the weight of the corresponding element of the liquid's column in a channel going towards the pole at the same distance from the centre of the Earth. Hence the pressure at the centre of the Earth will be the same only if we admit that the channel going towards the equator is longer. This means that the Earth is not spherical in shape, but is flattened at the poles. The oblateness f is given by the formula

$$f = \frac{D_{eq} - D_{pol}}{D_{eq}}, \tag{45.1}$$

where D_{eq} is the equatorial diameter of the Earth, and D_{pol} is the polar diameter.

Using the arguments described above, Newton obtained the value $f = 1/298$. The results of his calculations were published in 1687. Various methods have been used so far to experimentally determine the flatness of the Earth at the poles. The results were found to be close to the value obtained by Newton, although slight differences were indeed recorded. Until the launching of the first satellites of the Earth, the most widely used estimate for the flatness of the Earth was 1/297.1. By observing the motion of satellites, a much more accurate and reliable value of this quantity was obtained in comparison with the methods described here. The observations resulted in a considerable variation of the estimate mentioned above.

!

The deviation of the shape of the Earth from spherical can conveniently be represented in the form of harmonics, each of which is responsible for a certain deflection in the orbit of a satellite moving around the Earth's sphere. An analysis of these deflections in the orbit provides information about the magnitude of harmonics causing them. The true shape of the Earth can be determined from the values of these harmonics.

If the shape of the Earth is nonspherical, its gravitational field cannot be reduced to the gravitational field of a point mass located at the centre of the Earth. If we assume the shape of the Earth to be known, we can calculate the gravitational field and the trajectory of a satellite. At present, such calculations can be made only by means of computers, and we shall simply describe the result. If we consider the flatness of the Earth, the plane of an orbit will no longer preserve its position relative to fixed stars and will turn about the Earth's axis in a direction opposite to the rotation of a satellite. For example, if a satellite moves about the Earth's axis eastwards (see Fig. 120), the orbital plane will turn in the western direction. If we reverse the direction of rotation of the satellite without changing the plane of its orbit, the latter will also turn in the opposite direction. The angle i between the planes of the orbit and the equator (see Fig. 120) remains constant. If the plane of a satellite's orbit passes through the axis of the Earth's rotation, i.e. if the orbit is strictly polar, it will retain its orientation relative to fixed stars. The velocity of rotation of the orbital plane depends on the oblateness of the Earth and on the orbital parameters. Hence, by measuring the orbital parameters and the velocity of rotation of its plane, we can calculate the oblateness of the Earth.

In addition to the rotation of the orbital plane, the oblateness of the Earth also leads to another effect: the perihelion of the orbit rotates in its plane and hence moves from the northern hemisphere to the southern hemisphere, and vice versa. The velocity of rotation of the perihelion depends on the oblateness of the Earth and the angle of inclination of the orbit. Measurement of the velocity of rotation of the perihelion also led to a numerical estimate for the oblateness of the Earth, which is found to be in agreement with the estimate for the oblateness obtained from the rotation of the orbital plane.

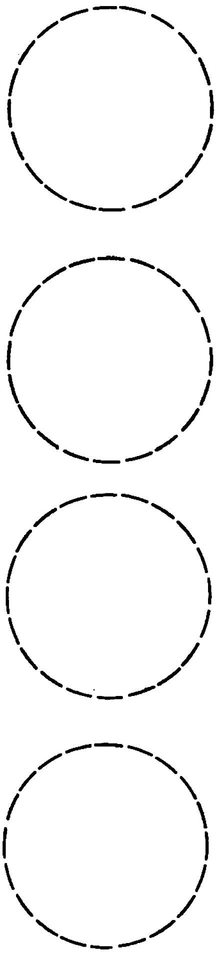


Fig. 121. The shape of the first harmonics characterizing the deviation of the Earth's surface from sphericity.

Measurements of the rates at which the orbital planes of the first satellites turned led to the conclusion that the oblateness of the Earth lies between 1/298.2 and 1/298.3. This means that the equatorial radius of the Earth is larger than the polar radius by 42.77 km and not by 42.94 km according to the estimate that existed before that time.

However, the oblateness of the Earth is not the only deviation from sphericity. The total deviation from sphericity can be represented mathematically as the sum of various regular deviations called harmonics. Oblateness corresponds to the second harmonic. The third harmonic takes into account the pear-like shape of the Earth, the fourth accounts for a square-shaped form, and so on. Figure 121 shows some of the first harmonics whose sum determines the actual shape of the Earth. The changes in the orbit of a satellite introduced by each harmonic can be calculated, and the role of each of them can be determined from actual observations of the shape of the Earth.

The third harmonic, which characterizes the pear-like shape of the Earth, is responsible for a change in the distance between the perihelion and the centre of the Earth depending on the hemisphere in which the perihelion is located. As the perihelion moves from one hemisphere to the other, its distance from the centre of the Earth changes. A study of the satellite's orbits showed that the pear-like shape of the Earth accounts for an elongation of about 40 m towards the north pole. This means that the level of the ocean water at the north pole is 40 m farther from the equatorial plane than at the Antarctic, where the sea level is buried under a 3-km thick layer of snow.

Subsequent harmonics also contribute towards a change in the orbital parameters of a satellite. Taking these harmonics into consideration, we can determine the shape of the Earth to a very high degree of precision. The most significant features of the Earth's shape are its oblateness and the pear-shaped asymmetry between its northern and southern hemispheres.

The next important result following from the observations of the satellite's motion is the establishment of the form of the equator. Even before the launching of satellites, indications existed that the equator was not exactly circular. These indications were based on the fact that the gravitational force changes slightly with longitude. However, this does not lead to an appreciable change in the orbit of a satellite since the satellite passes over all longitudes in view of the rotation of the Earth, and the changes in the gravitational force are averaged out over all longitudes. However, this averaging takes place through slight oscillations in the position of the satellite every day along its trajectory, the amplitude of oscillations being about a few hundred metres. These oscillations can lead to information about the change in the gravitational force along the longitudes and about the shape of the equator. In the first approximation, the equator is an ellipse whose major semiaxis is directed from the longitude 20° west to 160° east, and the minor semiaxis is directed from the longitude 110° west to 70° east. These two semiaxes differ in length by about 140 m. However, this is just an approximate picture. Subsequent refinements in the shape of the equator were carried out by following the motion of the satellites.

ATMOSPHERIC DRAG. The friction of satellites in the Earth's atmosphere serves as the second reason behind the deviation of satellites from the predictions of Kepler's laws. The density of air decreases nearly exponentially with altitude. This is a very rapid decrease, and yet the density of air up to altitudes of 160 km is such that it does not allow satellites to exist for a prolonged time. This is so because satellites at this altitude rapidly lose energy due to drag and fall to the ground.

The higher the altitude of a satellite, the longer the duration of its stay in the orbit.

The general nature of the change in an orbit due to drag can be explained as follows. The most significant losses in the energy of a satellite due to drag occur at the perihelion. The altitudes of the perihelion and aphelion decrease, but the altitude of the aphelion changes more rapidly than that of the perihelion, and hence the prolateness of the orbit decreases. The velocity of the satellite in the orbit increases, while the period of its revolution decreases. Some parts of the satellite's orbit may lie below 160-km altitude. In this case, the energy losses due to drag become quite significant and the satellite falls to the ground, following a trajectory that rapidly approaches the Earth. However, owing to the protective covering, the satellite does not burn out and can make a soft touch-down with the help of parachutes.

The trajectory of a satellite can be calculated by taking into account the change in the density of the atmosphere with altitude. Hence a knowledge of the trajectory helps in finding the density distribution in the atmosphere. Besides, the equipment mounted on the satellite enables the study of many other characteristics of the space near the Earth's surface.

The density of the atmosphere is mainly determined from the change in the period of revolution of a satellite. It was mentioned above that the period of revolution of a satellite decreases owing to drag in accordance with Kepler's laws, and its velocity increases. Of course, this conclusion does not contradict the energy conservation law. As a matter of fact, the total energy of a satellite is the sum of its positive kinetic energy and negative potential energy, and is a negative quantity. As a result of drag, the altitude at which the satellite moves decreases, and in accordance with the energy conservation law, its potential energy is spent on performing work against frictional forces and on increasing the kinetic energy of the satellite. The density of the atmosphere can be judged from the rate of decrease in the period of revolution of the satellite. At present, quite exhaustive data are available on the density of the atmosphere over a wide range of altitudes and on the dependence of density on various factors. The data have been obtained with the help of satellites.

GEOSTATIONARY ORBIT. The circular orbit of a satellite in the equatorial plane, the motion in which keeps the satellite's position over a point on the equator unchanged, is called a geostationary orbit. The biorbital parameters can be determined from the following two equations in two unknowns:

$$\frac{mv^2}{r} = G\frac{mM_E}{r^2}, \quad T = \frac{2\pi r}{v}, \tag{45.2}$$

where v is the velocity of the satellite, r is the radius of the synchronous orbit, $M_E = 6 \times 10^{24}$ kg is the mass of the Earth, $T = 1$ day $= 86{,}400$ s. From these equations, we obtain $v = 3.07 \times 10^3$ m/s $= 3.07$ km/s and $r = 4.22 \times 10^7$ m $=$ 42,200 km. Nearly half the surface of the Earth can be linked to a satellite on a synchronous orbit through linearly propagating high-frequency or optical signals. Hence synchronous-orbit satellites are of great importance for communication systems.

Sec. 46. THE TWO-BODY PROBLEM

The two-body problem is reduced to a one-body problem through coordinate transformation.

REDUCED MASS. In the problems on the motion under the action of gravitational forces considered above, it was assumed that the mass of the body which is the source of gravitational force is much larger than the mass of the body whose motion is being studied. Hence the heavier body can be treated as stationary, and the problem is reduced to the study of the motion of the lighter body in a given field. This is called a one-body problem.

However, such an approximation is not always possible, i.e. the errors introduced as a result of this approximation are not always negligible. For example, the components of a double star often have nearly equal masses and neither of the components can be treated as stationary. In a precise analysis of the motion of the Moon around the Earth, we must also consider the influence of the Moon on the motion of the Earth, and so on. Hence it becomes necessary to take into consideration the motion of both interacting bodies; such a problem is called a two-body problem.

Suppose that two bodies of mass m_1 and m_2 are drawn towards each other by gravitational forces. In the inertial reference frame, the equations of motion of these bodies have the following form (Fig. 122):

$$m_1 \frac{d^2\mathbf{r}_1}{dt^2} = G\frac{m_1 m_2}{r^2}\frac{\mathbf{r}}{r}, \qquad m_2 \frac{d^2\mathbf{r}_2}{dt^2} = -G\frac{m_1 m_2}{r^2}\frac{\mathbf{r}}{r}, \tag{46.1}$$

where $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ is the vector joining the interacting bodies and having a direction from m_1 towards m_2.

The general nature of the motion can be studied with the help of the ideas put forth in Sec. 21 on the motion of a reference frame of point masses. Obviously, the centre of mass, whose position is characterized by the radius vector

$$\mathbf{r}_{c.m} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{m_1 + m_2}, \tag{46.2}$$

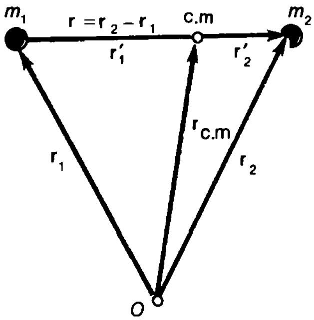


Fig. 122. To the solution of the two-body problem.

Point O is the origin of radius vectors.

moves uniformly and in a straight line, while the bodies of mass m_1 and m_2 move in such a way that their resultant momentum is zero in the centre-of-mass system. The angular momentum of these masses is conserved in an inertial reference frame, including the one fixed to the centre of mass.

However, it is more convenient to solve the two-body problem in a reference frame fixed to one of the bodies rather than in the centre-of-mass system since in this case the problem becomes equivalent to a one-body problem. For this purpose, we divide (46.1) by m_1 and m_2 respectively and subtract the second equation from the first. This gives

$$\frac{d^2}{dt^2}(\mathbf{r}_2 - \mathbf{r}_1) = \frac{d^2\mathbf{r}}{dt^2} = -\left(\frac{1}{m_1} + \frac{1}{m_2}\right) G \frac{m_1 m_2}{r^2}\frac{\mathbf{r}}{r}. \tag{46.3}$$

We denote the sum of the reciprocals of masses within the parentheses by

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}. \tag{46.4}$$

The quantity μ is called the reduced mass. We can now write (46.3) in the form

$$\mu \frac{d^2\mathbf{r}}{dt^2} = -G\frac{m_1 m_2}{r^2}\frac{\mathbf{r}}{r}. \tag{46.5}$$

This is the equation of motion in the one-body problem since vector $\mathbf{r}$ is the only unknown quantity in it. The solution of this type of equation was considered in detail in Secs. 44 and 45. The results obtained in these sections can directly be applied to (46.5). The only point that must be considered is that the force of interaction is determined by the masses m_1 and m_2 of the interacting bodies, while the inertial properties are determined by the reduced mass μ. While solving the problem, we treat one of the bodies (from which the radius vector is measured) as stationary and describe the motion of the other body with respect to this body.

The concept of reduced mass can be used in the two-body problems for other laws of interaction as well, irrespective of whether the interactions are classical or quantum-mechanical.

TRANSITION TO THE CENTRE-OF-MASS SYSTEM. After the variation of vector $\mathbf{r} = \mathbf{r}(t)$ has been obtained by solving Eq. (46.5), the trajectories of both masses can be found in the centre-of-mass system. Denoting the radius vectors of the masses m_1 and m_2 by $\mathbf{r}'_1$ and $\mathbf{r}'_2$ and taking the centre of mass as the reference point for measuring these radius vectors, we obtain, by definition of the centre of mass (see Fig. 122),

$$r'_1 = -\frac{m_2}{m_1 + m_2} r, \quad r'_2 = \frac{m_1}{m_1 + m_2} r. \tag{46.6}$$

Using these relations and knowing $r(t)$, we can plot $r'_1(t)$ and $r'_2(t)$. The trajectories of both bodies are homothetic relative to the centre of mass, the homothetic centre coincides with the centre of mass, and the homothetic ratio is equal to the ratio of masses.

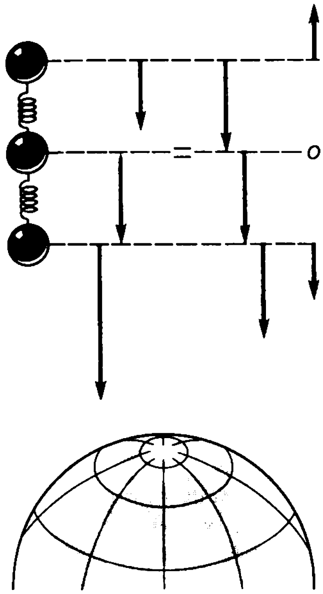


Fig. 123. The tide-producing force is caused by a change in the gravitational force with distance.

TIDES. The motion of bodies in a nonuniform gravitational field is accompanied by the emergence of forces tending to deform them. Accordingly, deformation appears in such bodies.

Suppose that three point masses, m each, connected through a weightless spring, fall freely in a nonuniform gravitational field along the straight line joining their centres (Fig. 123). The gravitational field in which the motion takes place is also produced by a point mass. The gravitational forces acting on these points are not equal: the upper point experiences a smaller gravitational force than the lower point. It can be seen from Fig. 123 that this situation is equivalent to the following: all the three points are subjected to the same force equal to the force acting on the central point, an additional force directed upwards acts on the upper point, while the same force directed downwards acts on the lower point. Consequently, the spring must become stretched. Thus, a nonuniform gravitational field tends to extend a body in the direction of nonuniformity.

In particular, the gravitational field of the Sun stretches the Earth along the line joining their centres. The Moon also exerts a similar effect on the Earth. The magnitude of the effect depends on the rate of change of this force, and not on the gravitational force.

The motion of a planet around the Sun is a free fall. The planet cannot fall onto the Sun just because of the tangential velocity perpendicular to the line joining the centres of the planet and the Sun. A celestial body subjected to the gravitational field of another body experiences the above-mentioned deforming force.

In the field of a spherical body, the gravitational force at a distance r from the centre is $F = -GMm/r^2$. Hence the rate of change of this force with distance is given by the formula $dF/dr = 2GMm/r^3$. For the gravitational fields of the Sun and the Moon, we obtain the following values (all quantities are referred to a unit mass): $2GM_S/r^3 = 0.8 \times 10^{-13}$ 1/s^2 and $2GM_M/r^3 = 1.8 \times 10^{-13}$ 1/s^2. Thus, the "deforming" force exerted by the Moon on the Earth is more than twice the corresponding force exerted by the Sun.

This "deforming" force does not significantly affect the shape of the Earth's crust since even small deformations in the crust

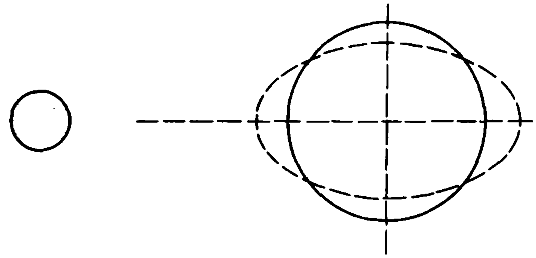


Fig. 124. The tides on the Earth caused by the gravitational field of the Moon.

The tides caused by the gravitational field of the Sun are several times weaker.

can compensate for the action of this force. However, the shape of the water surface in the oceans changes considerably: "humps" appear in the direction of the nonuniform gravitational field, while the ocean level drops in the perpendicular direction (Fig. 124). Each such pair of "humps" retains its position along the line joining the centre of the Earth to the centres of the Sun and the Moon respectively. In view of the rotation of the Earth, the "humps" and "valleys" move along the surface of the Earth and periodically raise and lower the level of water in the oceans. This phenomenon is manifested at the shores in the form of tides and ebbs. Calculations show that during lunar tides and ebbs, the level of water changes by a maximum of 0.56 m. It would be fair if the entire surface of the Earth were covered with water. In actual practice, the complex effect of the mass of the dry land during the displacement of the "humps" and "valleys" causes a change in the water level at different places from zero to 20 m (approximately). Obviously, there are two tides and two ebbs in a day at a given place.

Tides cause a motion of the water masses in the horizontal direction. This is accompanied by friction and a loss of energy in performing work against the frictional forces. This results in a tide-producing force, which leads to a decrease in the velocity of the Earth's rotation. The friction is not large, and so the velocity does not change significantly. Obviously, energy losses due to friction occur not only when liquid masses are displaced, but also when deformations travel along the surface of a body since a part of energy is always spent on deformation and its subsequent removal (there are no perfectly elastic bodies in nature). As a result of tides caused in the lunar matter by the gravitational forces of the Earth, the rotation of the Moon is slowed down to such an extent that it always faces the Earth from the same side. In this case, there is no tide-producing force.

The tide-producing force on the Earth decreases its period of revolution about its axis by 4.4×10^{-8} s per revolution. This is confirmed by astronomical observations. However, the angu-

!

The two-body problem is the simplest interaction problem and serves as the touchstone for the theory of interactions. This problem has an exact solution in many cases. The three-body problem is a lot more complicated and does not have a solution in a final analytical form. The advent of computers considerably simplified the numerical solution of this problem, which can be obtained without using the perturbation method.

?

Is the reduced mass of bodies larger, smaller or intermediate between their masses?

Under what conditions can one of the bodies in the two-body problem be assumed stationary?

What are the trajectories of interacting bodies in a centre-of-mass system?

In which (inertial or noninertial) reference frame does the equation of motion in a two-body problem contain the reduced mass?

lar momentum must be conserved in the Earth-Moon system. The Earth revolves about its axis in the same direction in which the Moon rotates around the Earth. Hence a decrease in the angular momentum of the Earth must be accompanied by an increase in the angular momentum of the Earth-Moon system moving around their common centre of mass. The angular momentum of the Earth-Moon system is

$$L = \mu v r, \tag{46.7}$$

where μ is the reduced mass of the Earth and the Moon, defined by (46.4), and r is the distance between them. Assuming their orbits to be circular, we can write

$$\frac{G m_E m_M}{r^2} = \frac{\mu v^2}{r}. \tag{46.8}$$

It follows from (46.7) and (46.8) that

$$r = \frac{L^2}{G m_E m_M \mu}, \quad v = \frac{G m_E m_M}{L}. \tag{46.9}$$

An increase in L due to the tide-producing force leads to an increase in the distance r between the Earth and the Moon, and to a decrease in the velocity of rotation of the Moon around the Earth. The rate of increase in the distance is about 0.04 cm/day at present. Although this is a small increase, the total increase in the distance between the Earth and the Moon over a period of several billion years will be comparable with the existing distance between them.

PROBLEMS

11.1. Find the ratio of the periods of time required for traversing the segments of an elliptic orbit bounded by the minor semiaxis.

11.2. Find the difference in the periods of time required to cover the parts into which the Earth's elliptic orbit is divided by the minor semiaxis. The eccentricity of the Earth is $e = 0.017$.

11.3. The period of revolution of a body in a circular orbit about a gravitational centre is T. Calculate the time taken by the body to fall from the orbit onto the gravitational centre if its instantaneous velocity became zero.

11.4. A body moves in a circular orbit of radius r about a gravitational centre. The period of revolution is T. At a certain point on the orbit, the direction of the velocity of the body is changed instantaneously by an angle smaller than π, the magnitude of the velocity remaining the same. What will be the new trajectory and the period of revolution of the body?

11.5. The minimum distance l to which a comet approaches the Sun is smaller than the radius r_E of the Earth's orbit ($l < r_E$). For how much time will the distance between the comet and the Sun be smaller than r_E?

11.6. During the motion of a body in a central gravitational field, the magnitude of its velocity is changed instantaneously by δv_1 at the perihelion (r_1) and by δv_2 at the aphelion (r_2) without changing the direction. The values of r_1 and r_2 will be different for the new orbit. Find the values of δr_1 and δr_2.

11.7. A shell is fired vertically upwards from the surface of the Earth at an initial velocity v. When it reaches the upper point of its trajectory, it explodes and splits into two equal parts, moving at a velocity u at the initial instant of time. What will be the maximum distance between the points at which these two parts fall onto the Earth?

ANSWERS

11.1. $(\pi + 2e)/(\pi - 2e)$. **11.2.** ~ 4 days. **11.3.** $T\sqrt{2}/8$. **11.4.** In an ellipse with axis $2r$; T. **11.5.** $\sqrt{2(1 - l/r_E)}[1 + 2l/r_E/(3\pi)]$ year. **11.6.** $\delta r_1 = 4\delta v_2 r_1/[(1 + e)v_2]$, $\delta r_2 = 4\delta v_1 r_2/[(1 - e)v_1]$. **11.7.** $2uv/g$ for $u \leqslant v$, $(u^2 + v^2)/g$ for $u \geqslant v$.

Chapter **12**

Motion in an Electromagnetic Field

Basic idea:

Electromagnetic forces have a decisive influence on the motion of charged particles over a range exceeding the nuclear dimensions but smaller than astronomical distances. Strong (nuclear) and weak interactions take place on the subatomic scale, while gravitational interactions occur between astronomical objects.

Sec. 47. PROPERTIES OF ELECTROMAGNETIC FORCES

The properties of electromagnetic forces are described.

LORENTZ FORCE. A point charge q in an electromagnetic field is acted upon by the Lorentz force

$$\mathbf{F} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B}, \tag{47.1}$$

where E, B and v denote the electric field strength, magnetic induction and the velocity of the point charge q respectively.

According to this relation, the force F consists of two components, each of which affects the motion of the charge in a different way. The force described by the second component, i.e.

$$\mathbf{F}_m = q\mathbf{v} \times \mathbf{B}, \tag{47.2}$$

which is called the magnetic force for the sake of brevity, is directed at right angles to the particle's velocity. This means that it is directed at right angles to the displacement of the particle and hence does no work:

$$dA_m = \mathbf{F}_m \cdot d\mathbf{r} = \mathbf{F} \cdot \mathbf{v}\, dt = q(\mathbf{v} \times \mathbf{B}) \cdot \mathbf{v}\, dt = 0,$$

where $d\mathbf{r} = \mathbf{v}\, dt$ is the displacement of the charge in time dt, and the scalar triple product is $(\mathbf{v} \times \mathbf{B}) \cdot \mathbf{v} = 0$ since two of its cofactors are collinear. *Thus, the magnetic force does not change the kinetic energy of a charge, i.e. does not change the magnitude of its velocity, but changes only its direction.* The force described by the first term in (47.1), i.e.

$$\mathbf{F}_e = q\mathbf{E}, \tag{47.3}$$

which is called the electric force for the sake of brevity, can be oriented at an arbitrary angle to the particle's velocity. The work done by this force in displacing a charge q by $d\mathbf{r} = \mathbf{v}\,dt$ is

$$dA_e = \mathbf{F}_e \cdot d\mathbf{r} = q\mathbf{E} \cdot \mathbf{v}\,dt.$$

The work done by an electric force may be positive or negative or zero. The work is zero if the electric field is directed at right angles to the particle's velocity. For $\mathbf{E} \cdot \mathbf{v} > 0$, the work is positive for a positive charge and negative for a negative charge. Consequently, the energy of a positive charge and the magnitude of its velocity increase in this case, while a decrease in the values of these quantities is observed for a negative charge. For $\mathbf{E} \cdot \mathbf{v} < 0$, the work is positive for a negative charge and negative for a positive charge. Hence the magnitude of the velocity of a negative charge increases and that of a positive charge decreases.

Thus, when a charge moves in an electromagnetic field, the magnitude of its velocity changes only due to the action of the electric force, while the direction of its velocity changes due to the action of both electric and magnetic forces.

POTENTIAL OF AN ELECTROSTATIC FIELD. The force of interaction of stationary point charges is described by Coulomb's law

$$F = \frac{1}{4\pi\varepsilon_0} \frac{q_1 q_2}{r_{12}^2}. \tag{47.4}$$

!

When a charge moves in an electromagnetic field, the magnitude of its velocity changes only due to the action of the electric force, while the change in its direction is caused both by the electric and magnetic forces.

?

Write down the relativistic equation in such a form that it explicitly contains the particle's acceleration. Using this relation, analyze the relation between the directions of the force and the acceleration.

It was explained in Sec. 25 (see (25.32)) that the forces which decrease in inverse proportion to the square of the distance are potential forces. Any electrostatic field can be represented as a superposition of the fields of stationary charges and is therefore a potential field. Hence the potential energy of a point charge q in a field can be represented in the form

$$E_p(\mathbf{r}) = q\varphi(\mathbf{r}), \tag{47.5}$$

where φ is the potential at the point where the charge q is situated. Hence, in accordance with (25.20), the electric force acting on this charge can be calculated from the formulas

$$F_x = qE_x = -\frac{\partial E_p}{\partial x} = -\frac{q\,\partial\varphi}{\partial x}, \qquad F_y = -\frac{q\,\partial\varphi}{\partial y}, \quad F_z = -\frac{q\,\partial\varphi}{\partial z}. \tag{47.6}$$

The methods of computing the potential, electric field strength and magnetic induction are considered in the theory of electricity and magnetism. In mechanics, fields are assumed to be known.

EQUATION OF MOTION. The equation of motion for an electromagnetic field has the form

$$\frac{d\mathbf{p}}{dt} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B} = \mathbf{F}. \tag{47.7}$$

This equation is valid for all velocities.

It was mentioned in Sec. 20 that in the relativistic case the direction of acceleration of a point mass does not coincide with the direction of the force acting on it. Hence, knowing the direction of the force and of the point's velocity, we can easily find how the momentum changes, but the variation of the velocity is not all that obvious. Hence we transform (47.7) as follows. Considering that $\mathbf{p} = m\mathbf{v} = m_0\mathbf{v}/\sqrt{1 - v^2/c^2}$, we can write

$$\frac{d\mathbf{p}}{dt} = \frac{dm}{dt}\mathbf{v} + m\frac{d\mathbf{v}}{dt}. \tag{47.8}$$

It follows from (33.3) that

$$\frac{d}{dt}\left(\frac{m_0c^2}{\sqrt{1 - v^2/c^2}}\right) = c^2\frac{dm}{dt} = \mathbf{F}\cdot\frac{d\mathbf{r}}{dt} = \mathbf{F}\cdot\mathbf{v}. \tag{47.9}$$

Consequently,

$$\frac{dm}{dt} = \mathbf{F}\cdot\frac{\mathbf{v}}{c^2}. \tag{47.10}$$

The equation of motion (47.7) then assumes the form

$$m\frac{d\mathbf{v}}{dt} = \mathbf{F} - \frac{\mathbf{v}}{c}\left(\mathbf{F}\cdot\frac{\mathbf{v}}{c}\right). \tag{47.11}$$

It can be seen that the acceleration coincides in direction with the force only if the force is perpendicular to the velocity. In other cases, the acceleration in the direction of the force is supplemented by an acceleration in the direction of the velocity. This acceleration is of the order of v^2/c^2 relative to the acceleration in the direction of the force and is negligibly small at nonrelativistic velocities.

Sec. 48. MOTION IN STATIONARY ELECTRIC AND MAGNETIC FIELDS

The most distinguishing features of the motion of charged particles in stationary electric and magnetic fields are described.

MOTION IN A UNIFORM MAGNETIC FIELD. While considering the motion of a charged particle in a magnetic field, it is convenient to represent the velocity v as the sum of the components $v_\perp$ and $v_\parallel$ perpendicular and parallel to the magnetic field respectively (Fig. 125):

$$v = v_\perp + v_\parallel. \tag{48.1}$$

The Lorentz force acting on the charge is

$$F = e(v_\perp + v_\parallel) \times B = ev_\perp \times B + ev_\parallel \times B = ev_\perp \times B. \tag{48.2}$$

Fig. 125. Decomposition of the velocity vector of the charge moving in a magnetic field into two components, one along the magnetic induction B and the other perpendicular to it.

It follows from here that the component of the Lorentz force parallel to the magnetic field is zero, i.e.

$$F_\parallel = e(v \times B)_\parallel = 0. \tag{48.3}$$

The component perpendicular to the magnetic field is given by

$$F_\perp = ev_\perp \times B \tag{48.4}$$

and depends only on the perpendicular component of the velocity.

The constant magnitude of the velocity of a charge moving in a stationary magnetic field indicates that the mass of the charge carrier, given by $m = m_0/\sqrt{1 - v^2/c^2}$, is also constant. Hence the equations of motion for the parallel and perpendicular components of the velocity have the form

$$m\frac{dv_\parallel}{dt} = 0, \tag{48.5a}$$

$$m\frac{dv_\perp}{dt} = ev_\perp \times B. \tag{48.5b}$$

It follows from (48.5a) that

$$\boxed{v_\parallel = \text{const.}} \tag{48.6}$$

In order to analyze the motion at right angles to the magnetic field, let us consider Eq. (48.5b). We have

$$v^2 = v_\perp^2 + v_\parallel^2 = \text{const.} \tag{48.7}$$

Taking (48.6) into consideration, we obtain

$$\boxed{v_\perp^2 = \text{const.}} \tag{48.8}$$

Let us look a little more closely at Eq. (48.5b) (for $v_\parallel = 0$). The angle between vectors $v_\perp$ and B remains constant and equal to $\pi/2$. The magnitudes of $v_\perp$ and B do not change. The force on the right-hand side of (48.5b) is perpendicular to the velocity and has a constant magnitude. Consequently, this equation describes the motion with a constant acceleration which is always perpendicular to the velocity. In other words, this equation describes the motion in a circle. The left-hand side of (48.5b) expresses the product of the particle's mass and the centripetal acceleration $v_\perp^2/r$, where r is the radius of the circle. The right-hand side of (48.5b) describes the force $|e|v_\perp B$ due to the centripetal acceleration. Hence we can write

$$\frac{mv_\perp^2}{r} = |e|v_\perp B. \tag{48.9}$$

This equation contains a complete description of the motion of a charged particle in a circle in a plane perpendicular to the uniform magnetic field.

The direction of rotation depends on the sign of a charge. It can be concluded from Eq. (48.5b) that the direction of rotation of a negative charge is associated with the direction of the magnetic field B through the right-hand screw rule, while the rotation of a positive charge is connected with the direction of the magnetic field through the left-hand screw rule (Fig. 126).

Equation (48.9) leads to the following expressions for the angular velocity of rotation and the radius of the orbit:

$$\omega = \frac{2\pi}{T} = \frac{2\pi}{2\pi r/v_\perp} = \frac{v_\perp}{r} = \frac{|e|B}{m}, \qquad r = \frac{v_\perp}{\omega} = \frac{mv_\perp}{|e|B}. \tag{48.10}$$

The complete motion of a charged particle in a constant uniform magnetic field is composed of a uniform motion along the field and rotation in a plane perpendicular to it. This means that the particle moves along a spiral with a lead

$$l = v_\parallel T = v_\parallel \cdot \frac{2\pi}{\omega}. \tag{48.11}$$

MOTION IN A TRANSVERSE NONUNIFORM MAGNETIC FIELD. The solution of this problem in the general form is quite complicated, and we shall confine ourselves to the case when a moving charged particle does not deflect appreciably from a rectilinear trajectory. The particle moves all the time nearly at right angles to the magnetic field that varies in magnitude but not in direction. Suppose that this field (Fig. 127) is defined by

$$B_x = B(z), \qquad B_y = B_z = 0. \tag{48.12}$$

We shall assume that the particle moves along the Z-axis. The magnitude of the magnetic field $B(z)$ also varies arbitrarily in this direction. At the instant $t = 0$, the particle is at the origin

!

A variation is called adiabatic if it takes place quite slowly in comparison with the variations characteristic of the phenomenon under consideration. Hence the same process may or may not be adiabatic under different conditions.

The drift of a charge in a uniform magnetic field is a consequence of the variation of the radius of curvature of its trajectory due to the variation of the particle's energy caused by a uniform electric field. The drift of a charge in a nonuniform magnetic field is due to the variation of the radius of curvature of its trajectory due to the variation of the magnetic field.

Particles are "reflected" from the region of a strong increase of the magnetic field, and their trajectories coil around the lines of force. This phenomenon is used for confining charged particles to a finite volume.

Charged particles rotate in the Earth's magnetic field around the magnetic field lines and are displaced in the meridional direction from north to south and from south to north, undergoing successive reflections from the regions of increase of the magnetic field near the poles. At the same time, the particles are displaced from one meridian to another, moving along parallels around the Earth.

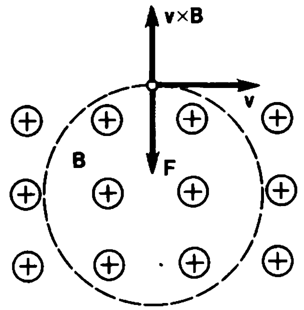


Fig. 126. Determining the direction of rotation of a negatively charged particle in a magnetic field B.

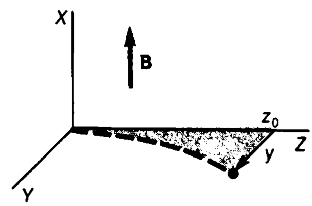


Fig. 127. Trajectory of a charged particle in a transverse magnetic field.

of coordinates and has a velocity v towards positive values of the Z-axis. It can clearly be seen that in this case the Lorentz force acts in the YZ-plane throughout, and hence the particle moves in this plane. Only small deviations of the particle from the Z-axis are considered. This means that the velocity along the Z-axis is much higher than the velocity along the Y-axis:

$$\frac{v_y}{v_z} \ll 1. \tag{48.13}$$

Hence the velocity v, which is constant in a magnetic field, can be represented in the form

$$v = \sqrt{v_z^2 + v_y^2} = v_z\left(1 + \frac{v_y^2}{v_z^2}\right)^{1/2} \approx v_z + \frac{1}{2}v_z\frac{v_y^2}{v_z^2} + \ldots, \tag{48.14}$$

where the square root has been expanded into a series and only the first term of the expansion in v_y^2/v_z^2 has been retained. Hence, to within small values of $v_y^2/v_z^2 \ll 1$, the particle's velocity along the Z-axis remains unchanged, i.e.

$$v = v_z = \text{const}. \tag{48.15}$$

Let us write down the equation of motion (47.7) in coordinate form by using the formula for vector product:

$$m\frac{d^2x}{dt^2} = 0, \quad m\frac{d^2y}{dt^2} = ev_zB, \quad m\frac{d^2z}{dt^2} = -ev_yB. \tag{48.16}$$

In view of the smallness of v_y in comparison with v_z, the force in the equation for the z-coordinate is much smaller than the force in the equation for the y-coordinate. Hence, taking (48.15) into consideration, the force on the right-hand side of the third equation in (48.16) can be put equal to zero, and the equation can be written in the form $m(d^2z/dt^2) = 0$.

Under the initial conditions

$$\begin{aligned} &x(0) = 0, \quad y(0) = 0, \quad z(0) = 0, \\ &\frac{dx(0)}{dt} = 0, \quad \frac{dy(0)}{dt} = 0, \quad \frac{dz(0)}{dt} = v, \end{aligned} \tag{48.17}$$

for $x(t)$ and $z(t)$, we obtain the following relations:

$$x(t) = 0, \quad z(t) = vt. \tag{48.18}$$

By means of formulas

$$\frac{dy}{dt} = \frac{dy}{dz}\frac{dz}{dt} = \frac{dy}{dz}v, \quad \frac{d^2y}{dt^2} = \frac{d^2y}{dz^2}v^2, \tag{48.19}$$

we can write the equation for y in the form

$$\frac{d^2y}{dz^2} = \frac{e}{mv}B(z). \tag{48.20}$$

The solution of this equation after two successive integrations has the form

$$y(z_0) = \left(\frac{e}{mv}\right) b, \tag{48.21}$$

where

$$b = \int_0^{z_0} d\xi \int_0^{\xi} B(\eta)\, d\eta = \int_0^{z_0} (z_0 - \eta) B(\eta)\, d\eta \tag{48.22}$$

is a constant determined by the configuration of the field through which the charged particle passes. In our problem, this constant is assumed to be known.

Measuring the deviation $y(z_0)$ and knowing the velocity v of the particle, we can determine the ratio e/m. This method was used to find the charge-to-mass ratio for electrons in one of the earliest measurements of this quantity.

MOTION IN A LONGITUDINAL ELECTRIC FIELD. Suppose that the z-axis is parallel to the force acting on a charge due to an electrostatic field. The particle's velocity is also directed along the z-axis, and its value at each point can be determined by the energy conservation law. In other words, the relation $v = v(z)$ is known. This enables us to find the time dependence $z(t)$ of the particle's position since

$$\frac{dz}{dt} = v(z). \tag{48.23}$$

The function on the right-hand side of this equation (velocity) is known. Hence if a point has a z_0-coordinate at the instant t_0, it will have a z-coordinate at the instant t; moreover, from (48.23), we obtain

$$\int_{z_0}^{z} \frac{dz}{v(z)} = \int_{t_0}^{t} dt = t - t_0. \tag{48.24}$$

Evaluating the integral on the left-hand side, we obtain the dependence of z on t in implicit form. For example, in the nonrelativistic case, the energy conservation law can be written in the form

$$\frac{mv^2}{2} + e\varphi = E_0 = \text{const}. \tag{48.25}$$

Hence

$$\int_{z_0}^{z} \frac{dz}{\sqrt{2[E_0 - e\varphi(z)]/m}} = t - t_0. \tag{48.26}$$

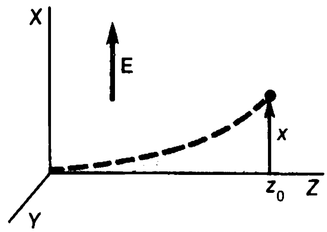


Fig. 128. Trajectory of a charged particle in a transverse electric field.

?

Which property of forces exerted by a magnetic field on a charge ensures the invariability of the magnitude of the velocity of the charge moving in this field?
What are the factors determining the parameters of the helical trajectory of a charge moving in a uniform magnetic field?
What does the direction of rotation of a charge in a magnetic field depend on?
What is the difference between relativistic and nonrelativistic motions of a charge in a uniform magnetic field?
Describe the mechanism responsible for the drift of charged particles in crossed electric and magnetic fields.
Does the direction of drift in crossed fields depend on the sign of the charge?
Write down the energy conservation law for a charged particle moving in a stationary electric field. How does the motion of one type of charge differ from that of the opposite charge in a given field?
What is an electron volt? Can it be used as a unit of energy in SI system? What are out-of-system units?

The sign of the root should be chosen in such a way that it corresponds to the sign of the velocity for a chosen positive direction of the z-axis.

The motion in the relativistic case in analyzed in a similar manner, the only difference being that while calculating the velocity, we should use the energy conservation law in the following form:

$$\frac{m_0 c^2}{\sqrt{1 - v^2/c^2}} + e\varphi = \text{const.} \tag{48.27}$$

This does not introduce any fundamental changes in the solution of the problem.

MOTION IN A TRANSVERSE ELECTRIC FIELD. Suppose that the initial velocity of a particle is directed along the Z-axis and the electric field is applied along the X-axis. As a result, the particle describes a certain trajectory in the XZ-plane.

In this case, the motion of the particle in the relativistic case is quite different from that in the nonrelativistic case.

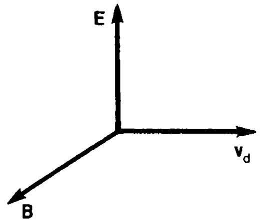
In the nonrelativistic case, the motion can be visualized as a combination of two independent motions: (1) along the Z-axis at a constant velocity equal to the initial velocity v_0, and (2) along the X-axis under the action of the electric field with zero initial velocity in this direction. Thus, at the instant t, the coordinate of the particle is $z = v_0 t$, while the x-coordinate can be found from the formulas obtained above for the motion in a longitudinal electric field. This is so because the motion along the X-axis is not connected in any way with the motion along the Z-axis.

In the relativistic case, it is impossible to represent the motion as a combination of two independent motions in mutually perpendicular directions. This is due to the fact that the directions of the force and the acceleration do not coincide (see Sec. 20). Hence the force acting along the X-axis causes an acceleration along the Z-axis as well, and the motions along the X- and Z-axes are mutually connected. The formulas describing the motion become quite complicated, and we shall confine ourselves to the remark made on the basic aspects of the relativistic motion.

SMALL DEVIATIONS. Suppose that the trajectory of a particle does not differ appreciably from a straight line, i.e. the radius of curvature of the trajectory is much larger than its length. Let the electric field be applied along the X-axis, the magnetic field being zero (Fig. 128):

$$E_y = E_z = 0, \quad E_x = E_z. \tag{48.28}$$

The magnitude of vector E generally changes along the Z-axis, i.e. $E = E(z)$. The equations of motion and the initial con-



ditions have the following form:

$$m_0\frac{d^2x}{dt^2} = eE(z), \quad m_0\frac{d^2y}{dt^2} = 0, \quad m_0\frac{d^2z}{dt^2} = 0, \tag{48.29}$$

$$x(0) = y(0) = z(0) = 0, \quad \frac{dx(0)}{dt} = 0, \quad \frac{dy(0)}{dt} = 0, \quad \frac{dz(0)}{dt} = v.$$

This problem can be solved by using the same arguments and transformations of variables that were described for Eqs. (48.16). The only difference is that we now obtain the following equations instead of (48.20):

$$\frac{d^2x}{dz^2} = \frac{e}{mv^2}E(z), \quad \frac{dx(0)}{dz} = 0, \quad x(0) = 0. \tag{48.30}$$

The solution of these equations is similar to Eq. (48.21):

$$x(z_0) = \frac{ea}{mv^2}, \tag{48.31}$$

where

$$a = \int_0^{z_0} d\xi \int_0^{\xi} E(\eta)\,d\eta = \int_0^{z_0} (z_0 - \eta)E(\eta)\,d\eta \tag{48.32}$$

depends only on the configuration of the electric field.

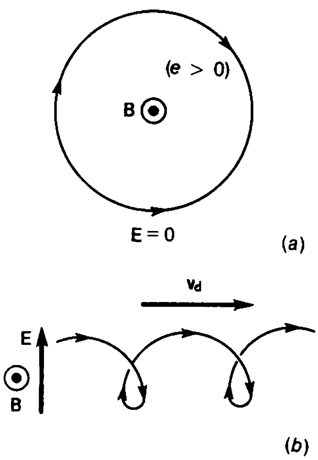


Fig. 129. Drift of a charged particle in crossed electric and magnetic fields.

DRIFT IN CROSSED ELECTRIC AND MAGNETIC FIELDS. If an electric and a magnetic field are applied simultaneously, the motion becomes much more complicated. Let us consider the simplest case when these fields are mutually perpendicular and their magnitude is such that the radius of curvature of a particle's trajectory is much smaller than the linear dimensions of the region in which the particle moves, i.e. the magnetic field is quite strong. As a result, the moving particle describes a large number of revolutions in the region of motion. Under these conditions, charged particles are said to drift.

Let us consider uniform crossed electric and magnetic fields ($\mathbf{E} \perp \mathbf{B}$) shown in Fig. 129. The general nature of motion can be determined with the help of purely qualitative considerations without resorting to any solution of equations. For the sake of definiteness, we shall assume that the charge of the particle is positive ($e > 0$). In the absence of an electric field, the particle moves in a uniform magnetic field in a circle at a constant velocity v (Fig. 129*a*). When an electric field is applied at right angles to the magnetic field, the particle's velocity becomes variable. For displacements along the electric force $\mathbf{F}_e$, the velocity increases, and the radius of curvature of the trajectory becomes larger (upper semicircles in Fig. 129*b*). When the direction of the velocity is reversed, the particle moves against the electric force, and hence its velocity and the radius of curvature of its trajectory decrease (lower semicircles in Fig. 129*b*). The motion with a small radius of curvature takes place in a smaller region of the trajectory, and hence during one complete cycle the particle moves in a direction perpendicular to both the electric and the magnetic field. This motion is called **drift**.

The drift in crossed electric and magnetic fields is independent of the sign of a charge.

If the particle has a negative charge ($e < 0$), the direction of rotation of the particle in the magnetic field is reversed, i.e. the rotation of the particle shown in Fig. 129*a* should be counterclockwise and not clockwise as shown. The direction of the electric force is also reversed, i.e. the electric force points downwards in Fig. 129*b*. Hence the radius of curvature will be larger in the lower part of the trajectory than in the upper part, and the drift will have the same direction as for the positive sign of the charge.

The drift of a particle can be represented as a motion in a circle about a centre which moves with the drift velocity $\mathbf{v}_d$. In order to calculate the value of this quantity, we must solve the equation of motion

$$m_0 \frac{d\mathbf{v}}{dt} = e\mathbf{E} + e\mathbf{v} \times \mathbf{B} \tag{48.33}$$

in the given field. This solution is sought in the form

$$\mathbf{v} = \mathbf{v}' + B^{-2}\mathbf{E} \times \mathbf{B}, \tag{48.34}$$

where $\mathbf{v}'$ is an unknown variable velocity, and the quantity

$$\mathbf{v}_d = B^{-2}\mathbf{E} \times \mathbf{B} \tag{48.35}$$

represents a constant velocity. It will be shown that this is the drift velocity of the particle. Substituting (48.34) into (48.33), we obtain

$$m_0 \frac{d\mathbf{v}'}{dt} = e\mathbf{E} + e\mathbf{v}' \times \mathbf{B} + \left(\frac{e}{B^2}\right)[\mathbf{B}(\mathbf{E}\cdot\mathbf{B}) - \mathbf{E}B^2] = e\mathbf{v}' \times \mathbf{B}, \tag{48.36}$$

where we have used the formula for the vector triple product, i.e. $(\mathbf{E} \times \mathbf{B}) \times \mathbf{B} = \mathbf{B}(\mathbf{E}\cdot\mathbf{B}) - \mathbf{E}(\mathbf{B}\cdot\mathbf{B})$, and have taken into account that the scalar product of vectors $\mathbf{E}$ and $\mathbf{B}$ is zero in view of their orthogonality. Equation (48.36) for vector $\mathbf{v}'$

$$m_0 \frac{d\mathbf{v}'}{dt} = e\mathbf{v}' \times \mathbf{B} \tag{48.37}$$

coincides with Eq. (48.5b) describing a uniform motion in a circle. Hence vector $\mathbf{v}'$ represents the velocity of this motion. The radius of the circle and the frequency of rotation are given

?

What is the mechanism of emergence of drift in a nonuniform magnetic field? Does the direction of drift depend on the sign of a charge? How?
How does the curvature of magnetic field lines result in a drift of charged particles?
Under what circumstances does a charged particle moving around a line of force deflect from it?
On what energy does the rotational magnetic moment of a particle depend in a magnetic field?
What is meant by the adiabatic invariance of magnetic moment?
What are the arguments supporting the statement that particles move on the surface of a magnetic tube?
Explain the effect of magnetic mirrors from the conservation of magnetic moment and a direct consideration of the forces exerted by the magnetic field on the charge.
What is a loss cone?
How do charged particles move in the radiation belts of the Earth?
Which factors force the particles in the radiation belts of the Earth to move along a longitude around the globe?

by formulas of the type (48.10). The centre of the circle moves at the drift velocity $\mathbf{v}_d$ given by formula (48.35). Its absolute magnitude is given by

$$v_d = B^{-2}|\mathbf{E} \times \mathbf{B}| = \frac{E}{B}, \tag{48.38}$$

while its direction is perpendicular to E and B.

Formula (48.35) *shows that the drift velocity is independent neither of the sign of a charge nor also of its magnitude and the mass of the particle.*

This is a significant conclusion since the drift of heavy particles like protons is analogous to the drift of light particles like electrons which have the opposite charge. Hence if we have a plasma formed by protons and electrons whose charges are mutually compensated, it will move at drift velocity when placed in crossed electric and magnetic fields. This is not accompanied by the emergence of any forces striving to separate the negatively and positively charged components of the plasma.

DRIFT IN A NONUNIFORM MAGNETIC FIELD. Consider a magnetic field whose lines are parallel to one another and whose magnitude changes in a direction perpendicular to the field. If the field were uniform, a charged particle would move in a circle under the action of this field. However, in view of the nonuniformity of the field, the radius of curvature of the trajectory changes during the motion: the radius of curvature will be smaller in regions where the magnetic field is stronger, and larger where the field is weaker. Thus, the pattern is the same as in crossed fields, the only difference being that in this case the radius of curvature of the trajectory changes not because of a change in the energy of the particle, but rather as a result of a change in the magnetic field at different points of the trajectory.

The drift of a particle occurs in a direction perpendicular to both the magnetic field and the direction of nonuniformity of this field. The drift pattern is shown in Fig. 130. It can clearly be seen in the figure that particles with opposite charges drift in opposite directions.

An exact calculation of the drift velocity in this case is quite complicated, but we can carry out a simpler approximate calculation which gives reasonably precise results. We assume that the magnetic field does not increase continuously, but its magnetic induction varies abruptly along the lines AA_1 (Fig. 131) such that $B_1 > B_2$. In each half-plane, the particle moves in a circular trajectory, but the radii of these trajectories

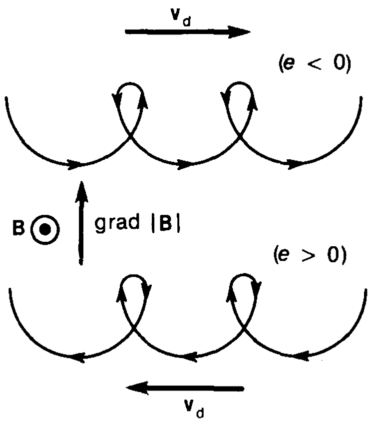


Fig. 130. Drift of a charged particle in a nonuniform magnetic field.

The arrow pointing upwards indicates the direction in which the magnitude of the magnetic field grad |B|) increases.

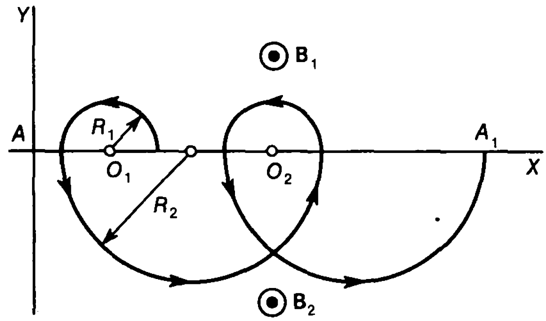


Fig. 131. Calculating the drift velocity in a nonuniform magnetic field.

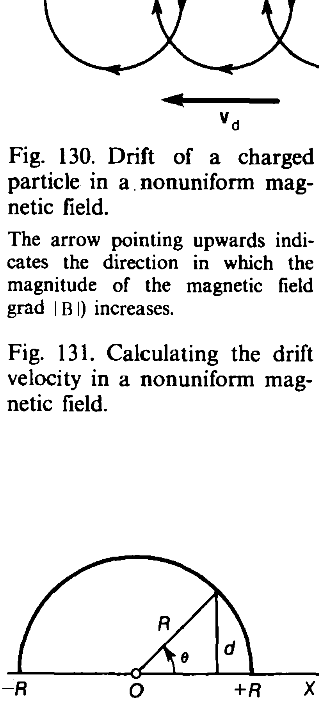


Fig. 132. Calculating the average distance between the points on a semicircle and the diameter.

are different ($R_2 > R_1$). It is obvious from Fig. 131 that for one revolution comprising two motions along semicircles of different radii the point about which the rotation takes place is displaced by $2(R_2 - R_1)$. If the total periods of revolution in circles of radii R_1 and R_2 are denoted by T_1 and T_2, we can write the following expression for the drift velocity:

$$v_d \approx \frac{2(R_2 - R_1)}{\frac{1}{2}(T_2 + T_1)}. \tag{48.39}$$

Expressing the periods T_1 and T_2 and the orbital radii R_1 and R_2 in terms of the magnetic induction with the help of formulas (48.10), we can transform (48.39) to:

$$v_d = \frac{2}{\pi} v \left(\frac{B_1 - B_2}{B_1 + B_2} \right). \tag{48.40}$$

Let us now calculate the average distance between the points on the semicircular trajectory and the diameter on which abrupt variations of the field occur. Obviously, we have (Fig. 132)

$$d = R \sin\theta, \qquad x = R\cos\theta,$$

$$\langle d \rangle = \frac{\int_{-R}^{+R} d \cdot dx}{\int_{-R}^{+R} dx} = \frac{R}{2}\int_0^{\pi} \sin^2\theta \, d\theta = \frac{R}{4}\int_0^{\pi} (1 - \cos 2\theta)\, d\theta = \frac{\pi R}{4}. \tag{48.41}$$

Hence, to a first approximation, we can write

$$B_1 - B_2 \approx \frac{\pi}{4}(R_2 + R_1)\frac{\partial B}{\partial y} \approx \frac{\pi}{4}\frac{m_0 v^2}{e}\frac{B_1 + B_2}{B_1 B_2}\frac{\partial B}{\partial y}, \tag{48.42}$$

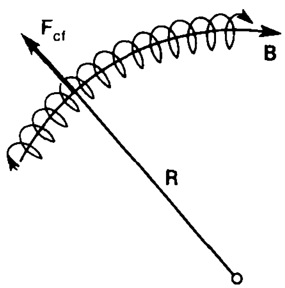


Fig. 133. Drift of a charged particle due to the curvature of the magnetic field lines.

The coordinate system fixed to the centre of rotation of the particle is noninertial and a centrifugal force F_{cf} emerges in it.

where B is the magnetic induction on the midline if the magnetic field is varied smoothly and not abruptly. Under the same assumption, we can write the following approximations:

$$B_1 + B_2 \approx 2B, \quad B_1 B_2 \approx B^2. \tag{48.43}$$

Substituting (48.43) and (48.42) into (48.40), we finally obtain the following formula for the drift velocity:

$$v_d = \frac{E_k}{eB^2} \frac{\partial B}{\partial y}, \tag{48.44}$$

where $E_k = m_0 v^2/2$ is the kinetic energy of the particle. The drift is perpendicular to the magnetic field vector B and to the direction of maximum variation of the magnitude of the magnetic induction. In vector form, Eq. (48.44) can be written as

$$v_d = \frac{E_k}{eB^2}(\mathbf{b}_1 \times \operatorname{grad}|\mathbf{B}|), \tag{48.45}$$

where $\mathbf{b}_1 = \mathbf{B}/B$ is a unit vector along the magnetic field, and grad $|\mathbf{B}|$ is a vector directed towards the maximum increase in the magnitude of vector B and equal to the derivative of $|\mathbf{B}|$ in this direction.

Formula (48.45) has been derived in the first approximation. This means that the variation of the magnetic field at distances of the order of the orbital radius must be small in comparison with the magnetic induction. Mathematically, this condition can be written as follows:

$$R\frac{|\operatorname{grad} B|}{B} \ll 1. \tag{48.46}$$

DRIFT DUE TO THE CURVATURE OF THE MAGNETIC FIELD LINE. In the general case, the magnetic induction of a nonuniform magnetic field cannot be represented by straight lines. The field lines are curves, each point on which corresponds to a definite radius of curvature. A charged particle rotates about the centre of the trajectory, which is as if fixed to the line, and moves along it. Hence it is called the leading centre. The particle's trajectory is helical, wound on the magnetic field line (Fig. 133). We attach the coordinate system to the leading centre. In this coordinate system, the particle is acted upon by an inertial centrifugal force $\mathbf{F}_{cf}$ which is equivalent to the action of an electric field of strength $\mathbf{E}_{eff} = \mathbf{F}_{cf}/e$. Thus, the particle as if moves in crossed fields. This case has just been analyzed by us. The particle must drift in a direction perpendicular both to **B** and to $\mathbf{F}_{cf}$, i.e. perpendicular to the plane (see Fig. 133). The drift velocity can easily be calculated. It is well known that the centrifugal force is defined by the formula

$$F_{cf} = \frac{m_0 v_\parallel^2}{R} = eE_{eff}, \tag{48.47}$$

where $v_\parallel$ is the projection of the particle's velocity onto the direction of the magnetic field. Substituting E_{eff} from (48.47) into (48.38), we obtain the following expression for the drift velocity due to the curvature of a magnetic field line:

$$v_d = \frac{m_0 v_\parallel^2}{eBR} = \frac{2E_{k\parallel}}{eBR} = \frac{v_\parallel^2}{\omega R}, \tag{48.48}$$

where $E_{k\parallel} = m_0 v_\parallel^2/2$ is the kinetic energy of motion along the field line, and ω is the cyclic frequency of the particle.

This drift is added up to the drift caused by the nonuniformity of the magnetic field, whose velocity is given by (48.45).

In view of the above digression, it can be stated that the motion of a particle in a magnetic field consists of three components:

(1) rotation around the field line;

(2) motion of the leading centre along the field line;

(3) the drift of the leading centre in a direction perpendicular to the magnetic field vector **B** and the gradient of the magnetic induction, i.e. in a direction perpendicular to the plane containing the magnetic field line near the given point.

MAGNETIC MOMENT. In many practically important cases, the magnetic field does not vary appreciably over distances of the order of the radius of a particle's trajectory. By analogy with the magnetic moment of a ring current, we can speak of the magnetic moment of the particle moving in a magnetic field. The expedience of such a concept is justified by the fact that this magnetic moment preserves its value in slowly varying magnetic fields and its introduction considerably simplifies the analysis of motion of the particle.

The magnetic moment p_m of a ring current of strength I is, by definition,

$$p_m = IS, \tag{48.49}$$

where S is the area of the region around which the current flows. The charge $|e|$, moving in a circle of radius R and having a period of revolution T, is analogous to the ring current of strength $|e|/T$. Consequently, the magnetic moment of a charged particle can be written in accordance with (48.49) in the form

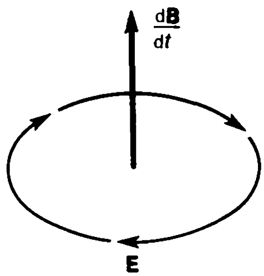


Fig. 134. The existence of a magnetic field generates a vortex electric field.

The directions of variation of electric and magnetic fields are connected through the left-hand screw rule.

$$p_m = \frac{|e|}{T}\pi R^2. \tag{48.50}$$

Considering that

$$T = \frac{2\pi R}{v_\perp}, \quad R = \frac{mv_\perp}{|e|B}, \tag{48.51}$$

we obtain the following expression for the magnetic moment of the particle:

$$p_m = \frac{mv_\perp^2}{2B} = \frac{E_{k\perp}}{B}, \tag{48.52}$$

where $E_{k\perp} = mv_\perp^2/2$ is the kinetic energy corresponding to the velocity component in a plane perpendicular to the magnetic field. •

ADIABATIC INVARIANCE OF THE MAGNETIC MOMENT. **Adiabatic invariance of the magnetic moment** means that its value is preserved in magnetic fields that vary slowly in time or in space.

Let us first consider the case of a slowly varying magnetic field in time (Fig. 134). Suppose that the magnetic field **B** increases in the direction shown in the figure. Then, in accordance with Faraday's law of electromagnetic induction, a particle moving in a circle is subjected to a vortex electric field **E** directed along the particle's trajectory

$$E = \frac{1}{2\pi R}\frac{d\Phi}{dt} = \frac{R}{2}\frac{dB}{dt}, \tag{48.53}$$

where $\Phi = \pi R^2 B$ is the magnetic flux (as stipulated, the magnetic induction varies insignificantly over distances of the order of the orbital radius, and hence the field can be assumed to be uniform). During one revolution of the particle, the field imparts it an energy equal to

$$\Delta(mv_\perp^2/2) = 2\pi RE|e| = |e|\pi R^2\frac{dB}{dt}. \tag{48.54}$$

The slow time variation indicates that during one revolution of the particle in a circle, the magnetic induction does not vary appreciably. The particle's energy varies insignificantly over one revolution, and hence we can divide both sides of (48.54) by T. Taking (48.50) into account, we can write

$$\frac{\Delta(mv_\perp^2/2)}{T} \approx \frac{d}{dt}(mv_\perp^2/2) = \frac{|e|\pi R^2}{T}\frac{dB}{dt} = p_m\frac{dB}{dt}. \tag{48.55}$$

Expressing the energy $mv_\perp^2/2$ through (48.52), we can represent

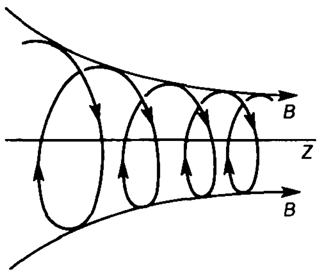


Fig. 135. When a charged particle moves in a region with an increasing magnetic field, its velocity along the field decreases, while the linear velocity of rotational motion increases.

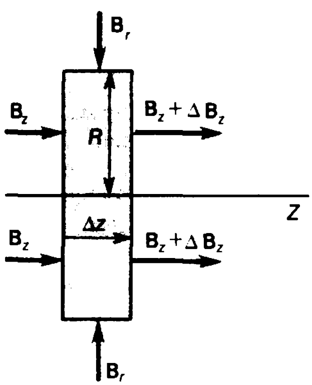


Fig. 136. Computing the radial component of the magnetic field.

(48.55) in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}(p_{\mathrm{m}} B) = p_{\mathrm{m}} \frac{\mathrm{d}B}{\mathrm{d}t},$$

whence

$$\boxed{\frac{\mathrm{d}p_{\mathrm{m}}}{\mathrm{d}t} = 0, \qquad p_{\mathrm{m}} = \mathrm{const},} \tag{48.56}$$

Q. E. D.

Let us now consider the variation of magnetic field in space. Suppose that a particle moves in the direction of variation of the magnetic field (Fig. 135). If the field is intensified along the Z-axis, the magnetic field lines become denser in this direction. In the present case, these lines have a component $\mathbf{B}_r$ along the radius R. In view of the velocity $\mathbf{v}_\perp$, the radial component $\mathbf{B}_r$ gives rise to the Lorentz force

$$\mathbf{F}_\parallel = e\mathbf{v}_\perp \times \mathbf{B}_r, \tag{48.57}$$

which acts along the Z-axis against the direction of crowding of the field lines, i.e. towards the decreasing magnetic field. This force decelerates the moving particle. In order to calculate the decelerating force (48.57), we must know the value of B_r. Since the magnetic field lines have neither beginning nor end, the number of such lines entering a certain volume is equal to the number of lines emerging from it. In other words, the magnetic flux entering a certain volume is equal to the magnetic flux leaving it. We choose this volume in the form of a cylinder of radius R and thickness Δz, its axis coinciding with the Z-axis (Fig. 136). Equating the flux entering the left base and the lateral surface of the cylinder to the flux leaving through the right base, we obtain

$$B_z \pi R^2 + B_r 2\pi R\, \Delta z = (B_z + \Delta B_z)\pi R^2. \tag{48.58}$$

Hence

$$B_r \approx \frac{R}{2}\frac{\Delta B_z}{\Delta z} \approx \frac{R}{2}\frac{\partial B_z}{\partial z}. \tag{48.59}$$

Consequently, the force (48.57) acting along the Z-axis is

$$F_\parallel = |e| v_\perp B_r = \frac{|e| 2\pi R}{T}\frac{R}{2}\frac{\partial B_z}{\partial z} = p_{\mathrm{m}}\frac{\partial B_z}{\partial z}, \tag{48.60}$$

where $v_\perp = 2\pi R/T$, and the definition (48.50) of the magnetic moment has been taken into consideration. It can be seen from Fig. 136 that the direction of this force is opposite to the direction in which the magnetic field increases, i.e. to the positive direction of the Z-axis in the present case. Hence the equation for the velocity component v_z can be written in the form

$$m\frac{\mathrm{d}v_z}{\mathrm{d}t} = -p_\mathrm{m}\frac{\partial B_z}{\partial z} = -p_\mathrm{m}\frac{\partial B}{\partial z}, \tag{48.61}$$

where we have considered that $\partial B_z/\partial z \approx \partial B/\partial z$ in view of the slow variation of the magnetic field. In other words, the component B_z of the magnetic field is replaced by the total magnetic field. This means that the magnetic field lines do not crowd very strongly, i.e. their slope towards the Z-axis is not so large. The minus sign in (48.61) indicates the direction in which the force acts.

Multiplying both sides of (48.61) by v_z and taking into account the equalities

$$\frac{\mathrm{d}v_z}{\mathrm{d}t}v_z = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{v_z^2}{2}\right), \quad \frac{\partial B}{\partial z}v_z = \frac{\mathrm{d}B}{\mathrm{d}t},$$

we can write (48.61) in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{mv_z^2}{2}\right) = -p_\mathrm{m}\frac{\mathrm{d}B}{\mathrm{d}t}. \tag{48.62}$$

Since the total velocity of the particle remains unchanged during its motion in a magnetic field, we can write

$$\frac{mv_z^2}{2} + \frac{mv_\perp^2}{2} = \frac{mv^2}{2} = \text{const},$$

and formula (48.62) assumes the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{mv_\perp^2}{2}\right) = \frac{\mathrm{d}}{\mathrm{d}t}(p_\mathrm{m}B) = p_\mathrm{m}\frac{\mathrm{d}B}{\mathrm{d}t}, \tag{48.63}$$

where we have used the substitution $mv_\perp^2/2 = p_\mathrm{m}B$ in accordance with (48.52). This equation is identical to (48.56), whence it follows that

$$\boxed{p_\mathrm{m} = \text{const}.} \tag{48.64}$$

In other words, the magnetic moment is also conserved during a slow (adiabatic) variation of the magnetic field in space.

Thus, we have proved that the magnetic moment p_m, *defined by Eq.* (48.52), *remains unchanged for a moving particle during a slow variation of the magnetic field in space or in time.*

It should be recalled that the criterion of slow spatial variation of the magnetic field indicates that its magnitude does not change appreciably over distances of the order of the radius of the circular orbit and the displacement of a particle during one revolution, while the criterion of slow time variation of the magnetic field indicates that its value does not change appreciably during one revolution. The constancy of the magnetic moment during slow variations of the magnetic field is termed the adiabatic invariance of the magnetic moment.

Adiabatic invariance means that particles move over the surface of a magnetic tube, i.e. tube formed by the magnetic field lines (see Fig. 135). In order to verify this, we consider that, by definition of a field tube, the magnetic flux penetrating the cross section of the tube does not change along it. The magnetic flux across the cross section of the field tube can be represented in the form

$$\Phi = \pi R^2 B = \frac{2\pi m}{e^2}\frac{E_{k\perp}}{B} = \frac{2\pi m}{e^2} p_m = \mathrm{const}\cdot p_m. \tag{48.65}$$

It is obvious from this formula that the constancy of the magnetic flux along a tube is equivalent to the constancy of the magnetic moment of a particle moving over the surface of the field tube. But since the constancy of the magnetic moment during the motion of a particle has already been proved independently, we can conclude that the particle indeed moves over the surface of a field (see Fig. 135). In order to completely describe the motion of a particle, we must take into account its drift. As a result of drift, the particle moves from one field tube to another in a nonuniform magnetic field in such a way that the magnetic flux enclosed in the tubes remains the same.

MAGNETIC MIRRORS. The decelerating force (48.57) decreases the velocity $v_\parallel$ of a particle moving in the direction of increasing magnetic field. If the increase in the field is quite large, the velocity $v_\parallel$ vanishes as a result of deceleration, after which the particle begins to move in the opposite direction. Thus, a region of increasing magnetic field acts on a particle as a "mirror" that reflects it. Hence an increasing magnetic field is referred to as a "magnetic mirror".

The reflection from a magnetic mirror can also be considered from the point of view of the conservation of a magnetic moment. Since $p_m = mv_\perp^2/(2B)$, the conservation of the magnetic moment in the direction of increasing values B of the magnetic field means that in this case the value of $v_\perp^2$ increases. However, the square of the total velocity $v^2 = v_\perp^2 + v_\parallel^2$ must also remain unchanged. Hence, when a particle moves towards an increasing magnetic field, the value of $v_\parallel^2$ must decrease. In other words, the particle is decelerated.

Let us find the region of field in which the particle is reflected. Suppose that at the initial instant of time the total velocity $\mathbf{v}_0$ of the particle forms an angle θ_0 with the direction

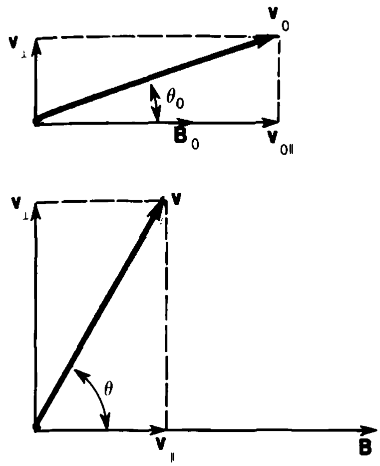


Fig. 137. The velocity of a charged particle in a magnetic field does not change. Hence if the component of velocity perpendicular to the magnetic field increases, the component along the field will decrease.

of the magnetic induction $\mathbf{B}_0$ (Fig. 137). At some other instant of time, when the particle has moved to another point in the field with a magnetic induction $\mathbf{B}$, its velocity remains unchanged, but the angle θ between $\mathbf{B}$ and $\mathbf{v}_0$ changes. This means that the velocity components $v_\parallel$ and $v_\perp$ parallel and perpendicular to the field change. In accordance with (48.52), the conservation of the magnetic moment can be expressed in the form of the equality (see Fig. 137)

$$\frac{\sin^2 \theta_0}{B_0} = \frac{\sin^2 \theta}{B}. \tag{48.66}$$

The particle will be reflected at the point where $\sin \theta = 1$, i.e. where the magnetic induction is given by

$$B = \frac{B_0}{\sin^2 \theta_0}. \tag{48.67}$$

In this field, all particles whose velocity vector at the initial instant of time lies outside a cone of angle θ_0 will be reflected. The reflection criterion is independent of the magnitude of the particle's velocity. All particles the directions of whose velocities lie within a cone of angle θ_0 will not undergo reflection and will penetrate the region of stronger magnetic fields. They can be reflected at the points of the field having a large value of B. However, there exists a certain limiting value B_{max}, and all particles the directions of whose velocites lie outside a cone of angle θ_{min} given by

$$\sin^2 \theta_{min} = \frac{B_0}{B_{max}} \tag{48.68}$$

will be reflected from this region. All particles the directions of whose velocities lie inside this cone will pass through the region with maximum value of the field and will leave the region under consideration. Hence the cone of angle θ_{min} is called the loss cone.

The reflection of particles from magnetic mirrors is used in the equipment for confining charged particles to a limited region of space, for example, in thermonuclear devices. As an example, we can mention the double-necked magnetic bottle in which the role of "corks" is played by magnetic mirrors (Fig. 138). The general nature of motion of particles in a bottle is clear from what has been stated above: the particles move in helical trajectories around the magnetic field lines from one magnetic mirror to the other. As a result of the drift, the particles move from one magnetic field line to another, slowly circumventing the Z-axis. If there were no collisions between particles the directions of whose velocities lie within the loss

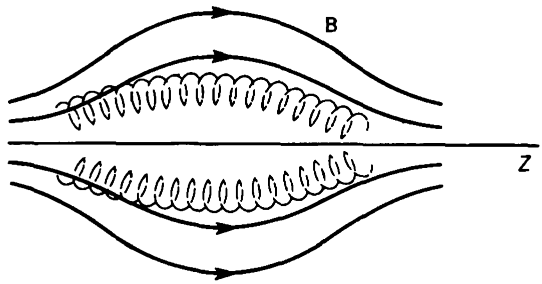


Fig. 138. The mirrors of a magnetic bottle are formed by crowding the magnetic field lines, i.e. in the regions where the magnetic field is intensified.

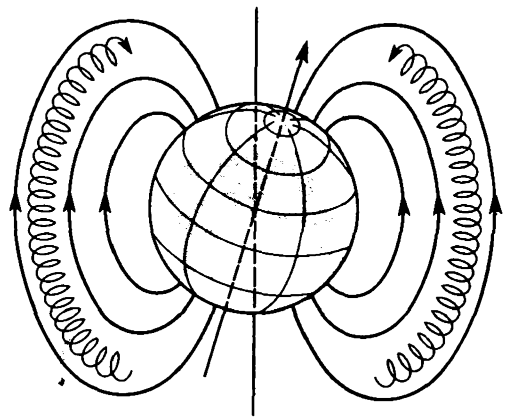


Fig. 139. Radiation belts of the Earth.

cone, all the particles would leave the bottle as a result of a single reflection from magnetic mirrors. The reflected particles the directions of whose velocities lie outside the loss cone would be confined within the bottle for an infinitely long time. In actual practice, however, the particles interact with one another. As a result of collisions, new particles appear in the loss cone and, in turn, quickly leave the bottle. The problem of confining particles within a limited region of space for quite a long time is one of the most important problems of controlled fusion. However, this problem has not been solved so far since the particles always manage to leave the region of the impending thermonuclear reaction in a much shorter time than what the physicists would like them to spend in this region.

RADIATION BELTS OF THE EARTH. The peculiar nature of the motion of charged particles in a magnetic field is responsible for the existence of radiation belts of the Earth. It is well known that a magnetic field exists in space near the Earth. The lines of the Earth's magnetic field start from the north magnetic pole and terminate at the south pole (Fig. 139). The magnetic field lines are crowded at the magnetic poles, i.e. the field is stronger at the poles.

Hence the regions near the poles serve as magnetic mirrors for charged particles. A charged particle moves in a helical trajectory around the field line in the meridional direction from one magnetic pole to the other. On reaching the pole, it is reflected and its direction is reversed. As a result of drift, the particle passes from one line to another, i.e. it changes its longitude by crossing all possible meridians. Consequently, charged particles are confined for a long time near the Earth by its magnetic field. This results in the formation of radiation belts, discovered during the flights of satellites. The radiation belts influence a number of phenomena occurring on the Earth and play a significant role in space flights.

Sec. 49. MOTION IN VARIABLE ELECTROMAGNETIC FIELDS

The most significant aspects of the motion of charged particles in variable electromagnetic fields are described.

MOTION IN THE FIELD OF A PLANE ELECTROMAGNETIC WAVE. In a plane electromagnetic wave, the electric and magnetic fields are perpendicular to each other and to the velocity of propagation which is equal to the velocity of light in vacuum. If the Z-axis coincides with the direction of wave propagation, the electric and magnetic fields of the wave can be represented in the form (Fig. 140)

$$\begin{aligned} E_x &= E_0 \sin(\omega t - kz), \quad E_y = E_z = 0, \\ B_y &= B_0 \sin(\omega t - kz), \quad B_x = B_z = 0, \end{aligned} \tag{49.1}$$

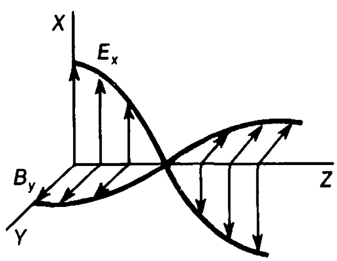


Fig. 140. A plane electromagnetic wave at a certain instant of time.

where $\omega = 2\pi/T$ is the cyclic frequency, and T is the period of the electromagnetic wave. The quantity $k = 2\pi/\lambda$ is called the wave number, and $\lambda = cT$ is called the wavelength. In a plane electromagnetic wave, the amplitudes E_0 and B_0 are connected through the relation $E_0 = cB_0$ which has been proved in the theory of electromagnetic waves.

The electromagnetic wave acts on a charged particle through its electric and magnetic fields. For a plane electromagnetic wave, the projections of the Lorentz force

$$\mathbf{F} = e\mathbf{E} + e\mathbf{v} \times \mathbf{B} \tag{49.2}$$

on the coordinate axes can be written in the form

$$\begin{aligned} F_x &= eE_x + e(v_y B_z - v_z B_y) \\ &= eE_0 \sin(\omega t - kz) - e\dot{z}B_0 \sin(\omega t - kz), \\ F_y &= eE_y + e(v_z B_x - v_x B_z) = 0, \\ F_z &= eE_z + e(v_x B_y - v_y B_x) = e\dot{x}B_0 \sin(\omega t - kz). \end{aligned} \tag{49.3}$$

Hence the equations of motion of a charged particle have the form

$$m\frac{d^2x}{dt^2} = F_x = eE_0\left(1 - \frac{\dot{z}}{c}\right)\sin(\omega t - kz),$$

$$m\frac{d^2y}{dt^2} = F_y = 0, \tag{49.4}$$

$$m\frac{d^2z}{dt^2} = F_z = eE_0\frac{\dot{x}}{c}\sin(\omega t - kz),$$

where $E_0 = cB_0$. If the particle's velocity is small as compared to the velocity of light ($\dot{z}/c \ll 1$), we obtain

$$kz = \omega\int_0^t \frac{\dot{z}}{c}\,dt \ll \omega t, \tag{49.5}$$

where $k = 2\pi/\lambda = \omega/c$. Hence we can neglect $\dot{z}/c$ in comparison with unity and kz in comparison with ωt in the equations of motion (49.4), which now assume the form

$$\ddot{x} = \left(\frac{eE_0}{m}\right)\sin\omega t, \qquad \ddot{z} = \left(\frac{eE_0}{mc}\right)\dot{x}\sin\omega t. \tag{49.6}$$

Integrating the first equation twice, we obtain

$$\dot{x} = -\left(\frac{eE_0}{m\omega}\right)\cos\omega t + \dot{x}_0,$$

$$x = -\left(\frac{eE_0}{m\omega^2}\right)\sin\omega t + \dot{x}_0 t + x_0, \tag{49.7}$$

where $\dot{x}_0$ is the x-projection of the particle's velocity at the instant $t = 0$, and x_0 is the coordinate of the particle at this instant. Substituting the solution (49.7) into the second of Eqs. (49.6), we obtain

$$\ddot{z} = -\frac{1}{2}\left(\frac{eE_0}{m}\right)^2\frac{1}{\omega c}\sin 2\omega t + \frac{eE_0}{mc}\dot{x}_0\sin\omega t. \tag{49.8}$$

Integrating this equation, we get

$$z = \frac{1}{8}\left(\frac{eE_0}{m}\right)^2\frac{1}{\omega^3 c}\sin 2\omega t - \frac{eE_0}{mc\omega^2}\dot{x}_0\sin\omega t + \dot{z}_0 t + z_0. \tag{49.9}$$

The following conclusions can be drawn from the solutions (49.7) and (49.9). If the particle is at rest at the initial instant of time ($\dot{z}_0 = 0$, $\dot{x}_0 = 0$), the electromagnetic wave will cause oscillations of the particle about its mean position, and no net displacement of the particle from its initial position will be observed. If the particle has a certain velocity $\dot{z}_0 \neq 0$, $\dot{x}_0 \neq 0$ at the instant $t = 0$, it will move away from its initial position at

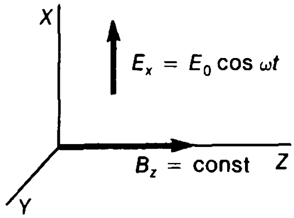


Fig. 141. Arrangement of a coordinate system relative to a constant magnetic field and a variable electric field in which the motion of a charged particle is considered.

this velocity as the average. Moreover, the particle will oscillate during its motion. *Thus, it can be stated that an electromagnetic wave does not change the average velocity of a particle, but causes the velocity to oscillate with the frequency of the electromagnetic wave.*

MOTION IN A VARIABLE ELECTRIC FIELD AND A CONSTANT MAGNETIC FIELD. Suppose that we have variable electric field with frequency ω and a constant magnetic field, which are directed as shown in Fig. 141 and described by the equations

$$E_x = E = E_0 \cos \omega t, \quad E_y = E_z = 0, \quad B_z = B_0, \quad B_x = B_y = 0. \tag{49.10}$$

The equations of motion have the form

$$\ddot{x} = \left(\frac{eE_0}{m}\right) \cos \omega t + \omega_0 \dot{y}, \quad \ddot{y} = -\omega_0 \dot{x}, \tag{49.11}$$

where $\omega_0 = eB_0/m$ is the cyclic frequency of the particle in the magnetic field B_0. This frequency is called the **cyclotron frequency**. We shall assume that at the instant $t = 0$ the particle is at rest at the origin of coordinates, i.e. $x_0 = y_0 = 0$, $\dot{x}_0 = \dot{y}_0 = 0$.

Integrating Eqs. (49.11) and taking these initial conditions into account, we obtain

$$\dot{x} = \left(\frac{eE_0}{m\omega}\right) \sin \omega t + \omega_0 y, \quad \dot{y} = -\omega_0 x. \tag{49.12}$$

Substituting the expression for $\dot{x}$ into the second of Eqs. (49.11), we get

$$\ddot{y} + \omega_0^2 y = -\frac{\omega_0}{\omega}\frac{eE_0}{m} \sin \omega t. \tag{49.13}$$

The nature of motion of the particle depends significantly on the relation between the frequency ω of the variable electric field and the cyclotron frequency ω_0. There can be four important cases: $\omega \ll \omega_0$, $\omega \gg \omega_0$, $\omega = \omega_0$, $\omega \approx \omega_0$. We shall consider each case separately.

Case 1. $\omega \ll \omega_0$. Under this condition, the electric field does not change appreciably during the period of revolution of the particle in the magnetic field. Hence the electric field can be treated as constant while calculating the motion of the particle. The quantity $\sin \omega t$ is a slowly varying function. Averaging both sides of Eq. (49.13) over a large number of periods of oscillations

$$\langle \ddot{y} \rangle = 0, \quad \langle \sin \omega t \rangle \approx \sin \omega t, \tag{49.14}$$

!

A plane electromagnetic wave does not change the average velocity of a charged particle. It only excites oscillations of the velocity about its mean position with the frequency of the wave, without changing the average energy of the particle.
In cyclotron resonance, a variable electric field coupled with a constant magnetic field causes an increase in the kinetic energy of a charged particle.

?

What is the relation between the electric and magnetic field vectors in a plane electromagnetic wave?
Under what condition can we neglect the spatial variation of the field in a wave in comparison with its time variation while solving the equations of motion?

we obtain the following equality:

$$\langle y\rangle = -\frac{1}{\omega\omega_0}\frac{eE_0}{m}\sin\omega t. \tag{49.15}$$

This leads to the following expression for the velocity of displacement of the mean position of the particle:

$$v_{\mathrm{d}} = \frac{\mathrm{d}}{\mathrm{d}t}\langle y\rangle = -\frac{1}{\omega_0}\frac{eE_0}{m}\cos\omega t$$
$$= -\frac{E_0}{B_0}\cos\omega t = -\frac{E}{B_0}. \tag{49.16}$$

This is nothing but ordinary drift in crossed electric and magnetic fields considered earlier for the case of a constant electric field. *The drift velocity changes with a change in the magnitude of the electric field E, i.e. oscillates with a frequency* ω.

Case 2. $\omega \gg \omega_0$. In this case, the electric field changes several times during one revolution of the particle in the magnetic field. Hence its rotation is a slow process, while the variation of the field is a rapid process. We carry out an averaging of (49.13) over many periods of oscillations of the electric field, which together form just a small part of the period of revolution of the particle. Obviously, in this case $\langle\sin\omega t\rangle = 0$, and (49.13) assumes the form

$$\langle\ddot{y}\rangle + \omega_0^2\langle y\rangle = 0. \tag{49.17}$$

Thus, the particle does not drift in this case. It oscillates with the cyclotron frequency ω_0.

Case 3. $\omega = \omega_0$. Under this condition, the phenomenon known as cyclotron resonance is observed. Equations (49.12) assume the form

$$\dot{x} = \left(\frac{eE_0}{m\omega_0}\right)\sin\omega_0 t + \omega_0 y, \quad \dot{y} = -\omega_0 x, \tag{49.18}$$

and instead of (49.13), we get

$$\ddot{y} + \omega_0^2 y = -\left(\frac{eE_0}{m}\right)\sin\omega_0 t. \tag{49.19}$$

The solution of this equation has the form

$$y = -\frac{1}{2}\frac{eE_0}{m\omega_0^2}(\sin\omega_0 t - \omega_0 t\cos\omega_0 t). \tag{49.20}$$

Hence, with the help of the second of Eqs. (49.18), we obtain

$$x = \frac{1}{2}\frac{eE_0}{m\omega_0}t\sin\omega_0 t. \tag{49.21}$$

Thus, the motion of the particle is not periodic in cyclotron resonance.

Let us calculate the kinetic energy of the particle:

$$E_k = \frac{m}{2}(\dot{x}^2 + \dot{y}^2)$$
$$= \frac{1}{8}\frac{e^2E_0^2}{m}\left(t^2 + \frac{\sin^2 \omega_0 t}{\omega_0^2} + \frac{t \sin 2\,\omega_0 t}{\omega_0}\right). \qquad (49.22)$$

The term proportional to t^2 indicates that the particle's energy increases indefinitely. The remaining terms are of no particular significance and characterize the oscillations of the particle's energy about the increasing value determined by the term containing t^2. *Thus, in cyclotron resonance, the energy is transferred from the variable electric field to the particle.*

Case 4. $\omega \approx \omega_0$. Under this condition, there is no exact cyclotron resonance. *The energy is transferred from the variable electric field to the particle only up to a certain maximum value. After this, the particle starts giving back its energy to the electric field, and so on.* This process of energy exchange is periodic and has a frequency

$$\Omega = |\omega - \omega_0|. \qquad (49.23)$$

Without going into detail, we simply mention that this formula expresses the beat frequency obtained as a result of superposition of two harmonic oscillations with comparable frequencies (see Chap. 13).

Let us calculate the maximum energy of the particle. The particle receives energy during half the period corresponding to frequency Ω, i.e. during time π/Ω. During this time, the particle is subjected to the action of an effective electric field E_{eff}, defined as the strength of a constant electric field in which during one period of oscillation the particle receives the same amount of energy as in the variable electric field. In other words, the energy received by the particle is given by $\langle Ev \rangle = E_{eff}v_0$, where v and v_0 are the velocity and the amplitude of oscillations of the velocity respectively. Considering that $\langle Ev \rangle = \langle E_0 \sin \omega t \cdot v_0 \sin \omega_0 t \rangle = E_0 v_0/2$ for $\omega \approx \omega_0$, we obtain $E_{eff} = E/2$. Hence during half a cycle, the particle acquires a momentum p_{max} which is given in accordance with Newton's equation of motion by

$$p_{max} = |e| \langle E_{eff} \rangle \frac{\pi}{\Omega} = \frac{\pi |e| E_0}{2|\omega - \omega_0|}. \qquad (49.24)$$

Consequently, the maximum energy of the particle is equal to

$$E_{max} = \frac{p_{max}^2}{2m} = \frac{\pi^2}{8}\frac{e^2E_0^2}{m(\omega - \omega_0)^2}. \qquad (49.25)$$

PROBLEMS

12.1. A point source emits particles of mass m and charge e in a sharp cone. The uniform magnetic induction vector B is directed parallel to the cone's axis. Assuming that the velocity components of all particles parallel to the induction vector are identical and equal to v, find the distance between the source and the point at which the particles are focussed.

12.2. What will be the mean free path of a relativistic particle of charge e before it comes to a halt in a decelerating uniform electric field E? The particle's velocity is collinear with vector E, its initial total energy is W, and the rest mass is m_0. Answer the question without solving the equations of motion and then solve these equations.

ANSWERS

12.1. $2\pi mv/(eB)$. **12.2.** $(E - m_0c^2)/(|e|W)$.

Chapter 13

Oscillations

Basic idea:
Oscillations are the most general form of motion of dynamic systems about the equilibrium position. For small deviations from the equilibrium position, the oscillations are harmonic and acquire a special significance.

Sec. 50. HARMONIC OSCILLATIONS

The properties of harmonic oscillations and their representation in complex form are described.

ROLE OF HARMONIC OSCILLATIONS IN NATURE. A large number of physical problems can be reduced to an analysis of the behaviour of systems that are slightly deflected from their equilibrium state. For example, suppose that a round ball is at rest in a round-bottomed cup (Fig. 142*a*). Let us find the motion of the ball when it is slightly deflected from its equilibrium position. In order to do so, we must know the force acting on the ball and solve the equation of motion. However, even in this simple case, the force has a very complex dependence on displacement and this complicates an analysis of the equation of motion. As another example, we can consider a ball fixed to a long elastic strip (Fig. 142*b*). In the equilibrium position, the strip is slightly bent and the ball is at rest in a certain position. We are interested in finding the motion of the ball in the vertical direction if it is deflected from the equilibrium position and then released. In this case, the force acting on the ball is expressed as a composite function of its deviation from the equilibrium position in the vertical direction, and the solution of the problem is fraught with the same difficulties that were encountered in the first case considered above.

In most cases of practical importance, however, we are interested in analyzing the behaviour of a system only at small deviations from the equilibrium position and not for all possible deviations. Under this condition, the problem be-

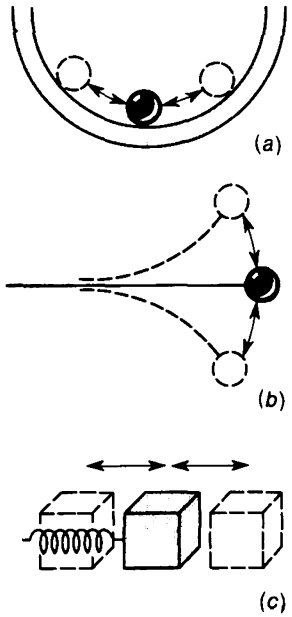


Fig. 142. Oscillations of various systems for small deviations.

comes much simpler. Irrespective of the complexity of the law of action of the force $f(x)$, this function can always be represented in the form of a Taylor series:

$$f(x) = f(0) + xf'(0) + \frac{x^2}{2!}f''(0) + \frac{x^3}{3!}f'''(0) + \ldots . \tag{50.1}$$

This is a purely mathematical statement, and the conditions under which such a function can be expanded into a series are considered in mathematics. For our purpose, it is sufficient to note that these conditions are usually satisfied for most of the laws of action of the force $f(x)$ encountered in physics. Obviously, $f(0) = 0$ in view of the fact that the point $x = 0$ is the equilibrium point, and hence the force at this point is zero. Two cases are possible here: either $f'(0) \neq 0$, or $f'(0) = 0$. In the former case, the term $xf'(0)$ is the principal term in the expansion (50.1). All subsequent terms of the series are proportional to x^2, x^3, etc. and are infinitely small in comparison with the first term for quite small values of x. Hence, while analyzing very small deviations x, we can take the force equal to $xf'(0)$. Since the point $x = 0$ must be the point of stable equilibrium, the force $xf'(0)$ must always be directed towards the point $x = 0$. This means that $f'(0) < 0$. If $f'(0) = 0$, we must consider the third term which is proportional to x^2. This term must be zero if the point $x = 0$ is the point of stable equilibrium. This is due to the fact that this term has the same sign for positive as well as negative values of x.

Hence when the point is deflected in one direction, the force corresponding to it tends to return the point to the equilibrium position, while when the point is deflected in the opposite direction, the force tends to remove the point farther away from the equilibrium position. Consequently, if this term were not zero, the point $x = 0$ would not be the point of stable equilibrium. Hence this term must be zero, i.e. $f''(0) = 0$.

Thus, the next nonzero term can be $x^3 f'''(0)/3!$ While analyzing small deviations in the case $f'(0) = 0$, we must use this term as the expression for the force. Although it is a little more complicated than the term $xf'(0)$, it is still quite simple as compared to the original function $f(x)$. In this case, the oscillations become much more complicated and nonlinear. The main features of these oscillations will be considered later.

In real physical systems, the term $xf'(0)$ is usually not zero, and the equation of motion for small deviations x from the equilibrium position has the following form:

$$m\frac{d^2x}{dt^2} = xf'(0) = -Dx, \tag{50.2}$$

where we have considered that $f'(0) < 0$ and have used the notation $D = -f'(0) > 0$.

This type of equation is obtained in the analysis of many physical phenomena. In the present example, x is the distance from the equilibrium position. However, x could mean, say, the charge of a capacitor in an LC circuit. If the physical factors are such that they strive to restore the zero value of the charge on the capacitor, the equation for small deviations of the charge from zero will have the form (50.2).

An equation of the type (50.2) is called the equation of harmonic oscillations, and the system in which these small oscillations take place is called a linear, or harmonic, oscillator. A familiar example of such a system is a body suspended on an elastic spring (Fig. 142*c*). According to Hooke's law, an extension or compression of the spring is accompanied by an opposing force that is proportional to the extension or compression. In other words, the expression for the force exerted by the spring has the form $F = -Dx$, and we arrive at the equation of a linear oscillator. Thus, a body oscillating at the end of a spring is a model of a linear oscillator.

If in addition to the term proportional to the first power of the deviation, we also retain in the expansion for the force the term proportional to x^2 or x^3 and leading to nonlinearity of oscillations, the oscillatory system obtained in this way is called an anharmonic oscillator. The main features of such an oscillator will be described in Sec. 51.

Other examples of a linear oscillator are a simple and a compound pendulum for quite small angles of deflection, which were considered in Sec. 34. For the model of a linear oscillator, we can take either a load suspended on a spring (Fig. 142*c*) or a pendulum.

The fact that most of the physical systems behave like linear oscillators for small angles of deflection makes an analysis of their motion extremely important in all branches of physics.

!

EQUATION OF HARMONIC OSCILLATIONS. Equation (50.2) describing the motion of a linear oscillator can conveniently be written in the form

$$\ddot{x} + \omega^2 x = 0, \tag{50.3}$$

where $\omega^2 = D/m > 0$. The time derivatives are marked by dots.

A real physical oscillation is represented by the real or imaginary part of the complex form of an oscillation. The complex form is more convenient for describing an oscillation because of the ease and clarity of mathematical operations under such a representation.

HARMONIC FUNCTIONS. A direct verification shows that $\sin \omega t$ and $\cos \omega t$ are the particular solutions of Eq. (50.3). This equation is linear. The sum of the solutions of a linear equation and the product of a solution by an arbitrary constant is also a solution of this equation. Hence the general

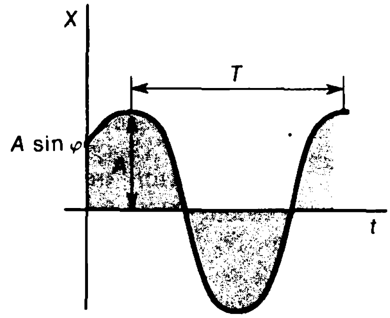


Fig. 143. Complex representation of harmonic oscillations.

solution of Eq. (50.3) has the form

$$x(t) = A_1 \sin \omega t + A_2 \cos \omega t, \tag{50.4}$$

where A_1 and A_2 are constants. A function of this type is called a harmonic function.

AMPLITUDE, PHASE AND FREQUENCY. It is expedient to transform the expression (50.4) as follows:

$$\begin{aligned} &A_1 \sin \omega t + A_2 \cos \omega t \\ &= \sqrt{A_1^2 + A_2^2}\left(\frac{A_1}{\sqrt{A_1^2 + A_2^2}} \sin \omega t + \frac{A_2}{\sqrt{A_1^2 + A_2^2}} \cos \omega t\right) \\ &= A(\cos \varphi \sin \omega t + \sin \varphi \cos \omega t) = A \sin(\omega t + \varphi), \end{aligned} \tag{50.5}$$

where we have put $\cos \varphi = A_1/\sqrt{A_1^2 + A_2^2}$ and $\sin \varphi = A_2/\sqrt{A_1^2 + A_2^2}$, and introduced the notation $A = \sqrt{A_1^2 + A_2^2}$. Thus, the equation (50.4) of harmonic oscillations can be represented in the form

$$\boxed{x = A \sin(\omega t + \varphi)} \tag{50.6a}$$

or

$$\boxed{x = B \cos(\omega t + \varphi).} \tag{50.6b}$$

The plot of this function with the notation introduced in (50.6a and b) is shown in Fig. 143. The quantities A and ω are respectively called the amplitude and the frequency of harmonic oscillations, while the quantity appearing in the argument of the sine (or cosine), i.e. $\omega t + \varphi$, is called the phase of the oscillations. The value of the phase φ for $t = 0$ is called the initial phase, or just the phase, of the oscillations. It can be seen from (50.6a and b) that the value of x is repeated after an interval $T = 2\pi/\omega$ of time. Such a function is called a periodic function, and T is called its period. Hence harmonic oscilla-

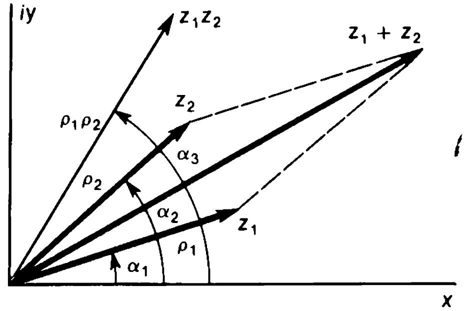


Fig. 144. Graphical representation of complex numbers and operations involving them.

tions are periodic oscillations. However, not every periodic function is a harmonic function. It will be harmonic only if it can be represented in the form (50.6a and b) with a definite frequency, phase and amplitude.

COMPLEX REPRESENTATION OF HARMONIC OSCILLATIONS. The analysis of harmonic oscillations involves their addition or decomposition into harmonics, the solution of more complicated equations than (50.3), and so on. All these operations can considerably be simplified with the help of the theory of complex numbers and the representation of harmonic oscillations in complex form.

In the Cartesian coordinate system, the real part of a complex number is laid off along the abscissa axis and the imaginary part along the ordinate axis (Fig. 144). We then use Euler's formula

$$e^{i\alpha} = \cos\alpha + i\sin\alpha \quad (i^2 = -1), \tag{50.7}$$

which makes it possible to express any complex number $z = x + iy$ in exponential form (see Fig. 144):

$$z = \rho e^{i\alpha}, \quad \rho = \sqrt{x^2 + y^2}, \quad \tan\alpha = \frac{y}{x}. \tag{50.8}$$

The quantity ρ is called the modulus of the complex number, and α is called its argument.

Every complex number z can be represented on a complex plane as a vector drawn from the origin of coordinates to the point with coordinates (x, y). Complex numbers are added in accordance with the parallelogram law. *Hence, for the sake of brevity, we can refer to complex numbers as vectors when speaking of their addition.*

It is more convenient to carry out the multiplication of

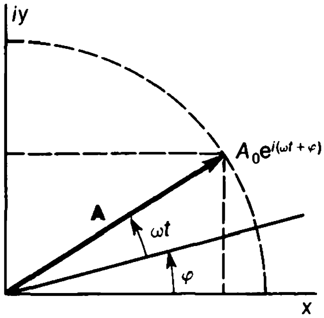


Fig. 145. Complex representation of harmonic oscillations.

complex numbers in complex form:

$$z = z_1 z_2 = \rho_1\rho_2 e^{i(\alpha_1+\alpha_2)},$$
$$z_1 = \rho_1 e^{i\alpha_1}, \quad z_2 = \rho_2 e^{i\alpha_2}. \tag{50.9}$$

Thus, when complex numbers are multiplied, their moduli are multiplied, while their arguments are added.

We shall not go into the detail of these purely mathematical questions. Readers desirous of a better understanding of these problems are referred to textbooks on the theory of complex numbers.

Instead of the real form of notation (50.6a and b) for harmonic oscillations we can also use the complex form:

$$\tilde{x} = Ae^{i(\omega t+\varphi)}. \tag{50.10}$$

The quantity $\tilde{x}$ in this equation is complex and cannot describe a real physical deviation like the real quantity x in (50.6a). However, the imaginary part of this quantity can be treated as the real harmonic oscillation (50.6a). On the other hand, the real part of (50.10), i.e. $A\cos(\omega t+\varphi)$, also represents a real harmonic oscillation. Hence a harmonic oscillation can be written in the form (50.10) and used for all calculations and discussions.

A transition to physical quantities can be made by taking the real or imaginary part of the expression in the final form. We shall illustrate this by considering several examples.

Figure 145 shows the plot of harmonic oscillations in the complex form (50.10). The values of various quantities appearing in (50.10) can be noted directly from the figure: A is the amplitude, φ the initial phase, and $\omega t+\varphi$ the phase of the oscillations. The complex vector A rotates about the origin of coordinates in the counterclockwise direction with an angular frequency $\omega = 2\pi/T$, T being the period of oscillations. The projections of the rotating vector A onto the horizontal and vertical axes are real physical oscillations in which we are interested.

?

Under what conditions is it impossible to reduce the effect of small deviations from the equilibrium position to the introduction of a linear term?
Define the frequency, amplitude and phase of a harmonic oscillation.
For a system in equilibrium at the point $x = 0$, why must $f''(0)$ be zero if $f'(0) = 0$? If $f'(0) \neq 0$, can $f''(0)$ be nonzero?

ADDITION OF HARMONIC OSCILLATIONS WITH THE SAME FREQUENCY. Suppose that we are given two harmonic oscillations with the same frequency but with different initial phases and amplitudes:

$$x_1 = A_1\cos(\omega t+\varphi_1), \quad x_2 = A_2\cos(\omega t+\varphi_2). \tag{50.11}$$

It is required to find the resultant oscillation $x = x_1 + x_2$. When represented in the form (50.10), the harmonic oscillations (50.11) express the real part of the resultant oscillations. Hence the required result of the addition of the oscillations

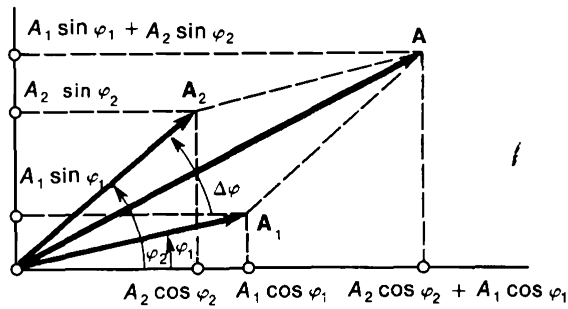


Fig. 146. Addition of harmonic oscillations in complex form.

(50.11) is the real part of the complex number

$$\tilde{x} = \tilde{x}_1 + \tilde{x}_2 = A_1 e^{i(\omega t + \varphi_1)} + A_2 e^{i(\omega t + \varphi_2)}$$
$$= e^{i\omega t}(A_1 e^{i\varphi_1} + A_2 e^{i\varphi_2}). \qquad (50.12)$$

The two quantities within the parentheses can easily be added in vector form (Fig. 146). It follows directly from Fig. 146 that

$$A_1 e^{i\varphi_1} + A_2 e^{i\varphi_2} = A e^{i\varphi}, \qquad (50.13)$$

$$A^2 = A_1^2 + A_2^2 + 2A_1 A_2 \cos(\varphi_2 - \varphi_1), \qquad (50.13a)$$

$$\tan \varphi = \frac{A_1 \sin \varphi_1 + A_2 \sin \varphi_2}{A_1 \cos \varphi_1 + A_2 \cos \varphi_2}. \qquad (50.13b)$$

Consequently, instead of (50.12), we obtain

$$\tilde{x} = \tilde{x}_1 + \tilde{x}_2 = A e^{i(\omega t + \varphi)}, \qquad (50.14)$$

where A and φ are defined by (50.13a) and (50.13b) respectively. Hence the sum of the harmonic oscillations (50.11) is given by the formula

$$x = x_1 + x_2 = A \cos(\omega t + \varphi),$$

the quantities A and φ having the same values as in (50.14).

The properties of the sum of harmonic oscillations can be determined from an inspection of Fig. 146. Obviously, owing to the presence of the common factor $e^{i\omega t}$ in (50.12), the entire pattern shown in Fig. 146 rotates about the origin of coordinates in the counterclockwise direction with an angular velocity ω. The amplitude of the oscillations attains its maximum value $A_1 + A_2$ at $\varphi_2 = \varphi_1$. The minimum value of the amplitude is attained at $\varphi_2 - \varphi_1 = \pm\pi$. In this case, the complex vectors expressing the oscillation components are directed against each other, and hence the minimum amplitude is $|A_2 - A_1|$. The variation of the phase φ can also be followed by looking at Fig. 146.

Thus, the sum of harmonic oscillations with identical frequency is a harmonic oscillation with the same frequency, its amplitude and phase being given by (50.13a) *and* (50.13b).

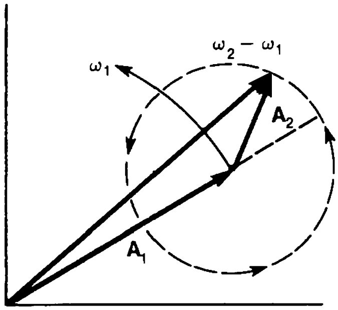


Fig. 147. Addition of harmonic oscillations with nearly equal frequencies ($\omega_1 \approx \omega_2$) in complex form.

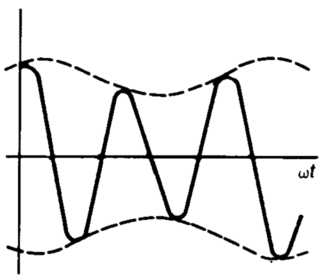


Fig. 148. Beats resulting from addition of oscillations with nearly equal frequencies.

Beat period $T = 2\pi/|\omega_2 - \omega_1|$.

?

What is the reason behind the complex representation of harmonic oscillations?

How are the argument and modulus of a complex number determined?

What is the relation between the addition of complex numbers and the addition rule?

What happens to the moduli and arguments of complex numbers upon their multiplication?

What are beats? Are beats harmonic oscillations?

ADDITION OF HARMONIC OSCILLATIONS WITH NEARLY EQUAL FREQUENCIES. BEATS. Let us denote the frequencies of the components being added by ω_1 and ω_2, and assume that $\omega_1 \approx \omega_2$, $|\omega_1 - \omega_2| \ll \omega_1 \approx \omega_2$. The equations for these oscillations have the form

$$x_1 = A_1 \cos(\omega_1 t + \varphi_1), \quad x_2 = A_2 \cos(\omega_2 t + \varphi_2). \tag{50.15}$$

Each of the oscillations (50.15) can be represented in the complex form (50.10), and their addition can be carried out according to the vector addition rule, taking the tip of the first vector as the beginning of the second.

For the sake of definiteness, let us assume that $A_1 > A_2$. The sum of vectors $\tilde{x}_1$ and $\tilde{x}_2$ at a certain instant of time can then be represented as shown in Fig. 147. With the passage of time, this pattern will change as follows. Vector $\tilde{x}_1$ will rotate about the origin of coordinates with an angular frequency ω_1, while vector $\tilde{x}_2$ will rotate with respect to vector $\tilde{x}_1$ about its tip with an angular frequency $\omega_2 - \omega_1$. If $\omega_2 > \omega_1$, the rotation of vector $\tilde{x}_2$ about the tip of $\tilde{x}_1$ will take place in the same direction as that of vector $\tilde{x}_1$ about the origin of coordinates as shown in Fig. 147. For $\omega_2 < \omega_1$, vector $\tilde{x}_2$ rotates in the opposite direction.

The time variation of this pattern can be described as follows: since $|\omega_2 - \omega_1| \ll \omega_1 \approx \omega_2 \approx \omega$, the entire oscillation pattern rapidly rotates about the origin of coordinates, and the mutual arrangement of vectors $\tilde{x}_1$ and $\tilde{x}_2$ changes very insignificantly during one revolution. Hence the resultant of the two oscillations is a harmonic oscillation with frequency ω and amplitude $\tilde{x}_1 + \tilde{x}_2$ over a large number of oscillations. However, the relative orientation of vectors $\tilde{x}_1$ and $\tilde{x}_2$ does change, albeit slowly. Thus, the amplitude of oscillations slowly changes with a frequency $|\omega_2 - \omega_1|$ from $A_1 + A_2$ to $|A_1 - A_2|$.

Consequently, the sum of two harmonic oscillations with nearly equal frequencies is an oscillation with a varying amplitude. This oscillation is only nearly harmonic with a frequency $\omega_1 \approx \omega_2 \approx \omega$, and its amplitude changes with a frequency $|\omega_2 - \omega_1|$ from its maximum value $A_1 + A_2$ to its minimum value $|A_1 - A_2|$. The real components of this oscillation have the form shown in Fig. 148. Oscillations of the amplitude with a frequency $\Omega = |\omega_2 - \omega_1|$ are called beats, and the frequency Ω is called the beat frequency. Beats appear as a result of addition of two harmonic oscillations with nearly equal frequencies. If the amplitudes of the component oscillations are nearly equal, i.e. $A_1 \approx A_2$, the amplitude of the resultant oscillation at the minimum is nearly zero, i.e. the oscillation is completely suppressed.

Sec. 51. NATURAL OSCILLATIONS

The methods of analysis of natural oscillations are described.

DEFINITION. Natural oscillations *of a system are oscillations produced only by the internal forces without any external influence on the system.* The harmonic oscillations considered in the previous section are the natural oscillations of a linear oscillator. In principle, natural oscillations can also be anharmonic. However, for quite small deviations from the equilibrium position, in most cases of practical importance, these can be reduced to harmonic oscillations as has been explained above.

INITIAL CONDITIONS. Harmonic oscillations are completely characterized by the frequency, amplitude and the initial phase. The frequency of oscillations depends on the physical properties of the system. For example, when a linear oscillator is a point mass oscillating under the elastic force exerted by a spring, the properties of the elastic spring are taken into consideration through the coefficient of elasticity D, and those of the point mass, through its mass m; $\omega = D/m$.

To determine the amplitude and the initial phase of oscillations, we must know the position and the velocity of the point mass at a certain instant of time. If the equation of oscillations is written in the form

$$x = A \cos (\omega t + \varphi), \tag{51.1}$$

and the coordinate and the velocity of the point mass at the instant $t = 0$ are equal to x_0 and v_0 respectively, we can write in accordance with (51.1):

$$x_0 = A \cos \varphi, \quad \dot{x}_0 = v_0 = \left.\frac{\mathrm{d}x}{\mathrm{d}t}\right|_{t=0} = -A\omega \sin \varphi. \tag{51.2}$$

These two equations can be used for determining the unknown amplitude and the initial phase:

$$A = \sqrt{x_0^2 + \frac{v_0^2}{\omega^2}}, \quad \tan \varphi = -\frac{v_0}{x_0 \omega}. \tag{51.3}$$

Thus, knowing the initial conditions, we can completely describe a harmonic oscillation.

ENERGY. The concept of potential energy is meaningful only if the forces are of a potential nature. In the one-dimensional motion between two points, there can be only one path along which the motion can take place. Consequently, the potential nature of forces is automatically ensured, and any force can be treated as a potential force if it depends only on the coordinates. The last stipulation is quite significant. For example, the frictional force is not a potential force in the one-dimensional

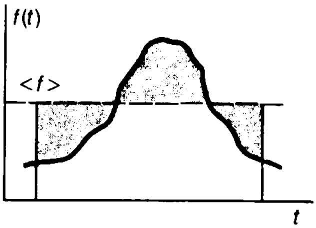


Fig. 149. Determination of time-average.

case either. This is so because (the direction of) force depends on (the direction of) velocity.

In the case of a linear oscillator, it is convenient to assume that the potential energy of a point is zero in its equilibrium position (at the origin of coordinates). In this case, considering that $\bar{F} = -Dx$ and taking into account formula (25.20) connecting the potential energy E_p with the force, we directly obtain the following expression for the potential energy of a linear oscillator:

$$E_p(x) = \frac{Dx^2}{2} = \frac{m\omega^2 x^2}{2}. \tag{51.4}$$

The energy conservation law has the following form in this case:

$$\frac{m\dot{x}^2}{2} + \frac{m\omega^2 x^2}{2} = \text{const}. \tag{51.5}$$

Of course, this law can also be obtained directly from the equation of motion (50.3) by multiplying both sides by $\dot{x}$ and then proceeding in the same way as in the transition from (25.1) to (25.5).

Two important conclusions can be drawn from the energy conservation law (51.5).

1. *The maximum kinetic energy of an oscillator is equal to its maximum potential energy.* This is obvious since an oscillator has its maximum potential energy when the oscillating point is deflected to its extreme position where its velocity (and hence the kinetic energy) is zero. The oscillator has its maximum kinetic energy at the instant when the point passes through the equilibrium position ($x = 0$) where its potential energy is zero. Hence, denoting the maximum velocity by V, we can write

$$\frac{1}{2}mV^2 = \frac{1}{2}m\omega^2 A^2. \tag{51.6}$$

2. *The average kinetic energy of an oscillator is equal to its average potential energy.*

First of all, let us define the mean of a quantity. If a certain quantity f depends on time, i.e. is a function of time, the mean value of this quantity over the interval of time between t_1 and t_2 is given by the formula

$$\langle f \rangle_t = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} f(t)\,\mathrm{d}t. \tag{51.7}$$

If $f(t)$ is represented in a graphic form (Fig. 149), the mean value $\langle f \rangle_t$ corresponds to the height of the rectangle whose area is equal to the area occupied between the curve $f(t)$ and the t-axis from t_1 to t_2. It should be recalled that the area under the t-axis is assumed to be negative.

Since the law of motion for a linear oscillator is described by the formula

$$x(t) = A\cos(\omega t + \varphi), \tag{51.8}$$

its velocity will be equal to

$$\dot{x} = -A\omega\sin(\omega t + \varphi). \tag{51.9}$$

The expressions for the kinetic and potential energies have the form

$$\begin{aligned} E_k(t) &= \frac{m\dot{x}^2}{2} = \frac{m\omega^2 A^2}{2}\sin^2(\omega t + \varphi), \\ E_p(t) &= \frac{m\omega^2 A^2}{2}\cos^2(\omega t + \varphi). \end{aligned} \tag{51.10}$$

The period of one oscillations is taken as the interval of time over which the mean value is determined. The computation of the mean values of $\langle E_k \rangle$ and $\langle E_p \rangle$ is reduced to the determination of the mean values of $\cos^2(\omega t + \varphi)$ and $\sin^2(\omega t + \varphi)$. These values can easily be found:

$$\begin{aligned} \langle \cos(\omega t + \varphi) \rangle_t &= \frac{1}{T}\int_0^T \cos(\omega t + \varphi)\,dt \\ &= \frac{1}{T}\int_0^T \frac{1}{2}[1 + \cos 2(\omega t + \varphi)]\,dt \\ &= \frac{1}{2}\frac{1}{T}\left[t + \frac{1}{2\omega}\sin 2(\omega t + \varphi)\right]_0^T = \frac{1}{2}, \end{aligned} \tag{51.11}$$

where T is the period of oscillations, $\omega T = 2\pi$. Similarly, we get

$$\langle \sin^2(\omega t + \varphi) \rangle_t = \frac{1}{2}. \tag{51.12}$$

Formulas (51.11) and (51.12) are quite important and must be committed to memory. Taking these expressions and (51.10) into consideration, we obtain

$$\boxed{\langle E_p \rangle_t = \langle E_k \rangle_t.} \tag{51.13}$$

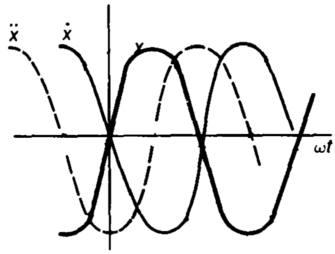


Fig. 150. Plots of displacement, velocity and acceleration in harmonic oscillations.

This means that the average kinetic energy of an oscillator is equal to its average potential energy. The angle brackets in this equation have been assigned a subscript t to emphasize that the averaging is carried out over time.

Whenever we refer to the mean value of a quantity, we must always clarify the variable over which the averaging is taken since the result of averaging over some other variable will generally be quite different. In most cases, however, the variable over which the averaging is taken is quite obvious and the subscript can be omitted.

RELATION BETWEEN DISPLACEMENT, VELOCITY AND ACCELERATION. The displacement and velocity can be obtained from the expressions (51.8) and (51.9), while the acceleration is

$$\ddot{x} = -A\omega^2 \cos(\omega t + \varphi). \tag{51.14}$$

Let us plot these quantities on the same diagram (Fig. 150). The ordinate axis represents the quantities of different dimensions. Hence we choose the amplitudes of the respective oscillations in such a way that their maxima have the same height, as shown in the figure. The displacement, velocity and acceleration are represented by identical curves displaced relative to one another in the direction of the ωt-axis. It can clearly be seen that the velocity curve is displaced relative to the displacement curve by $\Delta(\omega t) = \pi/2$ to the left, while the acceleration curve is displaced relative to the velocity curve by the same amount in the same direction.

Consequently, in harmonic oscillations, the velocity leads the displacement by $\pi/2$ in phase, while the acceleration leads the velocity by $\pi/2$ in phase. Thus, the acceleration leads the displacement by π in phase. Of course, we can also state that the displacement lags behind the velocity by $\pi/2$ in phase, and so on.

!

The frequency of oscillations is determined by the physical properties of a system, while the amplitude and phase are determined by the initial conditions.
In harmonic oscillations, the velocity leads the displacement by $\pi/2$ in phase, while the acceleration leads the velocity by $\pi/2$ in phase.
The most significant feature of nonlinear oscillations is the emergence of higher harmonics. The type of harmonics that emerge depends on the nature of nonlinearity of the force.

NONLINEAR OSCILLATIONS. If in addition to the linear term $xf'(0)$ in the expansion (50.1) for the force, the next term, say, $x^2f''(0)/2!$, is also significant, we must consider the following equation of motion instead of (50.2):

$$m\frac{d^2x}{dt^2} = xf'(0) + \frac{x^2}{2!}f''(0). \tag{51.15}$$

While discussing the expansion of the force into the series (50.1), we mentioned that if the system oscillates about the position $x = 0$ of stable equilibrium, then $f'(0) = 0$ means that $f''(0) = 0$ as well. If this is not so, the point $x = 0$ cannot be the point of stable equilibrium. Obviously, if $f'(0) \neq 0$, we must have $f'(0) < 0$, the derivative $f''(0)$ need not be zero and may have a positive or a negative sign. This is the situation that was considered in (51.15). Moreover, it is assumed that the quantity

?

What is the relation between the kinetic and potential energies in harmonic oscillations?
What is the relation between the amplitudes of velocity and displacement in harmonic oscillations?
How does the frequency of natural oscillations change with an increase in the mass of an oscillating point?

$f''(0)$ is very small, and hence the last term on the right-hand side of (51.15) is small in comparison with other terms. We divide (51.15) by m and write it as follows:

$$\ddot{x} + \omega_0^2 x = \varepsilon\omega_0^2 x^2, \tag{51.16a}$$

where, by analogy with (50.3), the following notation has been used:

$$\omega_0^2 = -\frac{f'(0)}{m}, \quad \varepsilon = \frac{f''(0)}{2m\omega_0^2} = -\frac{f''(0)}{2f'(0)}. \tag{51.16b}$$

The quantity ε is called the parameter of smallness of the term proportional to the square of displacement. It can be seen directly from (51.16a) that this parameter has dimensions of reciprocal length and can therefore be represented in the form $\varepsilon = 1/L$, where L is a large length. We can now explain the meaning of smallness of the quantity ε more clearly: if the displacements x are quite small and satisfy the relation $x \ll L = 1/\varepsilon$, the term on the right-hand side of (51.16a) can be treated as a small term. In this case, this term is called a perturbation, and the method used for finding an approximate solution of the equation of motion is called the perturbation method or perturbation theory. Let us consider the essence of this theory and the basic properties of nonlinear oscillations on the basis of Eq. (51.16a).

For $\varepsilon = 0$, i.e. in the absence of perturbation, the system performs harmonic oscillations. Suppose that these oscillations have the form

$$x_0(t) = A_0 \sin \omega_0 t. \tag{51.17}$$

Such oscillations represent unperturbed motion. To consider the right-hand side of (51.16a) as a perturbation, the amplitude A_0 must not be too large. It must satisfy the condition $\varepsilon A_0 \ll 1$. If this condition is not satisfied, we cannot use the perturbation theory. The solution in the presence of perturbations, i.e. for $\varepsilon \neq 0$, can be represented in the form

$$x = A_0 \sin \omega_0 t + x_1(t), \tag{51.18}$$

where $x_1(t)$ is a correction to the unperturbed motion. As $\varepsilon \to 0$, the quantity $x_1(t)$ must also tend to zero. Hence, $x_1(t)$ is small in comparison with the displacements in the unperturbed motion. In other words, the relation $|x_1| \ll A_0$ is satisfied. Substituting the expression (51.18) for x into (51.16a), we arrive at the following equation for $x_1(t)$:

$$\ddot{x}_1 + \omega_0^2 x_1 = \varepsilon\omega_0^2 (A_0^2 \sin^2 \omega_0 t + 2A_0 x_1 \sin \omega_0 t + x_1^2). \tag{51.19}$$

The second and third terms in the parentheses on the right-hand side are much smaller than the first term in view of the inequality $|x_1| \ll A_0$. Hence they can be neglected as compared to the first term. Equation (51.19) can then be represented in the form

$$\ddot{x}_1 + \omega_0^2 x_1 = \frac{\varepsilon\omega_0^2}{2} A_0^2 (1 - \cos 2\omega_0 t), \tag{51.20}$$

where we have used the formula $\sin^2 \omega_0 t = (1/2)(1 - \cos 2\omega_0 t)$. The solution of this equation is sought in the form

$$x_1 = a_1 + b_1 \cos 2\omega_0 t, \tag{51.21}$$

where a_1 and b_1 are constants. Substituting (51.21) into (51.20), we obtain

$$\omega_0^2 a_1 + b_1(-4\omega_0^2 + \omega_0^2)\cos 2\omega_0 t$$
$$= \frac{\varepsilon\omega_0^2}{2} A_0^2 - \frac{\varepsilon\omega_0^2}{2} A_0^2 \cos \omega 2\omega_0 t. \tag{51.22}$$

Since this equation must be satisfied for all instants of time, the coefficients of $\cos 2\omega_0 t$ on both sides of this equation must be equal. From this condition, we obtain

$$b_1(-4\omega_0^2 + \omega_0^2) = -\frac{\varepsilon\omega_0^2}{2} A_0^2, \tag{51.23}$$

$$b_1 = \frac{\varepsilon A_0^2}{6}. \tag{51.24}$$

For this value of b_1, the time-dependent terms in (51.22) cancel out. From the remaining terms, we obtain an equation according to which

$$a_1 = \frac{\varepsilon A_0^2}{2}. \tag{51.25}$$

Consequently, the solution (51.18) can be written in the following form by taking into account the first correction:

$$x = A_0 \sin \omega_0 t + \frac{\varepsilon A_0^2}{2} + \frac{\varepsilon A_0^2}{6} \cos 2\omega_0 t. \tag{51.26}$$

The most peculiar feature of this solution is that it contains a term with $\cos 2\omega_0 t$. *This shows that when the force contains a nonlinear term proportional to x^2, the oscillations acquire a term with double the frequency, i.e. with a frequency $2\omega_0$, called the* second harmonic. In the absence of the nonlinear term, the oscillations contain only a term with the fundamental frequency ω_0. Continuing the solution of Eq. (51.16a) and determining the next, smaller, corrections, we can verify that they contain higher frequencies $n\omega_0$ that are multiples of the fundamental frequency. In other words, the oscillations contain higher harmonics. *It can be stated therefore that the most* *significant consequence of the nonlinearity in the force is the emergence of higher harmonics in the oscillations.*

Further, it can be seen from (51.26) that both components of the oscillations with frequencies ω_0 and $2\omega_0$ appear near the point $x = (1/2)\varepsilon A_0^2$, and not about the point $x = 0$. *This means that the presence of the nonlinear term proportional to x^2 displaces the equilibrium point about which the oscillations occur.* This result becomes quite obvious if we consider that the force proportional to x^2 always acts in the same direction and, hence, inevitably displaces the point about which the oscillations occur.

In the same way, we can consider the case when the expansion (50.1) for the force does not contain a term with x^2 (i.e. when $f''(0) = 0$), and we have to take into consideration the term proportional to x^3. In this case, we arrive at the following equation instead of (51.15):

$$m\frac{d^2x}{dt^2} = xf'(0) + \frac{x^3}{3!}f'''(0). \tag{51.27}$$

This equation can also be represented in a form similar to (51.16a):

$$\ddot{x} + \omega_0^2 x = \eta\omega_0^2 x^3, \tag{51.28a}$$

where

$$\omega_0^2 = -\frac{f'(0)}{m}, \quad \eta = \frac{f'''(0)}{6m\omega_0^2} = -\frac{f'''(0)}{6f'(0)}. \tag{51.28b}$$

The quantity η plays the role of the parameter of smallness. As $\eta \to 0$, the solution (51.28a) must tend to harmonic oscillations with a frequency ω_0. The solution of this equation is obtained with the help of the perturbation theory in the same way as described above. In addition to the fundamental frequency ω_0, a higher harmonic appears in the first approximation, but the frequency of this harmonic is triple and not double the fundamental frequency. This is a consequence of the trigonometric formula

$$\sin^3 \omega_0 t = \frac{1}{4}(3\sin\omega_0 t - \sin 3\omega_0 t). \tag{51.29}$$

For the positive and negative values of x with the same magnitude, the force proportional to x^3 has the same absolute magnitude but opposite directions. This means that this force is either a force of attraction to the point $x = 0$ or the force of repulsion from this point. These forces are symmetric relative to the point $x = 0$. Hence, unlike the previous case, there is no displacement of the point about which the oscillations occur. The oscillations with frequencies ω_0 and $3\omega_0$ are performed about the point $x = 0$.

These examples show that the emergence of higher harmonics is the most significant feature of the nonlinear oscillations. The type of harmonics generated depends on the nature of nonlinearity of the force.

GENERAL CONDITION OF HARMONICITY OF OSCILLATIONS. In most cases, small deviations from the equilibrium position may lead to harmonic oscillations. *But this does not mean that only small oscillations can be harmonic.* The oscillations described by an equation of the type (50.3) are harmonic irrespective of the smallness of x. Equation (50.3) is obtained from the energy conservation law (51.5) by differentiating with respect to time and by taking into account that $\mathrm{d}(\dot{x}^2)/\mathrm{d}t = 2\dot{x}\ddot{x}$. *Hence it can be stated that if the total energy of a system, which is conserved, can be represented as a quadratic function of a certain variable and its time derivative, the free oscillations of this system are harmonic.*

As an example, let us consider a closed LC circuit. If the charge on the capacitor is denoted by Q, the current in the circuit will be $\dot{Q}$. The energies of the electric and magnetic fields are proportional to the squares of the electric field strength and magnetic induction respectively, which in turn are proportional to the charge Q and the current $\dot{Q}$. Consequently, the total energy of the system is

$$E = \alpha\dot{Q}^2 + \beta Q^2, \tag{51.30}$$

where α and β are constants depending on the configuration of the oscillatory circuit, i.e., on the capacitance and inductance of the circuit. Considering that $E = \text{const}$, we arrive at the following equation for Q:

$$\alpha\ddot{Q} + \beta Q = 0. \tag{51.31}$$

Thus, the natural oscillations of the current in the oscillatory circuit are harmonic oscillations with a frequency $\omega = \sqrt{\beta/\alpha}$; the harmonicity is independent of the smallness of oscillations and is due only to the fact that the total energy in the circuit is a quadratic function of charges and currents.

Example 51.1. COMPUTATION OF THE PERIOD OF OSCILLATIONS. In the one-dimensional case, any force that depends only on the coordinates is a potential force (see Sec. 27) and the energy conservation law has the form

$$\frac{m\dot{x}^2}{2} + E_p(x) = E. \tag{51.32}$$

Assuming that the potential energy $E_p(x)$ is minimum at the origin of coordinates and increases on either side of the minimum, we conclude that the region in which a particle moves depends on the condition under which the particle comes to a halt at the equilibrium position ($\dot{x} = 0$). In other words, the x_1- and x_2-coordinates defining the region of motion are the solutions of the equation $E_p(x) = E$. The time spent by the particle in moving from one turning point x_1 to the other point x_2 is half the period of oscillations. Hence, considering that $\dot{x} = dx/dt$, we obtain the following formula for the period T of oscillations from (51.32):

$$T = 2 \int_{x_1}^{x_2} \frac{dx}{\sqrt{(2/m)[E - E_p(x)]}}. \tag{51.33}$$

This formula can always be used for determining the period of oscillations. However, when numerical methods are used for obtaining the solution with the help of a computer, it must be borne in mind that the integrand turns to infinity at the boundaries of integration region and care must be exercised in choosing the points of division of the integration interval.

If the potential energy has an expression of the type (51.4), we obtain

$$T = 2\sqrt{\frac{m}{D}} \int_{-x_0}^{x_0} \frac{dx}{\sqrt{x_0^2 - x^2}} = 2\sqrt{\frac{m}{D}}\left[\arccos\frac{x}{x_0}\right]_{-x_0}^{x_0} = 2\pi\sqrt{\frac{m}{D}}. \tag{51.34}$$

As expected, this expression coincides with the period of harmonic oscillations with a frequency $\omega = (D/m)^{1/2}$. It can be seen that the period of oscillations is independent of the particle's energy. This is so because the potential energy increases with distance as x^2. For other types of potential energy, the period of oscillations depends on the energy.

Suppose that the potential energy is given by $E_p(x) = a|x|^n/2$. The boundaries of the region of motion are given by the equation $E = ax_0^n/2$. For the period of oscillations, we obtain the following equation instead of (51.34):

$$T = 2\sqrt{\frac{m}{a}} \int_{-x_0}^{x_0} \frac{dx}{\sqrt{x_0^n - |x|^n}} = 4\sqrt{\frac{m}{a}}\, x_0^{1-n/2} \int_0^1 \frac{d\xi}{\sqrt{1 - \xi^n}}. \tag{51.35}$$

Since $E = ax_0^n/2$, we have

$$T \sim E^{1/n - 1/2}. \tag{51.36}$$

This means that in a given force field, particles of different energies generally have different periods of oscillations. The period of oscillations is independent of energy only for the case $n = 2$, i.e. when the potential energy is proportional to the square of the distance, i.e. when the oscillations are harmonic.

Oscillations whose period is independent of energy are called isochronous oscillations. *It was shown that such oscillations are generated, in particular, for a quadratic dependence of potential energy on distance.* Isochronous oscillations are also possible for other forms of dependence of potential energy. These dependences can be obtained from the quadratic dependence by deforming it along the X-axis in such a way that the distance between the points corresponding to different energies on the curve remains unchanged. The only constraint on this deformation is the requirement that $E_p(x)$ be a single-valued function, i.e. a straight line perpendicular to the X-axis must intersect the E_p vs. x curve only at one point.

Example 51.2. Two identical loads of mass m are hung from the lower ends of a spring and a rubber string whose upper ends are rigidly fastened. The spring and the rubber string are stretched by Δl under the force of gravity of the loads. The extension is the same in both cases if no external forces are present. Find the period of oscillations of the loads if at the initial moment the string and the spring were stretched by $3\Delta l$ and if the loads were released without any initial velocity.

Let us consider the oscillations of the load attached to the spring. We make the origin of the coordinate system coincide with the equilibrium position in which the spring is stretched by Δl. The positive direction of the X-axis is taken as the vertical downward direction. Obviously, the force of elasticity of the spring is $T = D(\Delta l + x)$. Hence the equation of motion has the form

$$m\ddot{x} = mg - T. \tag{51.37}$$

From the condition of equilibrium of the load upon an extension of the spring by Δl it follows that $mg = D\,\Delta l$. Hence the equation of motion acquires the form

$$m\ddot{x} = -Dx = -\frac{mgx}{\Delta l}. \tag{51.38}$$

The general solution of this equation is given by the formula

$$x = A\cos\omega t + B\sin\omega t, \quad \omega = \sqrt{\frac{g}{\Delta l}}. \tag{51.39}$$

Consequently, the period of oscillations is

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{\Delta l}{g}}. \qquad (51.40)$$

From the initial conditions $x(0) = 2\Delta l$ and $\dot{x}(0) = 0$, we obtain $A = 2\Delta l$ and $B = 0$.

For the motion of the load suspended on the string, the initial part of the motion is identical to that of the spring, i.e. the motion is described by the formula

$$x = 2\Delta l \cos\omega t, \qquad \omega = \sqrt{\frac{g}{\Delta l}}. \qquad (51.41)$$

However, the motion takes place according to this law only up to the point $x = -\Delta l$, where $\cos\omega t = -1/2$; $\omega t = 2\pi/3$. Beyond this, there is no elastic force acting on the load from the string. The load's velocity at this instant is

$$\dot{x} = -2\Delta l\omega \sin\omega t = -\sqrt{3g\Delta l}. \qquad (51.42)$$

Consequently, the load moves freely upwards until it comes to rest in a time $(3g\,\Delta l)^{1/2}/g = (3\Delta l/g)^{1/2}$. Hence the half-period of motion can be written in the form

$$\frac{2\pi}{3\omega} + \sqrt{\frac{3\Delta l}{g}} = \sqrt{\frac{\Delta l}{g}}\left(\frac{2\pi}{3} + \sqrt{3}\right). \qquad (51.43)$$

Obviously, the time in which the point returns to its initial position will be equal to this very quantity. Consequently, the period of oscillations is given by the formula

$$T = 2\pi\sqrt{\frac{\Delta l}{g}}\left(\frac{2}{3} + \frac{\sqrt{3}}{\pi}\right). \qquad (51.44)$$

In this case, the oscillations are periodic, but not harmonic.

Sec. 52. DAMPED OSCILLATIONS

The damping of oscillations is discussed.

FRICTION. Natural oscillations of a linear oscillator take place in the absence of external forces. The energy of its oscillations is conserved, and so the amplitude of oscillations also remains unchanged. Natural oscillations are undamped oscillations.

In the presence of friction, which is an external force, the energy of oscillations of a linear oscillator decreases, and hence the amplitude of oscillations also decreases. The oscillations become damped in this case. It can easily be seen that the *frequency of oscillations must also change.* The frictional force acts against the velocity. Hence, for a linear oscillator, its action is equivalent to a decrease in the restoring force, i.e. the elasticity of a spring (i.e. to a decrease in the value of D). *Since* $\omega = D/m$, *the frequency of oscillations must decrease, and so the period of oscillations must increase.*

As the friction increases, the period of oscillations can increase to infinitely large values. When the friction is quite large, there will be no oscillations at all since all the energy of the oscillator will be spent in overcoming the frictional force over a very short path equal to just a fraction of one oscillation.

EQUATION OF MOTION. Let us consider liquid friction. This force appears on the right-hand side of the equation of motion which now assumes the form

$$m\ddot{x} = -Dx - \beta\dot{x}, \tag{52.1}$$

where β is the coefficient of friction. It is convenient to write this equation in the following form:

$$\ddot{x} + 2\gamma\dot{x} + \omega_0 x = 0, \tag{52.2}$$

where $\gamma = \beta/(2m)$ and $\omega_0^2 = D/m$.

FREQUENCY AND DAMPING DECREMENT. The solution of Eq. (52.2) can be sought in the form

$$\tilde{x} = A_0 e^{i\alpha t}. \tag{52.3}$$

Considering that

$$\frac{d}{dt}(e^{i\alpha t}) = -i\alpha e^{i\alpha t}, \qquad \frac{d^2}{dt^2}(e^{i\alpha t}) = -\alpha^2 e^{i\alpha t} \tag{52.4}$$

and substituting (52.3) and (52.2) into (52.4), we obtain

$$A_0 e^{i\alpha t}(-\alpha^2 + 2i\gamma\alpha + \omega_0^2) = 0. \tag{52.5}$$

The cofactor $A_0 e^{i\alpha t}$ is not zero. Hence the other cofactor must be zero:

$$-\alpha^2 + 2i\gamma\alpha + \omega_0^2 = 0. \tag{52.6}$$

This is a quadratic equation in α. The solutions of this equation can be expressed by the well-known formula

$$\alpha = i\gamma \pm \sqrt{\omega_0^2 - \gamma^2} = i\gamma \pm \Omega, \qquad \Omega = \sqrt{\omega_0^2 - \gamma^2}. \tag{52.7}$$

Substituting these values of α into (52.3), we arrive at the required solution:

$$\tilde{x} = A_0 e^{-\gamma t} \cdot e^{i\Omega t}, \tag{52.8a}$$

$$\tilde{x}' = A_0' e^{-\gamma t} \cdot e^{-i\Omega t}. \tag{52.8b}$$

The existence of two solutions indicates that (52.2) is a second-order equation and hence must have two independent solutions which are obtained for opposite signs of Ω.

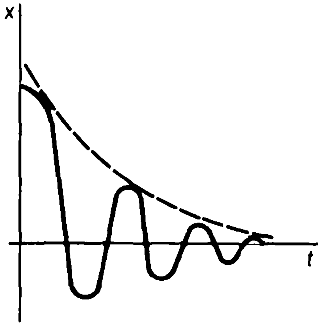


Fig. 151. Damped oscillations.

If the coefficient of friction is not very large, we have

$$\gamma = \frac{D}{2m} < \omega_0. \tag{52.9}$$

In this case, $\omega_0^2 - \gamma^2 > 0$, and hence Ω is a real quantity. Consequently, $e^{i\Omega t}$ is a harmonic function. The real part of the oscillation described by (52.8a) can be represented in the form

$$x = A_0 e^{-\gamma t} \cos \Omega t. \tag{52.10}$$

This is an oscillation whose amplitude decreases, while its frequency Ω remains unchanged. Such an oscillation is shown graphically in Fig. 151.

This oscillation is neither periodic nor harmonic. The period of harmonic (periodic) oscillations is defined as the time in which the oscillation is repeated. In the case of (52.10), the oscillation is not repeated, and hence the concept of period becomes meaningless. *Nevertheless, it is convenient to refer to the period of these oscillations, meaning thereby the interval of time in which the displacement becomes zero.* In the same context, we can also use the concept of the frequency $\Omega = 2\pi/T$ of oscillations. *The amplitude of the oscillations is given by* $A = A_0 e^{-\gamma t}$ *in accordance with* (52.10). *This quantity is roughly equal to the magnitude of the maximum deviations in successive oscillations.*

It can be seen from (52.10) that the amplitude of oscillations decreases by a factor of e = 2.7 in a period of time

$$\tau = \frac{1}{\gamma}. \tag{52.11}$$

The interval τ of time is called the die-away time of oscillations, while γ is called the damping decrement.

LOGARITHMIC DECREMENT. The damping decrement γ as such does not carry much information about the intensity of oscillation damping. For example, the amplitude decreases in a period of time Δt by a factor of $e^{\gamma \Delta t}$. However, depending on the period of oscillations, different numbers of oscillations can take place in this time. If a large number of oscillations has been performed in this period, the variation of the amplitude in successive oscillations will be insignificant. If, however, the number of oscillations is small, there will be a considerable variation of the amplitude after each oscillation. Obviously, it can be stated that the damping of oscillations takes place more slowly in the former case than in the latter case.

Hence damping must be reduced to a natural time scale of oscillations, i.e. to the period of oscillations. The intensity of damping is characterized by the damping of their amplitude during one period of oscillations, and hence it is expedient to use the logarithmic decrement instead of the damping decrement γ.

Let us find the amplitudes of oscillations in two successive time intervals differing by the period T of oscillations:

$$A_1 = A_0 e^{-\gamma t_1}, \quad A_2 = A_0 e^{-\gamma(t_1+T)}. \tag{52.12}$$

Hence

$$\frac{A_1}{A_2} = e^{\gamma T}. \tag{52.13}$$

Consequently, the variation of the amplitude of oscillations during one period of oscillations is characterized by the quantity $\theta = \gamma T$, *called the* logarithmic decrement. From (52.13), we get

$$\theta = \ln \frac{A_1}{A_2}. \tag{52.14}$$

!

The reciprocal of the logarithmic decrement is equal to the number of periods during which the amplitude decreases by a factor of e. The higher the logarithmic decrement, the stronger the damping of oscillations.

Damped oscillations are neither periodic nor harmonic. The period of such oscillations is taken to be the interval of time in which the displacement vanishes.

The logarithmic decrement can also be interpreted in another way. Let us consider the decrease in the amplitude of oscillations over N periods, i.e. in time NT. Instead of (52.12), we can write

$$A_1 = A_0 e^{-\gamma t_1}, \quad A_{N+1} = A_0 e^{-\gamma(t_1+NT)}. \tag{52.15}$$

Hence the ratio of amplitudes over time periods separated by N periods of oscillations is

$$\frac{A_{N+1}}{A_1} = e^{-\gamma NT} = e^{-N\theta}. \tag{52.16}$$

For $N\theta = 1$, the amplitude decreases by a factor of e. Hence it can be stated that the logarithmic decrement

$$\theta = \frac{1}{N} \tag{52.17}$$

is the reciprocal of the number of periods over which the amplitude of oscillations is reduced by a factor of e.

Such an interpretation provides a clear visual representation of the damping intensity:

The amplitude decreases by a factor of e over a number of oscillations, equal to the reciprocal of the logarithmic decrement. If, for example, $\theta = 0.01$, the oscillations will be damped only after about 100 oscillations. Over ten oscillations, the

?

What is the meaning of the period of damped oscillations in spite of the fact that they are aperiodic?

What leads to the conclusion that the frequency of damped oscillations must be lower than the frequency of the corresponding undamped natural oscillations?

What is logarithmic decrement? Which important properties of the damping of oscillations are characterized by the damping decrement?

amplitude will vary only insignificantly, i.e. by about 10% of its initial value. Hence for processes occurring over a small number of periods of oscillations, we can assume the oscillations to be undamped in the first approximation.

The situation is quite different if the logarithmic decrement is large. If $\theta = 0.1$, the oscillations will be completely damped over ten oscillations. The damping will be significant even after a few oscillations. Hence, the oscillations can only be treated as damped even while considering processes occurring over a few periods.

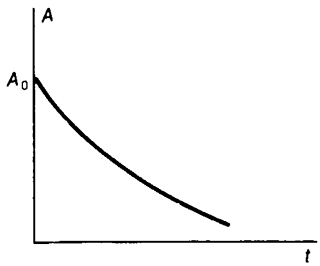


Fig. 152. The case of very strong friction.

No oscillations appear in this case.

LARGE FRICTION ($\gamma \gg \omega_0$). As the friction increases, the period of oscillations increases. For high friction, the motion is no longer oscillatory. This happens under the condition

$$\gamma = \omega_0, \quad \beta = 2\sqrt{Dm}. \tag{52.18}$$

Upon a further increase in friction, $\gamma > \omega_0$. Assuming that $\sqrt{\omega_0^2 - \gamma^2} = \pm i\delta$, where $\delta = \sqrt{\gamma^2 - \omega_0^2}$ is real, we can represent (52.3) in the form

$$x = A_0 e^{-(\gamma \pm \delta)t}. \tag{52.19}$$

Obviously, $\gamma \pm \delta = \gamma \pm \sqrt{\gamma^2 - \omega_0^2} > 0$. This simple exponential function does not contain any oscillation. The plot of such a function is represented in Fig. 152.

All these phenomena can clearly be demonstrated by considering the oscillations of a pendulum suspended in liquids having different viscosities. If the viscosity is very high (as, for example, in glycerine), the pendulum slowly returns from its deflected position to the equilibrium position, and the motion is by no means reminiscent of oscillatory motion.

ANALYSIS OF DAMPING BASED ON ENERGY LOSSES DUE TO FRICTION. It was mentioned above that the energy of oscillations of an oscillator is spent in overcoming the frictional forces and therefore decreases with time. Hence the law of decrease in the amplitude can be found by directly considering the work done by the frictional forces. The work done by these forces in one period of oscillations is given by

$$\Delta E_k = -\beta \int \dot{x}\, dx = -\beta \int_0^T \dot{x}^2\, dt$$
$$= -\beta \int_0^T V^2 \sin^2 \omega t\, dt = -\frac{\beta V^2}{2} T, \tag{52.20}$$

where we have considered that damping is small. Hence the change in the amplitude V of the velocity oscillations can be neglected over one period. On the other hand, the energy spent in doing work against the frictional forces during one period is the difference between the kinetic energies of a particle in two successive periods, i.e.

$$\Delta E_k = \frac{m}{2}(V_1^2 - V_2^2)$$

$$\approx \frac{m}{2}(V_1 - V_2)(V_1 + V_2) \approx \frac{m}{2} 2V\Delta V, \qquad (52.21)$$

where we have considered that the change in the amplitude over one period of oscillations is quite small. Equating the right-hand sides of (52.20) and (52.21), we get

$$-\frac{\beta V^2}{2} T = mV\Delta V, \text{ or } \frac{\Delta V}{T} = -\frac{\beta}{2m} V. \qquad (52.22)$$

For a weak damping, the period T is the small interval of time in comparison with the time period for the case of significant damping. During the time T, the amplitude of the velocity oscillations changes by a small amount ΔV. Hence we can assume in (52.22) that $\Delta V/T \approx \mathrm{d}V/\mathrm{d}t$. This leads to the following expression for the time variation of the amplitude of velocity oscillations:

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -\gamma V, \qquad (52.23)$$

where $\beta/(2m) = \gamma$ is the damping decrement. It is well known that the solution of Eq. (52.23) has the form

$$V = V_0 e^{-\gamma t}. \qquad (52.24)$$

This damping of the velocity amplitude is completely in accord with the damping of the displacement amplitude given by (52.10), which was derived from a rigorous solution of the equation of motion. *Hence the calculations carried out above show that the energy of an oscillator is indeed spent on overcoming the frictional force.*

Example 52.1. DAMPING IN THE CASE OF DRY FRICTION. According to an exponential law, the damping of the amplitude of oscillations occurs when the frictional force is proportional to the velocity. For other types of frictional forces, the amplitude of oscillations decreases according to different laws.

If the motion takes place against dry friction, the equation of motion has the form

$$m\ddot{x} + \left(\frac{\dot{x}}{|\dot{x}|}\right)F_0 + Dx = 0, \qquad (52.25)$$

where $(\dot{x}/|\dot{x}|)F_0$ is a constant quantity directed against the velocity. The quantity $\dot{x}/|\dot{x}| = \mathrm{sgn}\,\dot{x}$ determines the sign of the force. Using the change of variables $\xi = x + (\dot{x}/|\dot{x}|)F_0/D$, we can transform (52.25) to the form

$$m\ddot{\xi} + D\xi = 0, \tag{52.26}$$

which is characteristic of harmonic oscillations. Thus, between the instants of time when the velocity vanishes, the oscillations are harmonic with a frequency $\omega = \sqrt{D/m}$, although the oscillations take place about an equilibrium point which is displaced in the direction of deviation by an amount $\Delta x = F_0/D$. Hence, during one period of oscillations, the point of maximum deviation approaches the equilibrium position by $4F_0/D$, i.e. the amplitude decreases by $\Delta A = -4F_0/D$. This means that the amplitude of oscillations decreases in direct proportion to time and not according to an exponential law. •

Example 52.2. DAMPING IN THE CASE OF ARBITRARY FRICTIONAL FORCES. In order to find the law according to which the amplitude of oscillations decreases, we must solve the equation of motion. This is not always a simple task. However, with the help of the energy concepts, we can directly compute the energy losses due to friction and compare them with the total energy. This enables us to draw conclusions about the nature and rate of attenuation of oscillations in the same way as in the derivation of Eq. (52.24).

Let us apply this "energy" approach for analysis of damping in the case of dry friction. It is convenient to carry out computations by using the displacement amplitude A instead of the velocity amplitude that was used for the derivation of Eq. (52.24). During one period of oscillations, the force F_0 is directed against the velocity, the path traversed being equal to $4A$. Consequently, the energy spent in overcoming the frictional forces is $\Delta E = -4AF_0$. On the other hand, the energy of oscillations is $E = DA^2/2$, and hence $\Delta E = DA\,\Delta A$. Equating the last two expressions for ΔE, we obtain $\Delta A = -4F_0/D$, which is in accord with the exact solution of Eq. (52.25).

Let us use this method for determining the nature of damping for other dependences of the frictional forces on velocity. It has been shown above that if the frictional force were independent of the velocity, the energy loss in one period would be proportional to the amplitude, i.e. $\Delta E \sim A$. If $F_{\mathrm{fr}} \sim v$, then in accordance with (52.20), $\Delta E \sim A^2$. Similarly, calculating the work done against the frictional forces in one period, we find that if $F_{\mathrm{fr}} \sim v^n\,(n > 1)$, then $\Delta E \sim A^{n+1}$. Hence

$$\frac{\Delta A}{A} \sim \frac{\Delta E}{E} \sim A^{n-1}.$$

Consequently,

$$\frac{dA}{dt} \sim \frac{\Delta A}{T} \sim A^n. \tag{52.27}$$

This means that the law of variation of amplitude with time has the form

$$A \sim (t + b)^{1/(1-n)}, \tag{52.28}$$

where b is a constant. For example, for moderate velocities in air, the frictional force is proportional to the square of the velocity ($\sim v^2$). The amplitude of oscillations of a point must in this case decrease according to the law $1/(t + b)$. In other words, if the amplitude of oscillations at the instant $t = 0$ is A_0, the amplitude variation will take place according to the law

$$A(t) = \frac{bA_0}{t + b}. \tag{52.29}$$

If the amplitude decreases by a factor of $1/\gamma$ in one period, i.e. if $A(T) = \gamma A_0$, we obtain the following equation for constant b in (52.29):

$$\gamma A_0 = \frac{bA_0}{T + b}. \tag{52.30}$$

It follows from this equation that $b = \gamma T/(1 - \gamma)$. According to the law (52.29), the amplitude decreases until the frictional force depends linearly on the velocity as a result of a decrease in the velocity. Beyond this point, the amplitude decreases according to an exponential law.

Sec. 53. FORCED OSCILLATIONS. RESONANCE

The methods of investigating forced oscillations are discussed and the concepts associated with them are described.

EXTERNAL FORCE. Besides friction, other external forces can also act on a linear oscillator. *In this case, the nature of motion of the linear oscillator changes, depending on the properties of the forces acting on it.*

The most important case is that of a harmonic external force. It will be shown that more complex cases of variation of the external force with time can be reduced to this simple case. Hence we shall assume that the external force acts on the linear oscillator according to the following law:

$$F = F_0 \cos \omega t, \tag{53.1}$$

where F_0 is the amplitude of the force, and ω is its frequency.

EQUATION OF MOTION. Instead of (52.2), we can write the equation of motion in the following form:

$$m\ddot{x} = -Dx - \beta\dot{x} + F_0 \cos \omega t. \tag{53.2}$$

Dividing both sides by m, we obtain an equation similar to (52.2):

$$\ddot{x} + 2\gamma\dot{x} + \omega_0^2 x = \frac{F_0}{m}\cos\omega t, \tag{53.3}$$

where γ and ω_0 have the same meaning as in (52.2).

TRANSIENT CONDITIONS. Assuming that an external periodic force has started to act on a linear oscillator at a certain instant of time, *the subsequent motion of the oscillator over a certain interval of time will depend on the motion at the instant when the force began to act on it.* However, the effect of the external forces is weakened with time, and the oscillator starts performing steady-state harmonic oscillations. Irrespective of the conditions under which the external force begins to act on the oscillator, the latter will perform the same type of steady-state harmonic oscillations after a certain interval of time. *The process of establishing steady-state harmonic oscillations is called* transient conditions.

The most important aspect of the transient conditions is its duration. *It is determined by the die-away time of oscillations at the instant when the external force began to act. We know that this period is* $\tau = 1/\gamma$. This is the interval of time after which we can forget about the initial oscillations and consider only the steady-state oscillations established under the action of the external force. *On the other hand, if there were no initial oscillations, the forced oscillations could not have attained the steady state instantaneously. It can be shown that the time of establishing the steady state of forced oscillations after the initial force has started acting is again* $\tau = 1/\gamma$.

STEADY-STATE FORCED OSCILLATIONS. In this case, we must assume that the force $F_0 \cos\omega t$ has started acting a long time ago, i.e. at an instant of time in the infinite past. Thus, we consider that Eq. (53.3) is valid for all instants of time. To solve this equation, it is again expedient to use the complex form of harmonic oscillations. We write the expression for force on the right-hand side of (53.3) in complex form, after which the equation assumes the form

$$\ddot{x} + 2\gamma\dot{x} + \omega_0^2 x = \frac{F_0}{m}e^{i\omega t}. \tag{53.4}$$

The real part of the solution of Eq. (53.4) is the solution of Eq. (53.3). This solution is sought in the form

$$x = Ae^{i\omega t}. \tag{53.5}$$

Here, A is not a real quantity in general. Substituting this expression into (53.4), we get

$$Ae^{i\alpha t}(-\alpha^2 + 2i\gamma\alpha + \omega_0^2) = \frac{F_0}{m}e^{i\omega t}. \tag{53.6}$$

This equality must hold for all instants of time, i.e. the time t must be cancelled out. It follows from this condition that $\alpha = \omega$. We determine the quantity A from (53.6) and multiply its numerator and denominator by $\omega_0^2 - \omega^2 - 2i\gamma\omega$. This gives

$$A = \frac{F_0}{m}\frac{1}{\omega_0^2 - \omega^2 + 2i\gamma\omega} = \frac{F_0}{m}\frac{\omega_0^2 - \omega^2 - 2i\gamma\omega}{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2}. \tag{53.7}$$

The complex number (53.7) can conveniently be written in exponential form:

$$A = A_0 e^{i\varphi}, \tag{53.8a}$$

$$A_0 = \frac{F_0}{m}\frac{1}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2}}, \tag{53.8b}$$

$$\tan\varphi = -\frac{2\gamma\omega}{\omega_0^2 - \omega^2} = \frac{2\gamma\omega}{\omega^2 - \omega_0^2}. \tag{53.8c}$$

Consequently, the solution (53.5) can be written in complex form as follows:

$$\tilde{x} = A_0 e^{i(\omega t + \varphi)}, \tag{53.9}$$

while its real part, which is the solution of Eq. (53.3), can be written as

$$x = A_0 \cos(\omega t + \varphi), \tag{53.10}$$

where A_0 and φ are defined by (53.8b) and (53.8c), and ω is the frequency of the external force.

Thus, under the influence of an external harmonic force, an oscillator performs forced harmonic oscillations with the frequency of the external force. The phase and amplitude of these oscillations depend on the properties of the external force and of the oscillator itself.

Let us consider the change in the phase and amplitude of forced oscillations.

AMPLITUDE-FREQUENCY CHARACTERISTIC. The curve describing the dependence of the amplitude of steady-state forced oscillations on the frequency of the external force is called the amplitude-frequency characteristic. The analytic expression for this characteristic is given by formula (53.8b), and the plot of the characteristic is shown in Fig. 153.

The maximum value of the amplitude is attained when the frequency of the external force is close to the frequency of

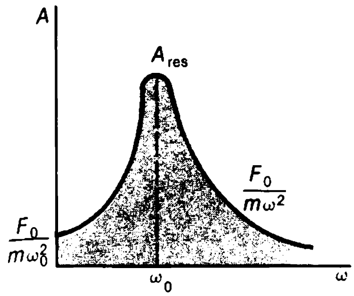


Fig. 153. Amplitude-frequency characteristic.

For small damping, the resonance frequency is close to the natural frequency.

natural oscillations of an oscillator ($\omega \approx \omega_0$). *Oscillations with the maximum amplitude are called* resonance oscillations *and the phenomenon of "building up" of oscillations to the maximum amplitude at* $\omega \approx \omega_0$ *is called* resonance. In this case, the frequency ω_0 is called the resonance frequency. When the frequency of the external force deviates from the resonance frequency, the amplitude of oscillations sharply decreases.

Let us consider the physical pattern of the phenomenon in various frequency ranges. Of most interest are oscillations at weak friction. Therefore we shall asume that $\gamma \ll \omega_0$.

Case 1. $\omega \ll \omega_0$. From (53.8b), we obtain the following expression for the amplitude:

$$A_{0\,st} \approx \frac{F_0}{m\omega_0^2}. \quad (53.11)$$

The physical meaning of this result lies in the following. *At very low frequencies, the external force acts on a system as a constant static force.* Hence the maximum displacement (amplitude) is equal to the displacement (53.11) under the action of the static force F_0, i.e. $x_{max} = F_0/D = F_0/(m\omega_0^2)$, where $D = m\omega_0^2$ is the rigidity characterizing the restoring force. It follows from the condition $\omega \ll \omega_0$ that in Eq. (53.3) the term $\ddot{x}$ due to the acceleration and the term $2\gamma\dot{x}$ due to the velocity are much smaller than the term $\omega_0^2 x$ due to the elastic force since $\dot{x} \approx \omega x$ and $\ddot{x} = -\omega^2 x$. Hence the equation of motion is reduced to the following form:

$$\omega_0^2 x = \frac{F_0}{m} \cos \omega t. \quad (53.12a)$$

The solution of this equation can be written as follows:

$$x = \frac{F_0}{m\omega_0^2} \cos \omega t. \quad (53.12b)$$

This means that the displacement at each instant of time is the same as it should be in the case of a time-independent force and should be equal to its instantaneous value. Frictional forces are neglected.

Case 2. $\omega \gg \omega_0$. Formula (53.8b) leads to the following expression for the amplitude:

$$A_0 \approx \frac{F_0}{m\omega^2}. \quad (53.13)$$

The physical meaning of this expression can be explained as follows. *When the frequency of the external force is very high, the term $\ddot{x}$ due to the acceleration is much larger than the terms due to the velocity and to the elastic force since*

$|\ddot{x}| \approx |\omega^2 x| \gg |\omega_0^2 x|$ and $|\ddot{x}| \approx |\omega^2 x| \gg |2\gamma x| \approx |2\gamma\omega x|$. Hence the equation of motion (53.3) assumes the form

$$\ddot{x} \approx \frac{F_0}{m} \cos \omega t. \tag{53.14a}$$

Its solution can be represented in the form

$$x \approx -\frac{F_0}{m\omega^2} \cos \omega t. \tag{53.14b}$$

Thus, the elastic force and the frictional force do not play any significant role in the oscillations in comparison with the external force. *The external force acts on an oscillator as if there were neither elastic force nor frictional force.*

Case 3. $\omega \approx \omega_0$. This is the situation corresponding to resonance. At resonance, the amplitude has its maximum value for which we obtain the following expression from (53.8b) at $\gamma \ll \omega_0$:

$$A_{0\,\mathrm{res}} = \frac{F_0}{m} \frac{1}{2\gamma\omega_0}. \tag{53.15}$$

The physical meaning of this result can be explained as follows. The term due to the acceleration is equal to the term due to the elastic force, i.e. $\ddot{x} = -\omega^2 x = -\omega_0^2 x$. *This means that the acceleration is caused by the elastic force, while the external force and the frictional force cancel each other.* Equation (53.3) then assumes the form

$$2\gamma\dot{x} = \frac{F_0}{m} \cos \omega_0 t. \tag{53.16a}$$

Its solution can be represented as follows:

$$x = \frac{F_0}{2\gamma m\omega_0} \sin \omega_0 t. \tag{53.16b}$$

Strictly speaking, the maximum amplitude is attained not exactly at $\omega = \omega_0$, but near this value. The exact value of frequency can be obtained from the general rule by equating the derivative of A_0 with respect to ω in (53.8b) to zero. However, when friction is not very strong, i.e. when $\gamma \ll \omega_0$, the displacement of the maximum from the position $\omega = \omega_0$ is quite insignificant and can be neglected.

Q-FACTOR. An important characteristic of the properties of an oscillator is an increase in its amplitude at resonance in comparison with its static value, i.e. in comparison with the displacement under the action of a constant force. From (53.11) and (53.15), it follows that

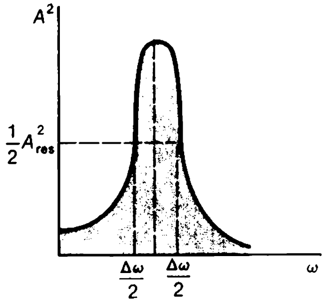


Fig. 154. Resonance curve for the square of the amplitude.

The resonance width $\Delta\omega$ is determined from this curve.

$$Q = \frac{A_{0\,\text{res}}}{A_{0\,\text{st}}} = \frac{\omega}{2\gamma} = \frac{2\pi}{2\gamma T} = \frac{\pi}{\theta}, \tag{53.17}$$

where θ is the logarithmic decrement. The quantity Q is called the Q-factor of a system. The Q-factor is an important characteristic of the resonance properties of the system.

It can be seen from (53.17) that the lower the damping of an oscillator, the more vigorous its resonance build-up since, in accordance with (53.17), $A_{0\,\text{res}} = A_{0\,\text{st}}Q = A_{0\,\text{st}}(\pi/\theta)$.

An important feature of the resonance properties is not only an increase in the amplitude at resonance, but also the rate of this increase. In other words, it is important to know not only the value of the resonance amplitude, but also the rapidity with which this amplitude decreases as a result of departure from the resonance frequency. This property is characterized by the width of a resonance curve. *However, this quantity is referred not to the amplitude of oscillations, but rather to the square of the amplitude.* This is so because the energy of a linear oscillator, which is a vital characteristic of the linear oscillator, is defined in terms of the square of the amplitude of displacement. The shape of the resonance curve for the square of the amplitude is similar to that shown in Fig. 153. This curve is represented in Fig. 154, where the half-width of the resonance curve is also indicated. The half-width of a resonance curve is the distance $\Delta\omega/2$ on the frequency scale from the resonance frequency ($\omega = \omega_0$) to the frequency at which the square of the amplitude is reduced to half its value. The value of the half-width can be calculated without any difficulty.

Near the resonance $\omega = \omega_0$ we can assume that

$$\begin{aligned} A_0^2 &= \left(\frac{F_0}{m}\right)^2 \frac{1}{(\omega_0^2 - \omega^2)^2 + 4\gamma^2\omega^2} \\ &= \left(\frac{F_0}{m}\right)^2 \frac{1}{(\omega_0 - \omega)^2(\omega_0 + \omega)^2 + 4\gamma^2\omega^2} \\ &\approx \left(\frac{F_0}{m}\right)^2 \frac{1}{\omega_0^2(\Delta\omega)^2 + 4\gamma^2\omega_0^2}, \end{aligned} \tag{53.18}$$

where we have considered frequencies close to the resonance frequency for which $\Delta\omega \ll \omega_0$ and $\omega \approx \omega_0$. Since $A^2_{0\,\text{res}} = (F_0/m)^2/(4\gamma^2\omega_0^2)$ at the resonance, the condition of a decrease in the amplitude to half its value at the resonance assumes the form

$$\frac{1}{2}\frac{1}{4\gamma^2\omega_0^2} = \frac{1}{\omega_0^2(\Delta\omega)^2 + 4\gamma^2\omega_0^2}. \tag{53.19}$$

Consequently, the width of the resonance curve is given by

$$\Delta\omega = 2\gamma, \tag{53.20}$$

i.e. the width of the resonance curve is equal to twice the damping decrement: the weaker the damping, the smaller the width of the resonance curve and hence the sharper the resonance.

Formula (53.20) can be represented in a more convenient form in terms of the logarithmic decrement and the Q-factor. We divide both sides of (53.20) by ω_0 and take (53.17) into account:

$$\frac{\Delta\omega}{\omega_0} = \frac{2\gamma}{\omega_0} = \frac{\gamma T}{\pi} = \frac{1}{2}\frac{1}{Q}, \tag{53.21}$$

$$\Delta\omega = \frac{\omega_0}{Q}. \tag{53.22}$$

Thus, the width $\Delta\omega$ of a resonance curve is equal to the resonance frequency divided by the Q-factor.

As the Q-factor increases, the resonance amplitude increases and the width of the resonance peak decreases. However, in accordance with (53.17) and with what has been stated above about the transient conditions, *a larger Q-factor means a longer stabilization time for the forced oscillations.*

PHASE CHARACTERISTIC. Another important characteristic of forced oscillations is the relation between their phase and the phase of the external force. In formula (53.10) for displacement, this relation is expressed through the quantity φ since the dependence of force on time is given by the function $\cos \omega t$. If $\varphi < 0$, the displacement lags behind the external force in phase. The dependence (53.8b) of phase φ on frequency is called the phase characteristic (Fig. 155).

At very low frequencies $\omega \ll \omega_0$ the phase φ is small and negative. This means that the displacement lags behind the force in phase by a very small amount. With increasing frequency, the phase lag of the displacement from the force increases. At resonance, the displacement lags behind the force by $\pi/2$ in phase. This means that when the force attains its maximum value, the displacement is zero, and conversely, when the force is zero, the displacement attains its maximum value. Upon a further increase in the frequency, the phase of the displacement continues to lag behind the phase of the force and approaches the value π for very high frequencies $\omega \gg \omega_0$. In other words, it can be stated that the displacement and the force are nearly opposite to each other since $\cos(\omega t - \pi) = -\cos \omega t$. Hence, when the force attains, say, its maximum positive value, the displacement has its maximum negative value. After this, the directions of the force and the displace-

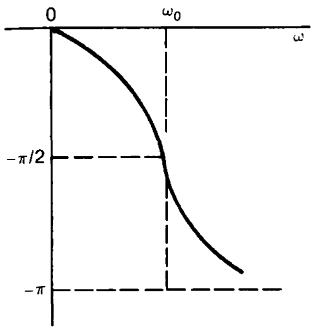


Fig. 155. Phase characteristic.
For small damping, the phase changes rapidly over a very small frequency interval near the resonance frequency, from the value 0 to the value π. In other words, a phase "reversal" takes place at the resonance frequency.

!

The Q-factor, which is equal to the reciprocal of the logarithmic decrement multiplied by π characterizes the build-up of oscillations at resonance. The Q-factor indicates the factor by which the amplitude at resonance exceeds the amplitude of the static displacement for the same amplitude of the force.
The width of a resonance curve is defined not with respect to the amplitude of oscillations, but to the square of the amplitude.
Resonance occurs when conditions are most favourable for an effective transfer of energy from an external source to an oscillating system.

ment are reversed and they pass the zero point almost simultaneously.

These phase relations between the displacement and the force permit a deeper understanding of the resonance phenomenon. It was mentioned in Sec. 50 that the velocity leads the displacement by $\pi/2$. On the other hand, at resonance the force also leads the displacement by $\pi/2$. Consequently, the force and the velocity oscillate in phase, i.e. the direction of the force always coincides with that of the velocity. Hence the work done by the external force attains its maximum value. If there is no resonance, the force coincides in direction with the velocity for a part of the period of oscillations, and hence the energy of an oscillator increases. For another part of the period, the force acts in a direction against the velocity, and hence the energy of the oscillator decreases. Hence resonance is characterized by the most favourable conditions for transferring the energy from the external source to the oscillator. The most unfavourable conditions for transferring the energy from the external source to the oscillator exist at $\omega \ll \omega_0$ and $\omega \gg \omega_0$, when the phases of the force and the velocity differ by about $\pi/2$. This means that the force is directed against the velocity about half the time and coincides with it for the remaining half. Thus, on the average, an insignificant amount of energy is transferred from the external source to the oscillator during one period of oscillations. Hence the amplitude of oscillations is very small in this case.

PERIODIC ANHARMONIC FORCE. If the external force $F_0 f(t)$ acting on an oscillator is periodic with a period T, in accordance with the formulas of mathematical analysis, it can be represented as a Fourier series, each of whose terms is a harmonic function:

$$F_0 f(t) = F_0 \sum_{n=0}^{\infty} (a_n \cos n\omega t + b_n \sin n\omega t), \tag{53.23}$$

where $\omega = 2\pi/T$. This force acts on the oscillator instead of the force (53.1) and appears on the right-hand side of Eq. (53.3).

In order to find the result of the action of this force, there is no need to carry out any new computations. It is sufficient to take into account that (53.3) is a linear equation, and hence its solution can be represented as the sum of the solutions of equations whose right-hand sides contain one of the terms from the sum (53.23). In other words, each of the terms of the harmonic forces in (53.23) acts independently on a linear oscillator. This action of the force has already been studied. The total oscillation is the sum of the oscillations caused by individual harmonic forces in (53.23).

The strongest influence on the oscillator is exerted by the terms in the sum (53.23) whose frequencies are close to the resonance frequency, i.e. for which $n\omega \approx \omega_0$. If such frequencies do not exist, the periodic force $F_0 f(t)$ does not cause a sharp rise in the amplitude of oscillations. If, however, such frequencies do exist, resonance is observed. The resonance amplitude, the width of the resonance line and the phase shift are determined with the help of the formulas considered above. The absolute value of the resonance amplitude depends on the coefficients a_n and b_n of the appropriate terms in the sum (53.23). If these terms are very small, even a hundredfold increase in the resonance amplitude does not cause a significant increase in the resultant amplitude of oscillations. In this case, the resonance terms in (53.23) are insignificant.

If the coefficients a_n and b_n in resonance terms are not very small, the resonance amplitudes corresponding to them play a decisive role in the action of the force $F_0 f(t)$ on the oscillator.

It was mentioned above that most physical systems behave like linear oscillators for small deviations from the equilibrium position. For example, the tops of structures (towers, houses), bridges of different designs, etc. oscillate like linear oscillators. The rotating shafts of engines perform torsional vibrations that can also be treated as the oscillations of a linear oscillator (the angular acceleration $\ddot{\alpha}$ produced upon a deviation from the equilibrium position is proportional to the angle of deflection, i.e. $\ddot{\alpha} \sim \alpha$). Moreover, these systems are often subjected to the action of periodic forces. For example, the shaft of an engine experiences periodic forces exerted by the piston as a result of combustion of the fuel in the cylinder, various parts of a bridge are subjected to nearly periodic pressure from the succession of motor vehicles following one another more or less regularly, as well as from the steps of pedestrians, and so on. In order to analyze the result of these periodic forces, we must carry out their spectral analysis, i.e. represent the forces in the form (53.23) and find the coefficients a_n and b_n in this expansion that are associated with various harmonic components of the force. After this, we must find the natural frequencies ω_{0i} with which a system can oscillate. In general, a system has several, and even an infinite number of, natural frequencies, and for small deviations we cannot always represent the system as a single linear oscillator. It may so happen that for small deviations a system may behave as an aggregate of linear oscillators with different natural frequencies. Each of these oscillators can initiate resonance oscillations under the action of appropriate harmonic components. For example, a bridge may perform vertical vibrations, horizontal displacements across its length, vibrations along its

?

What happens to the phase near resonance in the case of a weak damping?
How does the phase act for a slightly strong damping?
Under what condition can the analysis of the action of a periodic anharmonic force be reduced to the simple application of the results of analysis for a harmonic force?
What factors determine the nature of action of an aperiodic force on a system?

length, and so on. The natural frequencies of these vibrations are different, and each type may have several natural frequencies. All the natural frequencies must be taken into consideration in the analysis of the action of an external periodic force. One of the tasks facing the designers is to avoid the resonance action of the external forces on a system. In other cases, it may be equally important to create conditions ensuring the resonance action of the external forces on the system. For example in radioengineering, the reception of radio signals makes it imperative to attain their resonance action on the oscillatory circuits of the radio receiver. In both cases, the problem is reduced to the investigation of the forced oscillations of a linear oscillator under the action of an external periodic force.

The possibility of mutual coupling between various linear oscillators must also be taken into consideration. This will be discussed in the analysis of oscillations of coupled systems.

APERIODIC FORCE. The periodic force, whose action on a linear oscillator was considered above, is an idealized concept which is never realized in practice. For a force to be periodic in the strict sense of the word, its action must be periodic over an infinitely long time. If the action of the force has a beginning and an end, the force cannot be periodic. However, real forces having a periodic nature and acting over a finite interval of time can be treated as periodic forces. For this purpose, the force must act for "a sufficiently long time". In order to get an idea of the criterion of "sufficiently long time", let us analyze harmonic forces.

After the harmonic force (53.1) has started to act, a time $\tau = 1/\gamma$ is required for establishing steady-state forced oscillations. If the force continues to act for a much longer time than τ and the system performs a large number of oscillations in this time, the result will be the same as if the force had been acting for an infinitely long time. Consequently, the force is assumed to be harmonic under this condition, and its boundedness in time need not be taken into consideration.

A periodic force also has a beginning and an end and is not periodic in the strict sense of the word. However, in the same way as for the case of a harmonic force, a force can be considered to be periodic if the time τ for establishing forced oscillations is much smaller than the time for which the force acts. After the passage of time τ, the oscillations acquire their steady-state characteristic and the situation is just the same as if they had existed for an infinitely long time. In other words, the force can be assumed to be strictly periodic.

By an aperiodic force we mean a force during whose existence no periodic variation can be established. The action

?

What are transient conditions and how is their duration defined?
Define the frequency of forced oscillations under the action of a harmonic external force.
What are the distinguishing features of the amplitude resonance curve?
What property of the resonance curve determines the Q-factor?

of such a force can be studied with the help of the considerations put forth above. Suppose that the time T during which a force acts on a system is much longer than the time τ in which oscillations set in. Then after the passage of time τ certain steady-state conditions will be established in the system, and no significant changes in the system will take place in the subsequent interval of time $T-\tau$. Hence it is natural to consider the process to be periodic with a period T. Let us represent this force in the form (53.23). Obviously, the components of the force corresponding to terms with $n \gg 1$ will perform a large number of oscillations during the time T, while the steady-state conditions will be established during the first few oscillations. Hence all the conclusions concerning the action of periodic forces are fully applicable to these components. If the frequencies belong to the resonance region, the amplitude of the oscillations will increase significantly. Since it is possible that $\omega \ll \omega_0$ in this case ($\omega_0 = 2\pi/T$), many terms of (53.23) may have their frequencies close to the resonance value $n\omega = \omega_0$. The resulting nearly resonance oscillations are superimposed on one another. On the other hand, the first terms in the sum (53.23) with $n = 0, 1, 2, \ldots$ have frequencies much lower than the resonance frequency. Such frequencies satisfy Eq. (53.12) in which the force is followed almost instantaneously by the displacement. Thus, if an aperiodic force exists for a much longer time than the time during which the oscillations set in or the period of resonance oscillations, the process is considered in exactly the same way as for a periodic force. Strictly speaking, a certain amount of error will creep in under such an assumption since the motion of the oscillator at the beginning and end of the action of the force will not be the same. Hence we should add the die-away time τ to the period T so that the second "imaginary" period begins in the same way as the first, i.e. without any oscillations before the force begins to act. However, $\tau \ll T$, and this correction does not introduce any significant changes. From mathematical point of view, a more rigorous solution of the problem can be obtained by going onto a continuous spectrum, i.e. by assuming that the time for which the force acts is given by $T \to \infty$.

In this case, instead of the expression (53.23) for force as a sum over frequencies, the force can be represented as an integral over frequencies, called the Fourier integral. In this case, the frequencies assume continuously varying values instead of discrete values. Forced oscillations are also composed of all possible frequencies the densities of whose amplitudes are connected with the density of the force amplitudes of the same frequency. The components of force with frequencies lying in the resonance region cause a sharp increase in the displacement amplitudes. The physical nature of the phenomenon in the case of continuous spectrum is the same as in the case of discrete spectrum.

If the time T during which the external force acts is smaller than the time $\tau = 1/\gamma$ in which forced oscillations set in, the arguments based on the pattern of steady-state forced oscillations are not applicable. In this case, the oscillations must be investigated under the transient conditions.

RESONANCE OF NONLINEAR OSCILLATIONS. The most important feature of nonlinear forced oscillations is the resonance at combination frequencies. It was mentioned in Sec. 51 that in addition to the fundamental frequency ω_0, nonlinear oscillations also contain higher harmonics with frequencies $n\omega_0$. Under the action of an external harmonic force with a frequency ω, resonance occurs not only at the fundamental frequency, when $\omega \approx \omega_0$, but also at frequencies corresponding to higher harmonics, when $\omega \approx n\omega_0$. The spectrum of an arbitrary periodic force contains, besides the fundamental frequency ω, higher harmonics with frequencies $m\omega$ as well.

Hence resonance may occur at frequencies satisfying the condition $m\omega = n\omega_0$, i.e. for various combinations of the fundamental frequencies. Of course, the role of any resonance depends on its amplitude which in turn depends on the characteristics of a nonlinear system and the properties of the force. If the amplitude is small, there is no need to take resonance into account.

Sec. 54. SELF-EXCITED OSCILLATIONS AND PARAMETRIC OSCILLATIONS

The methods of obtaining self-excited oscillations and parametric oscillations are qualitatively described.

DEFINITION. Natural oscillations are gradually attenuated due to energy losses by friction. If energy is supplied from a source of an external harmonic force to an oscillator, it oscillates with the frequency of the force. This frequency is generally different from the natural frequency of the oscillator.

However, it is possible to construct devices in which the oscillator itself regulates the supply of energy from the external source in such a way as to compensate for the energy losses due to friction. The energy acquired by the oscillator from the external source during one period of oscillations is equal to the energy consumed by it in overcoming friction. As a result, the oscillator performs undamped oscillations. *Such self-sustained oscillations are called* self-excited oscillations. *If the friction is not strong, only a small fraction of the total energy of the* *oscillator is supplied to the system during one period. In this case, the self-excited oscillations are harmonic to a very high degree of accuracy, and their frequency is very close to the frequency of natural oscillations. If, however, the friction is quite strong, a considerable fraction of the total energy of the oscillator is acquired from the external source, and hence the oscillations differ significantly from the harmonic oscillations even though they are periodic oscillations. The period of these oscillations differs from the period of natural oscillations of the oscillator.*

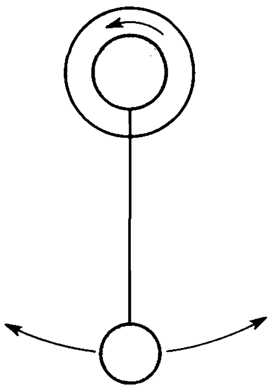


Fig. 156. A pendulum suspended from a rotating axle is the simplest example of a self-oscillating system.

SELF-EXCITED OSCILLATIONS OF A PENDULUM. Let us consider the oscillations of a pendulum suspended from the axle of a rotating bush (Fig. 156). We shall consider the transformation of the energy of the pendulum under different conditions. Suppose that the pendulum is at rest. Then the rotating bush slips about its axle and performs work in overcoming the frictional force. This work is completely transformed into internal energy as a result of which the bush and the axle are heated. The energy transformed into the internal energy is supplied by the motor turning the bush.

Let us suppose that the pendulum is set in oscillation. In the half-period of the pendulum's oscillations, when the direction of rotation of the pendulum's axis and the bush's axle coincide, the frictional forces have the same direction as that of the surface points of the axis and the axle. Hence these forces cause an enhancement of the oscillations of the pendulum. On the other hand, the energy transformed into the internal energy during the half-period of oscillations will be lower than for the case when the pendulum is at rest in view of the fact that the relative displacement of the surfaces rubbing against each other (the outer surface of the axis and the inner surface of the bush) decreases. Hence only a fraction of the energy supplied by the motor for rotating the bush is transformed into the internal energy, while the rest of the energy is spent in enhancing the oscillations of the pendulum.

In the other half-period of the pendulum's oscillations, when the direction of rotation of its axis is opposite to that of the bush's axle, the frictional forces act against the direction of motion of the pendulum. Hence they retard its motion, and the energy of oscillations of the pendulum is transformed into the internal energy. The energy supplied by the motor to rotate the bush is also completely transformed into the internal energy. The net result of the energy transformations over a complete period of oscillations is determined by the dependence of the frictional forces on velocity.

If the frictional force is independent of velocity, the energy acquired by the pendulum in the half-period, when the directions of rotation of its axis and of the bush's axle coincide, is equal to the energy spent by the pendulum in overcoming the work against the frictional force in the remaining half-period. In this case, the rotation of the bush does not cause any change in the oscillations of the pendulum in comparison with the case when the bush is not rotating.

If the frictional force increases with velocity, the energy acquired by the pendulum in the half-period, when the directions of rotation of its axis and of the bush's axle coincide, is lower than the energy spent by it in overcoming the work against the frictional force in the remaining half-period. This is so because the relative velocities are higher in the second half-period, and hence the frictional force is also stronger than in the first half-period. In this case, the rotation of the bush increases the damping of the pendulum's oscillations.

If the frictional force decreases with increasing velocity, the energy acquired by the pendulum in the half-period, when the directions of rotation mentioned above coincide, is stronger than the energy consumed by it in overcoming the work against the frictional force in the remaining half-period. This is so because the relative velocities are higher in the second half-period, and hence the frictional force is weaker than in the first half-period. Thus, the rotation of the bush causes an increase in the amplitude of oscillations of the pendulum. However, in this case, the energy losses due to friction of the pendulum against the air increase. When the energy supplied to the pendulum in a period becomes equal to the energy spent by it in overcoming friction, the amplitude and the frequency of oscillations acquire constant values, and the pendulum is said to oscillate in a self-excited mode. If the energy losses during one period are not large in comparison with the total energy of oscillations of the pendulum, and the amplitude of oscillations is quite small, the oscillations are harmonic, and their frequency is equal to the natural frequency of oscillations of the pendulum.

Self-excited oscillations are frequently employed in engineering. A familiar example is the pendulum clock. In this case, the energy is supplied to the pendulum in jerks following the application of force to the pendulum from a spring or a suspended weight at the instant of time determined by the oscillations of the pendulum itself. In an electric bell, the vibrations of the hammer switch on and off the electric current which supplies the energy to the bell for maintaining the self-excited oscillations of the hammer.

RELAXATION OSCILLATIONS. These oscillations are a particular case of self-excited oscillations, but the nature of variation of their parameters with time is quite peculiar:

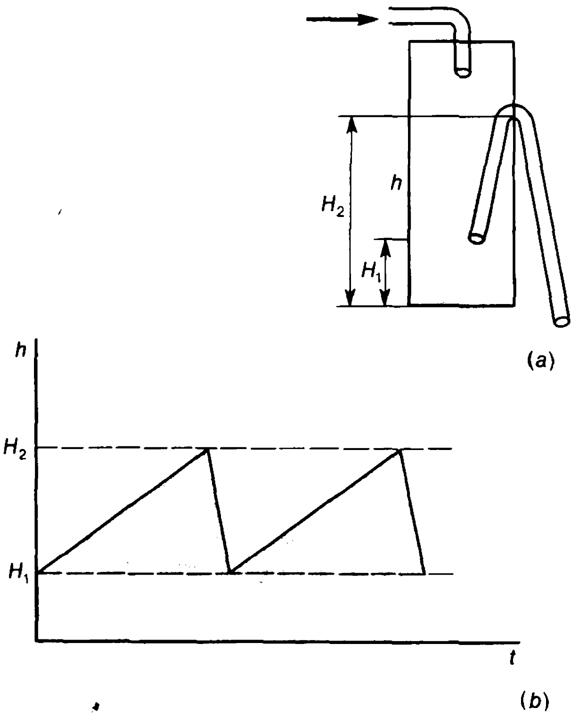


Fig. 157. Relaxation oscillations of the height of a liquid column.

!

An external source of energy is necessary for self-excited oscillations. An oscillating system itself takes the energy from this source at the required rate so that the oscillations are not damped.

?

What is the difference between self-excited oscillations and forced oscillations?
Can you prove by direct computations that the energy conservation law is obeyed in the parametric excitation of oscillations like the build-up of oscillations on a swing?

variations are slowly accumulated in a system over a rather long time. This is followed by a sudden, almost abrupt, variation in the state of the system which thus returns to its original state. After this, variations are again accumulated slowly in the system, and so on.

An example of such oscillations, which is well known since ancient times, is shown in Fig. 157*a*. A vessel is equipped with a wide siphon through which water can flow out of it. The vessel is filled from a tap from which a thin stream of water flows. As a result, the water level in the vessel slowly rises. When the level reaches the lower wall of the bend in the siphon (height H_2), water begins to overflow, expelling the air and filling the entire cross section of the siphon in the upper part. After this, water flows from the entire cross section of the siphon quite rapidly since this cross section is quite large. The water level in the vessel falls quickly to the lower end of the siphon inside the vessel (height H_1). After this, the cycle of filling begins again. The variation of the water level in the vessel is plotted in Fig. 157*b*. It can be seen that these oscillations are discontinuous: the rate of change of h at the upper and lower points abruptly reverses its sign from the plus sign for increasing h to the minus sign at the upper point when the liquid begins to flow out of the siphon.

PARAMETRIC EXCITATION OF OSCILLATIONS. The properties of oscillating systems are defined by quantities called parameters. For example, a simple pendulum is characterized by a single parameter, viz. its length. When this parameter changes, the nature of oscillations of the pendulum also changes, i.e. the natural frequency of its oscillations changes. If this parameter changes in step with the oscillations, we can supply energy to the pendulum and hence increase the amplitude of its oscillations or simply sustain the oscillations if they are damped. *Such an excitation and maintenance of oscillations is called* parametric.

A well-known example of excitation and maintenance of oscillations is a swing. When the swing reaches the uppermost point in its motion, the person rocking it squats down, while at the lowest point in the motion of the swing, the squatter gets up again. As a result of squatting, the magnitude of the work performed at the uppermost point is smaller than the work done in raising the swing from its lowest position. According to the energy conservation law, the difference in these two works is equal to the difference in the energies of the swing, and the oscillations of the swing increase. If this energy is completely spent in overcoming the work against friction, the swing will continue to rock in undamped mode.

Sec. 55. OSCILLATIONS OF COUPLED SYSTEMS

The meaning of terms used in the description of coupled systems is clarified.

SYSTEMS WITH MANY DEGREES OF FREEDOM. If a system has several degrees of freedom, small deviations from the equilibrium position may induce oscillations for all degrees of freedom simultaneously. For example, in the case of vibrations of a bridge considered above, one of the degrees of freedom of the bridge is its vibration in the vertical plane, while another degree of freedom is its vibration in the horizontal direction. Of course, other degrees of freedom also exist. A simple pendulum may oscillate in two mutually perpendicular vertical planes passing through the point of suspension. Hence it has two degrees of freedom. *If the oscillations corresponding to each degree of freedom are independent of each other, i.e. if they cannot exchange energy, the motion of the system with several degrees of freedom can be analyzed in a purely kinematic form: knowing the motion for each degree of freedom, we carry out a kinematic summation of the motion.* Although the resultant motion in this case may be quite complicated, it does not involve any new physical laws from the point of view of

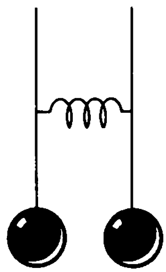


Fig. 158. Oscillations of coupled systems.

dynamics. *Only the coupling between various degrees of freedom introduces new physical regularities in an oscillatory system with many degrees of freedom.*

COUPLED SYSTEMS. *A* coupled system *is a system with many degrees of freedom which are mutually coupled and can therefore exchange energy with one another.*

As an example, let us consider two pendulums connected through a spring which serves as the coupling (Fig. 158). This system may oscillate in the vertical plane in which the pendulums and the spring are located in a state of equilibrium, as well as in directions perpendicular to this plane. In all, there are four degrees of freedom coupled with one another. If one of the pendulums is drawn out of the equilibrium position by deflecting it simultaneously in the plane of the pendulums, as well as in a direction perpendicular to this plane, the second pendulum will also begin to oscillate over its degrees of freedom after the onset of the oscillations of the first pendulum. The oscillations of the pendulums have varying amplitudes. On the whole, a quite complicated pattern of motion of the pendulums and of energy transfer between them is observed.

NORMAL OSCILLATIONS OF COUPLED SYSTEMS. In spite of the complicated motion of two coupled pendulums, we can always represent the motion as a superposition of four harmonic oscillations whose frequencies are called the normal frequencies of the coupled system. *The number of normal frequencies is equal to the number of the degrees of freedom.* In the case of the coupled pendulums, we have four normal frequencies. We shall give the definition of these frequencies and describe the method of their determination.

!

In a coupled system, energy is transferred between the parts of the system through the coupling.

To begin with, let us write down the oscillations of the pendulums in the vertical plane perpendicular to the line joining the points of their suspension. Each pendulum may occupy a certain position in this plane. The state of the system is characterized by the position of both pendulums. Let us consider the simplest cases: (1) both pendulums are deflected from the equilibrium position in the same direction by the same angle; (2) the pendulums are deflected in opposite directions by the same angle. These simple deviations are called normal deviations. Any possible deviation of the pendulums can be represented as the sum of their identical deviations in the same direction and in opposite directions. In other words, any state of the system in the above-mentioned sense can be represented as a superposition of states (1) and (2). The proof of this statement can easily be obtained with the help of the plot shown in Fig. 159. The dashed line shows the mean equilibrium line. The quantities a and b denote the

?

Which factor makes a system with many degrees of freedom a coupled system?
What is meant by normal oscillations of a coupled system?
How many such oscillations are possessed by a coupled system?
How can we represent an arbitrary oscillation of a coupled system in terms of normal oscillations?

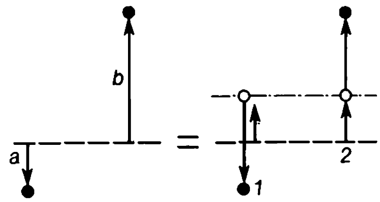


Fig. 159. Representation of an arbitrary deviation of two pendulums as the sum of two normal deviations.

deviations of the pendulums from the equilibrium position ($b > a$). On the right-hand side of the equality we have the combinations of deviations *1* and *2* whose sum gives the initial deviations of the pendulums.

If the pendulums are deflected in the same direction by the same angle and then released, they will oscillate with a certain frequency ω_1, called a normal frequency. The frequency of oscillations of the pendulums deflected in the opposite directions by the same angle is another normal frequency ω_2. In accordance with the decomposition shown in Fig. 159, an arbitrary oscillation of the two pendulums in the given directions can be represented as the sum of two harmonic oscillations with normal frequencies.

In the same way, we can analyze the oscillations of the pendulums in the vertical plane passing through the line joining their points of suspension. The normal oscillations in this plane are the oscillations of the pendulums deflected at the same angle in the same and opposite directions. The line of reasoning adopted in this case is the same as in the previous case. Hence, in this case, the oscillations of the two coupled pendulums can also be represented as the sum of two oscillations with normal frequencies equal to the frequencies of the corresponding normal oscillations.

The total motion of the two pendulums with four degrees of freedom is a superposition of four normal oscillations with the corresponding normal frequencies. In the present case, not all four normal frequencies are different, but this by no means alters the situation.

Thus, the problem of investigating coupled systems is reduced to finding their normal oscillations and normal frequencies. Sometimes, normal oscillations can be determined with the help of simple arguments, as was shown in the previous case. Two of the normal frequencies are just the natural frequencies of oscillations of the pendulums (taking into account the mass of the spring and the height of its suspension, or without considering these parameters). The remaining two normal frequencies are the frequencies of the pendulum's oscillations when an additional elastic force is exerted by the spring in the case of a symmetric deviation of the pendulums in opposite directions from their equilibrium position.

In most cases, the problem is found to be a lot more complicated. The normal frequencies can be determined with the help of some general methods which will not be described in this book.

Let us now carry out a detailed mathematical description of the oscillations of coupled systems by considering the example of coupled pendulums. We shall confine ourselves to the case of two degrees of freedom. It will be assumed that the pendulums oscillate in the same plane coinciding with the vertical plane and passing through the points of their suspension and the equilibrium position of the point masses of the simple pendulums (Fig. 160). For small oscillations, we can neglect the vertical deviations of the points, and their motion can be considered along a straight line. The position of the oscillating points is characterized by their deviations x_1 and x_2 from their equilibrium positions denoted by O_1 and O_2 respectively. If the points are simultaneously in the positions of equilibrium, the spring joining them will be undeformed and will not exert any force on the points.

Let us denote by ω_1 the frequency of normal oscillations of the pendulums when they oscillate synchronously (in the same phase), and by ω_2 the same frequency when the oscillations of the pendulums are in opposite phases. Obviously, $\omega_2 > \omega_1$. The resultant oscillation of the system is the superposition of two normal oscillations. According to the method described above for the decomposition of an arbitrary motion of coupled pendulums, we can write

$$\begin{aligned} x_1 &= A\sin(\omega_1 t + \varphi_1) + B\sin(\omega_2 t + \varphi_1), \\ x_2 &= A\sin(\omega_1 t + \varphi_2) - B\sin(\omega_2 t + \varphi_2). \end{aligned} \tag{55.1}$$

The four unknown constants A, B, φ_1 and φ_2 are determined from the initial conditions expressing the values of the deviations x_{10} and x_{20} and the velocities $\dot{x}_{10}$ and $\dot{x}_{20}$ at the initial instant of time, say, $t = 0$:

$$\begin{aligned} x_{10} &= A\sin\varphi_1 + B\sin\varphi_2, \\ x_{20} &= A\sin\varphi_1 - B\sin\varphi_2; \\ \dot{x}_{10} &= A\omega_1\cos\varphi_1 + B\omega_1\cos\varphi_2, \\ \dot{x}_{20} &= A\omega_2\cos\varphi_1 - B\omega_2\cos\varphi_2. \end{aligned} \tag{55.2}$$

Having determined the values of the constants A, B, φ_1 and φ_2 from these conditions, we can completely describe the motion with the help of Eqs. (55.1).

Let us now solve the same problem by applying the dynamic laws of motion directly. We can write the equations of motion

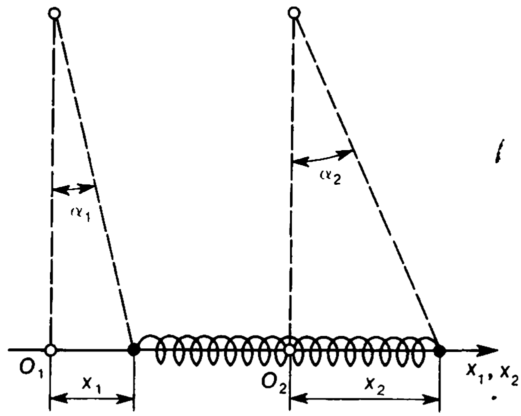


Fig. 160. Calculating the deviations of the oscillations of coupled systems.

of the given simple pendulums in the following form assuming that they have the same length l:

$$\ddot{\alpha}_1 = -\frac{g}{l}\alpha_1, \quad \ddot{\alpha}_2 = -\frac{g}{l}\alpha_2, \tag{55.3}$$

where α_1 and α_2 are the angles of deflection of the pendulums from the vertical position. The deviations from the equilibrium position are connected with the angles α_1 and α_2 through the obvious relations (see Fig. 160): $x_1 = \alpha_1 l$ and $x_2 = \alpha_2 l$. Hence the equations of motion of the point masses have the following form if we neglect their coupling through the spring:

$$\ddot{x}_1 = -\frac{g}{l}x_1, \quad \ddot{x}_2 = -\frac{g}{l}x_2. \tag{55.4}$$

The deformation of the spring gives rise to forces proportional to elongation (Hooke's law). The elongation of the spring is $x_2 - x_1$, and hence the forces acting on the point masses are

$$F_1 = -F_2 = D(x_2 - x_1), \tag{55.5}$$

where D is the proportionality factor. Hence, taking into account the coupling forces introduced by the spring, we can write the equations of motion as follows:

$$\begin{aligned} \ddot{x}_1 &= -\frac{g}{l}x_1 + \frac{D}{m}(x_2 - x_1), \\ \ddot{x}_2 &= -\frac{g}{l}x_2 - \frac{D}{m}(x_2 - x_1), \end{aligned} \tag{55.6}$$

where m is the identical mass of the point masses. These two coupled equations can easily be solved as follows. The addition and subtraction of the right- and left-hand sides of these equations gives

$$\ddot{x}_1 + \ddot{x}_2 = -\frac{g}{l}(x_1 + x_2),$$

$$\ddot{x}_1 - \ddot{x}_2 = -\frac{g}{l}(x_1 - x_2) - \frac{2D}{m}(x_1 - x_2).$$

Thus, the equation for the sum and the difference of the pendulum's deviations has the same form as the equations for natural harmonic oscillations:

$$\begin{aligned}(\ddot{x}_1 + \ddot{x}_2) + \omega_1^2(x_1 + x_2) &= 0,\\ (\ddot{x}_1 - \ddot{x}_2) + \omega_2^2(x_1 - x_2) &= 0,\end{aligned} \tag{55.7}$$

where

$$\omega_1 = \sqrt{\frac{g}{l}}, \qquad \omega_2 = \sqrt{\frac{g}{l} + \frac{2D}{m}}. \tag{55.7a}$$

The solution of these equations is well known:

$$\begin{aligned}x_1 + x_2 &= A_0 \sin(\omega_1 t + \varphi_1),\\ x_1 - x_2 &= B_0 \sin(\omega_2 t + \varphi_2).\end{aligned} \tag{55.8}$$

Hence adding and subtracting the left- and right-hand sides, we obtain for the deviations x_1 and x_2:

$$\begin{aligned}x_1 &= \frac{A_0}{2}\sin(\omega_1 t + \varphi_1) + \frac{B_0}{2}\sin(\omega_2 t + \varphi_2),\\ x_2 &= \frac{A_0}{2}\sin(\omega_1 t + \varphi_1) - \frac{B_0}{2}\sin(\omega_2 t + \varphi_2).\end{aligned} \tag{55.9}$$

As expected, these formulas are the same as (55.1) if we put $A = A_0/2$ and $B = B_0/2$. Hence the quantities ω_1 and ω_2, defined by (55.7a), are normal oscillation frequencies of the coupled system under consideration with two degrees of freedom.

PROBLEMS

Two springs having the same length l and rigidity D_1 and D_2 may form a combination of two springs of the same length l (parallel connection) or of length $2l$ (series connection). Find the circular frequencies of oscillations of a load of mass m suspended at the lower end of the combinations of the springs if their upper ends are rigidly fastened.
13.1. A weightless pulley is suspended at the lower end of a weightless spring of rigidity D_1. The upper end of the spring is rigidly fastened. A weightless elastic string of rigidity D_2 passes over the pulley. One end of the string is rigidly fixed to the ground, while a load of mass m is suspended at the other end. In the state of equilibrium, both rectilinear
13.2. sections of the string and the spring are in the vertical position. The

frictional forces can be neglected. Find the circular frequency of small harmonic oscillations of the load under which the string is always stretched in the vertical position.

13.3. A small load, suspended at the lower end of an elastic string, stretches it in the equilibrium position by Δl. The upper end of the string oscillates along the vertical, its deviation from the equilibrium position being described by the formula $A \sin \omega_0 t$. The positive value of deviations is measured downwards. Assuming that the string is always stretched, find the equation of motion of the load, denoting its deviation from the equilibrium position by x.

13.4. A ring of mass m_1 can slide without friction over a horizontal rod. A point mass m_2 is suspended from the ring on an unstretchable weightless string of length l. Find the frequency of small harmonic oscillations of the system.

13.5. A hole in which a point mass can move is drilled along the diameter of the Earth which is assumed to be a homogeneous sphere. Find the period of oscillations of the point in the hole near the centre of the Earth. The acceleration due to gravity at the surface of the Earth and the radius of the Earth are denoted by g and r respectively.

13.6. A hydrometer of mass m having a cylindrical tube of diameter d floats in a liquid of density ρ and is set into motion by a push in the vertical direction. Find the frequency of small oscillations of the hydrometer. The motion of the liquid and its resistance to the motion of the hydrometer can be neglected.

13.7. A homogeneous rod is suspended horizontally with the help of two vertical strings of length l each tied to its ends. In the state of equilibrium, the strings are parallel. Find the period T of small oscillations excited as a result of a twisting of the rod about a vertical axis passing through its centre.

13.8. A circular ring of radius r rotates about its vertical diameter at an angular velocity ω. A sphere of mass m can slide over the ring without friction. The position of the sphere is defined by the angle θ between the downward vertical direction and the direction of the radius vector of the sphere drawn from the centre of the ring. For what value of the angle θ will the sphere be in a stable equilibrium and what will be the frequency of its small oscillations about the equilibrium position?

ANSWERS

13.1. $\sqrt{(D_1 + D_2)/m}$, $[(D_1^{-1} + D_2^{-1})m]^{-1/2}$. 13.2. $\{D_1 D_2/[m(4D_2 + D_1)]\}^{1/2}$. 13.3. $\ddot{x} + (g/\Delta l)x = (gA/\Delta l)\sin \omega_0 t$. 13.4. $[(m_1 + m_2)g/(m_1 l)]^{1/2}$. 13.5. $\sqrt{g/r}$. 13.6. $d\sqrt{\pi \rho g/m}/2$. 13.7. $2\pi\sqrt{l/(3g)}$. 13.8. $\arccos\,[g/(r\omega^2)]$, $\omega\sqrt{1 - g^2/(a^2\omega^4)}$.

Appendix 1

SI Units Used in the Book

Quantity			Unit	
name of quantity	dimensions	basic symbol	name of unit	symbol
Base units				
Length	L	l	metre	m
Mass	M	m	kilogram	kg
Time	T	t	second	s
Electric current	I	I	ampere	A
Thermodynamic temperature	Θ	T	kelvin	K
Amount of substance	N	ν	mole	mole
Luminous intensity	J	I	candela	cd
Derived units				
Area	L^2	S	square metre	m^2
Volume	L^3	V	cubic metre	m^3
Plane angle	dimensionless	α, φ	radian	rad
Solid angle	dimensionless	Ω	steradian	sr
Period of oscillations	T	T	second	s
Frequency of periodic process	T^{-1}	ν	hertz	Hz
Circular frequency	T^{-1}	ω	inverse second	s^{-1}
Angular velocity	T^{-1}	ω	radian per second	rad/s
Angular acceleration	T^{-2}	α	radian per second squared	rad/s^2
Velocity	LT^{-1}	v, u	metre per second	m/s
Acceleration	LT^{-2}	a	metre per second squared	m/s^2
Density	$L^{-3}M$	ρ	kilogram per cubic metre	kg/m^3
Momentum	LMT^{-1}	p	kilogram-metre per second	kg·m/s
Force	LMT^{-2}	F	newton	N
Pressure	$L^{-1}MT^{-2}$	p	pascal	Pa
Moment of force	L^2MT^{-2}	M	newton-metre	N·m
Angular momentum	L^2MT^{-1}	L	kilogram-square metre per second	$kg\cdot m^2/s$
Moment of inertia	L^2M	J	kilogram-square metre	$kg\cdot m^2$
Work	L^2MT^{-2}	A	joule	J
Energy	L^2MT^{-2}	E	joule	J
Potential energy	L^2MT^{-2}	E_p	joule	J
Kinetic energy	L^2MT^{-2}	E_k	joule	J
Power	L^2MT^{-3}	P	watt	W
Phase of harmonic oscillations	dimensionless	φ		
Damping factor	T^{-1}	γ	inverse second	s^{-1}
Logarithmic decrement	dimensionless	θ		
Q-factor	dimensionless	Q		
Rigidity	MT^{-2}	D	kilogram per second squared	kg/s^2
Coefficient of friction	MT^{-1}	β	kilogram per second	kg/s
Electric field strength	$LMT^{-3}I^{-1}$	E	volt per metre	V/m
Magnetic induction	$MT^{-2}I^{-1}$	B	tesla	T

Appendix 2

Physical Constants Encountered in the Book

Quantity	Symbol	Numerical value
Velocity of light in vacuum	c	2.99792458×10^{8} m/s
Acceleration due to gravity	g	9.80665 m/s^2
Electron charge	e	$1.6021892 \times 10^{-19}$ C
Electron rest mass	m_e	9.109534×10^{-31} kg
Proton rest mass	m_p	$1.6726485 \times 10^{-27}$ kg
Neutron rest mass	m_n	$1.6749543 \times 10^{-27}$ kg
Gravitational constant	G	6.6720×10^{-11} m^3/(kg·s^2)

Conclusion

The basic concepts of mechanics were worked out in the course of a prolonged evolution rooted many centuries back. The ideas of space and time formed in human consciousness were inseparably linked with these concepts. These concepts and ideas served as the foundation for the development of classical physics. Mechanics has always held primary in the system of physical concepts.

The basic concepts of classical mechanics are associated with the ideas about a body, a point mass, motion of a point mass along a certain trajectory and force as the cause of peculiarities of motion of a body or a point mass. The interrelation between the concepts of space and time and the concepts of mechanics was refined during the evolution of science, but these concepts have always formed the basis of classical physics. The most important result of this evolution is the establishment of an inseparable link between space, time, matter and motion. These ideas are laid down in a clear philosophical form in the teachings of dialectical materialism. In dialectical materialism space and time are forms of existence of matter and are therefore inconceivable without matter. Motion is a mode of existence of matter.

Although modern concepts of classical mechanics were developed since the time of Newton, Aristotle was the first to put forth the main ideas associated with the description of motion of bodies. These ideas were retained in their entirety by Newton who formulated new laws of motion differing from Aristotle's laws. Aristotle divided all kinds of motion into two categories: natural and forced. Natural motion takes place on its own without any external influence, and it is senseless to raise the question of the cause behind natural motion. Forced motion does not take place on its own and is carried out under the action of external factors described by the concept of force. Aristotle considered natural motion to be the upward motion of light bodies, the downward motion of heavy bodies and the motion of heavenly bodies in the celestial sphere. All other types of motion are forced and can be explained only as being the result of the action of force. Force is imparted to a body from the surrounding space which according to Aristotle is never empty but is always filled with a certain medium. Aristotle's law of mechanics is similar in nature to Newton's second law of motion and can be reduced to the statement that velocity is proportional to force.

The approach towards the problems of motion contained in Aristotelian mechanics contained all basic elements of the theory which subsequently formed the basis of classical mechanics. The formulation of the basic laws of classical mechanics is based to a considerable extent on the analysis of the physical content of these original basic elements of the theory with the help of experimental investigation of objects, phenomena and processes which are studied in classical mechanics. The physical meaning of the fundamentals of classical mechanics has undergone significant changes, but Aristotelian mechanics has always remained the first fundamental stage in the development of mechanics.

Newtonian mechanics was the second fundamental stage in the development of classical mechanics. The first step towards the creation of this stage of classical mechanics was taken by Galileo who formulated the relativity principle in mechanics, the laws of inertia, the law of free fall of bodies and the law of summation of motion.

In Newtonian mechanics, natural motion (including the natural motion as understood by Aristotle) is the motion by inertia (first law). It does not require the action of any force for its description. The action of a force is not a result of the velocity of a "forced" motion, but rather the acceleration of bodies (second law). The third law of motion establishes the general property of forces in the interaction of bodies. The question of force in this context was not raised in Aristotelian mechanics.

Although the general approach towards the problem of motion was retained in Newtonian mechanics without any significant changes, the theory of motion put forth by Newton marks a fundamental change from Aristotle's theory. The questions related to the introduction of coordinate systems, including the relativity principle, the properties of interaction of bodies, a new equation of motion and a further development of the concepts of space and time are among the main aspects of Newton's theory.

During the evolution of Newtonian mechanics, its basic concepts were subjected to a deep analysis. One such basic concept is that of mass. Newton defined this concept as follows: the amount of matter is its measure determined by its density and volume together. Later, Newton indicated that the amount of matter will be called mass, thus defining mass in terms of density. However, he never gave an independent definition of density. Thus, mass was defined in terms of mass per unit volume, i.e. the definition of mass was tautological.

For this reason, the concept of mass was subjected to thorough analysis, which resulted in the statement that mass is a measure of inertial properties of bodies. This definition of mass is widely accepted at present.

Another important concept of classical mechanics is that of force. According to Newton, the applied force is the action on a body to change its state of rest or of uniform motion in a straight line. The force is manifested only in its action and does not remain in the body once the action ceases. Newton did not analyze the nature of forces. He treated the concept of force phenomenologically, assuming it to be something given. Subsequently, however, much attention was paid to determining the nature of force, which seemed enigmatic to many. But one thing was clear: the force is due to the properties of matter. According to Euler, the essence of force lies in the fundamental properties of matter, i.e. inertia and impenetrability. The mysterious nature of the concept of force led to attempts at constructing mechanics without forces. Such efforts, however, proved to be futile, and the concept of force continued to exist in classical mechanics as a basic concept. After the advent of quantum mechanics which does not use the concept of force, it was opined that the force need not have an objective nature in classical mechanics either. This conclusion is erroneous since quantum mechanics does not use other concepts of classical mechanics like a point mass moving along a given trajectory, acceleration, velocity and position in space in successive intervals of time, which impart physical meaning to the concept of force. Hence, force is one of the basic concepts of classical mechanics, and attempts to divest mechanics of this concept must be rejected.

Considerable attention was paid to investigations into the physical nature of Newton's laws and to the substantiation of mechanics in the course of its evolution. The idea that Newton's second law is a definition of force and not a law of nature is refuted by physical, historical and logical considerations presented in this book. Hence it is unacceptable to substantiate the concepts of mechanics and the definition of mass without using force as a basic concept.

The nature of inertial forces is fundamental in classical mechanics. In the framework of Newton's concepts about relative and absolute space, inertial forces are caused by accelerated motion relative to the absolute space and can easily be described, especially if the absolute space is assumed to be filled with ether. However, once the concepts of absolute space and ether had been rejected, the problem of inertial forces became important again in the framework of new concepts.

In addition to the basic concepts and ideas of classical physics, the latter also contains the description of the methods of physical investigations and formulation of physical theories. These concepts, ideas and methods were successfully used for creating the special and general theory of relativity and for investigating the field form of matter, especially in classical electrodynamics.

The creation of the general and special theory of relativity was the third culminating stage in the evolution of classical mechanics. At this stage, the inseparable link between space and time, motion and matter was established to the fullest possible extent. The requirement of relativistic invariance of the theory became an important heuristic principle of physical investigation. A transition was made from the formulation of the theory in the four-dimensional space-time manifold. The geometry of this manifold has become an important element of the theory. One of the important results preceding the evolution of the physical theory was the conclusion that the connection between physical quantities must be expressed in the form of local relations.

The study of phenomena in the microcosm revealed that the classical concepts are inadequate for explaining such phenomena. On the other hand, it is obvious that human intellect is not armed with any concepts except classical ones for constructing a theory to explain the phenomena of the microcosm, while the human experience and existence were, are and, apparently, will continue to be macroscopic. This means that the properties of objects of the microcosm and the laws of their motion must be described by using the concepts of classical physics. This problem was solved by quantum mechanics and quantum physics, based on it. The concepts, ideas and methods of classical mechanics has also retained their significance in quantum physics.

Mechanics has always been of prime importance in the life and activity of man since it is associated with all phenomena and processes involving macroscopic bodies. Hence of all the physical sciences, mechanics was the first discipline which was developed as a science and applied not only for cognition of the world surrounding us, but also for practical utilization of its laws. Later, other branches of physical sciences were also developed, but, as a rule, a practical application of these disciplines was impossible without mechanics. Hence, mechanics has continued to play a significant role in the life and activity of man, and in the growth of creative power.

At present, mechanics has not lost its importance in the advancement of science and engineering since it is linked with practically all aspects of human life and activity. Mechanics is vital for designing all kinds of machines and mechanisms, in construction, transport, space technology, rocketry, robotics, etc. Even the progress in electronics is closely connected with mechanics. For example, the motion of an electron beam which forms an image on a display screen is calculated with the help of the laws of classical mechanics. Data processing from magnetic tapes, floppy discs or rigid discs require the design of high-precision mechanical elements, and so on. If we take into account the fact that the design and construction of any instrument or equipment inevitably involve classical mechanics at least for the optimization of material consumption, there is hardly any branch of science and engineering that could do without mechanics.

Subject Index

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make.

Our address is:
Mir Publishers
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820

By the Same Author:

A. N. Matveev, D. Sc. (Phys.-Math.)

Electricity and Magnetism

This is a part of a course on general physics. The first book (*Molecular Physics*) was published in 1985.
The book starts with the experimental verification of the theory of electricity and magnetism, and is based on relativistic concepts with which the reader is supposed to be familiar. The relation between the electric and magnetic fields has been derived at the very beginning. Besides the traditional questions, certain modern questions like the fluctuations of current in electric circuits, anomalous skin effect, waveguides and resonators have also been discussed.
Intended as a textbook for university students.

A. N. Matveev, D. Sc. (Phys.-Math.)

Optics

This is a part of a course on general physics. The first two books (*Molecular Physics* and *Electricity and Magnetism*) were published in 1985 and 1986.

The material presented in this book is developed on the basis of the electromagnetic theory. Fourier transformations have been used for monochromatic, nonmonochromatic and random radiation. The description of traditional aspects is supplemented by a discussion of problems from the branches that have been developed intensively over the last 15-20 years. These include Fourier optics, Fourier analysis of random signals, matrix methods in geometrical optics, holography, lasers, nonlinear phenomena, etc.

ERRATA

p. 59 Equation (8.19) should be written as:

$$\mathbf{a} = \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{\tau}\mathbf{v}) = \frac{\mathrm{d}\boldsymbol{\tau}}{\mathrm{d}t}v + \boldsymbol{\tau}\frac{\mathrm{d}v}{\mathrm{d}t}.$$

p. 67 The notation in Fig. 22 should be written as:

$\mathbf{a}_\tau$ $\mathbf{a}_n$

p. 75 The answer to Problem 2.13 should be written as:

2.12. $\mathbf{i}_x + \mathbf{i}_y - \mathbf{i}_z$, ...

p. 264 The first line should be written as:

moment of force $\mathbf{F}_1$ ...

p. 344 The third line from below should be written as:

where $v_\perp = \ldots$

p. 365 The first line of Eq. (51.11) should be written as:

$$\langle \cos^2(\omega t + \varphi)\rangle_t = \frac{1}{T}\int_0^T \cos^2(\omega t + \varphi)\,\mathrm{d}t$$

...

p. 400 The number of Problem 13.1 should be put at the level of the first line of the brevier and that of Problem 13.2 at the level of the sixth line.