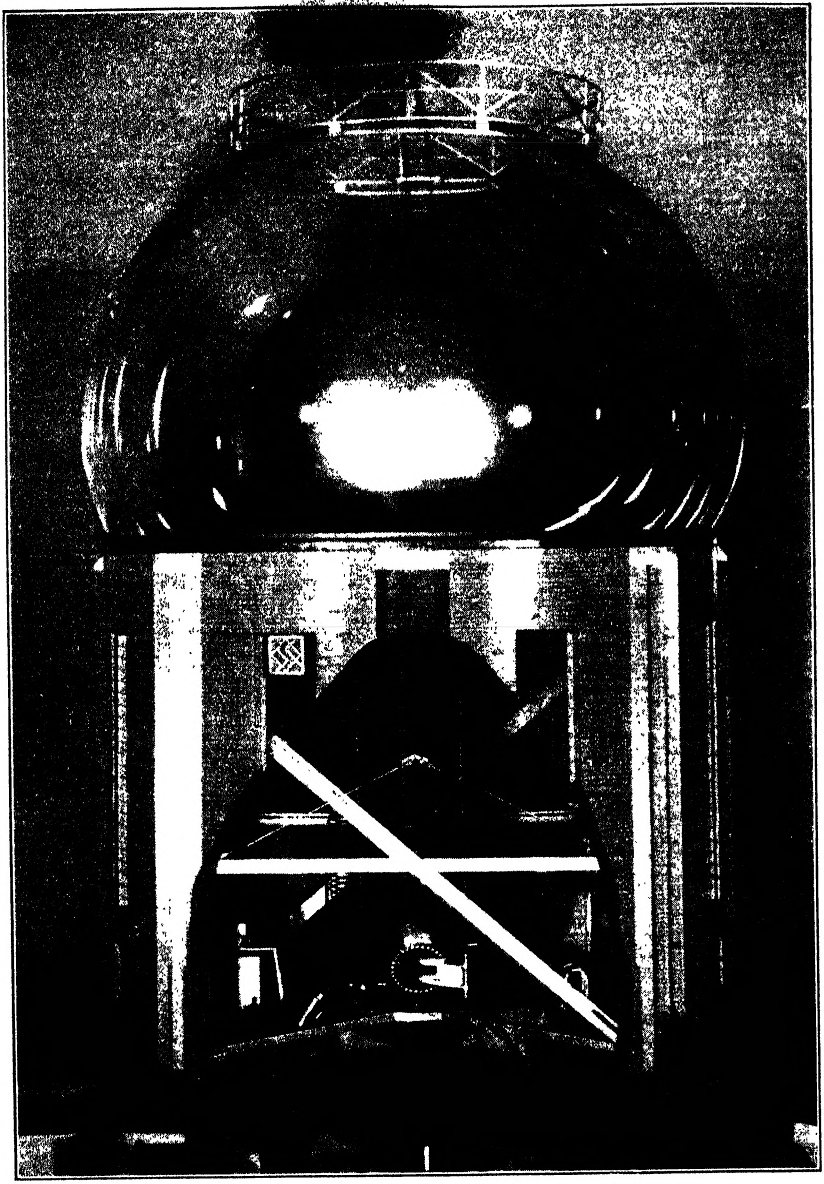


WORKING MODEL OF THE ATOMIC PHYSICS LABORATORY OF THE CARNEGIE INSTITUTION OF WASHINGTON, D. C. (Photograph used through the courtesy of Dr. J. A. Fleming, Director of the Department of Terrestrial Magnetism of the Carnegie Institution.)

FOUNDATIONS OF

Modern Physics

THOMAS B. BROWN

Professor of Physics

The George Washington University

Second Edition

New York · JOHN WILEY & SONS, Inc.
CHAPMAN & HALL, LIMITED
London

SECOND EDITION
Second Printing June, 1949

PRINTED IN THE UNITED STATES OF AMERICA

TO

Floyd Karker Richtmyer

"The Prof"

Preface to the Second Edition

Since the first edition of this book appeared, in 1940, the world has entered upon a period of political, social, and economic change which affects every phase of human thought and endeavor. The science of physics is not excepted. In many ways, indeed, it has played a leading part in these changes, both in its contributions to the conduct of war and in its influence upon everyday life. The importance of physics to the progress of civilization is recognized now as never before.

The science of physics could not have attained this importance without undergoing extensive changes within itself, and much has been added to its content in the past nine years. A large share of these additions may be called applied physics and depends, for the most part, upon fundamental principles which were well known long before this period began. Thus even the "atom" bomb depends upon principles which were public knowledge in time to be included in our first edition. Even so, these applications add greatly to the interpretation of many parts of physics. It is perhaps even more significant that this period has included also a surprisingly large amount of fundamental research, even when military need put the emphasis upon applications. While most of this research was necessary to support the practical developments, it represents also a certain innate characteristic of the scientific mind; a true scientist never can divorce his work from fundamental research. This has been so far a most noteworthy period in the history of science.

To the teacher of physics these accomplishments of physics have a dual significance. First, they have supplied much new material with which to illustrate and vitalize many fields of his teaching. The developments in microwaves provide interesting examples of this. Second, they have created interest on the part of students, and inquiries from the public at large, which are both challenge and inspiration to better teaching in these fields. Furthermore, although the developments in fundamental research have been few in number, they have made important contributions to physical theory.

These are the factors which have inspired revision of this book. A comparison with the first edition will show that the major changes appear in the sections dealing with electronics, with electromagnetic waves, especially microwaves, and with nuclear physics and the theory of elementary particles. Minor revisions have been made throughout, wherever they seemed desirable. Finally, an "Introduction" and an "Epilogue" have been added to explain the objectives and the philosophy of the book. These revisions have been greatly facilitated by the generous help of many people. One of the pleasures in writing a book in science is supplied by the willingness with which other scientists, even complete strangers to the author, will respond to requests for permission to use copyrighted material, for information in advance of publication, and for illustrative material such as photographs. Such help is gratefully acknowledged here, as well as in those places where such material appears.

The contents have now outgrown the scope of the one-term course at George Washington University for which it was written originally. It is still possible to employ it in a one-term course by making *study* assignments to the more important sections, treating the less fundamental details as reading assignments. On the other hand, there is now enough material for a year course, especially when reading assignments are made to the references at the end of each chapter. It is strongly recommended that any course be accompanied by demonstration lectures. Many ideas for such demonstrations are to be found in *Demonstrations in Physics*, by R. M. Sutton, and in the *American Journal of Physics*. Individual laboratory work could be included very profitably in a year course.

THOMAS B. BROWN

The George Washington University
October 3, 1948

Preface to the First Edition

What is meant by "modern physics"? Certainly it means those fields in the realm of physics which are at present greenest with new growth. But from the point of view of a book such as this one it includes also those older parts of physics which are necessary to the understanding of the work in these frontier fields, and it also properly includes the more important and more fundamental of the practical applications which have followed discoveries whose primary importance is to science itself.

For some years now a course in modern physics has been given at the George Washington University as the fourth unit of a two-year general course in physics, the scope and content of this course being essentially as defined above. This book grows out of the experience of the author in teaching this course, and is the successor to a briefer lithoprinted book which has been used in it for several years. Emphasis throughout is laid upon the experimental aspects of modern physics, and upon the evidence which these experiments give in support of the new theories. Emphasis is given also to the close relationships which exist between modern physics and the older physics out of which the recent advances have grown. The development of each topic follows the order which may be interpreted as logical from our present vantage point, rather than the historical order. Although at times this may seem incongruous to those already familiar with the topics involved, it simplifies the approach for the beginner.

It is hoped that this book may assist students in all fields of learning to obtain some acquaintance with the fundamental discoveries of modern physics, together with some understanding of the theories whereby these discoveries are explained and interpreted. To further this purpose, technical terms and phraseology have been avoided except where such terms may be fully explained, and only elementary mathematics is used. This does not imply that the treatments are only qualitative but, rather, that the quantitative developments are made in simple terms, with careful interpretations of the mathematical processes involved, instead of by means of the advanced mathematical methods so necessary to advanced theory. At the same time every effort is made to keep these treatments exact. This is accomplished, where necessary, by limiting their scope to specific cases. The problems assist by further illustrating the principles and by providing concrete examples of the magnitudes involved in the various phenomena.

ACKNOWLEDGMENTS

The author takes great pleasure in acknowledging here his deep indebtedness to all those who have assisted him in so many ways. First, he wishes to thank all those who have kindly sent him photographs and other materials from which illustrations have been made, as is noted elsewhere. Many thanks for assistance and advice in the preparation of the manuscript are then due to his colleagues at the George Washington University, Dr. G. Gamow, Dr. E. Teller, and Dr. W. L. Cheney, and likewise to Dr. T. H. Johnson, of the Bartol Research Foundation of the Franklin Institute, who assisted him similarly with the chapter on cosmic rays. Finally, he wishes to acknowledge a debt of longer standing to Dr. P. Scherrer, who gave a series of demonstration lectures on modern physics at Massachusetts Institute of Technology, in the winter of 1930–31. These lectures first revealed to the author the possibility of presenting modern physics in both an exact and an elementary manner.

THOMAS B. BROWN

THE GEORGE WASHINGTON UNIVERSITY
March 4, 1940

Contents

INTRODUCTION

i. Modern Physics. *Modern physics* is an elastic term, having different meanings for different people. To the research physicist, for example, it may mean only the very frontiers of science along which he and his fellow scientists are working. To most of us, however, it is a much broader term, including also those discoveries and developments which may have become "classic" to the research workers, but still may be considered new to scientists generally, as well as to the average well-educated man. This broader definition, which will be followed here, presents two general aspects. First of all comes fundamental science; this includes discoveries such as the electron, the neutron, and the meson, and all other new developments which extend our scientific knowledge of the physical world, together with the laws and concepts which interpret them. Along with these fundamental developments, however, we find many discoveries and inventions which are also of practical value, such as X-rays, radar, and nuclear energy sources; these practical applications of modern physics constitute the second aspect. This latter aspect, indeed, is all that is known about modern physics by many people, and to all of us it is of great interest. In so far as such applications illustrate the fundamental principles of physics, they also have place in a book such as this one.

ii. The Unity of Physics. Modern physics cannot be understood properly without first understanding its relation to the rest of physics. A student just beginning the study of physics finds his textbook divided into sections on mechanics, heat, light, sound, electricity, and magnetism, which at first appear to have little in common. However, he does not progress far beyond the most elementary considerations in any section before he discovers it necessary to possess greater or less knowledge from several, if not all, of the other sections. The various sections of physics did not come together by chance, but are closely related and essential components of the whole science of physics. Indeed, physics itself is only one large component of the still larger body of physical science.

Modern physics is even more strongly dependent upon the older sections of physics than these sections are dependent upon one another. This will be apparent from the beginning to the end of this book. The student will find it necessary to keep his general-physics textbooks near at hand, for frequent reference and review, if he is to understand properly the newer developments. Contrary to certain popular notions, nothing in modern physics contradicts or sets aside any of the facts or laws of the older physics, within the range of experience for which they have been found valid. Whenever new concepts appear to be at variance with earlier ideas, the disagreements are with *extrapolations* of the earlier laws and concepts, beyond the range of experience for which they were established. Now a good scientist never assumes that any extrapolations of the laws of science are valid until they have been tested by experiment. Nor does he carry out explorations in new regions for the purpose of *proving* the validity of any extrapolations. It is a fundamental principle of science to search for the truth, then to make laws which fit the discoveries. If the extrapolations do not fit the new experiments, new laws are required, although the old laws remain valid over the range of experience for which they were originally established.

What modern physics has done concerning the older laws and concepts is to reveal the limitations of many, and to set up new ones to cover the ranges of experience for which the older ones are inadequate. In most cases, indeed, these newer laws are valid over the entire range of experience, such that the older ones must be considered as approximations to the newer. This does not, however, discredit the older laws. Within the range for which they were originally established they still are valid, and generally more convenient to use than the newer ones. Many examples of this may be found, even in the older parts of physics. Thus the "rays" of geometrical optics are only approximations to the paths followed by light waves, but they provide the easiest way to solve many optical problems. Boyle's law is another example. This law has definite limitations (see Sec. 156) but is always used when these limitations are not exceeded.

iii. Experimental Evidence. One of the greatest differences between modern physics, in the broader sense in which this term is employed in this book, and the older parts of physics is their different points of view. The earlier physics boasted of being an experimental science, and, for the most part, it consisted of fairly direct generalizations derived from experimental data. We may say that the experimental data provided *direct evidence* to establish and support the laws of this earlier part of physics. The physical existence of the entities of this earlier physics was unquestioned because those entities were directly observable. Modern physics also boasts of being an experimental science, but the entities with which it is most concerned are, with few exceptions, hopelessly invisible. For the most part, the experimental data from which all our knowledge concerning atoms, electrons, photons, and the like is derived, provide *indirect* or *circumstantial evidence*, rather than direct evidence. This in no wise invalidates the deductions from such data. Circumstantial evidence can be the most reliable kind of evidence provided there is enough of it and observers know how to interpret it. Modern physics exists because physicists have learned how to ascertain the validity of such evidence, and how to interpret it.

iv. Beginnings of Modern Physics. From this point of view the beginnings of modern physics were made nearly 100 years ago, when the kinetic theory of gases first was introduced. This theory encountered stormy opposition, which was not silenced until near the end of the nineteenth century. Even earlier (about the beginning of the nineteenth century) the atomic theory for matter met determined opposition in the field of chemistry, on similar grounds; it was insisted that the existence of atoms could not be proved unless they could be made visible. The opposition did not trust circumstantial evidence! The experimental proof for the wave theory of light, during the early years of the nineteenth century, must also be included among these early examples of indirect or circumstantial evidence. Curiously enough, this theory was accepted far more readily than was the kinetic theory of gases nearly fifty years later. In part this may be accounted for by the nature of the discovery; it was not unreasonable for waves to be invisible. A more important factor, however, was the genius of the experimenters, whose methods serve today (see Chapter III) as an outstanding example of how to obtain and analyze such data.

v. Foundations of Modern Physics. What has been said above makes clear how important it is that the student base his study of modern physics upon a clear understanding of its *foundations.* At this point in his studies it is far less important that he know all the details of the quantum theory, or the properties of the elementary particles, than that he understand how that theory ever was established, or how the existence and properties of those particles have been discovered and measured. Many of these foundations go back to the basic physics of many years ago, and all represent natural outgrowth from the beginning to the present time. Modern physics cannot be divorced from the rest of physics because it is an integral part of it.

One purpose of this book is to survey the present knowledge in the various fields of modern physics, in so far as this may be done on an elementary level. An even more important purpose is to set forth these *foundations* of modern physics, and to acquaint the reader with the methods employed by scientists in their search for more knowledge. With this purpose in mind the book includes several chapters which belong also in a general-physics textbook, such as Chapters III and IV, covering the wave properties of light, and Chapter XI, on the kinetic theory. In addition to these fundamental considerations, it includes also sufficient reference to practical discoveries and applications to illustrate the principles involved and to indicate the significance to the world generally of scientific work in these fields.

vi. Outline. The content of this book may be divided roughly into four parts.

The first part, which includes Chapters I to VI, concerns what may be referred to as particles which behave as particles, and waves which behave as waves. The particles include electrons, atoms, and molecules, and here we are concerned with their physical properties, such as mass, volume, and electric charge, and with methods for measuring these properties. Waves include the entire electromagnetic spectrum, from radio waves through X-rays, and we are concerned with methods for identifying these radiations, measuring their wavelengths, and studying their properties. The historic proofs for the wave character of light are included as necessary background and support for the same methods when applied to the newer portions of the spectrum. The applications related to this section include a large part of electronics, radio, and X-rays.

Part two, which includes Chapters VII to X, relates generally to the dual wave-particle aspects of matter and energy, or to the particle characteristics of waves and the wave characteristics of particles. Here are encountered the fundamental concepts of the quantum theory, beginning with Einstein's explanation for the photoelectric effect. Here belongs also an elementary consideration of atomic structure, as revealed by the data of spectroscopy. Applications include electron lenses and the electron microscope.

Part three, Chapters XI to XIV, is concerned chiefly with the kinetic theories for atoms and molecules in solids, liquids, and gases, electrons within metals, and photons in black-body radiation. It gives brief consideration also to molecular forces and crystal structure.

The fourth and concluding section surveys nuclear physics and the physics of elementary particles, including the important topic of nuclear or "atomic" energy. It begins with Chapter XV, on radioactivity, and runs through Chapter XIX, on cosmic rays. Although this field of physics is the latest to be explored and includes most of the present frontiers of physical research, there is an even more important reason for leaving it until last. The developments in this field have grown out of developments made in all the fields considered earlier, and may be understood only to the extent that these earlier developments have been comprehended. In this most recent field of physics the essential unity of the science is again demonstrated.

vii. Additional Reading. This book will be most successful if it creates interests on the part of its readers which they cannot satisfy within these pages. It then will serve as a guidebook to further reading and study. The references listed at the end of each chapter will assist in this. They have been chosen carefully and are divided into two groups, according to the preparation of the reader in physics and mathematics. One group of references may be read by students having the degree of preparation required by this book. Students with greater preparation may read with profit both groups of references.

viii. Problems. Finally, the student is urged to work as many problems as possible. There is no better way to check one's knowledge in any field than by trying it out on specific examples—in other words, by working problems. The author has attempted to include among the problems found at the end of each chapter those which serve in this manner to point up the work of the text. In numerous cases they supplement the text by bringing out additional points of interest. For the most part the problems are relatively easy to solve, although simple ones of the substitution-in-formula type have been avoided as far as possible. The student will obtain the greatest value from the problems when he first thinks out the *physical* solution in terms of the physical processes involved, then proceeds to obtain a mathematical solution.

CHAPTER I

THE ELEMENTARY ELECTRIC CHARGE AND THE ELECTRON

1. The Granular Nature of Electric Charges. One of the cornerstones of the structure of modern physics is the existence of an elementary electric charge, such that all electric charges, of either sign, are composed of greater or less numbers of these elementary charges. This idea is now a familiar one, even to those having only a casual acquaintance with physics. But if it is not to be accepted without question, merely because it is familiar, or because it has the authority of textbook writers behind it, how may its verity be proved? The existence of an elementary or smallest quantity of electric charge was made evident by the researches in electrolysis of Faraday and was firmly established before Townsend and H. A. Wilson made the first rough measurements of the value of this charge. The outstanding features of this earlier experimental evidence will be presented later on in this chapter. Disregarding the historical order, the evidence which will first be considered is that given by Millikan's method for measuring the elementary quantity of electric charge.

2. Millikan's Oil-Drop Experiment. The ordinary laboratory instruments for measuring electric charge, such as the ballistic galvanometer or even the most sensitive electroscope, are none of them sensitive enough to give any evidence that electric charges are not infinitely divisible. To measure the elementary electric charge a new method had to be developed, and the one which Millikan found to be sensitive enough proves also to be the simplest and the most direct method possible.

This method is to measure the force exerted upon the charge by a known electric field. It is illustrated in Fig. 1. *A* and *B* represent two horizontal metal plates, seen in cross-section; in the original apparatus of Millikan they were about 10 in. in diameter and several centimeters apart. In the laboratory apparatus shown in the figure they are smaller and spaced about 5 mm apart. A uniform electric field may be produced in the space between these plates by a potential difference of several hundred volts. Any charged particle in this space will then be acted upon by this field, with a force which is proportional to the charge on the particle. The most suitable particle is a tiny oil droplet. A fog of oil droplets is sprayed into the cup, C, and ultimately one of these slowly settling droplets may fall through the small hole made for that purpose in the center of the top plate and come into view in the field of the telescope.

If this droplet then possesses an electric charge, q, the electric field, E, will exert upon it a force equal to Eq, which may be made to act upward by throwing the reversing switch in the proper direction. (If the droplet does not acquire a “frictional” charge as a consequence of being sprayed from the atomizer, it may be given a charge with the aid of a bit of radioactive material.) The droplet may be observed through the short-range telescope, T, and the electric field may then be varied in strength by varying the potential difference across the condenser until the electric force just balances w, the effective weight of the droplet, so as to hold it stationary. (The effective weight equals the true weight less the buoyant force of the displaced air.) When the droplet is stationary

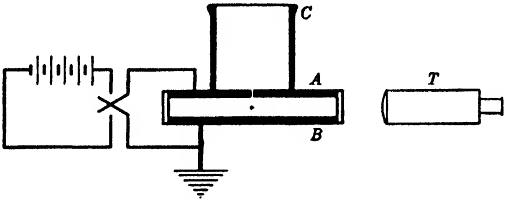


FIG. 1. MILLIKAN'S OIL-DROP EXPERIMENT.

$$Eq = w, \quad \text{or} \quad q = \frac{w}{E} \tag{1·1}$$

Actually, it is found more expedient to make the field stronger, so that the drop moves slowly upward when the field is acting, and then falls slowly downward with zero field. The mathematical reasoning is longer, but the principles are the same.

3. Weighing the Oil Drop. Before this charge may be computed, the oil droplet must be weighed; and no ordinary means of weighing will suffice for so small a mass, which is rarely greater than 10^{-8} mg. A new method of weighing had to be devised, and again a very fundamental physical principle was used. Any body moving through the air is opposed by forces due to the internal friction or *viscosity* of the air; if the body is of spherical shape, such as a balloon, soap bubble, or fog droplet, and is moving slowly, the force is that given by Stokes's law

$$F = 6\pi\eta vr \tag{1·2}$$

See Fig. 2. Here v is the speed of the sphere, r is its radius, and η is the coefficient of viscosity for the air. A soap bubble or a fog droplet will soon reach a steady speed where the forces acting upon it are balanced, i.e., such that the friction force equals the effective weight. Since the effective weight for an oil droplet may be computed from its volume and the densities of the oil and of the displaced air (d and d', respectively),

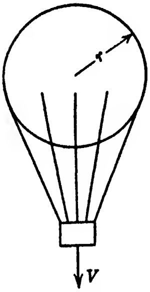


FIG. 2. STOKES'S LAW.

$$6\pi\eta v r = w = \tfrac{4}{3}\pi r^3(d - d')g \qquad (1\cdot3)$$

for an oil droplet. This equation, when solved for r, the radius of the oil droplet, gives[1]

$$r = \sqrt{\frac{9\eta v}{2(d - d')g}} \qquad (1\cdot4)$$

The droplet may thus be weighed by first observing its speed of fall under gravity (zero electric field), next computing its radius by equation (1·4), and then computing its effective weight from equation (1·3). Once this weight is obtained, the charge on the droplet may be computed as indicated by equation (1·1).

4. Value of the Elementary Charge, *e*. The smallest charge ever measured by this method has a value of 4.80×10^{-10} stat-coulomb (electrostatic unit of charge), and all other charges measured in this manner are found to be exact multiples of this amount, within the limits of experimental error. This smallest quantity of electric charge is always designated by e. It is possible to measure e by a variety of other methods, and all measurements agree very closely. The best available average for all measurements (see Appendix II) is

$$e = 4.8025 \times 10^{-10} \text{ stat-coulomb}$$

with an uncertainty indicated by a "probable error" of $\pm 0.0010 \times 10^{-10}$ stat-coulomb.

All of the evidence is in agreement in indicating that **this is the ultimate unit of electric charge, and that all charges, both positive and negative, are integral multiples of this charge.**

5. Cathode Rays. Although this experiment of Millikan's is commonly referred to as measurement of the electron charge, it does of itself give proof only for the existence of elementary charges, and provides no evidence as to how those charges are related to material parti-

---

[1] Millikan discovered that, for such small spheres as these oil droplets, Stokes's law requires a small correction term to give sufficiently accurate results. See reference at the end of the chapter to his work. For simplicity it is neglected here.

cles such as the electron. The discovery of the electron had been made earlier, of course, and is attributable in considerable part to the improvements in vacuum pumps which were made about 1860. At the lowest pressures which were producible before that time (about 10^{-4} atmosphere) an electrical discharge in a gas appears principally as a glow throughout the tube. This type of discharge is exemplified by the glowing tubes of luminous signs.[2] The improved vacuum pumps could reduce the pressure to values a hundred times or so lower than this, and at these lower pressures the phenomena of electrical discharge are essentially different. There is now very little glow visible in the residual gas, and the most noticeable phenomenon is a fairly bright glow which appears on the inside of the glass walls of the evacuated vessel.

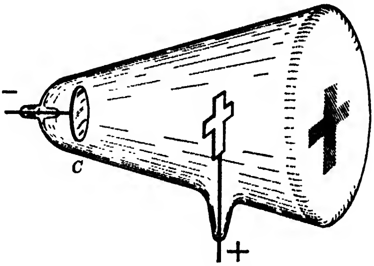


Fig. 3. Cathode-Ray Tube.

If a discharge tube such as is shown in Fig. 3 is used, the flat electrode, *C*, being made the negative terminal, or **cathode,** this glow will be confined to the end of the tube opposite to *C*, and any obstacle set up in the tube between *C* and the end of the tube will cause a sharp shadow to appear in the midst of the glow. These phenomena strongly suggest that this glow in the glass wall is due to some kind of rays which start out from *C* in a direction perpendicular to its surface and proceed in more or less straight lines across the tube to the opposite wall; if the pressure within the tube is not too low, and the discharge is strong enough, the path of this beam of rays may be visible, marked out by a faint glow in the residual gas. Since the negative terminal is called the cathode, these rays, whatever they may be, are called **cathode rays.** (Curiously enough, the position in this tube of the anode, or positive terminal, is relatively unimportant, and it may be put in a side tube as shown. The cause of this curious behavior of the discharge may be traced to the presence of the residual gas. The explanation will be given later on, in Sec. 137.)

6. Cathodoluminescence. A good many substances will luminesce, or glow as does glass, under the impact of cathode rays. This phenomenon

---

[2] Although usually called "neon" signs, these tubes may contain a variety of gases, the color of the glow depending chiefly upon the nature of the gas in the tube. Neon gives the red glow: helium, yellow; mercury vapor, blue; and mercury vapor in a yellow glass tube, green. The phenomena involved in this stage of the discharge will be discussed in Chapter IX.

has acquired a great deal of importance in recent years, in connection with cathode-ray-tube oscillographs (Sec. 13) and television (Sec. 74). Willemite (a variety of zinc silicate, containing a trace of manganese), which gives a brilliant green luminescence, is commonly used for cathode-ray-tube screens, although calcium tungstate, which luminesces a bright blue, is also sometimes used for this purpose, especially when the patterns produced on the screen are to be photographed. The luminescent properties depend upon the presence of traces of impurities of suitable kinds (manganese, in willemite), the color as well as the brightness being dependent upon the nature of the impurity. This is shown most remarkably by zinc sulfide, which may luminesce with a number of colors from blue to orange-red when "activated" by different impurities.

Beautiful effects may be produced by powdering luminescent materials and using them to paint designs upon metal or mica plates which serve as targets in cathode-ray tubes. Sometimes the glow occurs only while the cathode rays are falling upon the target; it is then called fluorescence. Sometimes the glow persists for a greater or less time after the cathode rays are stopped (this is especially true of some of the sulfides), and is then called phosphorescence. These same materials also luminesce under the influence of ultraviolet light (Sec. 87) and X-rays (Sec. 92).

7. Deflection of Cathode Rays. The physical nature of cathode rays is revealed by the deflections which they undergo upon passage through electric and magnetic fields. Deflection by an electric field may be seen by means of a tube of the form shown in Fig. 4. If the plates *A* and *B* are uncharged, the narrow beam of cathode rays which comes through the hole in the anode, *D*, will continue straight across the tube to the opposite end, where it will make a spot of fluorescent light. (If the gas pressure is high enough, a small fraction of the rays will be stopped by molecules of the residual gas, which will then emit light, and the path of the beam through the tube will appear as a faint line of light.)

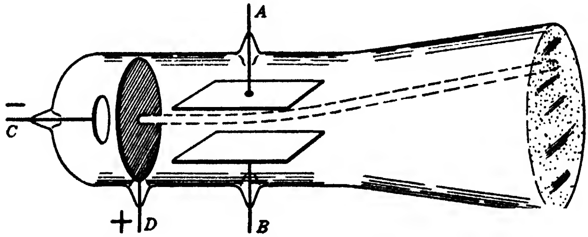


FIG. 4. ELECTROSTATIC DEFLECTION OF CATHODE RAYS. The plates *A* and *B* are charged + and − respectively by connection to a high voltage source.

If, however, a potential difference is set up between the plates A and B, by connecting them to a suitable voltage source, the resultant electric field between them will deflect the cathode-ray beam as shown. The direction of this deflection, away from the negative plate and towards the positive one, is evidence that the beam consists of *negative electric charges;* and the fact that the beam is not carried directly to the positive plate, but is only deflected somewhat out of its original path, is equally good evidence that the rays have inertia, i.e., that they consist of particles possessing *mass.*

The same tube may be used to show the deflection produced by a magnetic field. If a horseshoe magnet is held astraddle of the tube, with the north pole in front and the south pole behind, so that the lines of force of its magnetic field are perpendicular to the beam of cathode rays, the beam will be deflected downward, at *right angles* to the direction of the magnetic field. If a larger tube is used, such as the one shown in Fig. 5, together with a magnetic field which is uniform over the whole tube, the beam may be bent into a complete circle as shown. (Here the anode is a cylinder, A, with a short slit in one side, and the cathode, C, is a hot tungsten wire. The advantage of using a hot cathode will be explained in Sec. 11. The magnetic field is perpendicular to the plane of the figure and directed inward, or away from the reader. Enough gas is left in the tube to reveal the path of the cathode-ray beam by the luminescence of the residual gas.)

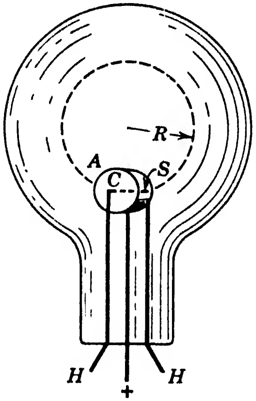


FIG. 5. MAGNETIC DEFLECTION OF CATHODE RAYS:[3] A—Anode cylinder, with slit S through which electrons emerge. C—Cathode—incandescent wire along axis of cylinder; heated by current from battery connected to H-H.

The deflection is in exactly the direction in which a conductor would be pushed when a stream of negative electric charges is flowing through it; and a free stream of tiny fast-moving particles, each carrying a negative electric charge, would be bent by a magnetic field into just such a circular path.

8. The Electron. Sir William Crookes in 1879 advanced the hypothesis that cathode rays are streams of negatively charged particles of

---

[3] This form of apparatus was devised by K. T. Bainbridge for use in the student laboratory. It is more fully described in *American Physics Teacher*, Vol. 6, pp. 35–36, February, 1938.

matter traveling at very high speeds. Then, in 1897, Sir J. J. Thomson showed that this hypothesis is correct, and that all these cathode-ray particles are identical in kind, no matter what the metal of the cathode, or what the gas in the tube. He named these particles *electrons*. To show that they were all alike he "weighed" them by means of their deflections by electric and by magnetic fields.

9. Magnetic Force on a Moving Charged Particle. Before Thomson's method of weighing electrons may be explained, we must first obtain the equation for the force exerted upon a moving electric charge by a magnetic field. As is well known, if a current-bearing conductor is placed at right angles to a magnetic field, each unit length of it is acted upon by a force, $F = BI_m$, which is perpendicular both to the field and to the conductor. See Fig. 6(*a*). B is the flux-density of the magnetic field,[4] and I_m is the current strength in *electromagnetic units*. Or, if electric charge and current are measured in *electrostatic units* (as will be the practice throughout this book; see Appendix D), the force equation becomes

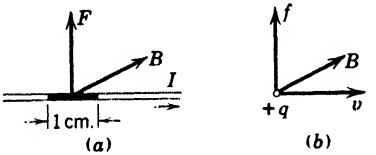


FIG. 6. (*a*) Force on a current-bearing conductor in a magnetic field. (*b*) Force on a moving charge in a magnetic field.

$$F = \frac{BI}{c} \tag{1·5}$$

where I is the current in stat-amperes, and c is the ratio between the two systems of units.

If now we think of the current as being a stream of electrically charged particles, each having the same charge, q, and speed, v, the current is then $I = nqv$, where n is the number of such charged particles per unit length of the conductor, and the force equation then becomes

$$F = \frac{Bnqv}{c} \tag{1·6}$$

It follows that the force on each charged particle is

$$f = \frac{F}{n} = \frac{Bqv}{c} \tag{1·7}$$

---

[4] Sometimes the symbol H, representing magnetic field intensity, will be found in these equations in place of B. This will give *numerically* correct results when the magnetic field is in air or in a vacuum, since the numerical values of B and H are equal in these media; in other media B must be used.

10. Mass of the Electron. The mass of the electron may be measured by apparatus such as is shown in Fig. 5. The principles used are much the same as those employed by Thomson's method, although the procedure is somewhat simpler. First we shall assume that the charge on the electron is the elementary negative charge, $e = 4.80 \times 10^{-10}$ statcoulomb. The validity of this assumption is usually assumed without question. Indeed, as has been stated earlier, Millikan's measurement of the elementary charge is usually referred to as measurement of the charge of the electron.

It is certain that the electron charge cannot be smaller than e, or the oil-drop apparatus would sometimes measure this smaller value. Likewise, if the electron charge were two or more times e, we should expect the charge on the oil drop to change frequently by this multiple value, and this is not observed. It is nevertheless an assumption, and its justification comes from the consistency of all the many results which follow from its use.

Since the electrons travel through the magnetic field, in a circle of radius R, the magnetic force must produce an acceleration of v^2/R, or

$$f = \frac{Bev}{c} = \frac{mv^2}{R} \tag{1·8}$$

Since both the mass, m, and the speed, v, are unknown, this equation is insufficient to give us the value of either quantity.

A second equation involving these two quantities may, however, be obtained from the fact that the electrons acquire their speed as a result of the action upon them of the potential difference, U, between the cathode, C, and the anode, A. Hence the kinetic energy of each electron equals the work, Ue, done upon it by this potential difference, i.e.,

$$\tfrac{1}{2}mv^2 = Ue \tag{1·9}$$

Solving these equations simultaneously, one obtains

$$v = \frac{2U}{BR}c \tag{1·10}$$

and

$$m = \frac{B^2R^2e}{2Uc^2} \tag{1·11}$$

A large number of different methods have been used to determine m (or e/m; see below), all of them now being in substantial agreement with each other. From all these measurements the best value for the mass of the electron may be taken to be

$$m = 9.11 \times 10^{-28} \text{ gram}$$

As a matter of history, Thomson performed his experiments before any experiments had yet been devised with which to measure the elementary charge; all that he then could compute was the specific charge, or the ratio of the electron charge to its mass. In terms of the best values obtainable today, this is

$$\frac{e}{m} = 5.2737 \times 10^{17} \text{ stat-coulombs per gram}$$

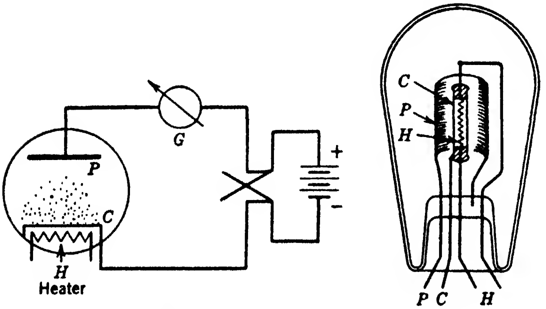
with an uncertainty indicated by a probable error of $\pm 0.0015 \times 10^{17}$ stat-coulombs. See Appendix II.

The enormous size of this figure emphasizes the insignificance of the mass of the electron as compared to its charge. Indeed, if the electron is considered to be a little sphere about 4×10^{-13} cm in diameter, electromagnetic theory shows that *all* its mass might be attributed to its electrical field! And it cannot be smaller than this, since its electrical field would then account for *too much* mass.

It is interesting to note that, although masses of bodies of more familiar size are most easily compared by comparing their weights on a balance, we do not, except figuratively speaking, "weigh" a particle of atomic dimensions. Instead, we determine its mass by direct application of Newton's second law—that is, by applying a known force to it and observing the *change of motion* which results. It must also be noted that, when electrons have been set into motion by more than a few hundred volts potential difference, the simple expression, $\frac{1}{2}mv^2$, does not give an exact value for its kinetic energy, and another expression must be used. This is explained in Sec. 212.

11. Hot-Cathode Electron Tubes—Thermionic Emission. In the tubes described above, in which the cathode is cold (at room temperature), some residual gas is required to maintain the discharge. Edison discovered that, on the other hand, if the cathode is heated to a fairly high temperature (bright incandescence for pure tungsten), electrons will flow from this cathode to a cold anode for even the smallest potential differences. Even when all the air has been removed from the tube, this will take place. (Although this is not literally possible, the pressure can be made so low with modern pumps that the residual gas is of negligible effect.) Figure 7(*a*) illustrates schematically a tube in which this *thermionic* emission of electrons may be studied. The cathode, *C*, is shown as a metal box which is heated electrically by a heater coil, *H*, placed inside it. In actual construction both electrodes may be made cylindrical, with the cathode placed on the inside, as shown in Fig. 7(*b*), where the parts are correspondingly labeled. An even simpler arrangement is to have the heater coil itself act as cathode in place of the box. Both types of cathodes are used in commercial tubes.

The high temperature of the cathode *evaporates* electrons from the metal in much the same manner as water vapor is evaporated from the surface of water in an open dish, and an electron cloud forms in similar fashion above the surface of the cathode. Since, however, the electrons are negatively charged particles which repel each other, but are attracted to the metal surface of the cathode by the positive charges which they induce in that surface, this electron cloud is much denser close to the cathode surface than it is anywhere else. Some electrons are returning to the surface while others are escaping; and, when *P* is insulated or is negatively charged, just as many are returning as are escaping, on the average. If, however, *P* is connected to the positive terminal of a battery as shown in Fig. 7(*a*), the cathode *C* being connected to the negative terminal, electrons are attracted to *P* by the positive charge upon it, and an *electron* stream flows through the tube from *C* to *P* and around through the rest of the circuit, including the meter, *G*. (This is equivalent to a conventional positive-charge current from *P* to *C*.) No current will exist if the battery is reversed, since no electrons are emitted by *P*.



(*a*) (*b*)

FIG. 7. HOT-CATHODE ELECTRON TUBE. (*a*) Schematic diagram. (*b*) Commercial tube with the front half of the plate, *P*, cut away. *C*—cathode. *P*—anode. *H*—heater coil.

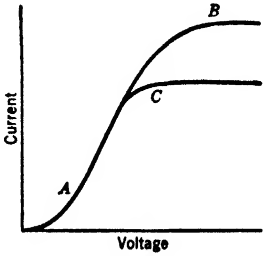


FIG. 8.

The manner in which this current depends upon the potential of *P* is shown in Fig. 8. Low voltages (point *A* on this curve) affect only the outermost electrons of the electron cloud, and the current is small. Higher voltages reach deeper into the electron cloud, and eventually a voltage will be reached (point *B*) which will draw all the electrons over to *P* as rapidly as they are emitted from *C*. Further increase in voltage produces very little increase in the current, which then is said to be saturated. (There would be no increase in current at all, were it not for a small effect of voltage upon the evaporation rate.) At a lower cathode temperature the saturation current is lower in value, and saturation is reached at a lower voltage, as shown by the point *C* on the lower curve.

This *thermionic* emission depends also upon the nature of the surface of the cathode. Tungsten was first used, primarily because it could be heated to the high temperatures required without disintegration. Then it was discovered that tungsten in which is dissolved (while still molten) a small amount of thorium oxide will emit electrons copiously at much lower temperatures. Later surfaces coated with barium oxide were found to be even better electron emitters. By proper preparation these barium oxide surfaces may be made to give copius electron emission at temperatures under 1000° C (dull red heat). In thoriated tungsten filaments, the emission has been shown to be due to a monomolecular layer of thorium atoms covering the surface of the tungsten filament; in the oxide coatings, a similar layer of barium atoms seems to be the primary source of electrons. Oxide cathodes are common in radio tubes, but they are less rugged than the pure metal cathodes, which are consequently still used for high-voltage tubes such as X-ray tubes. The cathode is always heated electrically, by a separate low-voltage battery or transformer. It may be in the form of a filament through which the heating current flows directly, or it may be a metal box or cylinder with heating coils placed inside, as shown in Fig. 7. The cathode-heating circuits have been omitted from most of the circuit diagrams, to simplify the circuits. Their positions are obvious.

The discovery of the electron and the invention of the electron tube, especially the hot-cathode electron tube, mark the beginnings of the science of electronics. The developments in this field have been amazing and far reaching, with applications to be found throughout all fields of science and engineering; many of these applications will be mentioned in later chapters of this book. Four basic principles underlie most of these applications, and these principles are illustrated by four special types of tubes—the cathode-ray oscillograph tube, the rectifier tube, the amplifier tube, and the photocell. The first three of these will be described here, the fourth in Chapter VII.

12. Rectifier Circuits. If the switch shown in Fig. 7(*a*) is reversed, so as to make the hot terminal the positive one, no current will flow, even though the voltage is raised to a very high value. And if an alternating voltage is applied to the terminals of the tube, current will flow only during the half-cycle when the hot terminal is the cathode, and will flow always in the same direction. This is shown by the curve of Fig. 9(*b*). The current reaches a maximum value for the peak of the half-cycle which makes *P* positive, and is zero for the other half-cycle.

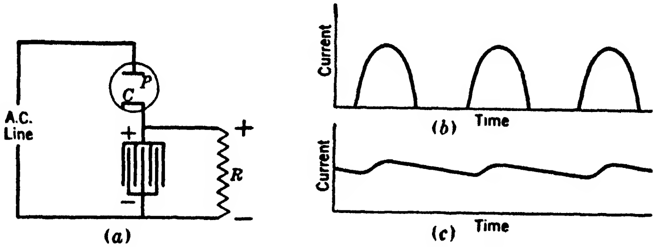


FIG. 9. SIMPLE RECTIFIER CIRCUIT. (*a*) Circuit. (*b*) Oscillograph record of current flowing when condenser is removed. (*c*) Oscillograph record of current with condenser present.

If a more nearly steady current is desired, the circuit shown in Fig. 9(*a*) may be used. When the capacitance of the condenser is large enough, the current through the load resistance, *R*, will then be represented by a curve such as shown by Fig. 9(*c*). This is explained as follows: When the upper terminal of the a-c. line is positive, the tube is conducting—current flows through *R*, and also into the condenser to charge it up as indicated. That is, electrons flow out of the upper plate of the condenser, through the tube and the a-c. line, and into the lower plate. During the reverse half-cycle the tube does not conduct; no current flows through the a-c. line, and the condenser can discharge only through the load *R*. This discharge maintains the current through *R* during the reverse half-cycle. If the capacitance of the condenser is large enough in comparison with the load resistance, the flow of charge from the condenser during this half-cycle will lower only slightly the voltage across it, and the current through *R* will be quite steady.

If two tubes are used (or a double tube) with a special transformer such as is shown in Fig. 10, the condenser is recharged every half-cycle, and the currents drawn from it are correspondingly steadier. It is charged during one half-cycle through one tube, and through the other tube during the other half-cycle. This figure shows also a "filter" circuit (two condensers separated by a choke coil) in place of the single condenser of Fig. 9. This still further steadies the current. The transformer charges the first condenser through the rectifier tubes, as explained above. The first condenser keeps the second condenser charged, and the charge which flows out from the second condenser maintains a steady current through the load connected to the d-c. terminals. The choke coil is a coil of many turns of wire wound on an iron core, so as to have many lines of magnetic induction running through this core when current flows in the coil. Since any change in the current through such a coil is opposed by a comparatively large induced e.m.f., this coil steadies the flow of charge from the first condenser into the second, keeping it nearly constant in spite of the pulsations of voltage at the first condenser caused by the rectifier current.

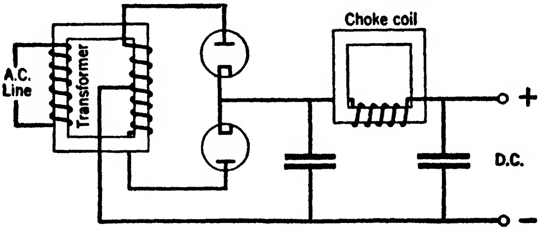


FIG. 10. STEADY D-C. POWER FROM AN A-C. LINE. Both half-cycles are rectified.

Small rectifier circuits of this kind provide d-c. power to a-c.-operated radio receivers, and much larger ones serve radio transmitters in like manner. Well-made rectifier circuits will produce potential differences and currents which are as steady as those obtainable from storage batteries and find many important uses in scientific laboratories. Many other applications exist.

Another type of rectifier tube is described in Sec. 139.

13. Cathode-Ray Oscillograph Tubes. The device illustrated in Fig. 11 is an adaptation of the tube shown in Fig. 4. The electrode *Q* is an "electron gun" containing a small hot surface which emits electrons because of its temperature, and a high potential difference (several thousand volts) between *Q* and the anode *P* "shoots" these electrons through the hole in *P* and towards the fluorescent screen *F*. Auxiliary electrodes, not shown, control the intensity of this electron beam and focus it to a fine spot upon the screen *F*. If now a small potential difference is applied to the plates *A* and *B*, the electric field produced between these plates by means of that potential difference will deflect the beam up or down, depending upon the polarity of the terminals. The point of fluorescent light will correspondingly move up or down on the screen. Likewise, an a-c. voltage applied to these terminals will cause the luminous point to move up and down repeatedly; its motion will be so rapid that, owing to persistence of vision, its path will appear as a stationary line.

Voltages applied to the plates *C* and *D* will in similar manner deflect the luminous spot in a horizontal direction, and, if a-c. voltages are applied to both sets of deflecting plates simultaneously, the motion of the luminous spot will be the geometric resultant (vector sum) of the two separate motions. Thus, if the two voltages have the same frequency and are *in phase* (reach their maximum values simultaneously) the spot will trace a diagonal line, as is shown in Fig. 12(*a*); if they are out of phase (e.g., if the maximum of the vertical deflection occurs after the horizontal deflection has passed its maximum value), the path traced will be an ellipse, as illustrated by Fig. 12(*b*). If the two voltages are a quarter of a cycle out of phase and equal in amplitude, this ellipse becomes a circle.

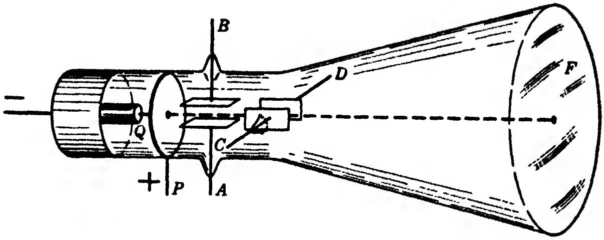


FIG. 11. CATHODE-RAY OSCILLOGRAPH TUBE.

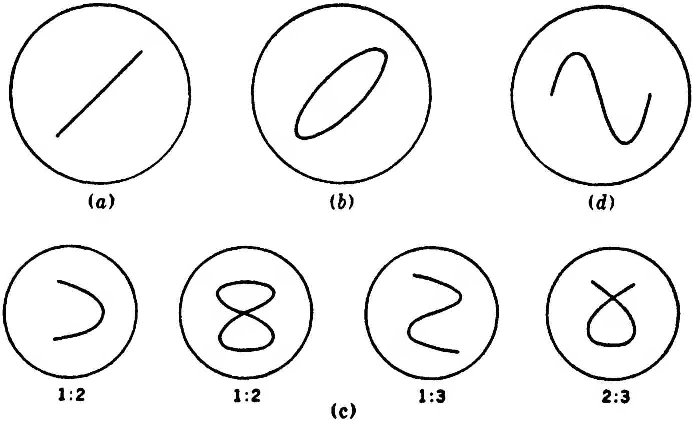


FIG. 12.

When the two voltages differ in frequency, the patterns traced are more complicated. A few of the patterns traced when the frequencies are commensurable are shown in Fig. 12(*c*); the ratio given under each figure is the ratio of the frequency of the vertical deflection to that of the horizontal one. Similar patterns may be traced by combining the motions of two pendulums of different length. They are named Lissajous figures after the man who first studied them.

In most commercial cathode-ray tubes 50 volts or more are required to deflect the luminous spot to the edge of the screen. If smaller voltages are to be studied, as for example those produced by sound waves falling upon a microphone, an amplifier (see Sec. 15) may be used to obtain the requisite deflecting voltages.

Frequently it is desirable to obtain a voltage-time plot upon the oscilloscope screen—that is, a curve in which vertical deflections represent voltage and the horizontal deflection is proportional to time—and an ingenious circuit has been devised to supply this need. This "sweep" circuit, as it is called, deflects the spot horizontally across the screen at a steady rate until it reaches the end of its path, then jumps it back to the starting point again in so short a time that the motion appears always forward. This motion, which repeats itself time after time, may be made slow or fast, and furthermore it may be synchronized with the vertical motion by a modification of the circuit so that the pattern will appear stationary upon the screen. Figure 12(*d*) shows the pattern produced by an alternating voltage applied to *A* and *B*, with the sweep circuit connected to *C* and *D*.

The cathode-ray oscilloscope has become an essential instrument in all fields of science and industry. In the electrical laboratory it is invaluable to the study of a-c. phenomena from a few cycles per second up to high radiofrequencies. When combined with auxiliary equipment, such as microphones, photocells, and amplifiers, it has become almost as important in nearly all physical, chemical, engineering, biological, and medical laboratories. For example, the physiological laboratory finds it invaluable to the study of muscle and nerve actions, such as heartbeats, which produce small a-c. voltages that may be amplified and traced out on the oscilloscope screen. It serves television in both the receiver and the transmitter, and most recently it has become an essential part of all radar equipment. (See Secs. 74 and 80.)

14. Amplifier Tubes. If a third electrode, in the form of a grid or net of fine wires, is placed between the cathode and the anode of a tube such as is illustrated in Fig. 7, this grid will act somewhat like a valve to control the flow of electrons between the cathode and the anode. This arrangement is indicated in Fig. 13, which shows also a cross-section through a commercial radio tube having such a grid, *G*. When this grid is at the potential of the cathode, which we may conveniently call zero potential, it more or less screens the electron cloud from the electric field of the anode, *P* (commonly called the "plate"). Only a small part of the field due to the potential of the anode reaches through the spaces between the grid wires to the cathode, and the current to *P* is quite small unless the voltage on *P* is quite large as compared to that which would have been required to produce the same current with the grid absent. A negative potential on the grid reduces still further the electric field in the neighborhood of the cathode, thus making the electron current still smaller.

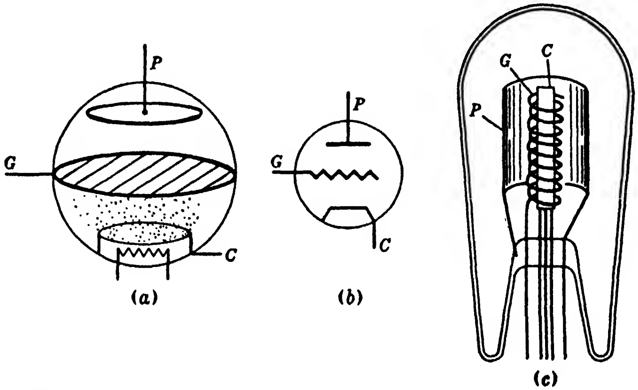


FIG. 13. AMPLIFIER TUBE. (*a*) Schematic diagram. (*b*) Conventional symbol. (*c*) Commercial tube—front half of the plate, *P*, cut away.

Customarily this tube is operated as shown in Fig. 14, with a small negative potential, or "bias" (produced by the battery, *C*), on the grid, and the potential of the anode high enough to cause appreciable current to flow to the anode. A small positive voltage between *a* and *b*, Fig. 14, then will cause the anode current to increase proportionally, whereas a negative voltage will cause it to decrease. Furthermore, the changes of current will be as large as would be produced by much larger changes of anode voltage. For example, a 1-volt change in grid potential, for a type 6C5 tube, will produce as much change in current as would a 20-volt change in anode potential.

Such a tube is an *amplifier* tube, and the ratio of the anode and grid potential changes which produce the same change in anode current (when each change of voltage acts alone) is called its *amplification factor*. In the example cited, the amplification factor equals 20. The value of the amplification factor is determined by the size and spacing of the grid wires and their distance from the anode; in commercial tubes it may vary from 2 or 3 for some tubes to as much as 100 or more for others.

As long as the grid is at a negative potential, it can collect no electrons, and the current to it is zero.[5] On the contrary, when the grid potential is positive, electrons will flow towards it, but unless the anode potential is low (or zero) only a few such electrons stop at the grid; most of them go through the spaces between the grid wires and on to the anode, thus causing an increase in the anode current.

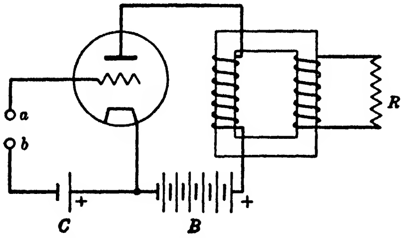


FIG. 14. AMPLIFIER CIRCUIT.

15. Amplifier Circuits. By means of circuits using amplifier tubes, very feeble alternating voltages may be enabled to produce large results, such as much amplified voltages or large alternating currents. A simple amplifier circuit is shown in Fig. 14. The small alternating voltage is applied to the grid of the tube by connecting the source of this voltage to the terminals, *a–b*, and the resulting variations of the plate current cause an alternating current to flow in the load *R*, which may be a resistance as indicated, or may be some device to be operated by the alternating current, such as a loud speaker. This is explained as follows:

The alternating voltage in the grid circuit causes corresponding variations in the anode current, as has been explained in Sec. 14, these variations being of enhanced amplitude because of the amplifying properties of the tube. The plate current thus has a pulsating value which is equivalent to an alternating current superimposed upon the original steady value, the alternating component being proportional to the alternating voltage in the grid circuit. The transformer serves to separate these alternating and steady components of the anode current: only the alternating components will induce an electromotive force in the secondary windings, and there will be only an alternating current in the load *R*. To insure that no current flows in the grid circuit, a negative "bias" voltage is supplied to it by means of the battery *C*; this bias voltage must be as great as the maximum voltage to be amplified. As long as the grid potential remains negative, an almost negligible amount of power drawn from the original a-c. source may cause a very considerable amount of power to be developed in the load *R*.

---

[5] Owing to a number of secondary causes, such as residual gas in the tube, an exceedingly small current may flow, but usually this may be ignored.

This extra power comes, of course, from the battery *B* which supplies the anode current; the grid circuit acts, as has been stated earlier, somewhat like a valve alternately to increase and decrease the flow of power from this battery.

Several amplifier circuits of the type shown in Fig. 14 may be lined up so as to produce a "cascade" amplifier which will bring about enormous gains in voltage and power; the gain of the first is further amplified by the second, then by the third, and so on down the row. The details of these circuits may be found in the references given at the end of the chapter. Public address systems employ such cascade amplifiers to build up voltage and power sufficient to operate large loud speakers, starting from the feeble output of a microphone, and similar amplifier systems are employed in radio sets, electric phonographs, talking-picture apparatus, long-distance telephone lines, and many other applications.

Modifications of the circuit shown in Fig. 14 may be utilized to amplify very small steady currents. Although the amplifier tube is essentially a voltage-amplifying device, these very small currents may be amplified by sending them through resistances of very high value and then amplifying the voltage which they develop across those resistances. With tubes especially designed for this purpose, currents as low as 10^{-16} ampere are easily measured.

16. Special Electron Tubes. Many of the applications of electron tubes require special types. Thus the tiny amplifiers which are the heart of electronic "hearing-aid" devices for deaf people require very small tubes, some less than an inch high. Even smaller, unbelievably rugged tubes were developed for the "proximity fuze" circuits which enabled artillery shells to explode upon mere approach to their targets, thus increasing their effectiveness many times. Other types have their electrodes very close together, so that they can operate effectively at ultrahigh frequencies; the grid may be placed only one one-thousandth of an inch from the cathode, with the anode only a few times that distance farther away. Some of these tubes have picturesque names, such as "acorn," "door-knob," and "lighthouse" tubes, these names being derived from their unusual appearance. On the other hand, tubes have been made large enough to handle many kilowatts of power, as rectifiers, amplifiers, and oscillators.

Modifications are not limited to size alone. Quite early in the development of electronics it was found that, for many purposes, it was advantageous to add a second grid, between the first grid and the plate. Later a third grid was added to meet additional requirements, and thus a five-electrode tube, or *pentode,* was created. The pentode has proven a very versatile tube, with many applications. It is beyond the scope of this book to consider the functions of its several grids. It must suffice to say that the pentode, in its many applications, functions in essentially the same manner as the simple tube having a single grid, but with higher values of amplification. Tubes having as many as seven grids between anode and cathode are made for special purposes.

Uses for electron tubes have multiplied also, although almost every use may be analyzed in terms of the basic functions of rectification and amplification. Thus the way in which an electron tube is able to generate electrical *oscillations* (see Sec. 63) may be explained in terms of its *amplifying* properties. Other applications for electron tubes, including very special types such as magnetrons and klystrons, will be mentioned in later chapters.

PROBLEMS

1. The condenser plates of an apparatus such as is shown in Fig. 1 are 7.0 mm apart and charged to a potential difference of 600 volts. Compute the force in dynes on a water droplet 0.0030 mm in diameter and carrying 5 excess electrons, when placed between these condenser plates.

2. Will the droplet in problem 1 rise or fall if the top plate is the + plate? Compute the potential difference which will hold the drop in equilibrium between the plates.

3. Compute the speed at which the droplet of problem 1 will fall when the potential difference between the plates is zero. Use 0.000182 (c.g.s. units) as the coefficient of viscosity of the air.

4. Compute the force acting between a milligram of electrons at Washington and another milligram of electrons at Baltimore, 40 miles away.

5. How many excess electrons need be added to a mercury droplet 0.00220 mm in diameter, to give it a potential of 1 volt?

6. If 1 mg of electrons were added to the earth from some outside source, how much would it raise the potential of the earth, assuming it to be a sphere 8000 miles in diameter?

7. Compute the velocity of an electron which has been set into motion by a potential difference of 100 volts.

8. Compute the strength of magnetic field required to bend 300-volt electrons (electrons which have been set into motion by 300 volts potential difference) into a circle 80 mm in diameter.

9. An electron passes between two condenser plates which are 30.0 mm apart and have a potential difference of 3000 volts between them. The electron is not deflected, because of the presence in this same region of a magnetic field of 100 gauss. (*a*) Show by diagram the relative directions of the two fields. (*b*) Compute the speed of the electron.

10. How many electrons per microsecond flow from the cathode to the plate of a radio tube when the current is 5.0 ma (milliamperes)?

11. The current through a radio tube is 125 ma with a potential difference of 800 volts. Compute the power used by the tube. Explain what becomes of this power.

12. A half-wave rectifier circuit [Fig. 9(a)] is used to supply a d-c. current of 5.0 ma from a 60-cycle a-c. source. What capacitance is needed in the condenser, if the potential difference across the resistor, R, is not to vary more than 1 per cent from 250 volts? (*Suggestion:* The charge which flows through R, in each cycle, must be but 1 per cent of the average charge stored in the condenser.)

REFERENCES

For Additional Reading

R. A. Millikan, *Electrons, Protons, Photons, Neutrons, and Cosmic Rays*, Chapters I to V.
Physics Staff, University of Pittsburgh, *Atomic Physics*, Chapter III.
J. B. Hoag, *Electron and Nuclear Physics*, Chapters 1, 2, 5, and 6.
J. J. Thomson, *Conduction of Electricity Through Gases*, Vol. II, Chapter I.
Max Born, *The Restless Universe*, parts of Chapter II (popular).
G. F. Hull, *Modern Physics*, parts of Chapter 2.
J. D. Stranathan, *The "Particles" of Modern Physics*, Chapters 2 and 4.
R. G. Kloeffler, *Principles of Electronics*.

Much practical information concerning electron tubes, including circuits, is to be found in the booklets issued by the manufacturers of radio tubes.

For Advanced Reading

Harnwell and Livingood, *Experimental Atomic Physics*, Chapters III and IV.
L. R. Koller, *Physics of Electron Tubes*.
D. G. Fink, *Engineering Electronics*.
Keith Henney, *Electron Tubes in Industry*.
W. G. Dow, *Fundamentals of Engineering Electronics*.
H. J. Reich, *Principles of Electron Tubes*.
L. B. Arguimbau, *Vacuum Tube Circuits*, Chapters II, III, and VI.

CHAPTER II

DIMENSIONS OF ATOMS

17. Chemical Atomic Weights. As soon as chemists began to make accurate chemical analyses of compounds, they discovered that the elements always combine in the same proportions to produce the same compounds; when they had analyzed a large enough number of compounds, they were able to make up tables of combining weights for the elements which can be used for many compounds. The exactness of this evidence suggested very strongly that, if the substance of an element could be magnified sufficiently, one would find it made up of separate "indivisible" particles, or *atoms*, all alike in every particular, including mass. In compounds, integral numbers of these atoms would be found grouped together to form *molecules*.

Thus the atomic hypothesis of chemistry was born, and the table of combining weights, when refined from the point of view of this hypothesis, became a table of atomic weights. The first tables of atomic weights were worked out on the more or less obvious basis of making that of the lightest element, hydrogen, equal to unity. Later on it was observed that, if these atomic weights were all increased so as to make that for oxygen equal to 16, many of the other atomic weights were then either whole numbers or just a little larger, and this revision was made; although this was done largely as a matter of convenience, it seemed possible even then that there might be some significance in so many integral values.

It is to be emphasized that chemical atomic weights are purely *proportional* values and give no information as to the absolute mass (mass in grams) of the atoms. It will now be shown that the determination of the elementary electric charge, together with Faraday's laws for electrolysis, gives the additional information which makes possible the determination of atomic masses in absolute value.

18. Charges Carried by Ions. When an electric current is passed through a solution of a chemical salt such as NaCl, the component elements of the salt are liberated at the electrodes. The phenomenon is called electrolysis. The laws of electrolysis, which were discovered by the great experimental scientist, Michael Faraday (both the chemists and the physicists claim him as their own), may be summarized as follows:

> **The same quantity, F, of electric charge will liberate a mass of A grams of any univalent chemical element, and nF units of electric charge are required to liberate A grams of any element whose valence is n.**

A = atomic weight. (A grams is called a gram atom.)
F = 96,488 coulombs = 9649 ab-coulombs = 2.8925×10^{14} stat-coulombs. It is called the **faraday.**

For example, copper has an atomic weight of 63.6 and, in a solution of copper sulfate in water, a valence of 2. Then 63.6 grams of copper will be deposited out of a copper sulfate solution by passage through the solution of $2 \times 2.89 \times 10^{14}$ stat-coulombs of electric charge.

The most rational explanation for these phenomena of electrolysis assumes that the component parts of the salt molecule possess independent existence in the solution. They differ from free atoms of the same elements by virtue of possessing electric charges, and are called *ions*. In NaCl, the example already cited, the Na ion carries a positive charge, the Cl ion, a negative charge. Because they possess these charges of opposite sign, they remain uniformly distributed throughout the solution, equal numbers of both kinds of ions being everywhere present. For this reason also the charged electrodes cause them to drift through the solution and finally to be liberated. An ion may also consist of a group of atoms, such as the negatively charged SO_4 group in $CuSO_4$.

Faraday's laws indicate that all univalent ions carry the same charge, a divalent ion carries twice as much charge, and in general an ion whose valence equals n carries n times the charge carried by a univalent ion. If the charge carried by the univalent ion is the elementary charge e, as it seems most natural to conclude, then an ion whose valence is n carries a charge equal to ne.[1]

19. Atomic Mass in Absolute Measure. If this be so, then the average mass, m, of any atom may be computed by a simple proportion: since A grams of an element in the form of ions having a valence n carry an

---

[1] This explanation for electrolysis was at first an hypothesis, just as was the assumption that the electron carries the elementary charge, e. See Sec. 10. The validity of this hypothesis has been established by the consistency of all the other phenomena which may be described by it, and by agreement with experiment for all conclusions deduced from it.

electric charge of nF units through an electrolytic solution, and each ion carries ne units of electric charge,

$$nF:ne = A:m$$

or

$$m = \frac{e}{F}A = \frac{4.802 \times 10^{-10}}{2.892 \times 10^{14}}A = 1.660 \times 10^{-24}\, A \text{ gram}^2 \quad (2\cdot1)$$

For example, for copper, $m = 1.660 \times 10^{-24} \times 63.6 = 105.6 \times 10^{-24}$ gram.

It may now be seen how light the electron is in comparison with atoms. Since the mass of the electron is but 9.11×10^{-28} gram, its "atomic weight" is 0.000549.

20. Avogadro's Number. The number of atoms in a gram atom is called Avogadro's number. It will be represented by N_0. Once the absolute masses of atoms are known, this constant may be determined by dividing the mass of the gram atom of any element by the mass in grams of its atom. Or it may be obtained directly from the electrical data, as follows:

Since one gram atom of an element in the form of ions having a valence n carries an electric charge of nF units through an electrolytic solution, and since it consists of N_0 atoms, each of which carries ne units of electric charge,

$$nF = N_0 ne$$

or

$$N_0 = \frac{F}{e} = \frac{2.8925 \times 10^{14}}{4.802 \times 10^{-10}}$$

$$= 6.023 \times 10^{23} \text{ atoms per gram atom}^{2}$$

An alternative method for computation of the average mass of an atom is then

$$m = \frac{A}{N_0} \quad (2\cdot2)$$

and this is perhaps the easier one to remember.

Although this electrical evidence gives the first accurate value for N_0, it should be mentioned that some approximate values were obtained earlier by means of the kinetic theory and particularly by the Brownian movements (this evidence will be discussed in Chapter XI), and that these earlier values are in substantial agreement with this later one. This agreement is important, since it verifies the assumption

---

[2] More precise values for all these constants are found in Appendix II.

which has been made, that a univalent ion carries a single elementary charge.

21. The Mass Spectrograph. The absolute masses of atoms may also be measured directly, by the same methods used to measure the mass of the electron. The apparatus which does this is called a *mass spectrograph*. A much stronger magnetic field is required, of course, because even the lightest atom is nearly two thousand times more massive than the electron, but this field may be produced by a strong electromagnet. The most troublesome difficulty is obtaining a suitable beam of electrically charged atoms, or *ions*. Different methods are required for different kinds of atoms, and several different methods are now possible, depending upon the kind of atom and the type of mass spectrograph used. Details of these are given in the references listed at the end of the chapter. If the atoms are easily obtained in gas or vapor form, the method described below may be employed.

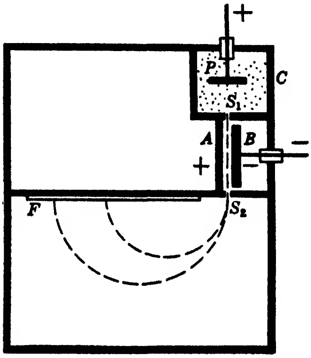


Fig. 15. Mass Spectrograph with a Bainbridge Velocity Filter.

Figure 15 shows schematically a mass spectrograph which measures the masses of gas ions in much the same manner that the apparatus illustrated in Fig. 5 measures the mass of an electron. The heavy outline represents a strong metal box which is thin enough in the direction perpendicular to the figure to go between the poles of a powerful electromagnet. The gas or vapor of the element whose atomic mass is to be determined is introduced into the smaller chamber, C, at a suitable low pressure. The rest of the apparatus is kept highly evacuated by powerful pumps. When an electrical discharge is set up within C by a high potential difference between the lower wall of C and the electrode, P, the positively charged ions which are produced by this discharge will be accelerated downward by the electric field, and some of them will pass through the fine slit, S_1, into the evacuated space below.

These ions will emerge from S_1 with many different speeds, and the next part of the apparatus is a velocity selector, a device to sort out from this heterogeneous beam of ions those having the same speeds. As an ion (carrying a charge q) passes between the condenser plates, A and B, it is pushed towards the right by the force Eq exerted upon it by the electric field E between A and B; at the same time it is urged to the left by the magnetic force Bqv/c, due to the magnetic field. If either of these forces is the greater, the ion will be deflected so as miss the second slit, S_2. But if they balance, that is, if

$$Eq = \frac{Bqv}{c} \tag{2·3}$$

the path of the ion between the condenser plates will be a straight line through S_2 into the analyzing chamber, where it will be bent into a semicircle by the magnetic field as shown. From (2·3) it follows that, in this case,

$$v = \frac{E}{B}c \tag{2·4}$$

This means that all ions which pass through this selector have the same speed, v, regardless of their mass or charge.

Their path in the analyzing chamber will be determined by equation (1·8), which is

$$\frac{mv^2}{R} = \frac{Bqv}{c} \tag{1·8}$$

whence

$$m = \frac{Bq}{cv}R \tag{2·5}$$

Since B, c, and v are now constants of the apparatus, all ions having the same mass and charge will follow the same semicircular path. *For ions having the same charge, the radius of path,* R, *will be proportional to the mass,* m. The radius of path is determined experimentally by allowing the beams of ions to fall upon a photographic plate, *F*, where they produce an effect which appears as a blackened line when the plate is developed in the usual manner. The use of paths which are just semicircles has in this case a focusing action that makes these lines sharp images of the slit S_2. More commonly some electrical means is employed to locate the focus spot, rather than the photographic one described here.

If several different kinds of ions happen to be present in the ion beam, then a *mass spectrum* will appear on the photographic plate, a different line appearing for each kind of ion. Since ions of the same mass may carry different numbers of elementary charges, several lines may appear for the same atomic mass. Thus, a doubly charged atom is deflected as if it were a singly charged atom of half the mass. In addition, molecular ions may appear in the beam, producing still other lines on the plate. These complications are easily interpreted by the experienced observer, who indeed finds in them additional checks on the measurements made for the singly charged ions.

Although the mass spectrograph may be used to determine the absolute masses of atoms, it is most commonly employed to obtain their relative masses—that is, their atomic weights—by direct comparison of the positions of the lines produced on the photographic plate by the ion under measurement and by a known ion, for example, oxygen. Modern mass spectrographs are capable of determining atomic weights in this manner with greater precision than most chemical determinations, sometimes to better than one one-thousandth of one per cent. The mass spectrograph has now become a very important laboratory tool, even in the chemical laboratory, where it makes many chemical analyses more quickly and more precisely than is possible by straight chemical means.

A very considerable number of other types of mass spectrographs have been devised; they are described in the references given at the end of the chapter. The first one was invented in 1912 by the same J. J. Thomson who first measured the mass of the electron, and the first precise one was built ten years later by Aston.

22. Isotopes. Some atomic weights as determined by the mass spectrograph agree quite well with those obtained by chemical analysis, but others do not. A notable example is neon. The chemical value for the atomic weight of neon is 20.2, but the mass spectrograph shows no line corresponding to this value. Instead, it shows *two* lines, one at 20.0 and the other at 22.0, the first line being about nine times the stronger of the two. Both these lines are due to neon ions, and the average weight, taking into account the greater abundance of the ion of weight 20 (as indicated by the strength of the lines), is the chemical value, 20.2. These two neon atoms are identical chemically, differing from each other only in mass.

The name **isotope** has been invented to designate one from another of several atoms which differ in mass but are identical in other respects. The mass spectrograph has shown that all the elements possess two or more isotopes. For some elements this is easily discoverable, as it is for neon, while for many others, for example oxygen, one isotope is so very much more abundant than all the rest that only the most sensitive mass spectrographs will detect the others. Because oxygen does possess isotopes 17 and 18 as well as the abundant 16, atomic weights determined by the mass spectrograph are very slightly larger than chemical atomic weights: Mass-spectrograph values are based upon exactly 16 for the most abundant oxygen isotope, while chemical atomic weights assign 16 to the average. The ratio is 1.000272 to 1.

The best-known isotopes are probably the two principal isotopes of hydrogen, whose atomic weights (on the "physical" scale) are 1.0081 and 2.0147, respectively. The heavy isotope (now called deuterium) is very rare but differs so much in mass from the other that it can be separated out (concentrated) by a number of means and is now available commercially for experimental work. It is most commonly obtainable as deuterium oxide, or "heavy" water. There has also been found still a third and very much rarer isotope of hydrogen, whose atomic weight is 3.0171. It is called tritium.

23. Atomic Diameters—Solids and Liquids. We shall now consider such evidence as may tell us something concerning the space occupied by each atom. Before anything definite was known concerning the shapes of atoms, it was customary to assume for them the easiest imaginable shape, namely, that of a sphere, and hence to talk about atom diameters. It is interesting to note here that later evidence indicates that in general this assumption is a fairly accurate one. A rough estimate of the diameters of atoms may be made by assuming that they are touching each other in the solid or liquid state. Consider, for example, mercury, with an atomic weight of 200.5 and a density of 13.6. One gram atom of mercury will weigh 200.5 grams and occupy a volume equal to $200.5/13.6 = 14.7$ cm^3. Since this volume contains 6.02×10^{23} atoms, each atom "occupies" $14.7/(6.02 \times 10^{23}) = 24.4 \times 10^{-24}$ cm^3. That is to say, if the atoms were regularly arranged in parallel rows in three perpendicular directions, each one would be surrounded by a little cube of this volume. The length of the side of this cube is obtained by taking the cube root of this volume; it equals 2.9×10^{-8} cm. This is the diameter of the atom, considered to be a sphere which will just fit into this cube. Spheres can actually be packed more advantageously than this (see Sec. 186), so that they could be somewhat larger, say 3×10^{-8} cm as a rough estimate, and still occupy the same volume. Other liquids and solids considered in this same manner give values of the same *order of magnitude* as this.

To avoid the necessity of always writing the factor 10^{-8}, it is convenient to use the unit of length which was originally invented for the measurement of wavelengths of light, but which proves ideally suited for atomic dimensions. This is the angstrom unit (named after its inventor); it equals 10^{-8} cm, and is represented by the symbol A.

$$\mathbf{1A = 10^{-8}\ cm}$$

For atoms in crystals, where the arrangement is known from X-ray observations, the diameters (assuming that they just touch) can be figured with considerable exactness. See Sec. 186. The diameter of one of the largest atoms, potassium, is found in this manner to be 4.61A.

24. Atomic Diameters—Gases. The frequency of collisions between the molecules of a gas gives an estimate of their size. Suppose that you are shooting a gun blindfolded in a forest. Your chance of hitting a nearby tree will be the greater, the larger the trees; and the smaller, the farther the trees are apart. If the trees are small and quite well separated your bullets may, on the average, shoot past many trees before they strike one. In like manner one may estimate, from the average distance between atoms as compared to their diameter, their **mean free path**—that is, the average distance which one atom can travel before hitting another. See Sec. 167. Or working backwards, using mean-free-path values obtained by various experimental means, one may get an estimate for the diameter of the molecule. Since gas molecules are usually made up of two or more atoms, these results cannot be directly compared with those obtained by the method described in Sec. 23, but the *order of magnitude* is the same. Molecular diameters come out to be a small number of angstrom units.

25. Alpha-ray Tracks. Quite contrary evidence is obtained by shooting alpha rays through gases and solids. Alpha rays are shot out by radium and other similarly radioactive substances, and their origin and other matters concerning them will be discussed in Chapter XV. At present we are concerned with them only as very useful bullets to shoot at atoms. When alpha rays are deflected by electric and magnetic fields in the usual mass-spectrograph manner, it is found that the alpha-ray particles are doubly charged helium ions, and that they travel with speeds as high as 5 per cent of the speed of light. This is all we need to know concerning them at this time.

The paths which they follow through a gas can be revealed by means of a very ingenious apparatus called a Wilson cloud chamber. This apparatus, invented in 1911 by C. T. R. Wilson, is one of the most useful of the new tools of physics. Figure 16(*a*) shows a cross-section of this apparatus, and Fig. 16(*b*), a top view. The chamber itself is a shallow cylindrical box with glass top and sides, having a bottom in the form of a movable piston, *P*. It contains air (or some other gas) at approximately atmospheric pressure, plus a little water vapor—enough to saturate the space with water vapor. When the bottom is slowly pushed up to the level indicated by the dotted line, so as to compress the gas somewhat, and then quickly dropped down again to its former level, the sudden expansion cools the gas, and a fog forms in the chamber.

Fog droplets usually form about dust particles. The dustier the air, the thicker the fog. Indeed, if the gas in the chamber is very clean, and there is nothing else in the chamber, no fog will form, because it is very difficult to form a fog without having something such as dust particles upon which condensation may start. If, however, a bit of radium or some other source of alpha rays is placed in the chamber (this is indicated at S in the figure), the fog will form very readily upon the ions which are produced by the passage of the alpha rays through the gas.[3] The electric field across the chamber, which is produced by a potential difference of several hundred volts between the piston, P, and the metal ring, R, at the top of the chamber, now serves a most essential function. Without this field the chamber would fill up with ions, and the expansion would fill it with fog. But when the field is acting, the ions are driven to the electrodes very soon after they are formed; only the ions which have just been formed are present when the expansion occurs, and condensation upon them takes place

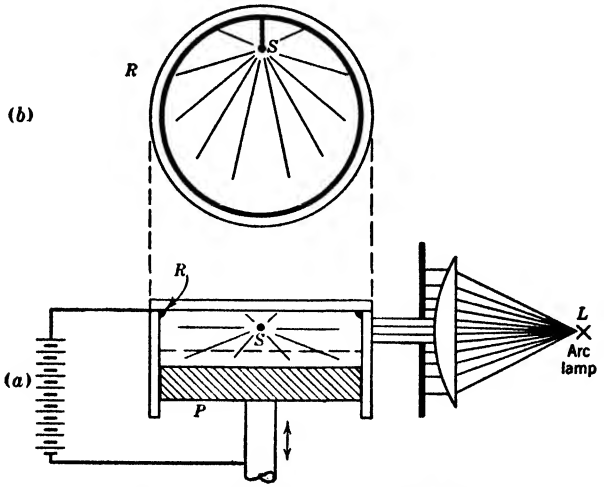


FIG. 16. WILSON CLOUD CHAMBER.

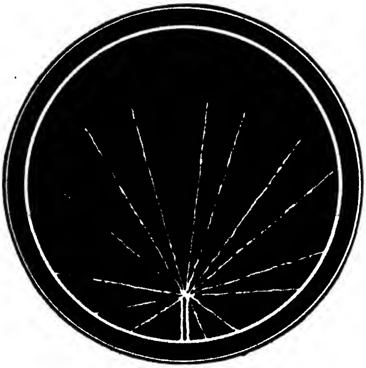


FIG. 17. ALPHA-RAY TRACKS.

---

[3] An explanation for this ionization is given at the end of Sec. 210.

while they are still in the positions which they occupied as the alpha ray passed by. The fog particles then mark out visibly the tracks of the alpha rays, as is shown in Fig. 17. They are best seen if they are illuminated from the side and viewed from the top, as shown in Fig. 16.

26. Nucleus of the Atom. The most significant thing about these alpha-ray tracks is their straightness. Almost without exception all of them are straight lines from start to finish, a distance of several centimeters. All this is entirely contrary to expectations. If the gas in the chamber is air, each atom is several times heavier than the alpha particle, so that collision of an alpha particle with a single atom in the gas should deflect it sharply from its original path. And if the atoms are several angstrom units in diameter, as the evidence presented in the previous sections seems to indicate, an alpha particle should meet up with one hundred thousand or so of them for each centimeter of path. Yet every alpha-ray track is a straight line! It must be that collisions *do not occur.*

Out of thousands of cloud-chamber photographs, one or two do indeed show branched tracks, evidence for collisions which occur very infrequently. However, when these collisions do occur, they take place exactly like collisions between perfectly elastic particles of ordinary size. The infrequency of occurrence shows that the colliding particles—both the alpha particle and the gas atom—are exceedingly tiny: thousands of times smaller than the previous estimates of the size of the atom. These two kinds of evidence concerning the size of the atom may be reconciled only by postulating that the *massive* part of the atom forms a very tiny *nucleus* at its center, and that the rest of the atom consists of nothing heavier than electrons. The alpha particle itself, according to this picture, is the nucleus of the helium atom, without any electrons. See Sec. 210.

27. Rutherford's Experiments. Collisions appear in cloud-chamber photographs too infrequently to give a very good estimate of the size of the nucleus. More frequent collisions occur when alpha rays are shot through solids; and Rutherford, by shooting a pencil-like beam of alpha rays through thin metal foils and making a statistical study of those which were scattered out of their original path by collisions with the atoms of the foils, estimated that nuclei are not bigger than about 10^{-12} cm (1/10,000A) in diameter. His experiments indicated further that the nuclei are positively charged, so that most "collisions" are really repulsions which take place between the positively charged nucleus and the positively charged alpha particle when they come very close to each other. They also indicated that the charge on the nucleus increases with the atomic weight of the atom, the number of elementary positive charges on a nucleus appearing to be roughly half the chemical atomic weight of the atom.

As will be seen in Chapters IX and X, spectra, and especially X-ray spectra, give exact evidence as to the numbers of elementary charges on atomic nuclei, which may be summarized as follows: If Z is the **atomic number** of any atom (the number which gives its position in the periodic table—see Appendix III), then its nucleus has a charge equal to $+Ze$. For example, zinc, which is element number 30 in the periodic table, has a nuclear charge of $+30\ e$.

28. "Outline" of an Atom. We may now visualize a normal (uncharged) atom as a very tiny nucleus (of the order of magnitude of 10^{-4}A diameter) having a positive charge of Ze, and surrounded by a cloud of Z electrons which in some fashion account for the rest of the volume of the atom (of the order of magnitude of 1A diameter), although they comprise a negligible part of the mass. For further information concerning this electron part of the atom we must turn to the experimental evidence given by the light which is emitted by it under suitable conditions. But before we may do this, we must first consider the fundamental facts concerning the nature of light itself which are included in the next few chapters.

PROBLEMS

1. How many coulombs of electric charge are required to deposit electrolytically 5.00 grams of a metal whose atomic weight is 65.4 and whose valence is 2?

2. Compute the number of atoms in 1 mm^3 of oxygen gas at standard pressure and temperature, if its density is 1.43 grams per liter.

3. Compute the number of atoms in 1 mg of lead (atomic weight = 207).

4. Compute the masses in grams of the hydrogen isotopes, the chlorine isotopes, and the principal isotope of oxygen.

5. (*a*) Compute the speed given to an hydrogen ion by a fall of potential of 1500 volts. Compute also the speed given to (*b*) a deuterium ion, (*c*) a singly charged helium ion, and (*d*) an alpha particle, all for the same fall of potential of 1500 volts.

6. Compute the magnetic field intensity required to bend each of the ion beams described in problem 5 into a circle of 20-cm diameter.

7. An ion is deflected into a circle of 13.10-cm diameter in a Bainbridge mass spectrograph (Fig. 15). The magnetic field is 8000 gauss, and the electric field in the velocity selector is 4500 volts per cm. What might the mass of the ion be? (There are several possibilities.)

8. Compute the average distance between centers of the atoms in copper, assuming them to be piled up in regular rows, as was done in Sec. 23 for mercury. Atomic weight = 63.6; specific gravity = 8.9. (In Sec. 186 it is explained that a more advantageous packing is possible in the copper crystal, so that the distance between centers is actually 2.56A.)

9. If the atom is considered as a target for alpha rays, with the nucleus as the bull's-eye, what proportion of the alpha particles "hitting" an atom will be likely to hit the nucleus?

10. An alpha particle shoots to within 10^{-11} cm of the nucleus of an atom whose atomic number, Z, is 30. Compute the repulsion force between the nucleus and the alpha particle at this distance.

REFERENCES

FOR ADDITIONAL READING

SIR WILLIAM BRAGG, *Concerning the Nature of Things*, Chapter I.
PHYSICS STAFF, UNIVERSITY OF PITTSBURGH, *Atomic Physics*, Chapters II and III.
J. B. HOAG, *Electron and Nuclear Physics*, Chapter X.
F. W. ASTON, "Atomic Weights," *Annual Report of the Smithsonian Institution for* **1921**, pp. 181–196.

FOR ADVANCED READING

F. K. RICHTMYER, *Introduction to Modern Physics*.
F. W. ASTON, *Isotopes*.
WALTER BLEAKNEY, "Mass Spectrograph and Its Uses," *American Physics Teacher*, Vol. 4, p. 12, February, 1936.
J. D. STRANATHAN, *The "Particles" of Modern Physics*, Chapter 5.

CHAPTER III

THE WAVE CHARACTER OF LIGHT

29. Light. Newton's classic experiments with light and color showed that *pure* light always produces sensation of *color.* White light is always composite, and a prism spreads it out into a spectrum of its component parts. But color is a result of sensation (sight) and is not an attribute of light in the physical sense of the term. What then is the *physical* difference between different parts of a spectrum?

Newton came close to the answer to this question, but is was left to other physicists nearly a century later (just at the beginning of the nineteenth century) to obtain the first good answer. Thomas Young, an Englishman, and a Frenchman named Fresnel share the honor of producing the first irrefutable evidence that light is wavelike in character, and thus presumably banishing forever the rival theory of the particle nature of light. Just how difficult it is to do away with the particle theory will be seen in Chapter VII. For the present let us examine the evidence for the existence of light waves.

30. Interference of Waves. The reader will recall that the phenomena of interference are among the most distinguishing characteristics of wave motion of any kind, so that the appearance of such a phenomenon is reliable evidence for the existence of waves. Standing waves, produced by the over-running of two identical wave trains which are traveling the same path in opposite directions, provide the simplest example of wave interference. In standing waves, however, the distance between nodes, or places of maximum destructive interference, is only a half-wavelength, so that observation of this phenomenon is limited to waves long enough for this distance to be observable. Since it is obvious that light waves cannot be much longer than the smallest distance which may be made visible by them (cf. Sec. 41), some other phenomenon must be employed for light waves.

An interference phenomenon of much larger scale may be produced by means of two *identical* wave trains traveling in nearly the same direction, as is shown diagrammatically in Fig. 18. It may be seen that, along the lines marked "max.," crest always coincides with crest (solid lines), and trough with trough (dotted lines). Everywhere along these lines the resultant effect is a maximum. On the contrary, along the lines labeled "zero," a crest of one wave everywhere coincides with a trough of the other, so that the two wave trains neutralize each other's effects along these lines. If these were water waves, the water would lap up and down to a maximum extent along vertical stakes placed at *A*, *C*, and *E*, but would remain at rest alongside stakes placed at *B* and *D*. It these were sound waves, a loud sound might be heard at *A*, *C*, and *E*, but no sound would be heard at *B* and *D*.

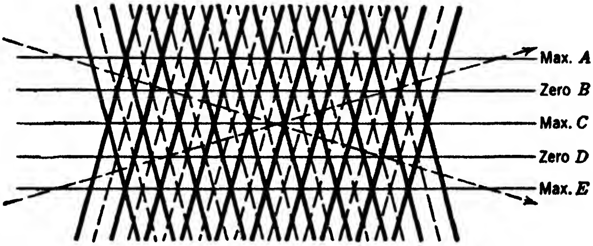
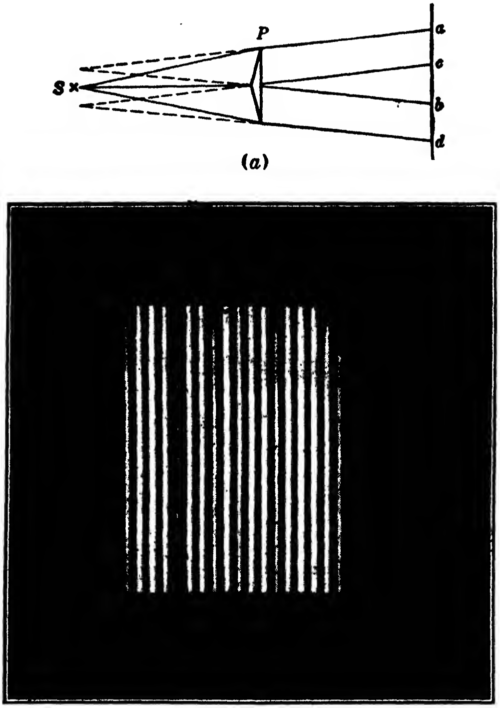


FIG. 18. INTERFERENCE OF WAVES. Two wave trains crossing each other at a small angle. Crests are represented by solid lines, troughs by broken lines.

31. Interference of Light Waves. Since the best microscopes enable us to see details of somewhere between 10^{-4} and 10^{-5} cm in size, it is obvious that the light waves by means of which we see these details cannot be much larger than this. If then we wish visually to demonstrate interference of light waves, the angle between the two interfering beams must be exceedingly small. Furthermore, both beams must be derived from the same source, since this is the only possible way to obtain two beams of light which are identical both in wavelength and in phase. Figure 19 shows one of the simplest of the many possible arrangements for demonstrating interference of light. The two halves of a beam of light coming from the very narrow slit, *S*, are caused to overlap by means of a double prism, which is equivalent to two prisms placed base to base. Since the prism angles are very small (less than one degree) the angle between the waves in the two overlapping beams is very small, and visible "fringes," such as those shown in Fig. 19(*b*), appear upon the screen between *b* and *c*, where the two beams overlap. The bright fringes correspond to the places labeled *A*, *C*, *E*, etc., of Fig. 18, while the dark fringes correspond to *B*, *D*, etc., and the explanation is the same as that given in Sec. 30. **This experiment demonstrates the existence of light waves.**

The source slit must be very narrow, and for best results the light should be nearly pure—that is, include only a narrow portion of the spectrum. If the experiment is repeated with light from different parts of the spectrum, the fringes are found to broaden progressively from the violet to the red end of the spectrum, showing that the *physical difference* between different parts of the spectrum is purely a difference of *wavelength,* the waves at the red end of the spectrum being the longest. Indeed, the experiment may be used to measure these wavelengths, which are found to range from around 8000A at the extreme red end of the spectrum to about 4000A at the violet end. The more fundamental difference is in the *frequency* of vibration, which may be computed from these wavelength values and from the speed of light, 3×10^{10} cm/sec. From the red end to the violet end of the spectrum the frequences range from about 38×10^{13} up to 75×10^{13} vibrations per second. If white light is used in this experiment, multicolored fringes appear, due to the overlapping of the fringes of different widths from different parts of the spectrum. Consequently, after the first five or six fringes they blur out entirely.



(*b*)

Fig. 19. **Fresnel's Bi-Prism.** (*a*) Arrangement of apparatus. (*b*) Photograph of the interference fringes which appear between *b* and *c*.

Many other arrangements will produce similar interference fringes. Young, who gave the first proof for the interference of light, used two fine slits placed close together. Since this experiment of Young's involves *diffraction* as a means of causing the beams to overlap, it will be described later on (Sec. 38).

32. Thin Films. Although the experiment just described requires precise apparatus and delicate adjustment to produce good results, many interference phenomena occur naturally every day. Such, for example, are the phenomena that produce the colors seen in thin films of oil on a wet pavement, in thin soap films, or in a thin air film formed between two glass plates pressed close together. Such a film may be formed between a perfectly flat glass plate and one having a very slightly convex surface, such as the surface of a very weak lens. If white light is used, the plates being held so that light is reflected to the eye by both surfaces of the air film, a series of multicolored rings, known as Newton's rings, will be seen close to the point of contact between the two surfaces. If pure light is used, many light and dark rings are seen, getting closer and closer together as they grow larger. See Fig. 20. The explanation of these interference phenomena in thin films is as follows:

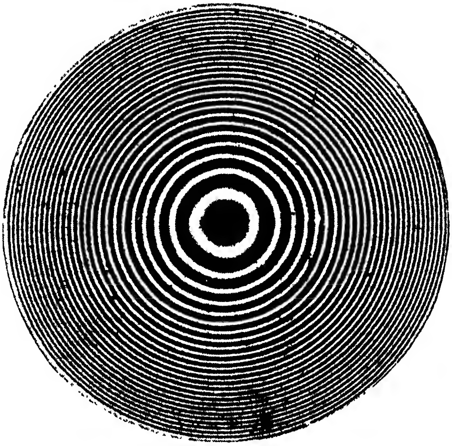


FIG. 20. NEWTON'S RINGS—MONOCHROMATIC LIGHT. Photograph by Professor F. A. Molby of West Virginia University. (For the method used, see his article in *American Physics Teacher*, Vol. 5, p. 178, 1937.)

Figure 21 shows a small portion of a thin film of glass, very much enlarged, with a beam of light passing through it from above. The ray, *a-b-c-d,* represents the path of a portion of this beam. A little light is reflected by each surface of the film, and since the film is so very thin, both these reflected beams enter the eye as if coming from the same spot. But whether they add up their effects or not—that is, whether this bit of the film appears bright or dark—will depend upon whether the two beams are in *phase* ("in step") or out of phase when they enter the eye. Let us consider that the thickness of this portion of the film represented in the figure is one-half wavelength of the light used, and that the light falls upon the film at nearly normal incidence. Then, when the light beam reflected from the lower surface of the film catches up with the one reflected from the upper surface, it will have traveled twice the thickness of the film, or one whole wavelength, farther than the beam reflected from the upper surface. Hence, *if the two reflections occur in exactly the same way,* the two beams would come together one whole wavelength out of phase, or *in phase* again, and this bit of the film would appear bright. Actually it appears dark, because of a difference in the way in which the two reflections take place at the two surfaces of the film. This may be explained most easily by first considering the analogous reflections of waves in a stretched cord.

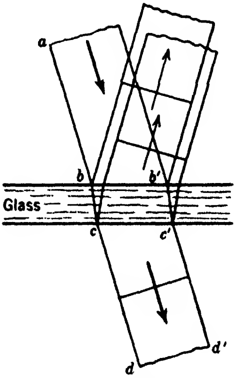


FIG. 21. REFLECTION OF LIGHT BY A THIN FILM OF GLASS.

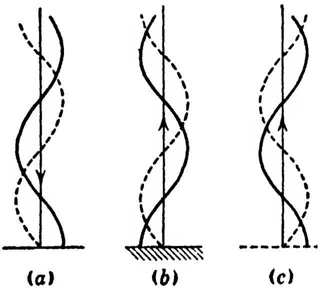


FIG. 22. REFLECTION OF WAVES IN A STRETCHED CORD. (*a*) Incident waves. (*b*) Waves reflected at a *fixed* end. (*c*) Waves reflected at a *free* end. Solid lines show these waves at the instant when displacements at the end are a maximum; dotted lines show them ¼ cycle later.

Figures 22(*a*) and (*b*) illustrate the reflection of waves at a clamped or "fixed" end. The solid line in (*a*) represents the incident wave at the instant when its displacement at the fixed end is a maximum. At this same instant the corresponding displacement for the reflected wave must be a maximum in the reverse direction, as shown by the solid line in (*b*), so that the resultant of both displacements will be zero at the fixed end. Had this reflection occurred at a point where a light cord was joined to a heavy one, the relative phases of the incident and the phase of the reflected waves in the light cord would be the same as for a fixed end, although then the amplitude of the reflected waves would be smaller and the point of reflection (the junction between the two cords) would not be at rest, but instead would be a point of *minimum* motion.

On the other hand, when reflection occurs at a free end (e.g., where the cord is joined to a very much lighter one[1]) conditions are just reversed; now, for the same phase of the incident wave which is represented by (*a*), the phase of the reflected wave must be as shown in (*c*), so that the resultant displacement of the free end is a *maximum.* Thus waves reflected at a free end are exactly opposite in phase to those reflected at a fixed end, providing the phase of the incident waves is the same in each case. This same difference in phase may be caused also by a path difference of one-half wavelength, and this is a convenient way to describe it. Let us now compare this with the reflection of light at a transparent surface.

At the top surface of the glass film, where the light is slowed up on going from air to glass, the conditions are comparable to reflection at the fixed end of a cord; at the lower surface, where the light passes from glass into air with a gain in speed, they are comparable to reflection at a free end. Here also it is easiest to describe the phase difference introduced by these two types of reflection as equivalent to increasing (or decreasing) by one-half wavelength the difference between the two paths traveled by the two reflected beams of light, just as it was interpreted above for the stretched cord. In the film represented by Fig. 21, the effective path difference is then the one wavelength, as figured for the actual thickness of the film, plus (or minus) a half-wavelength for the phase difference between the two types of reflection; the two reflected beams are thus out of step and the film appears dark. Likewise a film only a very small fraction of a wavelength thick would appear dark because of this difference in phase resulting from reflection, whereas a film a quarter of a wavelength thick would appear bright.

---

[1] The light cord is needed to maintain the tension.

In general:

> **Whenever the difference in lengths of the two paths is a whole number of wavelengths, the film appears bright; when it differs by half a wavelength from an integral number of wavelengths, it appears dark. The difference in path in each case must include the extra half-wavelength which represents the relative shift in phase upon reflection.**

Most thin films of these kinds will be of different thicknesses at different places over their surfaces. Thus two flat glass plates may touch at one edge and be slightly separated at the opposite edge, and in this case a series of parallel fringes, alternately bright and dark, may be seen in reflected monochromatic light. The dark fringes are located where the air film is an exact number of half-wavelengths thick, as explained above, while at the bright fringes the thickness differs by a quarter-wavelength from that at the adjacent dark fringes. In Newton's rings, the air film between the plane and the convex surfaces increases in thickness in all directions from the point of contact, and the alternate dark and bright rings are explained in exactly the same way.

The above discussion applies only to normal or approximately normal incidence. When the incident and reflected rays make appreciable angles with the normal, the extra path difference produced by the thickness of the film is less than twice that thickness, being in fact that distance multiplied by the cosine of the angle of incidence.

33. "Invisible Glass." When the refractive index of the film is intermediate between the refractive indices of the optical media above and below it, the reflections taking place are of the same kind at both surfaces of the film, and the difference in phase between the two reflected beams is only that due to the thickness of the film. If for example a glass surface is coated with a film of some substance whose index is between those of glass and air, light falling upon such a surface from the air side would be slowed down upon crossing each surface. Both reflections would then be of the same type. Further, if the film is a quarter-wavelength thick and the same amount of light is reflected by each surface, the surface will be invisible, since the two reflected beams will then be out of phase by one-half wavelength, and hence will cancel each other.

The two beams may be made of equal intensity by making the relative index of refraction from the film into the glass the same as that from air into the film. "Invisible glass" was first produced in this manner by K. Blodgett, who succeeded in coating glass with such trans parent films. The effectiveness of her process may be seen in Fig. 23. One half of the glass instrument-cover is coated on both sides with such a film; the other half is ordinary clean glass. The invisibility is almost perfect for normal incidence, and is still good at quite wide angles.

Unfortunately, these first films rubbed off very easily, so that the surfaces could not be cleaned. Since then films have been developed which are much more durable and may be utilized commercially for coating lenses, prisms, and other optical parts. The image produced by a complex optical instrument may be more than doubled in brightness by coating all its optical surfaces, and this process attained extensive application to military instruments. Both magnesium fluoride and cryolite are employed as coating materials. The commonest method of coating is by evaporation of the coating material in a high vacuum. The molecules of vapor fly in straight lines across the vacuum chamber to the clean glass surfaces placed there to be coated, forming a smooth, uniform coating whose thickness may be controlled by the time of evaporation. Magnesium fluoride produces the harder film, but cryolite

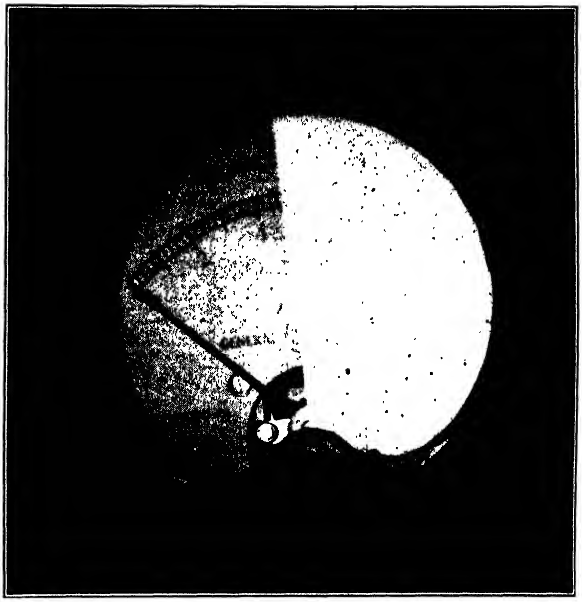


FIG. 23. INVISIBLE GLASS. This method for making glass invisible has been developed by Dr. Katharine Blodgett of the General Electric Company Research Laboratories. (This figure is reproduced with her permission from her article in *Physical Review*, Vol. 55, p. 391, 1939.)

yields somewhat lower reflection. Other coating materials and processes have been tried out, and even better methods may be developed in the future.

34. Applications. If the interference fringes produced by the air film between two flat glass plates are perfectly straight and uniformly spaced, and if one of the surfaces is optically flat, then the other surface is also optically flat: any imperfections of the surface are clearly revealed by lack of straightness in some of the fringes or by irregularities in their spacing. This is a standard method for testing the surfaces of prisms, mirrors, and other optical parts, and this test makes possible the production of optical surfaces which do not differ from true planes by more than a small fraction of a wavelength of light. The same test is employed in making the original optically flat test plate. Three plates are ground together and polished until any two of them (all combinations of two taken at a time) will show interference fringes which are straight and equally spaced to the degree of precision required. Three plates must be used rather than two, since two plates alone might fit together so as to show very uniform interference fringes, and yet both of them might be slightly curved, one being convex and the other correspondingly concave.

The same optical test is used to compare the gauge blocks employed as standards of length in precision-machine-tool testing. The ends of these gauge blocks are polished optically flat, and two such blocks are compared by placing them on end and side by side on one optical flat and placing a second optical flat across their top ends. Any deviations of the interference fringes as they cross the boundary between the ends of the two blocks show that the blocks differ in length. In this manner standards of length which are accurate within one one hundred thousandth of an inch may be produced for use in industrial work.

Interference fringes may also be utilized to detect and to measure very small changes of length, such as the thermal expansions of small blocks of material, or very slight deformations of supposedly rigid structures. The bending produced by the pressure of one finger at the center of a 3-foot length of standard railway rail, supported at its two ends, may easily be detected in this manner.

35. Interferometers. If the air film is thick, parallel rays must be used, the surfaces must be flat all over to within a small fraction of a wavelength of light, and the light must be very pure. When these conditions are satisfied, interference can be observed for long path differences between the interfering beams. Devices which employ such long interference paths are called interferometers, and the several types are described in books on optics, which also describe their many uses.

36. Standard Wavelengths. One of the most important uses for interferometers is the measurement of standard wavelengths of light. Using the interferometer which he invented, Michelson made a direct comparison of the wavelength of one of the red lines of the cadmium spectrum with the standard meter bar at Paris. This he did by counting the number of waves in the length of the standard meter, and thence computing the wavelength of the light. He chose this particular cadmium line because he found it to be the purest, or the nearest to just one frequency, of all lines which have ever been examined. This experiment has since been repeated by other observers, using other types of interferometers, and the results of all such measurements agree to within a fraction of a wavelength in the whole meter length. The average of all of them is

$$6438.4693 \text{ A}$$

This is probably the most accurate physical measurement ever made. Many other spectrum lines have been measured in similar manner, by comparison with this cadmium line, and these wavelengths serve as standards for the measurement of all other wavelengths.

Recently[2] it has been found possible to produce the 198 isotope of mercury in sufficient quantity to make a mercury lamp, and the 5461A green line from this source is far purer than the cadmium red line, and much brighter. Its wavelength is found, by comparison with the cadmium line, to be

$$5460.752 \text{ A}$$

This is a preliminary value, supplied by Dr. W. F. Meggers, who is making this comparison at the National Bureau of Standards. It is expected that this comparison, when completed, will be exact to within 1 part in 100 million, or to within 0.00005 A.

As this figure indicates, the precision with which two wavelengths of light may be compared exceeds the accuracy of measurement of any wavelength by comparison with the Paris meter bar, because of the *coarseness* of the lines which mark the ends of the meter on that bar. For this reason it is certain that some wavelength of light, probably the Hg^{198} green line, will be established eventually as the *primary standard of length.*

---

[2] See J. H. Wiens, *Physical Review*, Vol. 70, p. 910, December, 1946. The nuclear reactions are $Au^{197} + n \rightarrow Au^{198}$, which is radioactive and emits an electron, becoming Hg^{198}.

See Chapter XVI for a discussion of this type of process.

It is interesting to note that, to produce this isotope, *gold* is transmuted into mercury by neutron bombardment. It is one of the important peacetime products of the Manhattan Project, which developed the atomic bomb.

37. Diffraction of Light. When sound or water waves pass through a small opening, they spread out in all directions from that opening as if they originated at it. This is illustrated diagrammatically in Fig. 24. The wave fronts for the ripples which have passed through the opening are semicircles having their centers at the center of the opening. This is the simplest possible example of the phenomenon of diffraction, which is another essential characteristic of all wave motions. Indeed, as Huygens first demonstrated, every portion of every wave tends to spread in all directions in just this fashion.

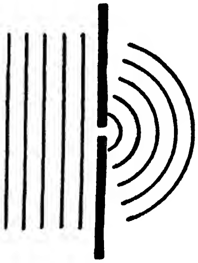


FIG. 24. DIFFRACTION OF WAVES BY A NARROW SLIT.

Some of the early objections to the wave theory of light arose from the observation that light waves apparently did not exhibit this phenomenon of diffraction. These observations were that light beams cast sharp shadows, and that beams of light may be narrowed down by passage through quite fine slits without any apparent spreading outside the boundaries determined by straight lines drawn from the source past the edges of the slit. The fallacy of such comparisons of light with sound or water waves lies in the discrepancy in size of the waves compared—a slit may be made quite fine by ordinary standards yet be many wavelengths wide, and Huygens showed that, when a beam is many wavelengths wide, the tendency of each little portion of this beam to diffract or spread out is checked by the same tendency in all its neighbors. Only straight forwards may they add up their effects; in all other directions (except at the edges of the beam) they overlap to cause destructive interference.

To demonstrate diffraction with light waves, first of all the source must be very small and very bright; usually it is a very narrow slit with a very bright light close behind it. If light from this source passes through a moderately narrow slit and then onto a screen some distance away, the bright bar of light which it produces upon the screen will appear quite sharply outlined, unless we examine it very closely. If for lack of a better term we call this bright bar a "bright shadow," then we may say that its boundaries are the geometric shadow, that is, the shadow determined by drawing straight-line rays as described above. For example, if the second slit is 5 mm wide and is placed 10 cm from the source, its bright shadow upon a screen 200 cm away will be 10 cm wide.

If the slit is narrowed, the bright shadow becomes narrower also, but not indefinitely. In the example cited, when the slit is about 0.4 mm wide, the bright shadow is at its narrowest: it is then actually considerably narrower than the geometric shadow, owing to secondary interference phenomena which do not need to be described here. If the slit is narrowed further, the band of light upon the screen begins to spread out wider and wider. Even before it reaches its narrowest width it is quite noticeable that the edges are far from sharp, and now the band of light just fades away into nothingness at the edges. Again there are some secondary interference effects which we may neglect here—which indeed we may probably not see at all unless the light source is exceedingly bright. When the slit is 1/100 mm (about 20 wavelengths) wide, the band of light is about 20 cm wide and very, very faint. To approach the extreme condition shown in Fig. 24 the slit must be made only a wavelength or so wide, and the light would be far too dim to produce any visible effect for a single slit. The intensity may be increased by using many identical slits together, as in a diffraction grating; then their combined effect shows diffraction nearly as wide as this. (See Sec. 39.)

38. Young's Experiment. Diffraction phenomena, other than the very simple example shown in Fig. 24, involve interference as well as diffraction. With the exception of the two or three examples which follow, it is not the purpose of this book to go into detail concerning them, since they are described and discussed in many excellent and readable books, some of which are listed at the end of the chapter. The simplest example of interference combined with diffraction is Young's famous double-slit experiment, the first experiment to give proof of the existence of light waves. To perform this experiment the single diffracting slit described in Sec. 37 is replaced by two very narrow slits, parallel to each other and quite close together. Each of these slits will diffract the light in the manner just described, and the two spreading beams of light will then overlap and produce interference fringes, just as did the two beams produced by the biprism.

One may view these fringes with the simplest of means: if two fine slits are scratched about half a millimeter apart in the opaque part of a camera negative, or in the silvered back of a piece of mirror glass, the fringes may be seen by holding this double slit close to the eye and looking towards a small bright source of light, such as a distant arc light. The fringes fall directly upon the retina of the eye, although of course they appear to be (as a virtual image) somewhere out in front of the eye. The spreading of light by diffraction through a single slit may similarly be demonstrated with a single fine slit held close to the eye. If the slit is only moderately fine, it may be seen that the broadened bar of light described in Sec. 37 is accompanied on either side by several faint secondary maxima. These secondary maxima are also due to interference, in this case between the light coming from the two sides of the single slit, as is explained in the books on optics.

Figure 25 explains the interference fringes produced by the double slit. In this figure *a* and *b* are the two slits, seen end on, and the circular arcs represent crests of successive wavelets spreading out from each slit. The dashed line, *a–c*, is drawn through the crest of a wavelet which is just coming through the slit, *a*, and tangent to the wavelet which has at the same instant traveled just one wavelength out from *b*. As these wavelets spread outward, this line travels outward with them, as is indicated by the other dashed lines drawn in the figure; it is evident that, when these expanding wavelets have traveled a long distance from the slits, they will merge together along this line so as to produce the first interference fringe beyond the middle one. The position of this fringe will be given by the angle, θ_1, which the dashed line, *a–c*, makes with the surface upon which the slits are scratched. In the triangle *a–b–c*, the short side, *b–c*, is the wavelength, λ, of the light used, while the hypotenuse, *a–b*, is the distance, *d*, between the slits. Hence

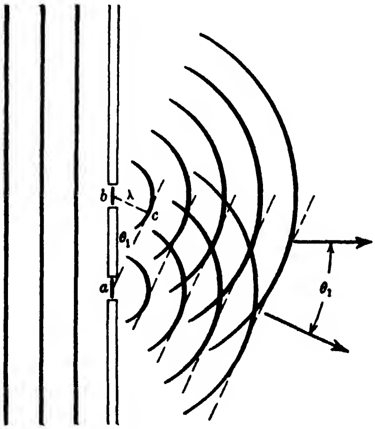


FIG. 25. YOUNG'S DOUBLE-SLIT EXPERIMENT.

$$\sin\theta_1 = \frac{\lambda}{d} \tag{3·1}$$

Successive fringes are formed where the difference in path, *b-c*, is two wavelengths, three wavelengths, etc., so that

$$\sin\theta_2 = \frac{2\lambda}{d}, \quad \sin\theta_3 = \frac{3\lambda}{d}, \text{ etc.}$$

and in general,

$$\sin\theta_n = \frac{n\lambda}{d} \tag{3·2}$$

Usually the separation of the slits is wide enough that the angles are all quite small and the fringes are practically equidistant.

39. Diffraction Grating. If a distant line source, such as a narrow slit or a straight-filament lamp, is viewed through a piece of fine wire gauze held close to the eye, the source will still be visible, although somewhat blurred; in addition, to either side of it will appear a number of bright interference fringes. If both the source and the wire gauze are quite fine, it may be seen that these interference fringes are really little spectra, with the red ends farthest from the original, or central, image. With light of a single wavelength, each will be a quite narrow line of the single color represented by that wavelength. Light from the source is diffracted by the slits formed between the wires of the gauze, and the many sets of wavelets interfere in much the same manner as do the wavelets from the two slits of the double slit.

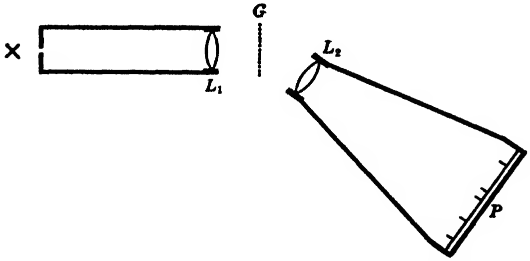


FIG. 26. GRATING SPECTROGRAPH. G is the grating. The light source is placed at ×. The collimator lens, L_1, makes parallel the light rays coming from the slit, and the camera lens, L_2, projects the spectrum image upon the photographic plate at P.

Such an assemblage of slits is called a diffraction grating. If an optical grating (see Sec. 40) is similarly placed in front of the eye, broad, well-separated spectra will be seen, several of them appearing to either side of the central undispersed image. The first spectrum to either side is called the *first-order* spectrum; the second, the *second-order* spectrum, etc. The high-order spectra are proportionally more dispersed, so that beyond the second order they begin to overlap. Effects somewhat similar to those observed with the wire gauze are produced by a piece of fine cloth, but the spacing of the threads in the cloth is too irregular to produce any sharp lines or clear spectra.

For all of the slits of a diffraction grating to be able to combine their effects it is necessary first that the incident light consist of parallel waves: either a distant source must be used, as in the examples described above, or a lens must be placed between the source slit and the grating, as shown in Fig. 26, with the slit at the principal focus of the lens so that parallel rays emerge from the lens towards the grating. Second, a lens must be used on the other side of the grating to focus the spectra. The eye lens may serve for this purpose, as it does in the examples given above, but better spectra are formed by a telescope focused for parallel rays. Where the spectra are to be photographed, this lens is the camera lens, focused for parallel rays, as is shown in Fig. 26.

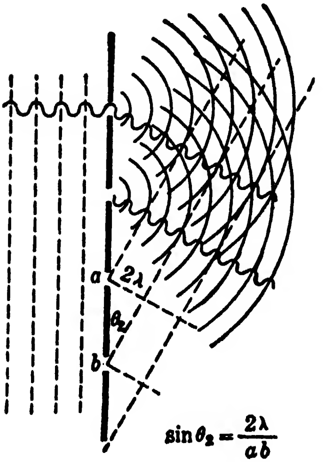


FIG. 27. THE PRINCIPLE OF THE DIFFRACTION GRATING.

Figure 27 shows schematically four elements of a diffraction grating. The dotted lines to the left represent wave fronts of a beam of plane waves of monochromatic light approaching the grating. The circular arcs represent the diffracted wavelets, just as in Fig. 25. The dotted lines on the right may be considered as the plane waves[3] which will form the spectrum line, the *second-order* line in this case. Consider, for example, the dotted line which starts at *a*. Along this line the wavelets from each successive slit are just two wavelengths behind those from its immediate neighbor just below; and where this plane wave front, from the several hundred or even several thousand slits of the whole grating, is brought to a focus by the camera lens, a bright sharp second-order spectrum line is formed. The formula giving the angular position of this spectrum line may be derived directly from the figure; it is the same as the one used for the double slit, namely, $\sin\theta_2 = 2\lambda/d$ for this second-order line, or for the nth order

$$\sin\theta_n = \frac{n\lambda}{d} \tag{3·2}$$

The reason for the sharpness of the spectrum lines produced by a diffraction grating is this: With only two slits, zero brightness in the fringes is not reached until the waves from the two slits are out of phase by all of ½ wavelength, and this does not occur until a point midway between the two fringes is reached. But with a grating of, say, 1000 slits, zero brightness occurs as soon as the waves from *adjacent* slits are

---

[3] At this distance from the grating the wavelets do not, of course, form a complete wave front; but, when they have expanded to many wavelengths in radius, they will coalesce along this line (which has moved along with them) into a continuous plane wave.

but 1/1000 of a wavelength out of phase, since then the waves from the 501st slit are 500/1000 or ½ wavelength out of phase with those from the first slit, and the same is true for the 502nd slit and the 2nd slit, the 503rd slit and the 3rd slit, etc., all across the grating. Thus the waves from one half of the grating cancel the effects of those from the other half, and the bright line appears 500 times narrower than the bright fringe produced by the double slit.

40. Construction of Diffraction Gratings. A piece of fine-mesh wire cloth (it may be obtained up to 200-per-inch mesh) makes a good demonstration grating, and the first gratings were not much more than this, being merely grids of fine wires stretched close together. Modern gratings are made by ruling the equivalent of opaque lines on glass or polished metal. The ruling tool is a diamond point which grooves the surface, and the ruling is done by automatic machinery. If a broad spectrum is to be obtained, the rulings must be very close together (gratings are made with as many as 25,000 rulings per inch), but if that spectrum is to have sharp narrow lines, it is even more important that the rulings be spaced equal distances apart with an accuracy comparable to the fineness of the rulings. Years have been spent in making a lead screw accurate enough to guide the ruling tool from one ruling line to the next with the precision required.

Gratings ruled on metal must of course be used as reflection gratings. The reflecting strips left between the rulings are so narrow that the light reflected by them spreads out as little circular overlapping wavelets identical to those shown in Fig. 27, but turned backward towards the direction from which the incident waves are coming. Spectra are formed by these overlapping wavelets in the same manner as for the transmission grating.

If a grating is ruled upon the surface of a concave metal mirror, it can be used without any lens: it will focus its own spectra. Such concave reflecting gratings have been made with a radius of curvature of about 21 feet, which make a first-order spectrum of visible light about 5 feet long! Concave gratings are particularly useful for studies in those extreme infrared and ultraviolet regions of the spectrum for which most if not all substances are opaque, and for which consequently it is difficult or impossible to provide lenses. See Chapter VI.

Original gratings are quite expensive. Fortunately, a process has been developed for molding gratings in transparent substances such as gelatine, using an original grating as a matrix, and gratings skillfully made by this process produce excellent spectra and are quite inexpensive. They may be obtained mounted on glass plates as transmission gratings, or on plane or concave mirrors as reflecting gratings.

41. Optical Images. Lest it be considered that the phenomena of interference and diffraction of light occur only under unusual circumstances, and that the formation of sharp spectrum lines by a grating is something out of the ordinary, it should be mentioned here that even the formation of an image by a lens is a phenomenon which depends upon interference and diffraction of light waves. In terms of wave theory, the function of a lens is so to slow up the middle portions of the beams of light which pass through it that all the light waves which come to the lens from any one point of the object will be in phase with each other when they arrive at the corresponding image point. In other words, the distance from the object point to the corresponding image point, along a ray passing through any portion of the lens, is the same for all rays when measured in wavelengths.

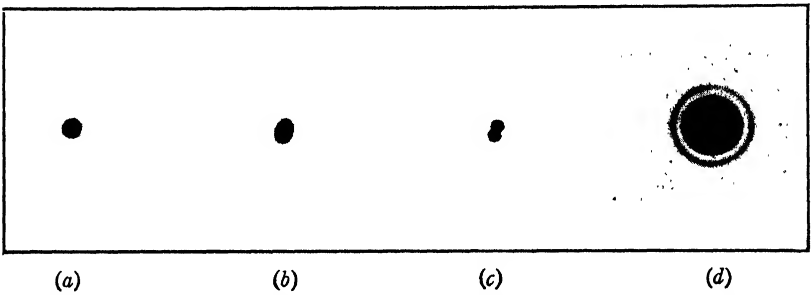


FIG. 28. HIGHLY MAGNIFIED "STAR" IMAGES. (a) Single point source. (b) Double point source, images unresolved. (c) Double point source, images resolved by using a larger aperture. (d) Single point source—longer exposure, showing diffraction rings.

No matter how perfect the lens surfaces may be, the image of a point object will never be a point, however. Rather, it will be a diffraction pattern consisting of a tiny central spot of light, somewhat brighter at its center, and surrounded by several faint rings of light which may not appear at all unless the source be very bright. Such a diffraction pattern is shown in Fig. 28(a), which is a much enlarged photograph of a fine point source, made with a good camera lens stopped down to its smallest aperture. Increasing the size of the aperture of the lens will make this diffraction pattern smaller, but even the telescope soon to be in use at Mt. Wilson observatory, having an aperture of 200 inches, or more than 16 feet, will not produce a perfect point image. (This telescope is a reflecting telescope, the image being formed by a mirror instead of by a lens, but the same principles apply to the formation of images by focusing mirrors as apply to the formation of images by lenses.)

The image of an object of finite size is slightly blurred by this diffraction effect, the blurring being entirely due to the imperfection, if so it may be called, of the light itself. All of the light waves coming through the lens from a point object are exactly in phase at the center of the diffraction spot, but they are also nearly in phase for a finite distance around this center, and this accounts for the finite size of this spot of light. The central spot does not fade to zero until it has spread to a point where the light from one half of the lens is exactly out of phase with the light coming from the other half; it may be shown that its diameter is approximately $2F\lambda/a$, where F is the focal length of the lens, a is the diameter of its aperture, and λ is the average wavelength of the light.

As a further example of the importance of this phenomenon, consider Figs. 28(*b*) and (*c*), which reproduce photographs taken of the same object through apertures of different sizes. The same aperture was used for making (*b*) as for making (*a*), while (*c*) was taken through a somewhat larger aperture. (*c*) shows clearly that this object consists of two fine points close together, the diffraction patterns for each point being small enough so that the two form a distinct double pattern. But (*b*) appears very much the same as (*a*), the fact that this object is double being entirely concealed by overlapping of the diffraction patterns from the two points. This illustrates one of the important reasons for building larger and larger telescope lenses and mirrors. Not only do larger lenses gather more light and thus detect fainter stellar objects, but also all stellar images are made clearer and sharper. Stars which appear to be single may be revealed as double; this is quite important, since double stars frequently are able to tell us much more about their size, distance, etc., than are single stars. More detail also may be seen in nebulae, those distant stellar universes from which we may hope to learn more about our own galaxy.

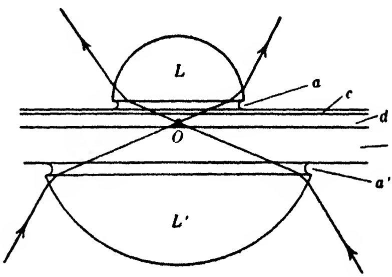


FIG. 29. VIEWING ANGLE FOR AN IMMERSION MICROSCOPE. *O* is the object; *s*, the slide; *c*, the cover glass; and *d*, the mounting medium. *L* is the lowest lens of the objective while *L'* is the uppermost lens of the condenser. *a*, *a'* is the immersion liquid.

The same principles set a fundamental limit to the smallest detail which may be seen with a microscope. Even with the best immersion objective, into which the light comes from almost all sides of the object at the same time (see Fig. 29), the smallest detail which may be seen is of the order of magnitude of half a wavelength of the light used. This means that, if two points are closer together than half a wavelength, they cannot be resolved (seen as separate points), no matter how great the magnification. On this basis the useful limit for magnification works out to be about 1500× for visual work.

> Taking the average wavelength as 0.0006 mm in air or 0.0004 mm in the mounting medium, lenses, etc., we find that the smallest detail which may be resolved is about 0.0002 mm; and 1500× will magnify this to 0.3 mm, so that it is clearly visible.

Greater magnification than this will only show a fuzzier image with no more detail. If the condenser lens is stopped down or an objective of smaller numerical aperture used, still less detail will be seen. The only way to increase the resolving power is to improve the *light* used, and this has actually been done. Microscope cameras have been made with lenses of quartz and fluorite (both of which are transparent to ultraviolet light), so that the intense mercury light of the single wavelength, 2536A, may be used; this doubles the resolving power. The really clever thing about this invention is the use of a combination of quartz and fluorite to construct a lens which is achromatic in the sense that the microscope may be adjusted and focused in visible light, and is then in focus for 2536A.

The electron microscope (see Sec. 117) makes possible much higher resolving powers.

42. Scattering of Light. When light falls upon a tiny particle of matter which is not very many times bigger than the wavelength of the light, the light which is reflected by that particle will spread out into a beam much wider than the beam intercepted, even if the particle is optically flat. This example of the diffraction of light has already been mentioned in connection with reflecting gratings (Sec. 40). If the particle is considerably smaller than the wavelength of the light falling upon it, then the light will be diffracted or "scattered" by it with about equal intensity in all directions, no matter what the shape of the particle may be. This accounts for the milky or smoky appearance of almost all translucent substances; unless the concentration of scattering particles is too great their existence may be discovered by microscopic examination. Tobacco smoke will illustrate this. If smoke from a cigarette is drawn into a glass box which is strongly illuminated from the side, and is then viewed from above through a microscope (see Fig. 118), the box will appear to be filled with a myriad of bright starry points. Each point is a smoke particle so much smaller than light waves that it scatters light in all directions and appears itself as a tiny source of light. The so-called ultramicroscope operates in a similar fashion to reveal the presence in colloidal suspensions of submicroscopic particles.

43. Blue Eyes and Blue Skies. When the scattering particles are very much smaller than the waves of visible light, the shorter waves are scattered very much more than the longer ones. This makes the scattered light predominantly blue in color. Even tobacco-smoke particles are small enough to make the smoke appear bluish against a black background, and with still smaller particles the blue color due to this cause may be quite intense. Blue eyes, the blue in a peacock's feather, and probably all the blue colors appearing in animal fibers are due to scattering of this kind. Likewise the blue of the sky is caused by this kind of scattering. It has been shown that the intense blue of the sky on a very clear day results from the scattering of sunlight by the molecules of the atmosphere.[4] On less clear days other submicroscopic particles contribute to the blue color. Since these particles are larger, the blue color is less intense.

44. Red Sunsets. Since scattering robs the direct beam of its shorter waves, the transmitted light will appear yellowish or even reddish. Tobacco smoke, which appears bluish against a black background, appears yellow or brown against a white background. The light from the setting sun is yellowish or reddish for the same reason. At the time of sunrise or of sunset the sunlight traverses a very thick layer of air, as compared to other times of day, and the loss from the transmitted beams of the short waves which are scattered on the way is very apparent in the sunset hues which appear.

PROBLEMS

1. Two glass plates are separated by an air film 0.00123 mm thick. Will this air film reflect the green light of the mercury arc (5460A) or not? Explain.

2. Compute the thickness of the thinnest soap film that will strongly reflect sodium light (5890A). Take the index of refraction for the soapy water to be 1.34.

3. A mixture of red light (6600A) and green light (5000A) falls normally upon an air film 0.00035 mm thick. What will be the color of the reflected light? Explain, making the necessary computations.

4. What wavelengths of *visible* light will be reflected most strongly by an air film 0.00150 mm thick?

5. Newton's rings are formed between a plate and a lens, in mercury green light (5460A). Compute the thickness of the air film between the lens and the plate at the place where the 15th bright ring appears.

6. Compute the radius of curvature of the lens of problem 5, if the diameter of the 15th bright ring is 22.7 mm.

---

[4] Wood, *Physical Optics*, Third Edition, p. 426.

7. If the space between the lens and plate of problem 5 is filled with water, will the diameter of the 15th ring be greater or smaller than before? Explain.

8. Prove that, if the light makes an angle θ with the normal *in the film*, the path difference for the reflected rays is $2d \cos \theta$.

9. The bright interference fringes produced by performing Young's double-slit experiment are found to be 3.25 mm apart, on a screen 200 cm away from the slits. The slits are 0.300 mm apart. Compute the wavelength of the light used.

10. How many waves of the cadmium red line (Sec. 36) are there in one meter? Michelson counted twice this number, since the light beam is doubled back upon itself in the interferometer. How big an error would have been made in the wavelength of this light, if he had made a counting error of one wavelength?

11. A diffraction grating has 2500 rulings (equally spaced slits) per cm. Light of 6000A wavelength falls normally upon this grating, as shown in Fig. 27. Find the angle at which the third-order spectrum image line appears. Do this graphically, by drawing a large-scale diagram and measuring the angle with a protractor.

12. A fine-mesh wire screen, when held in front of the eye, produces several orders of spectrum images of the 5460A mercury line. The fifth-order line is 10 degrees away from the central image. How fine is the screen?

13. In the spectrum formed by a diffraction grating, one of the first-order spectrum lines is formed by rays which make an angle of 20 degrees with the normal. The grating has 5000 rulings per cm, and the incident light falls normally upon it, as in Fig. 27. (*a*) Compute the wavelength of this light. (*b*) How many orders of spectra may be seen for this line? (Graphical solution is suggested.)

14. The higher orders of spectra formed by a diffraction grating overlap each other. What other wavelengths will produce image lines which will coincide with the third-order line of 5880A? Which of these are in the visible part of the spectrum? Explain why the answer to this question does not depend upon the grating spacing.

REFERENCES

FOR ADDITIONAL READING

F. A. SAUNDERS, *Survey of Physics*, Chapter 35.
P. I. WOLD, *Kimball's College Physics*.
EDWIN EDSER, *Light for Students*.
J. K. ROBERTSON, *Introduction to Physical Optics*.
A. A. MICHELSON, *Light Waves and Their Uses*.

FOR ADVANCED READING

R. W. WOOD, *Physical Optics*, Third Edition, Chapters I, VI, VII, XIII.
THOMAS PRESTON, *The Theory of Light*.
R. A. HOUSTOUN, *A Treatise on Light*.
A. C. HARDY and F. H. PERRIN, *The Principles of Optics*, Chapter XXVIII.
K. B. BLODGETT, "Use of Interference to Extinguish Reflection of Light from Glass," *Physical Review*, Vol. 55, p. 391 (1939).

CHAPTER IV

POLARIZED LIGHT

45. Longitudinal and Transverse Waves. The distinction between longitudinal and transverse waves is easy to make when the individual waves are visible, as for example are waves in stretched cords. But how may one be distinguished from the other when they are as small and as invisible as light waves? The answer to this question is given by the one-sidedness or *polarization* of transverse waves. Longitudinal waves in a horizontal stretched cord, for instance, may have motions of but one kind: forward and backward motions in the direction of the cord. But transverse waves in the same cord may consist of vertical motions, horizontal motions, diagonal motions, or even circular or elliptical motions, just as long as all of the motion is at *right angles to the cord.* In general, whereas the vibrations in longitudinal waves can occur in only the one direction determined by the direction of travel of the waves, in transverse waves the motions are limited only by the necessity of being at right angles to the direction of travel of the waves. When transverse vibrations are linear, so that both the line representing the direction of travel of the waves and the perpendicular line along which the vibrations are occurring may be drawn in a single plane, the wave is said to be **plane polarized.**[1] If the particles move in circular or elliptical paths, then the wave is said to be **circularly** or **elliptically** polarized.

46. Polarized Light. When the wave theory of light was first being considered, scientists did not conceive of the possibility of anything but longitudinal light waves: transverse waves as they knew them traveled in rigid media only, whereas light travels through liquids, gases, and even empty space. The phenomena which will now be considered as evidence for the transverse character of light waves were considered by them to be phenomena which could not be explained by the wave theory.

---

[1] Strictly speaking, the so-called plane of polarization for light waves is the plane through the line of travel which is perpendicular to the direction of vibration, this convention having been established before the direction of vibration was known. This idea is now a confusing one and will not be used here.

Light waves do not ordinarily exhibit any polarization characteristics, and might therefore easily be suspected of being longitudinal waves. This lack of any polarization characteristics, however, may be due to the composite nature of a beam of light. A beam of ordinary light is produced by countless light sources; even if all these sources are vibrating with the same frequency (monochromatic light) they may still be vibrating in all possible directions and in all possible phases, so that the average effect, even if the light waves are transverse, would be the same for all directions of vibration. However, if this is so, it might be demonstrated by filtering out from ordinary light a beam in which the vibrations are in one direction, namely, plane-polarized light. The various ways in which this may be done will now be described.

47. Polarization by Selective Absorption. Tourmaline is a mineral which cleaves into thin semitransparent colored plates (sometimes yellowish in color, and sometimes greenish). If two such plates are superimposed so that light travels through both of them, and if their axes are parallel as is shown in Fig. 30(*a*), then much of the light which passes through the first plate will pass through the second one. But if one of the plates is turned through 90 degrees, so as to be at right angles to the other one, as is shown in (*b*), then the second plate stops entirely (absorbs) all the light which has passed through the first one.

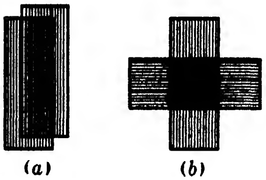

(*a*) (*b*)

FIG. 30. TOURMALINE PLATES.

The fact that light can be completely polarized in this way proves that light waves are transverse waves, and transverse waves only.

The tourmaline plate transmits light if its vibrations are in one certain direction, absorbs it if the vibrations are in the direction perpendicular to that one. Ordinary light may be considered a mixture of transverse waves vibrating in every possible direction perpendicular to the direction of travel of the light; and the first tourmaline plate polarizes this ordinary light by transmitting only those vibrations, and components of vibrations, which are in this certain direction. The transmitted light is then plane-polarized light. If the tourmaline plates have their axes parallel, these vibrations will pass through the second plate, but if their axes are perpendicular to each other, the vibrations passed by the first plate are absorbed by the second, so that no light gets through.

A few other crystals have properties similar to those of tourmaline, and the same properties are utilized in the Polaroid [2] polarizing mate-

---

[2] Produced by the Polaroid Corporation, Boston, Mass.

rials, which consist of plastic sheets in which are imbedded submicroscopic polarizing elements with *all their polarizing axes parallel.* In the Polaroid J polarizers the polarizing elements are needle-shaped crystals of quinine iodosulfate; in the more recent Polaroid H and K polarizers the dichroic sheets are respectively chains of polyiodine $(—\text{I}—\text{I}—\text{I}—)_n$ and chains of polyvinylene $(—\text{CH}{=}\text{CH}—)_n$. Polaroid transmits much more light than tourmaline, especially the newer forms; it may be obtained in large sheets; and it is neutral in color. Figure 31 shows a photograph of two Polaroid plates with their axes crossed; note that the light is completely extinguished by the overlapping portions.

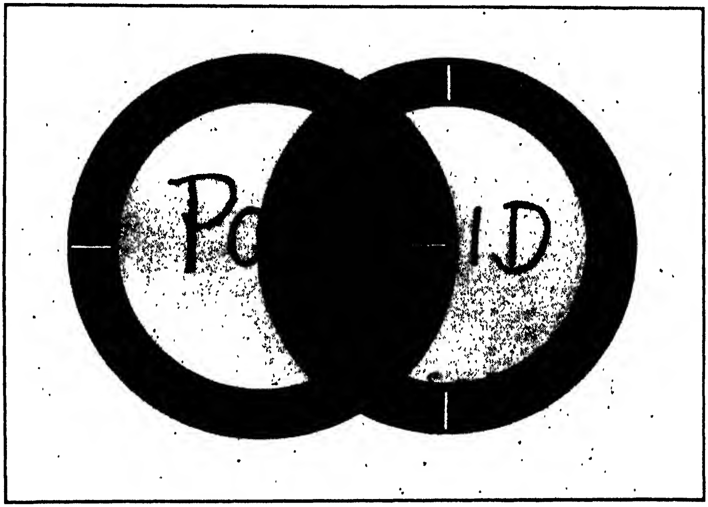


FIG. 31. POLAROID PLATES.

48. Polarization by Scattering. If the blue sky on a clear day is viewed through a tourmaline crystal or a Polaroid plate, looking in such a direction that the line of sight is at right angles to the sun's rays, a position for the polarizer may be found such that the sky appears almost black. Quite evidently the sky light which is emitted in this direction is polarized. Since light from the clouds is not polarized, a polarizing plate makes an excellent filter for cloud photography, especially for photography in color. Although most effective in the direction perpendicular to the sun's rays, it may be turned so as to produce appreciable darkening of the sky (corresponding to partial polarization of sky light) in almost any direction except straight towards the sun or straight away from it.

As has been explained in Sec. 43, sky light is light which has been scattered by exceedingly fine particles of matter. In general, light which is scattered by very fine particles in the direction perpendicular to the incident beam is plane polarized, as may be further demonstrated by passing a vertical beam of light into a tall jar of water containing a very dilute suspension of colloidal particles. (If some rosin is dissolved in alcohol and a few drops of this solution are stirred into the jar of water, a suitable suspension will result.) A photograph of this experiment is shown in Fig. 32. Because of the light scattered by the colloidal particles the path of the beam appears luminous; and the scattered light may be shown to be polarized, just as sky light is found to be, by viewing the beam through a polarizer. The lower of the two Polaroid plates shown in Fig. 32 is turned so as to transmit fully this scattered light, and through it the beam appears nearly as bright as if the plate were not present; but the beam is invisible through the upper Polaroid, which is turned so that its axis is perpendicular to that of the first one.

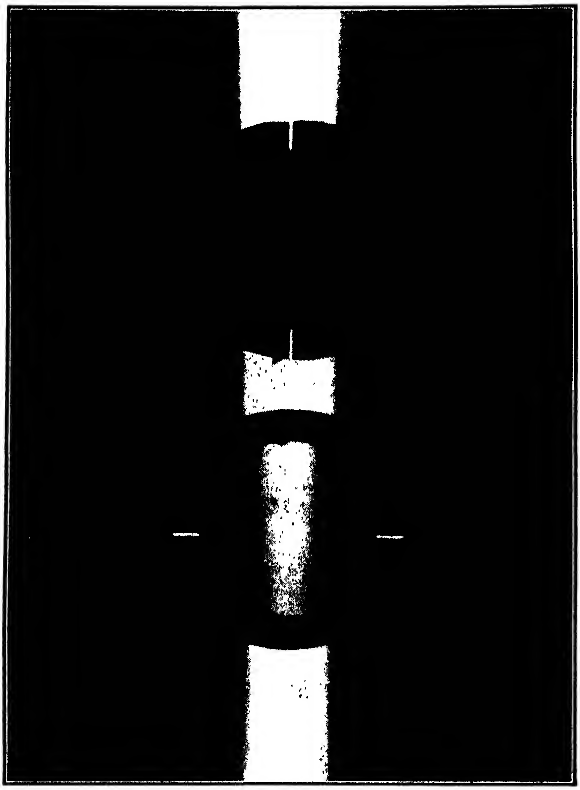


Fig. 32. Tyndall's Experiment.

Although this method of producing polarized light is not of much practical value, it is of fundamental importance to our understanding of all the phenomena of polarized light. Although all other methods of producing polarized light prove that light waves are transverse waves, none of them gives direct evidence as to the *direction* of the vibrations. Polarization by scattering, on the contrary, does identify this direction. Since the light vibrations in the primary beam lie all of them in a plane perpendicular to the beam, the vibrations of the light scattered in this plane must also be in the plane and consequently polarized. In the example illustrated in Fig. 32, the light scattered towards the camera must be polarized with its vibrations *horizontally* directed.

If now the direction of the vibrations passed by a Polaroid plate is found by viewing this scattered light through it, this Polaroid plate may then be used as an *analyzer* to determine the direction of polarization in other cases. The directions for the white lines which appear on the rims of the Polaroids in Fig. 32, and which show the directions of the vibrations passed by them, were located in this manner. Nicol prisms (Sec. 51) and other polarizing devices may be calibrated for use as analyzers in the same fashion.

As a modification of the light-scattering experiment, the incident beam of light may itself be polarized. It will now be discovered that the scattered light is invisible from two positions around the jar, positions which are on opposite sides of the jar and lie directly along the line which represents the direction of the vibrations of the incident light. This will be understood if it is remembered that only longitudinal waves may be produced by vibrations which are occurring in the same direction as that in which the waves are traveling. A simple way to picture what takes place is to consider the tiny rosin particles to be set into vibration by the incident waves which pass over them. The scattered light is then considered to be emitted as a result of these vibrations. According to this picture, both the rosin particles and the scattered waves would be vibrating in the direction of vibration of the incident light waves, and this consequently would be the direction of the line along which no light would be seen.

This is most direct evidence that light waves are transverse waves only.

49. Polarization by Reflection. A much more prosaic way to polarize light is to reflect it from the surface of a clean piece of window glass at an angle of incidence of about 56 degrees. If the light reflected at this angle is passed through an analyzer, such as a Nicol prism or a Polaroid plate, it will be found that it is polarized with its vibrations parallel to the surface of the glass. Likewise a clean surface of still water will polarize light incident upon it at an angle of about 53 degrees. In general, light will be polarized by reflection from the surface of a transparent substance when the angle of incidence is such that the reflected ray and the refracted ray are perpendicular to each other, as is shown in Fig. 33. This is Brewster's law. It may be expressed mathematically by $\tan\theta = n$, where n is the index of refraction for the transparent substance, and the angle θ, called the *polarizing angle*, is the angle represented by i in Fig. 33.

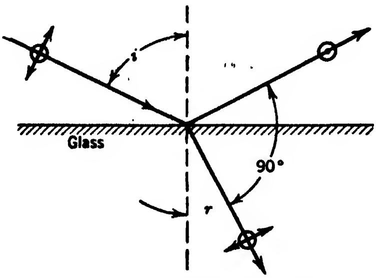
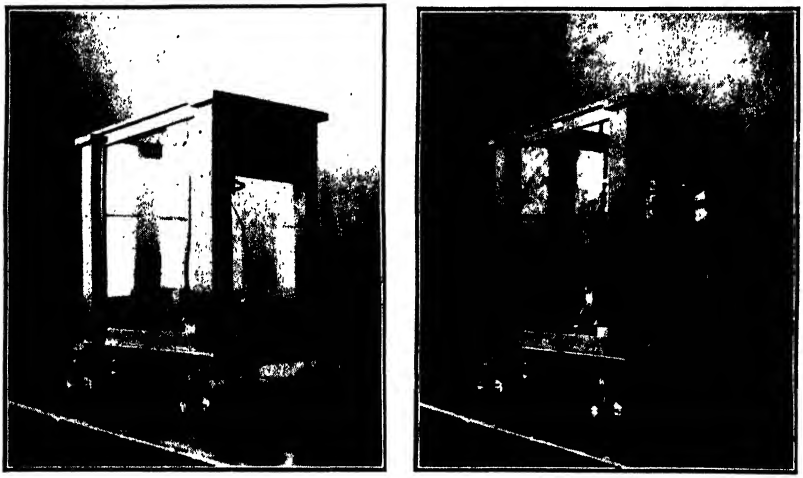


FIG. 33. BREWSTER'S LAW.

Since considerably less than half the light is reflected at these angles, the refracted beam of light is only partially polarized, in a direction at right angles to the polarization of the reflected beam. At other angles than the polarizing angle there is still a partial polarization of the reflected light; indeed, most reflected light, other than that reflected by metallic surfaces, is partly polarized. This accounts for the effectiveness of polarizing devices in reducing glare from water, wet streets, glossy paper, painted or varnished surfaces, etc. The polarizer may be used in front of the eye, turned to cut out the glare; or in front of the lamp, turned to polarize the light falling upon the surface in such a



(*a*) (*b*)

FIG. 34. (*a*) Photographed under ordinary conditions. (*b*) Photographed through a Polaroid plate with its axis horizontal. Note that the reflections are eliminated most completely in the middle of the picture. Why is this?

direction that a minimum of it is reflected. Likewise annoying reflections from show windows may be effectively eliminated from photographs of their contents if a suitable polarizer is placed in front of the camera lens and the window viewed at approximately the polarizing angle with the glass. This is illustrated by Fig. 34.

50. Iceland Spar. Iceland spar is a very pure form of calcite $(CaCO)_3$ which cleaves into rhombohedric crystals, that is, crystals in which opposite faces are parallel to each other, but adjacent faces intersect at oblique angles, as is shown in Figs. 35 and 36. Each face is an oblique parallelogram, as seen in the figures. Everything viewed through such a crystal appears doubled, as does the letter *T* in Fig. 36. Figure 35, which is a cross-section through the crystal along the line *a–c*, shows how this double image is produced. A narrow beam of light, *m–n*, incident normally upon the lower surface of the crystal, emerges as two separate but parallel beams, *o–p* and *q–r*. One of these, *o–p*, is a continuation of the incident beam, just as it would be for a thick plate of glass; it is called the "ordinary" beam. The other, or "extraordinary," beam *q–r*, although parallel to the first after emergence, "side-slips" through the crystal as shown, in a direction determined by the crystal. It follows that, as the crystal is rotated about the incident ray as an axis, the image formed by the ordinary beam remains fixed in position, but the image produced by the extraordinary beam rotates with the crystal. The two beams are equally intense, and examination with a polarizing plate shows that *both* are polarized, the directions of vibration being at right angles to each other, as shown in Fig. 35. **The incident beam of unpolarized light has been resolved by the Iceland spar into two equal plane-polarized beams with vibrations at right angles to each other.**

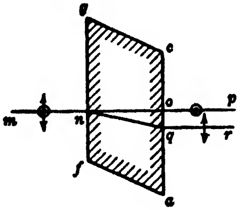


FIG. 35. CROSS-SECTION THROUGH ICELAND SPAR (Fig. 36) ALONG *a–c–g*.

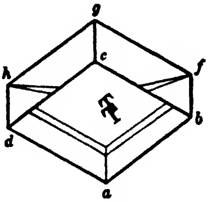


FIG. 36. ICELAND SPAR.

51. Nicol Prism. Various compound polarizing prisms, constructed from doubly refracting materials (Iceland spar or quartz), are described in the books on light. The best known of these, the Nicol prism, is represented by Fig. 37: (i) is a cross-section, (ii) an end view. It consists of two odd-shaped prisms of Iceland spar, *abd* and *cdb*, cemented together along the faces *bd* by a thin film of Canada balsam. The resultant prism has the shape of a long natural Iceland spar crystal and, except for practical difficulties, could have been made by cutting a long spar crystal in two along *bd* and cementing it together again. Its operation depends upon the curious fact that, although the index of refraction in the spar for the ordinary beam is greater than it is in the balsam, for the extraordinary beam the index is the same in both the spar and the balsam, so that the extraordinary beam goes straight through the prism as if it were a single piece of spar. The ordinary beam, on the other hand, strikes the spar-balsam surface at an angle greater than the critical angle and is *totally reflected*, as is shown. The emergent beam is then the extraordinary beam, and is vibrating in the direction indicated.

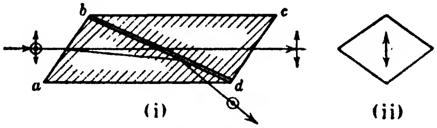


FIG. 37. NICOL PRISM.

These polarizing prisms polarize light more perfectly than do any other polarizing devices, but they are quite expensive and are not now obtainable with apertures much beyond a centimeter.

52. Double Refraction. A full discussion of the optical properties of doubly refracting materials such as Iceland spar is outside the scope of this book; only some of the simpler properties, especially those which have some practical applications, will be included here. Although Iceland spar is the best-known example of substances showing strong double refraction, there are many other doubly refracting substances. $NaNO_3$, for example, is even more strongly doubly refracting than Iceland spar. Quartz is less strongly doubly refracting; but, because its surface is hard and will retain a high polish, it is often used in making polarizing prisms and other optical parts for polarization apparatus. Many substances which are apparently not doubly refracting, in that they do not show double images, may be shown to be slightly doubly refracting by means of what is essentially an interference phenomenon which will now be described. The instrument used for this purpose is called a polariscope.

53. Polariscope. The polariscope consists of two polarizing devices, such as Nicol prisms or Polaroid plates, placed far enough apart that the optical substances being studied may be placed between them. The first of these devices serves to polarize the light before it passes through the specimen under study and is called the polarizer. The second, called the analyzer, serves to detect any change in the beam of light which may be produced by passage through the specimen. One or both of these polarizing devices may be rotated about the axis of the beam. The analyzer and polarizer are said to be "crossed" if they are turned so that the light which passes through the polarizer is cut off by the analyzer, and "parallel" if set to the position of maximum brightness. A very simple polariscope is shown in Fig. 38.

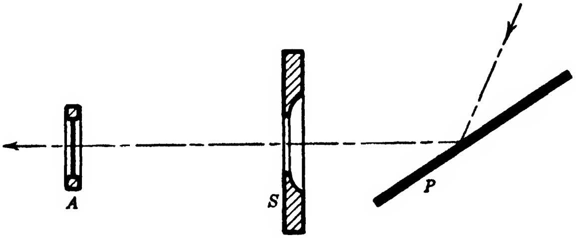


FIG. 38. A SIMPLE POLARISCOPE. The polarizer, *P*, is a plate of black glass which reflects light at the polarizing angle. The analyzer, *A*, is a polarizing plate. The specimen is placed on the stage, *S*.

Consider now a plate of Iceland spar placed in the polariscope, with the analyzer removed. In general, the plane-polarized light from the polarizer will be split into two beams, as is ordinary light. But these two beams will not generally be of equal intensity. Two positions now may be found, by turning the spar plate about the axis of the instrument, such that only one beam is visible, the other having vanished. These two directions, which are at right angles to each other, will here be called the "vibration-axes" of the crystal plate.[3] In one of these positions the vibrations in the plate are in the direction indicated by *a–a* in Fig. 39, and the beam is an "extraordinary" beam. The other position corresponds to vibrations at right angles to the first ones, as shown by *b–b*, and the beam is an "ordinary" beam. If the incident light is vibrating in any other direction, such as *c–c*, it will, upon entering the plate, split into two beams having vibrations in the directions of the two vibration-axes, as shown, the amplitude of these vibrations being the vector components in these directions of the amplitude of the original vibrations.

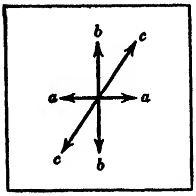


FIG. 39.

---

[3] These should not be confused with what are called the optical axes of a crystalline substance in books on optics, to which they are related in a manner which is, in general, rather complex.

If a plate of material other than Iceland spar is used, the double refraction may not be detected by the displacement of the extraordinary beam; but if the beam of light which has passed through the crystal plate is examined through the analyzer, it becomes very apparent. Unless the vibrations of the light from the polarizer coincide with the vibration-axes of the plate, the plate as viewed through the analyzer will probably appear brightly colored, the color changing to its complementary when the analyzer is turned from the "crossed" to the "parallel" position. It is easiest to see what takes place here if we consider first a crystal plate of a special thickness, called a "half-wave" plate, and monochromatic light.

54. Half-Wave Plate. The ordinary and extraordinary beams travel with different speeds in a doubly refracting substance, and in consequence, even though they may travel the same path, the waves of these two beams will get *out of step* as they travel through the crystal plate. A half-wave plate is one of such a thickness that the two beams get out of step by just half a wavelength (of the light used) as they pass through it. Such a plate may be made by splitting a sheet of clear mica to the requisite thickness. It is quite thin and is usually mounted between glass plates for protection.

This difference in phase is revealed by interference between the two beams. Strictly speaking, the waves of these beams cannot directly interfere with each other, since they are vibrating at right angles to each other. But if their components in any one direction are taken, and this is exactly what is done by the analyzer, then those components can interfere in the usual manner. Suppose the polarizer to be set so that the waves passing through it are vibrating in a vertical direction. If the half-wave plate is placed in the polariscope with its vibration-axes set at 45 degrees to the vertical, upon entering the half-wave plate the vertical vibrations will be resolved into two equal vibrations in the directions of the vibration-axes. This is indicated in Fig. 40, which shows the incident wave, *i*, entering the back of the crystal plate (here enlarged so as to appear as a thick block) and being resolved into two waves, an *ordinary* wave, *o*, and an *extraordinary* wave, *e*.[4]

Since these waves travel with different speeds through the crystal, the ordinary wave falls a half-wavelength behind the extraordinary wave, on the way through the plate. Consequently, when they emerge

---

[4] It is not necessary to know which is the ordinary beam, and which the extraordinary one; the essential fact is that there are two beams, polarized at right angles to each other and traveling at slightly different speeds. In some substances the extraordinary beam travels faster, as has been assumed in this case; in other substances, the ordinary beam may have the greater speed.

from the front surface and recombine to form a single wave again, this resultant wave is vibrating in a *horizontal* instead of a vertical direction. (In the actual crystal plate, the difference in the speeds of these two waves is much less, and the number of waves much greater, than is shown in the figure, which is exaggerated for clearness of representation.) Thus, if the analyzer is "crossed" or set to extinguish the light from the polarizer before the half-wave plate is inserted, it will transmit a maximum of the light which has passed through the half-wave plate. And in the "parallel" position it will cut out the light after it has passed through the half-wave plate.

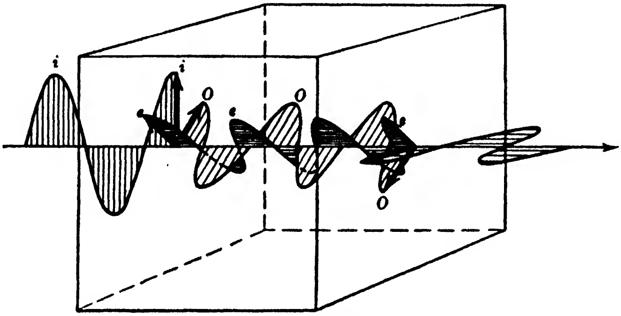


FIG. 40. HALF-WAVE PLATE.

A plate twice as thick as this would be a full-wave plate, and the resultant vibrations in the beam of light which passes through it will be in the same direction as before entering it. Plates producing retardations, or differences in path between the two component waves, of 3/2, 5/2, ··· wavelengths of monochromatic light would produce the same result in the polariscope as does the half-wave plate, while retardations of 2, 3, 4, ··· wavelengths would be identical with the whole-wave plate in effect in the polariscope. In white light, quite different results are obtained. Strictly speaking, a half-wave plate is such for one wavelength only, and the same is true for any other retardation plate. The difference is not so great for a half-wave plate, which may be made of a thickness that will function reasonably well for white light. In a thicker plate, however, the retardation may be a whole number of wavelengths for one end of the spectrum, and a half-wave more (or less) for the other end. As a result such a plate will appear of one color with the polarizer and analyzer parallel, and of the complementary color when they are crossed, as has been mentioned earlier.

55. Quarter-Wave Plate; Circularly and Elliptically Polarized Light. The above discussion of the polarization colors seen in a crystal plate be plane polarized, but its direction of vibration will have been turned, as is shown schematically by Fig. 42, through an angle which is found experimentally to be proportional to the concentration of the sugar (grams per cubic centimeter) and to the distance the light travels through it, i.e., to the thickness of the solution. The proportionality factor is called the **specific rotation,** and the experimental law may then be expressed by

Angle of rotation = specific rotation × concentration × thickness

The specific rotation depends upon the wavelength, increasing rapidly as the wavelength decreases, so that white light is twisted into a beam that presents different colors as the analyzer is rotated. In ordinary sugar the rotation is positive, or clockwise to an observer looking towards the source of light. Many substances are similarly *optically active,* some of them producing positive rotation and some negative or counterclockwise rotation. The polarimeter, an instrument devised to measure accurately such rotations, is an important tool in the chemical and biological laboratory. It consists of, first, a polarizer to polarize the incident light, next the tube containing the liquid under test, and then the analyzer to detect the amount of rotation. Various other parts are added to increase the precision of setting; these are described in references given at the end of the chapter.

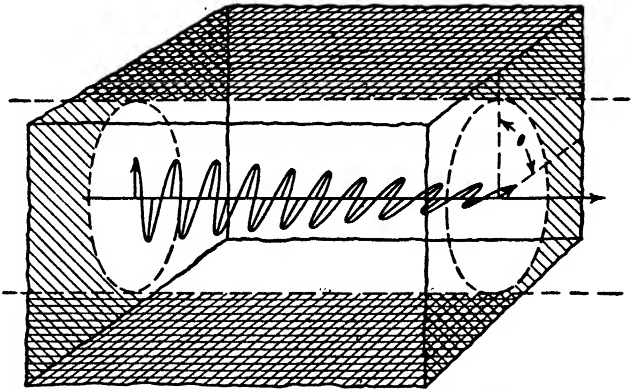


FIG. 42. ROTATION OF THE DIRECTION OF VIBRATION IN A BEAM OF POLARIZED LIGHT.

58. Strain Testing; Photoelasticity. Glass, celluloid, Bakelite, and other materials which are not normally doubly refracting become slightly so when strained mechanically or thermally; such strained regions appear bright, or even colored if the strain is large, when viewed between the crossed polarizer and analyzer of a polariscope. This is an extraordinarily sensitive test for strain and has many practical applications. It has long been employed as a very sensitive means for detecting residual strains left in chemical and optical glassware by too rapid cooling from the fabricating temperature (imperfect annealing). More recently it has served as a means of analyzing mechanical stresses in engineering structures, such as bridges and gear wheels, whose complex form makes ordinary computations and tests difficult if not impossible. A model of the structure is made out of celluloid, transparent Bakelite, or glass, and the stresses produced under conditions of loading which represent the normal use of the structure are studied by examining the model in a suitable polariscope. Quantitative measurements may also be made by elaborations of this simple *photoelastic* method of testing. Figure 43 shows in this manner the stresses produced in a bent glass beam.

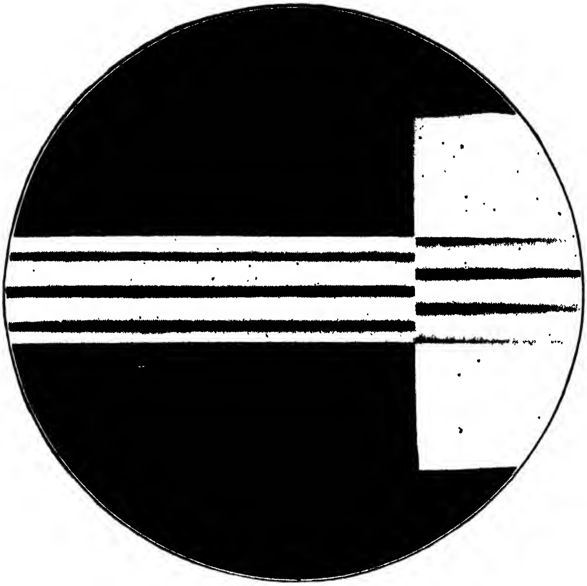


FIG. 43. GLASS BEAM UNDER BENDING STRESSES, BETWEEN CROSSED-POLAROID PLATES. The retardation is zero along the axis of the beam, which consequently appears as a dark horizontal line. The two other dark lines or bands occur where the stress-induced retardations are one wavelength. A half-wave plate of Cellophane covers the right-hand end of the beam: Note that it restores the background to full brightness, and shifts the positions of all the retardation bands.

59. Faraday Effect. Conditions of magnetic and of electric stress will also modify the optical properties of materials. The first discovery of a phenomenon of this kind was made by Faraday, who observed that very dense glass, when placed in a strong magnetic field, became optically active for light passing through it in the direction of the field. Since then many other substances have been found to exhibit this "Faraday effect" to some degree. It may be demonstrated with a tube of carbon disulfide placed along the axis of a large coil of wire through which a strong electric current is flowing. If plane-polarized light is passed through this tube (which is provided with flat glass ends) it will be found that the plane of polarization has been rotated by the magnetic field, the amount of rotation being proportional to the length of the tube and to the strength of the magnetic field, and the direction of rotation being, in this substance, in the direction of flow of the current around the coil. (In some substances the direction is the reverse of this.)

This optical activity, which resembles that of a sugar solution, differs from it in one significant way: In optical activity which occurs naturally, as in a sugar solution, the direction of rotation is determined by the direction of passage of the light, whereas this magneto-optical rotation occurs in a direction determined by the direction of the magnetic field. This experiment is of importance as indicating some fundamental connection between the magnetic properties of materials and their optical properties.

60. Kerr Effect. A phenomenon which indicates that there is also a fundamental connection between the electrical and the optical properties of substances was discovered by Kerr, who found that a piece of glass, placed in a strong electric field, becomes very slightly doubly refracting. Since this effect is also exhibited by liquids (nitrobenzene shows it best), it must be due to electrical stress. Figure 44 shows an arrangement by which this effect may be demonstrated. Two parallel metal plates, *C* and *D*, are immersed in a glass tank, *T*, containing nitrobenzene; when the terminals of a high-voltage source are connected to these plates, a strong electric field is set up between them. This part of the apparatus is called a *Kerr cell.* The beam of light which passes between these plates is polarized at 45 degrees with the surfaces of the

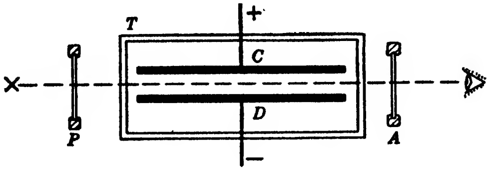


FIG. 44. KERR CELL.

plates by the polarizing plate P, and the second polarizing plate, A, is set so as to extinguish the light which comes through P when there is no field between the plates. When the electric field is set up between the plates, however, some light is passed by A. The effect is exactly the same as would be produced by a thin mica plate with one of its vibration-axes parallel to the electric field; the electric field has made the nitrobenzene doubly refracting to a slight degree.

The Kerr cell has had many practical applications as a *light valve*, because of the extreme rapidity at which it may be operated. An interesting example is its application to measurement of the speed of light.[5] A Kerr cell was "opened" and "closed" about 28 million times a second by means of a high-frequency radio oscillator, thus chopping the light beam which passed through it up into sections only a little more than 10 meters long. A very ingenious scheme was then used to time the travel of these sections of the beam, such that it was possible to measure the speed of light, inside the Harvard Physical Laboratory, to a precision probably as great as that of the measurement made over a distance of 25 miles by the rotating-mirror method.

PROBLEMS

1. When two Nicol prisms are placed in the "parallel" position, 50 per cent of the incident (unpolarized) light is transmitted through both of them; when they are "crossed," no light is transmitted. (*a*) How much light is transmitted when they are turned 45 degrees from the "crossed" position? (*b*) How much, when the angle is 30 degrees? Explain each case.

2. Prove that, when the tangent of the angle of incidence equals the index of refraction, the reflected and the refracted rays are perpendicular to each other (Brewster's law). (This may be done in general terms, or it may be computed for a special substance, e.g., glass having an index of refraction of 1.500.)

3. The experiment illustrated by Fig. 32 may be modified so that the path of the light through the jar will appear green from one direction, red from another. Figure out how this may be done.

4. (*a*) How may two quarter-wave plates be superimposed so as to serve as a half-wave plate? (*b*) If one of these plates is now turned through 90 degrees about the axis of the light beam, it completely neutralizes the effect of the other. Explain. (*c*) Describe how it is possible to turn one plate over (other side out) so as to produce the same result as that produced by turning it through 90 degrees about the axis of the beam.

5. Show that, if a half-wave plate is placed so that one of its vibration-axes makes an angle θ with the direction of vibration for incident plane-polarized light, the emergent light will also be plane-polarized, with the direction of its vibration rotated through an angle 2θ from the original direction. Draw vector diagrams to explain this. (Note that Fig. 40 shows the special case for which $\theta = 45$ degrees.)

---

[5] W. C. Anderson, *Review of Scientific Instruments*, Vol. 8, pp. 239–247, July, 1937.

6. The principal indices of refraction for quartz (for 5890A) are 1.544 for the ordinary ray, and 1.553 for the extraordinary ray. Compute the number of waves of this light in a quartz plate 0.100 mm thick for both the ordinary and the extraordinary beam. (This plate must be cut with its faces parallel to the "optic axis." If cut otherwise, the effective indices of refraction will not be the ones given above, for reasons explained in the books on optics.) Thence compute the "retardation," in wavelengths, of this light in this plate.

How thick is a quarter-wave plate for 5890A, made of quartz cut in this manner?

7. Plane-polarized sodium light is "rotated" through 66 degrees by passage through a 1-decimeter-thick layer of sugar containing 1 gram of sugar per cubic centimeter of solution. Another solution is made up by dissolving 50 grams of sugar in 100 cc of water. (This solution has a specific gravity of 1.143.) How much will sodium light be rotated by 40 cm of this solution?

REFERENCES

For Additional Reading

F. A. Saunders, *Survey of Physics*, Chapter 36.
P. I. Wold, *Kimball's College Physics*.
Edwin Edser, *Light for Students*.
J. K. Robertson, *Introduction to Physical Optics*.

For Advanced Reading

R. W. Wood, *Physical Optics*, Third Edition, Chapters IX, X, XVII, XXII.
Thomas Preston, *Theory of Light*.
R. A. Houstoun, *A Treatise on Light*.
A. C. Hardy and F. H. Perrin, *The Principles of Optics*, Chapter XXIX.

CHAPTER V

ELECTRICAL OSCILLATIONS AND ELECTROMAGNETIC WAVES

61. What Kind of Waves Are Light Waves? In previous chapters ample evidence has been presented for the existence of light waves, and evidence also that they are transverse waves. But they are waves traveling in what medium? Not in any material medium, certainly, since they can travel through space void of any material substance, such as the space between us and the sun, and between us and all the stars which send their light to us. If, however, this medium is not a material one, *what* kind of vibrations are these light vibrations which we have been talking about? We shall find answers to these questions in still another field of physics, namely, the field of electricity and magnetism.

62. Electrical Oscillations. First we must consider electrical oscillations, such as may be produced by charging a condenser up to a fairly high voltage and then allowing it to discharge through a coil of wire.

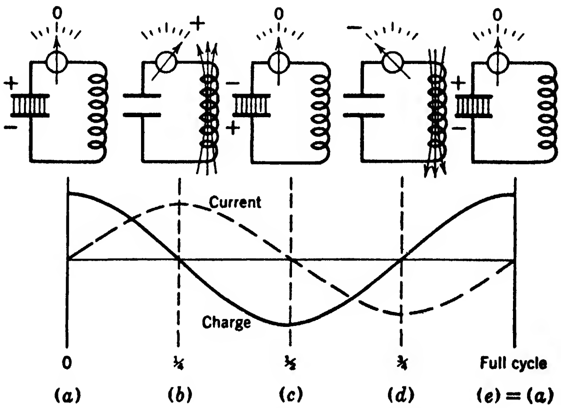


FIG. 45. ELECTRICAL OSCILLATIONS.

One cycle of such an electrical oscillation is represented in Fig. 45. At the instant that the condenser is fully charged the current is zero. This corresponds to the point (*a*) on the oscillogram and is also represented by the diagram directly above. As the condenser discharges, the current builds up to bigger and bigger values; when the condenser is fully discharged, as represented at (*b*), the current has reached its highest value. The current continues in the same direction, although its value is now decreasing, and does not cease until the condenser is again charged up, this time in the reverse direction, as shown by (*c*). If the resistance of the circuit is very low, this charge will be nearly as great in magnitude as the initial charge at (*a*). The process now proceeds as it did from (*a*), but in the reverse direction, as shown by (*d*) and (*e*), which brings it back to the initial conditions, ready to repeat the process all over again.

In this process the magnetic field produced by the current plays an essential part. When the current begins to build up, the magnetic field also builds up, and in doing so it cuts across the turns of the coil and induces in it an e.m.f. which *opposes the increase* (Lenz's law); this slows up the discharge. When, however, the condenser is entirely discharged, the current and hence the magnetic field are at their maximum values; they then both begin to decrease. The lines of magnetic force now cut across the turns of the coil in the opposite direction, and the induced e.m.f. now *keeps the current flowing* (with ever-decreasing value) until the condenser is again charged up.

From the above discussion it is easy to see that the period, P, or the time for one complete oscillation, depends upon the capacitance, C, of the condenser and the inductance, L, of the coil. It is proved in Appendix IV that

$$P \text{ (seconds)} = 2\pi\sqrt{L \text{ (henries)} \times C \text{ (farads)}}$$

Another way to look at these oscillations is from the point of view of the energy involved. At (*a*), (*c*), and (*e*), all the energy is stored in the condenser, as *electrostatic* energy, whereas at (*b*) and (*d*) it is stored in the region around the coil as *magnetic* energy. As the circuit oscillates, the energy is transformed from electrostatic energy into magnetic energy and back again time after time. If the resistance of the circuit were zero, this would keep up indefinitely; but if the resistance is not zero, but small, a little energy will be dissipated as heat in the resistance in each oscillation, until finally the oscillations cease. For example, if $L = 200$ microhenries and $C = 500$ micro-microfarads, then $P = 2$ microseconds, corresponding to 500,000 oscillations per second. If now the resistance equals 5 ohms, these oscillations will die down to 1 per cent of their initial value in about 0.0004 second, or after about 200 oscillations have occurred.

63. Electron-Tube Oscillators. The oldest and simplest way to set such a circuit into oscillation is shown in Fig. 46. The spark gap acts as an automatic switch which stays *open* until the transformer has charged the condenser up to a high enough potential difference to break down the gap. Once the gap is broken down, the hot gases between its terminals serve as a *closed* switch until the oscillations die down to a small value. The process is repeated each half-cycle of the power supply, or 120 times a second if 60-cycle a-c. power is used. Since the oscillations die away in a minute fraction of a second, the operation of such a circuit consists of a succession of isolated groups of oscillations, each one of which dies away before the next one starts.

The amplifying action of a three-electrode electron tube (Sec. 14) may be used to prevent such electrical oscillations from dying out. The electron-tube circuit, which enables the battery or other d-c. power supply to maintain steady electrical oscillations, may be compared to the mechanism of a clock, which enables the steady pull of a spring to maintain the mechanical oscillations of a pendulum. Figure 47 shows two suitable circuits. It is easier to explain circuit (*a*). Electrical oscillations take place in the part of the circuit formed by the coil, *L*, and the condenser, *C*, the rest of the circuit forming the driving circuit. When the oscillations are in such a direction as to make the grid of the tube, *G*, positive, then the current to the anode, *P*, increases as has been explained in Sec. 14. This increasing current, flowing in the coil *L'*, induces in *L* (through the interaction of its magnetic field with *L*) an e.m.f. which maintains the current in *L* a little longer and puts a little more charge into *C*. This process repeats itself cycle after cycle until the oscillations have been built up to a large, steady value. When this steady state is reached, the battery is supplying to the oscillation circuit just enough power to make up for the power lost because of the resistance in that circuit.

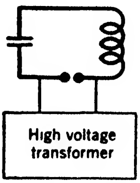
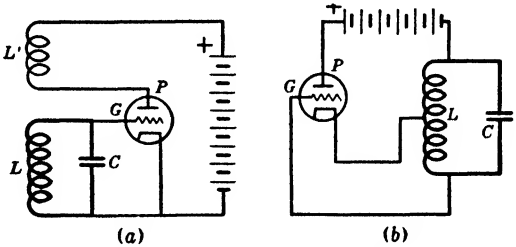


Fig. 46. Spark Oscillator.



(*a*) (*b*)

Fig. 47. Electron-Tube Oscillators.

Both of these circuits are simplified, additional details (which are explained in technical books) being required for efficient operation.

64. Electrical Resonance. Electrical oscillations occurring in one circuit, which will be called the *primary* circuit, will set up oscillations in another, or *secondary* circuit, even when the secondary circuit is some distance away, provided that the secondary circuit is tuned to the *same frequency* as the primary. This may be illustrated experimentally, driving the primary oscillator with an electron tube, and using a small electric lamp in series with the coil and condenser of the secondary circuit as a current indicator. See Fig. 48. As the capacitance of the secondary condenser is varied, a value will be found for which the lamp will glow brightly, although a slight turn of the condenser dial to either side of this setting will dim it to invisibility. The condenser dial of such a secondary circuit may be calibrated in terms of frequency, so that it may be used as a frequency meter.

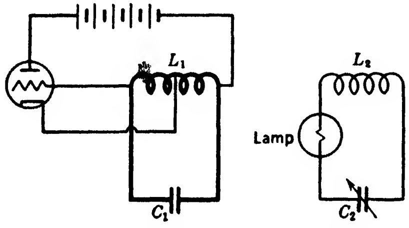


FIG. 48. COUPLED CIRCUITS.

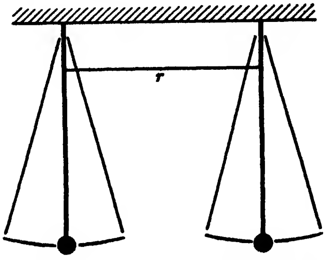


FIG. 49. COUPLED PENDULUMS.

The principle here involved is illustrated also by two tuning forks tuned to the same frequency and placed near each other, and by two pendulums of the same length, coupled together near their points of support by means of a weak rubber cord, indicated by r in Fig. 49. In the pendulum system, each swing of the primary pendulum transmits through the rubber cord a weak pull to the secondary pendulum. If the two pendulums are not of the same length, then these weak pulls will have little effect upon the secondary pendulum. If, on the contrary, they are of the same length, each feeble impulse will occur at just the right time to add its effect to those of all the other impulses, and a large swing of the secondary pendulum will result. In the electrical system the weak impulses are furnished by the magnetic field of the primary current, a small part of which cuts back and forth across the secondary coil. Here, also, the secondary current starts at a low value and builds up cycle by cycle to a high value. This principle of *resonance* is of fundamental importance in all high-frequency electrical circuits.

65. High-Frequency Heating. When a piece of metal is placed inside a coil in which an alternating current is flowing, it is heated by the currents which the alternating magnetic field induces in it. Such currents are called "eddy" currents. If enough power is available, the metal may be heated to incandescence or even melted; such an *induction furnace* is particularly useful where it is desirable to keep the metal inclosed, or even to keep it in a vacuum. Since this method of heating is much more effective at high frequencies, most induction furnaces utilize high-frequency oscillators of one kind or another. A very important use for induction furnaces is in the production of electron tubes. In the process of evacuation of such tubes it is necessary to heat the metal parts to incandescence to drive out all occluded gases and vapors, and this is accomplished most readily with an induction furnace.

This high-frequency electromagnetic induction will heat any substance which is a good enough conductor, and dozens of industrial applications exist for this kind of heating. For example, paint on metal surfaces may be dried quickly if the metal is warmed by high-frequency induction. If, on the other hand, the substance is a poor conductor, it may be heated with high-frequency power by placing it in the *electric* field between the plates of a large condenser which is supplied with high-frequency voltage. In this manner wood may be seasoned, glue dried, rubber cured, and even bread baked. Precooked foods which have been kept in cold-storage may be made piping hot in a minute or less. This method is unusually efficient because it concentrates the heat where the moisture is greatest. Since most oscillators employed for these purposes are driven by electron tubes, this is often called "electronic" heating.

66. Physiological Effects; Diathermy. The physiological effects of high-frequency electric currents are quite different from those produced by direct currents or by low-frequency alternating currents. Currents large enough to light commercial lamps may flow to electrodes held in the hand or otherwise connected to the body, not only without any fatal results, but even without any sensible results whatsoever unless the areas of contact are quite small. When very high frequencies are used, an appreciable rise of body temperature may be produced, and the utilization of electronic and other high-frequency equipment for such "diathermic" treatments has had a considerable development.

67. Electromagnetic Waves. In 1864 James Clerk Maxwell made public a new theory for electric and magnetic fields, a theory which was considered at the time to be as revolutionary as the relativity theory has been considered in more recent times. Maxwell's theory is still as fundamentally important to present-day physics as it was when it first appeared, although time and the development of scientific thought have brought it into the realm of the commonplace or "classical" physical theory, and recent discoveries have revealed its limitations and have modified its interpretations. One of the most significant predictions of this theory was that both electric and magnetic fields are propagated as waves. Thus, if an electric charge is suddenly placed upon a conductor, the electric field corresponding to it does not come into existence at all points simultaneously; instead, it begins at the conductor and spreads outward at a very high but finite speed. The same is true for a magnetic field produced by the sudden starting of a current in a wire. Furthermore, while the electric field is moving outward, it is always accompanied by a magnetic field which it sets up by its own motion, and a moving magnetic field likewise sets up an electric field. The *two moving fields together* constitute an *electromagnetic wave*, and Maxwell predicted that its speed would be equal to the ratio of the size of the electromagnetic unit of electric charge to that of the electrostatic unit of charge, or approximately 3×10^{10} cm/sec. The identity of this value with the value found experimentally for the speed of light (the agreement is within 1 part in 30,000) cannot be ignored, but consideration of its significance will be postponed to the next chapter.

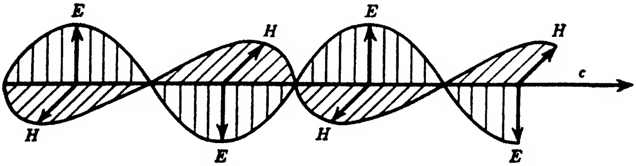


FIG. 50. ELECTROMAGNETIC WAVES. *E*—Electric field. *H*—Magnetic field.

According to this theory, electrical oscillations such as those discussed in Sec. 62 should produce continuous waves like those represented by Fig. 50. The vertical wave plot represents the electrical field, and the horizontal plot, the magnetic field, as they vary from place to place. The wave is represented as traveling towards the right, in the direction of the arrow, so that at any particular place the electric field alternates between up and down values as the wave passes through it, and the magnetic field alternates similarly in its direction perpendicular to that of the electric field. Whatever the direction of the electric field, the magnetic field is always perpendicular to it, and in such a sense that an imaginary rotation of the vector representing the electric field towards the one representing the magnetic field would move a right-hand screw in the direction in which the wave is traveling. Both fields are always exactly in phase.

68. Hertz's Experiments. All these predictions were purely theoretical, and it was not until twenty-three years later that Heinrich Hertz devised apparatus to produce and to detect such electromagnetic waves and thus to verify experimentally Maxwell's predictions and the theory upon which they were based. Hertz's original experiments have been described in many books; the experiments to be described here follow them in their essential characteristics, but utilize modern electronic equipment.

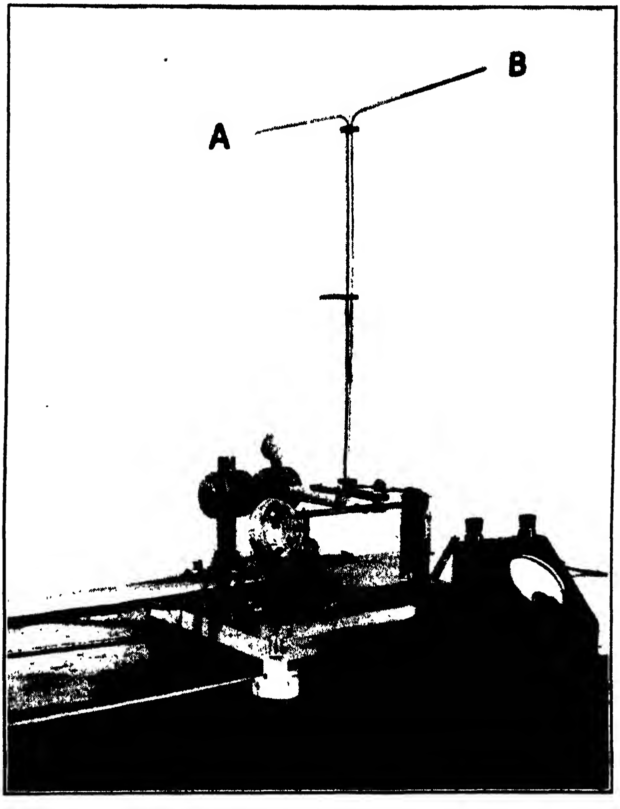


FIG. 51. 400-MEGACYCLE OSCILLATOR. (Western Electric 316-A "doorknob" tube used in the driving circuit.)

First of all, it is evident that waves short enough to study conveniently must be of very high frequency. Hertz first used waves about 25 feet long, corresponding to a frequency of about 40 megacycles (40×10^6 cycles) per second. Later he used waves as short as 2 feet, having a frequency in the neighborhood of 500 megacycles. Figure 51 shows a photograph of an oscillator of modern design which will radiate waves of about this same length. The problem which Hertz had to solve was that of designing an electric oscillator which would operate at such high frequencies and would also be an efficient radiator of waves; he solved this problem by using as an oscillator a straight rod with a spark gap at its middle. This rod corresponds to the horizontal portion, $A-B$, of the transmitter shown in Fig. 51. In this modern apparatus, however, the spark gap is replaced by more complicated but more effective electronic equipment for setting up and maintaining the oscillations.

Although this straight-rod oscillator does not at all resemble the coil and condenser type of oscillator, its electrical behavior is equivalent. Electric charges (in the form of electrons) flow back and forth along the rod, such that alternating currents appear at the central portions of it, and alternating accumulations of electric charge appear at its ends. The current is greatest at the center and zero at the ends, as is represented by the dotted-line plot in Fig. 52(*a*); the magnetic field produced by this current is represented by the circular lines of force also shown in this figure. The distributed inductance corresponding to this magnetic field is the inductance of this circuit. At the instant here represented the current is a maximum, and the charge is everywhere zero. A quarter of a cycle later the current is everywhere zero, and the accumulation of charge is a maximum, its distribution then being as shown by Fig. 52(*b*). The density of charge is greatest at the ends, being positive at one end, negative at the other. The figure also shows the lines of electric force due to this distribution of charge. Although no condenser appears in this circuit, the necessary capacitance is supplied by the distributed capacitance between the two ends of the rod. Both the inductance and the capacitance of this circuit change with the length of the rod, so that such an oscillator is tuned by simply changing its length. As a matter of fact, its length is close to half the length of the waves which it radiates.

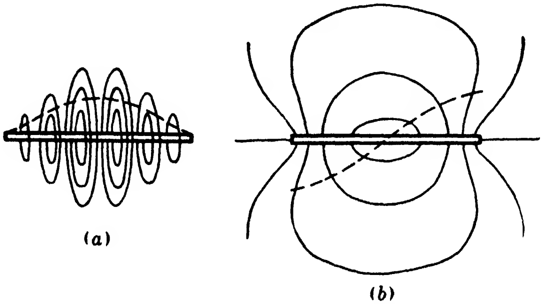


FIG. 52. (*a*) Magnetic field, and (*b*) electric field, around a half-wave antenna.

The electrical oscillator described above constitutes a short-wave radio transmitter. As a receiver circuit, we may use a second rod of equivalent length. If this receiver rod is held parallel to and some distance away from the transmitter rod, electrical oscillations will be set up in it, and since the two rods are tuned to resonance, the oscillatory current in the receiver may be large enough to light a little lamp inserted at the middle of the receiver rod. (This is done by cutting the rod at the middle and connecting the terminals of the lamp to the cut ends.[1]) This, of course, does not prove the existence of waves, but only that oscillations are taking place—the effect upon the receiver may be due to ordinary electromagnetic induction.

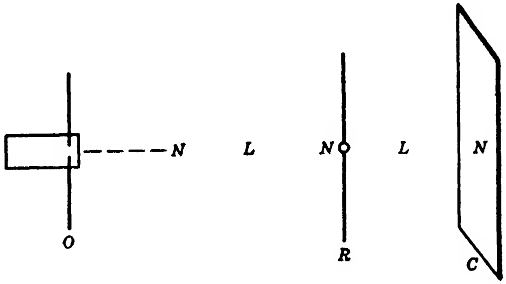


FIG. 53. STANDING ELECTROMAGNETIC WAVES. O—oscillator antenna. C—copper sheet reflector. R—receiver antenna.

To prove the existence of waves, some *interference* experiment must be demonstrated, and this may be accomplished by reflecting the waves, if waves there be, back upon themselves by means of a suitable mirror. A large sheet of copper placed a considerable distance away from the transmitter along the line perpendicular to the transmitter rod, as shown in Fig. 53, serves admirably. If now the receiver rod is held parallel to the transmitter rod and carried along the line between the copper reflector and the transmitter, a regularly spaced sequence of nodal points, N, will be found at which the lamp is dim or dark, while at the points L, midway between the nodal points, there will be found antinodes or loops at which the lamp will be extra bright.

The standing waves which are thus demonstrated by this experiment both prove the existence of electromagnetic waves and provide a means for measuring their wavelength. Hertz also showed, by measuring the frequency of the oscillations in his apparatus, that the speed of his waves was approximately the speed predicted by Maxwell. This has since been verified, with very high precision, for radio waves.

---

[1] This receiver circuit is shown lying on the table in front of the transmitter in Fig. 51.

69. The Nature of Electromagnetic Waves. If the receiver rod, instead of being held parallel to the transmitter rod, is held at right angles to that position, the lamp will not light up. Evidently the waves are *polarized,* just as is shown in Fig. 50, which represents the predictions of Maxwell's theory. When the receiver rod is parallel to the transmitter rod, the electric-field component of the waves, which is itself parallel to the transmitter rod, pushes back and forth upon the electrons in the receiver rod, setting up an oscillatory current in the rod. When the receiver rod is perpendicular to the electric field, this cannot occur and no current flows. A polarizing mirror may be made by covering a wood frame with a grid of many parallel wires. If the wires are parallel to the electric field, the grid will reflect the waves; but if they are perpendicular to the field, the waves will pass through the grid without reflection. In this respect the grid somewhat resembles a tourmaline crystal or a Polaroid plate.

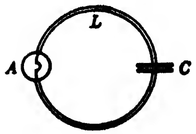


FIG. 54. LOOP ANTENNA.

As has been implied, the rod receiver is primarily a detector for the *electric* field. By using the "loop" receiver shown in Fig. 54, which responds primarily to the *magnetic* field, we may repeat Hertz's demonstration of the existence of the magnetic-field component of these waves. This loop receiver consists of a single turn of heavy wire, L, tuned to resonance by the small condenser C, together with some indicating device such as the lamp A. If this loop receiver is held so that its plane is perpendicular to the magnetic-field component of the waves, it will be found to respond most strongly at the points N (Fig. 53) at which the rod receiver showed the least response, and least at the points L where the rod receiver responded most strongly. This may be explained as follows:

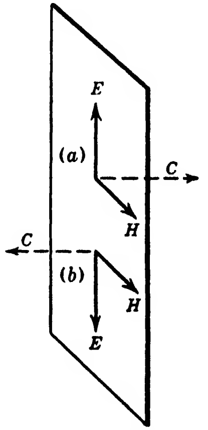


FIG. 55. REFLECTION OF AN ELECTROMAGNETIC WAVE AT A COPPER "MIRROR." (a) Incident wave. (b) Reflected wave.

When the wave is reflected at the copper mirror, the electrical component of the reflected wave is reversed so as to make the resultant electric field equal to zero at the surface of the mirror. Since the reflected wave is traveling in the reverse direction, the magnetic components for both incident and reflected waves must have the same direction at the surface of the mirror as shown in Fig. 55, and this makes the resultant magnetic field have a maximum value there, and at every other point labelled N in Fig. 53. The nodes for the electric field are thus antinodes for the magnetic field, and the current in the loop receiver at these antinodes is caused by the e.m.f. induced in it by the rapidly fluctuating magnetic field passing through it at these points. Thus all of Maxwell's predictions are completely verified.

In the electromagnetic wave itself (that is, in a vacuum) *nothing moves*. It is merely a wave of varying electric and magnetic force-fields traveling through space. As this wave passes over or through material substance, these oscillating fields will set into motion any charged particles within that substance. If, for example, the substance is a conductor, the free electrons within it will be set into motion and an oscillatory current will result, as has been described for the receiver rod. Likewise currents are set flowing in the copper sheet reflector, and this accounts for its ability to reflect the waves. When the substance is an insulator, the electrons are elastically bound so that they can move only very slightly as the waves pass through. These motions are enough, however, to react upon the waves so as to alter their speed through the substance—usually to slow them down.

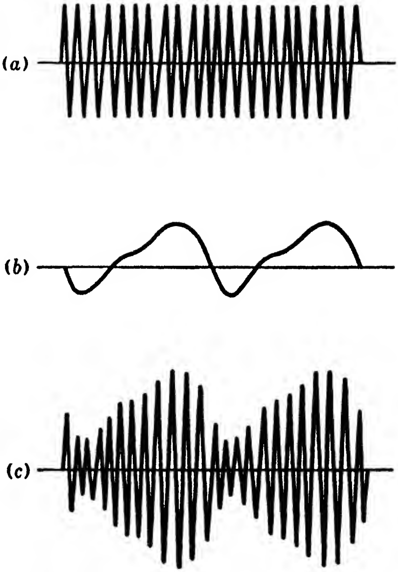


FIG. 56. AMPLITUDE MODULATION OF RADIO WAVES.

70. Radio Telephone. The use of electromagnetic waves for communication requires no introduction. For radio telegraphy all that need be done is to interrupt the transmission of waves at intervals representing the dots and dashes of the international Morse code.[2] For telephony, however, it is necessary to "modulate" the waves; the commonest practice is to vary their *amplitude* at a rate and to a degree determined by the frequency and the loudness of the sound to be transmitted. The oscillograms shown in Fig. 56 illustrate this modulation process. A steady radiofrequency oscillation would appear on an oscillograph screen as shown by (*a*), while a vibration of audible frequency might appear as shown by (*b*). If a radio wave of the frequency shown

---

[2] Even for telegraphy it is necessary further to modulate the waves in some fashion, either at the transmitter or at the receiver, in order to make the signals audible. This is explained in technical books on the subject.

by (*a*) is to carry the sound signal represented by (*b*), then the ***amplitude*** of its oscillations must vary in the manner shown by (*c*). Note that the crests of these radiofrequency oscillations, on both sides of the axis, outline a curve which is the same as (*b*). For example, WRC, in Washington, D. C., transmits at a frequency of 950 kilocycles. During every second of an interval of silence in its program it then sends out 950,000 waves of equal amplitude. In order to transmit a note of 500 vibrations per second, the amplitude of the waves must swing from maximum to minimum and back again 500 times per second, or once every 1900 waves. If the sound producing this note is feeble, the change in amplitude will be slight. Louder sounds will cause larger variations in amplitude, and the loudest sound which may be correctly transmitted will make the amplitude fall to zero value at the minimum (100 per cent modulation).

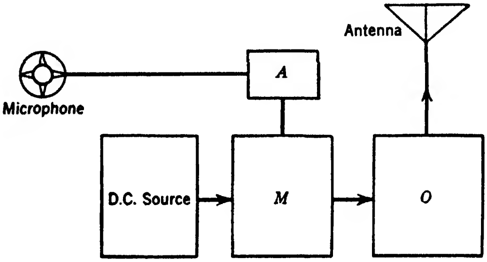


Fig. 57.

Figure 57 is a schematic diagram of a radio telephone transmitter. The oscillator, *O*, is a circuit similar to those shown in Fig. 47. It supplies the radiofrequency power, which is radiated from the antenna as radio waves. The d-c. power which is needed to operate this oscillator is supplied by a high-voltage transformer-rectifier circuit of the kind described in Sec. 12. This d-c. power flows to the oscillator through the modulator circuit, *M*, which functions as a kind of valve to control this flow of power in a manner that is determined by the currents produced in the microphone when sound waves fall upon it. These feeble microphone currents are amplified by the high-gain electron-tube amplifier, *A*, so as finally to produce voltages large enough to govern the flow of power through the modulator, and hence to vary the amplitude of the radio waves in the manner described above. The modulator itself is another electron-tube circuit, somewhat resembling a very powerful amplifier circuit, and the amplified voltages supplied to it by the amplifier, *A*, are applied to the grid of the modulator tube.

71. Radio Receivers. Figure 58 shows a diagram for a simple radio receiver. As the radio waves pass in succession across the antenna, A, they push up and down upon the electrons in that antenna, setting up a feeble current in the coil L'. This current then induces in the nearby coil L an e.m.f. which sets up a secondary current in the circuit formed by the coil L and the condenser C. If now this secondary circuit is tuned, by means of the variable condenser, so that its natural frequency is the same as the frequency of the radio waves, a comparatively large current will result. (Cf. Sec. 64.) The "detector" tube, T, the condenser, C', and the telephone, H, constitute a *rectifier* circuit precisely like the one shown in Fig. 9(a) and explained in Sec. 12. The solid lines in Fig. 59 represent the rectified current pulses through T when the modulated waves represented by Fig. 56(c) are received by the antenna, and the dotted line represents the filtered audiofrequency current in the telephone. It is evident, from comparison of this dotted line with Fig. 56(b), that the resultant vibrations of the telephone diaphragm will reproduce the original sound more or less exactly.

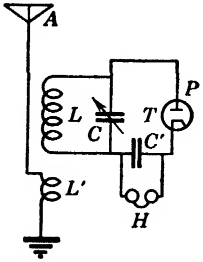


FIG. 58. RADIO RECEIVER.

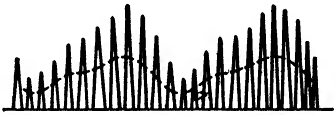


FIG. 59. DETECTOR CURRENT.

This, of course, is the simplest kind of circuit, which might serve satisfactorily within a few miles of a radio station. For greater distances, electron-tube amplifiers must be combined with this simple circuit to build the electric currents up to satisfactory amplitudes. If the power of the amplifiers is great enough, a loud speaker may be used.

72. Frequency Modulation. The modulation process which has been described in Sec. 70 is called *amplitude* modulation, to distinguish it from the newer system, *frequency* modulation. As the name implies, frequency-modulated radio waves remain constant in amplitude, while their frequency varies up and down from the normal or unmodulated frequency, the *change of frequency* at any instant being directly proportional to the value of the audiofrequency *voltage* at that instant. For example, if the unmodulated frequency is 100 megacycles, a weak audiofrequency signal might cause that frequency to swing from 100.001 megacycles at the peak of the audio signal to 99.999 megacycles at the trough, whereas a stronger audio signal would cause the frequency change to be greater, say from 100.010 to 99.990 megacycles. The amplitude remains always the same (see Fig. 60). Many advantages are claimed for the frequency-modulation (FM) system, including greater freedom from interference between stations, and from static. The principal disadvantage, the requirement for a wide frequency band as compared with the amplitude-modulation (AM) system, is met by operating at very high frequencies. At present a series of frequency bands between 44 and 300 megacycles has been assigned to FM. Waves modulated in this manner may not be reduced to intelligible sounds by the ordinary types of radio receivers, such as those described in Sec. 71, but require a special type.

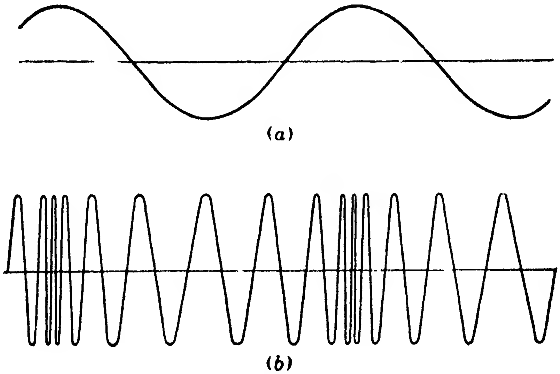


FIG. 60. FREQUENCY MODULATION. (*a*) Audiofrequency (modulating) signal. (*b*) Frequency-modulated radiofrequency oscillogram.

73. Radio Antennas. For quite short radio waves the antenna may be a horizontal or vertical rod about one-half wavelength long; it functions exactly like the transmitter and receiver rods described in Sec. 68. For longer waves the antenna may be a vertical rod or mast about a quarter-wavelength high, with its lower end connected to a large conducting sheet, called a counterpoise, placed flat on the ground. This counterpoise may consist of a network of heavy wires, or even the damp earth itself. The current is a maximum in the end connected to the counterpoise, and the electric and magnetic fields around it are just the same as the fields around one end of the half-wave antenna illustrated in Fig. 52. When, as is usual for the longer broadcast waves, the construction of a vertical structure one-quarter of a wavelength high presents practical difficulties, the antenna may be extended vertically as high as is practicable and then horizontally so as to make the overall length about one-quarter of a wavelength. Transmitting antennas are built so as to keep their resistance as low as possible, and are always tuned to approximate resonance for the frequency used. Receiving antennas are not usually tuned, except for very short waves. Loop antennas are sometimes used for receiving circuits; their operation is explained in Sec. 69.

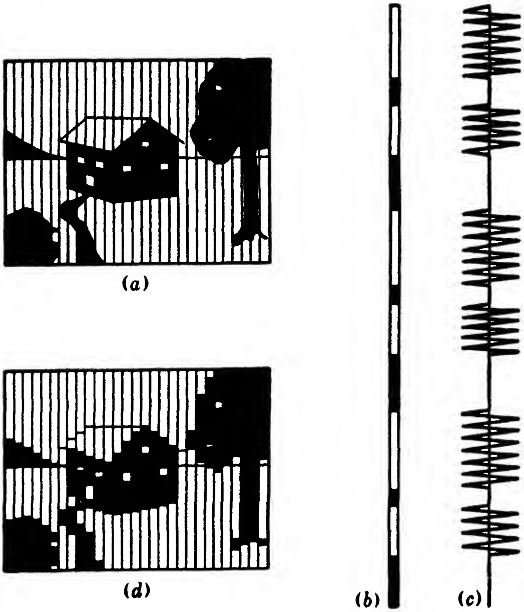


FIG. 61. TELEVISION IMAGES. (*a*) Image on iconoscope—to be scanned in vertical strips as indicated, from left to right. (*b*) Scanning strips laid end to end—read down. (*c*) Oscillogram of corresponding radio oscillations. (*d*) Reproduced image in receiver.

The waves radiated from a straight antenna are most intense in the directions perpendicular to the antenna, and zero in the direction of its length. The radiation is principally from the vertical portion of a flat-topped antenna. For very short waves multiple antennas, or antenna arrays, are used, and these may be arranged so as to send out quite narrow beams of radio waves. Cf. Sec. 77.

74. Television. The transmission of pictures by electric signals, whether by wire or by radio waves, depends upon the same principles. Since the electric signal can represent but one spot in the picture at a time, the picture must be "scanned" piece by piece, usually in close-spaced parallel strips, as shown in Fig. 61(*a*). In this figure the scanning is represented as taking place vertically downward, beginning at the upper left-hand corner. Figure 61(*b*) shows the first three scanned strips end to end, and (*c*) indicates the corresponding modulated radio oscillations. Various schemes have been devised in the past to accomplish this scanning and to translate the brightness of each point, as it is scanned, into an electric signal. The device used most extensively at present is called an iconoscope. In this iconoscope an image of the view to be transmitted is formed by a large-aperture camera lens directly upon a flat, photosensitive screen of a very special kind, which is at the same time being regularly swept over or scanned by a sharply focused cathode-ray beam. This screen consists of thousands of separate, infinitesimal photoelectric cells whose only connection with the rest of the circuit is that made by the cathode-ray beam. Thus the photoelectric current at any instant depends upon the brightness of the image at the point touched by the cathode-ray beam at that instant, and it varies in intensity as the beam sweeps over the picture and as the picture changes. This photoelectric current is fed into powerful electron-tube amplifiers, and the amplified voltages thus produced are used to modulate the oscillations of a radio transmitter. *The lights and shadows of the picture have been translated into fluctuations in amplitude of a radio wave.* Very short waves are required for this service, because of the complexity of the signals transmitted; and the range of transmission for television pictures is at present limited to little farther than the distance the waves can travel in a straight line without encountering obstacles to their passage.

In the receiver, which is in most respects the same as an ordinary short-wave radio receiver of high quality, the radio signals are detected and amplified, and the resultant amplified voltages are applied to the intensity-control grid of a cathode-ray tube similar to that shown in Fig. 11, so as to produce corresponding fluctuations in the intensity of the cathode-ray beam. At the same time auxiliary apparatus deflects the cathode-ray beam so that it scans or traces upon the fluorescent screen lines which are replicas of the scanning lines at the transmitter, and are in synchronism with the scanning at the transmitter. The picture then appears as shown in Fig. 61(*d*). One essential part of the apparatus is the means employed to synchronize the scanning of this beam with the scanning beam at the transmitter, and the perfection of the resultant image depends upon how exactly the synchronism is maintained, upon the fidelity of the amplifiers, and upon the fineness of the scanning pattern. In the simple block picture of Fig. 61, no variations of light and shade are included. In actual television apparatus, full variations of light and shade are reproduced, and the scanning lines are so close that the picture is as good as the usual newspaper picture or better. The present practice is to scan the picture with 525 horizontal lines, covering the picture once by scanning every other line, then again to fill in the gaps (this is called interlace). The picture is gone over thus from top to bottom 60 times a second, or 30 times per second for the complete picture.

75. The Radiofrequency Spectrum. Present-day radio communication utilizes ratio waves ranging in length from several kilometers down to a few meters, or in frequency from about 10 kilocycles up to about 100 megacycles per second. The longest waves are reserved for transoceanic and other long-distance, point-to-point toll service. Commercial broadcasting occupies a middle range, from 550 to 1500 kilocycles per second, while television and frequency modulation operate near the top of the radiofrequency spectrum. The longer waves are capable of bending around even large obstacles, and of following the curve of the earth far beyond the visible horizon. They are confined to the earth's atmosphere by the *ionosphere,* which consists of conducting layers of gaseous ions, high up in the atmosphere, and serves as a reflector for these radio waves. The very short waves at the other end of the spectrum, including the television and frequency-modulation ranges, behave more like light waves. They are diffracted around obstacles which are comparable to their wavelengths in size, but they cast noticeable shadows of larger ones; hence their range is limited to line-of-sight distance, or the distance to the visible horizon, and is broken by large obstructions such as buildings. One way for extending this line-of-sight distance is to carry the radio transmitter high above the earth in an airplane, and several radio laboratories are experimenting with this procedure. The ionosphere does not stop these very short waves. A most spectacular demonstration of this fact was made by Signal Corps scientists in January, 1946, when they sent radio waves of about 3-meter wavelength to the moon and back with a special radar set. The photographic record of this remarkable accomplishment is reproduced in Fig. 67(*b*).

76. Microwaves. Although 3-meter electromagnetic waves were once considered very short, wartime developments brought into wide-scale use 30-cm, 10-cm, even 3-cm waves, and waves less than one centimeter long have been produced and studied in the laboratory. Such *microwaves,* as they are called, are particularly interesting to the physicist because they resemble light waves so closely A physicist, because of his knowledge of optics, is better equipped to work with microwaves than the expert radio engineer! All the phenomena of reflection and refraction of light have counterparts in microwave phenomena. For example, microwaves may be concentrated into a "searchlight" beam by a concave parabolic mirror of sheet metal. (Lattice work made of metal strips or wire will serve instead of sheet metal for the longer microwaves.) Figure 62 shows such a microwave mirror. The source, which is placed at the focus of the mirror, just as for an optical searchlight, is a little half-wave antenna quite like the one described in Sec. 68, only smaller. Such a mirror may be employed also to pick up a parallel beam of microwaves and concentrate it upon a receiving antenna at its focus. Microwaves may also be refracted by lenses and prisms made of some dielectric material, such as paraffin or pitch, although such prisms and lenses have little practical value.[3]

Most phenomena of interference and diffraction of waves, such as those discussed in Chapter III for light waves, are demonstrated even more easily for microwaves. The maxima and minima must of course be located by means of a small microwave receiver in place of the eye. In addition, many interference experiments which are impossible with light waves may be performed with microwaves, since it is possible to have two or more *identical* sources of microwaves (cf. Sec. 31). For example, two identical half-wave antennas, or *dipoles* as they are often called, placed several wavelengths apart, will produce precisely the same wave pattern shown in Fig. 25 for two diffracting slits, with one difference: there will be overlapping waves on both sides of the plane through *a–b*, Fig. 25, whereas they appear on only one side for the slits. This difference is explained in advanced textbooks on optics. Since this wave pattern is far more intense than the one produced by slits, it is easier to demonstrate. Likewise, if quite a few identical dipoles are lined up in a row at equal distances (several wavelengths) apart, they will produce the grating wave pattern shown in Fig. 27. In this case the receiver must be kept a long distance away, or it must be placed at the focus of a parabolic mirror so as to be a receiver for parallel rays; the mirror serves the function of the lens shown in Fig. 27 and is needed for the purpose explained in Sec. 39.

77. Microwave Beams. If the dipoles described in the preceding paragraph are moved closer together, so that their separation is less than 1 wavelength, the radiation from them will be narrowed down into a beam which extends equally in all directions in a plane perpendicular to the row, but is concentrated close to that plane. The more dipoles, the narrower is the beam in the direction perpendicular to its plane. If now a whole sheet or *array* of dipoles is formed by placing a considerable

---

[3] The Bell Telephone Laboratory has developed a practical lens of quite a different type, whose operation depends upon the fact that electromagnetic waves travel faster through wave guides (see Sec. 77) than in free space. This lens is described by W. E. Kock in *Electronics*, Vol. 19, 1946, pp. 101–102, and in the *Proceedings of the Institute of Radio Engineers*, Vol. 34, p. 828 (1946).

FIG. 62. RADAR ANTENNA. Dipole at the focus of a parabolic metal mirror. It may be pointed in any direction, like a searchlight. U. S. Signal Corps photograph.

FIG. 63. RADAR "MATTRESS" ANTENNA. A plane array of 32 dipoles oscillating in phase, placed one-quarter wavelength in front of a wire-grid reflector. An antenna similar to this one, but having 64 dipoles, was used to obtain the radar echo from the moon shown in Fig. 67(*a*). U. S. Signal Corps photograph.

number of such rows side by side and less than a wavelength apart, the beam is narrowed in all directions to form a searchlight beam radiating in a direction perpendicular to the plane of the array. There is also a second searchlight beam radiating in the opposite direction and from the opposite side of the array. This reverse beam may be made to add to the effect of the other one by means of a flat reflector, such as a metal plate or sheet of wire screen. To bring the reflected beam into phase with the direct beam this mirror must be located parallel to and one-quarter wavelength behind the array; the forward intensity then will be quadrupled. Figure 63 shows a photograph of such a "mattress" array. Another manner of obtaining a searchlight beam of microwaves from a row of dipoles is to place the row along the focal line of a concave cylindrical mirror, as shown in Fig. 64. Other methods for producing searchlight beams have been described in Sec. 76. All these arrangements illustrate beautifully the fundamental principles of wave optics.

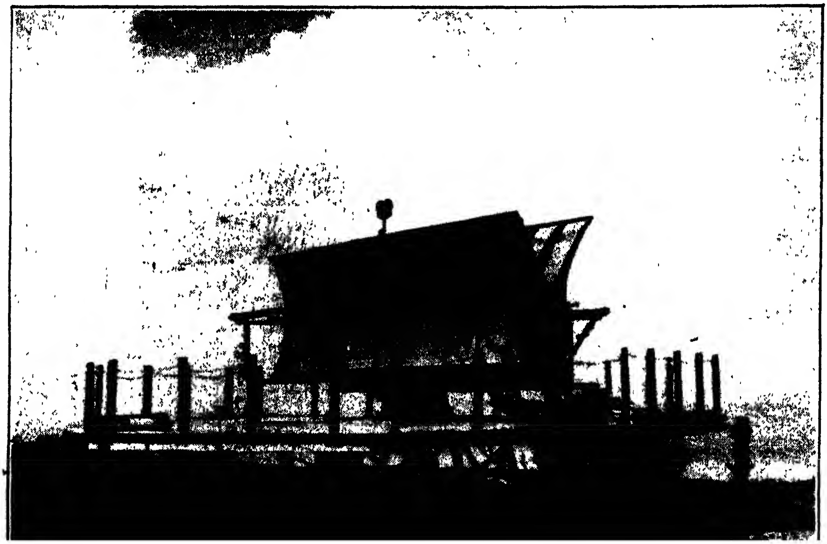
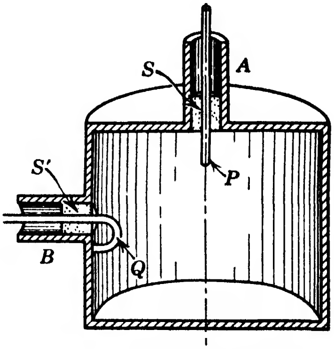


FIG. 64. RADAR ANTENNA. A row of in-phase dipoles placed along the line-focus of a parabolic cylinder. (The assembly shown in the photograph includes two such antennas, back to back. It rotates about a vertical axis to scan the horizon.) U. S. Signal Corps photograph.

Microwaves exhibit many other fascinating properties which may only be mentioned here. For example, they may be transmitted through hollow metal pipes, called "wave guides," and radiated from metal horns like sound waves from a megaphone. Even a solid bar of dielectric material will serve as a wave guide, the microwaves being totally reflected internally in the manner of light waves traveling through a glass rod. Other properties will be mentioned in later sections.

78. Electron Tubes for Microwaves. The very rapid oscillations required for microwaves cannot be generated with ordinary electron tubes. The primary difficulty is the time taken by electrons to travel from the cathode to the anode. If one tries to generate microwave frequencies with ordinary tubes, the electrons will arrive at the anode after its a-c. potential has reversed! They now tend to stop the oscillations rather than to maintain them. One obvious solution is to make smaller tubes, with the cathode and anode closer together. The "door-knob" tube shown in Fig. 51 is an example. The glass envelop is large, but the electrodes are tiny; the plate is only about 3 mm in diameter. The wavelength limit for this tube is about 50 cm. The same principle is utilized in the design of "acorn" and "lighthouse" tubes (so-called because of their superficial appearance), which will operate at still higher frequencies. The lighthouse tube will generate up to about 3 billion cycles per second, or 10-cm wavelength.

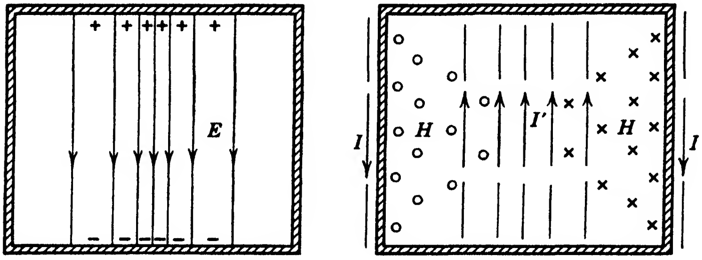
The power which may be generated with these small electron tubes is not very great, and becomes less and less as the wavelength becomes smaller and the tubes correspondingly more tiny. On the other hand, microwave applications such as radar require power measured in kilowatts and even in megawatts, and at the same time work best at wavelengths shorter than the shortest which may be generated with tubes of ordinary types. To meet these needs some entirely new types of electronic microwave generators were developed, such as the klystron, the magnetron, and the resnatron. The descriptions of these remarkable devices must be omitted here, however. Several references which do describe them are listed at the end of the chapter.



Front half is cut away to show the interior.
A, *B*: coaxial lines.
S, *S'*: insulating plugs.
P: straight probe.
Q: loop probe.

Fig. 65. Cylindrical Cavity Resonator.

79. Cavity Resonators. At microwave frequencies the coil and condenser type of resonant electric circuit is replaced by much more efficient electric resonators in which the oscillating electric and magnetic fields are almost entirely enclosed in a conducting cavity. A single example of such a *cavity resonator* will serve to illustrate this. Consider the closed cylindrical metal box shown in Fig. 65. One way in which oscillating electromagnetic fields may be set up in this box is shown in Fig. 66. These oscillations consist of (1) alternating electric charges appearing at the two ends of the cylinder, (2) alternating electric currents flowing up and down the cylindrical wall, and (3) alternating electric and magnetic fields accompanying these charges and currents. These fields are confined to the interior of the box, as shown in Fig. 66, and this accounts for the high efficiency of this resonator; none of the energy stored in these fields is lost by radiation. Power is put into or taken out of this cavity by *probe* electrodes such as shown in Fig. 65. The straight probe interacts principally with the electric field, the loop probe with the magnetic field.



(*a*) (*b*)

FIG. 66. OSCILLATIONS IN A CAVITY RESONATOR. (*a*) Charge Q and electric field E. (*b*) Currents I and I', and magnetic field H ($\frac{1}{4}$ cycle later). I—current flowing in cylindrical wall (along inside surface). I'—*displacement* current. H—lines of force are concentric circles whose intersections with the plane of the figure are indicated by the symbols O and ×. For the current directions shown, these lines point up from the paper at O and into the paper at ×.

The fact that the magnetic field is confined to the interior of the box has a very interesting explanation. The rapid alternations of the electric fields are equivalent to an alternating *current flowing up and down the middle of the box,* in the same direction as the lines of electric force. And this *displacement current,* as it is called, is at all times opposite in direction and equal in magnitude to the currents flowing down and up in the walls of the box. See Fig. 66. These two currents thus subtract their magnetic effects outside the box. The existence of such displacement currents was predicted by Maxwell and is the fundamental basis of his prediction for the existence of electromagnetic waves. See Sec. 67. In plane waves also the magnetic fields are produced by the *displacement currents* which correspond to the rapid variations of the electric fields.

Many other forms of cavities will serve as microwave resonators. Moreover, any one cavity has innumerable overtone modes of oscillation as well as several fundamental modes; in this respect these electrical resonators may be compared with acoustic resonators, such as stretched strings, bars, plates, or organ pipes and other air-filled cavities. The resonant frequency depends upon the dimensions of the cavity and the mode of oscillation. For the box and the mode shown in Fig. 66 the resonant wavelength would equal 1.30 times the diameter of the cylinder. The frequencies of the overtones may sometimes be harmonic multiples of the fundamental, as for a stretched string; more often, however, they are not. Cavity resonators, which become quite small at the highest microwave frequencies, form an integral part of microwave generators such as the klystrons, magnetrons, and resnatrons mentioned in Sec. 78.

80. Radar. The principal wartime application of microwaves was **radar.** The radar principle is beautifully and wonderfully simple. If a very narrow "searchlight" beam of microwaves falls upon a distant object, such as a ship or an airplane, a small part of this beam will be reflected back towards the source. If now this faint microwave "echo" can be picked up by a sensitive receiver, the position of the reflecting object may be determined. The direction in which the searchlight beam is pointing at the time the echo is picked up gives the *direction* of the reflecting object, and the *time* required for the signal to travel to the object and back measures the *distance.* The best radar sets now can locate an object as big as a ship at any distance up to 200 miles to within ½ degree of angle, and measure its distance to within a few yards. The unaided eye is roughly ten times better in determining angular position, but the eye cannot do more than guess at distance without the aid of elaborate optical range-finders. Furthermore, the precision of optical range-finders falls off rapidly as the distance increases, whereas the precision of radar remains nearly constant.

Since microwaves travel with the speed of 186,000 miles per second, the measurement of distance by radar requires the measurement of time intervals in millionths of seconds, or *microseconds.* For example, the echo will return from a ship 5 miles away in a time of only 54 microseconds. And to determine any distance to within 10 feet requires measuring time to a precision of $\frac{1}{50}$ microsecond. To accomplish such measurements it is necessary first of all to make the radar beam a series of very short, very powerful *pulses.* The instantaneous power in a radar pulse is measured in kilowatts, and may exceed a million watts. It may last less than one microsecond, and never more than a few microseconds. Pulses are repeated several hundred times a second. When the reflected pulse returns, it is picked up by the same antenna, which is automatically switched over from the transmitter to the receiver, and the time elapsed is shown on the screen of a cathode-ray tube (see

Sec. 13). In the simplest arrangement the cathode-ray beam starts to trace a horizontal line across the screen when the pulse starts out, and when the echo pulse arrives it causes a vertical mark, or "pip," to appear on this line. The length of this line, from start to pip, measures the time elapsed [see Fig. 67(a)]. Greater precision is made possible by electric circuits which delay the start of the cathode-ray beam until shortly before the arrival of the echo, so that only a comparatively short distance need be measured by the screen. The distance corresponding to the delay time is measured by the electric circuit and is added to the distance measured by the screen.

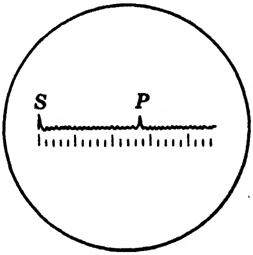


FIG. 67. RADAR A-SCOPE.

In the simplest type of radar the antenna, which may be any one of the types described in Sec. 77, is rotated steadily about a vertical axis while the observer watches the "scope," or cathode-ray screen, for pips which indicate the presence of ships or planes. Once such an object is located, the radar beam may be held steady upon that object by manual control or, in later models, by automatic control. Gun-fire control radar, which is a modification of this type, is so interconnected with the gun-training mechanisms as to point and elevate the guns automatically, and fire them at the instant they are correctly aimed.

FIG. 67(a). A-SCOPE RECORD OF RADAR ECHO FROM THE MOON. The pip which represents the moon-echo is indicated by the sharp rise at 238,000 miles on the scale. The higher pip at about 40,000 miles is of extraneous origin. U.S. Signal Corps photograph.

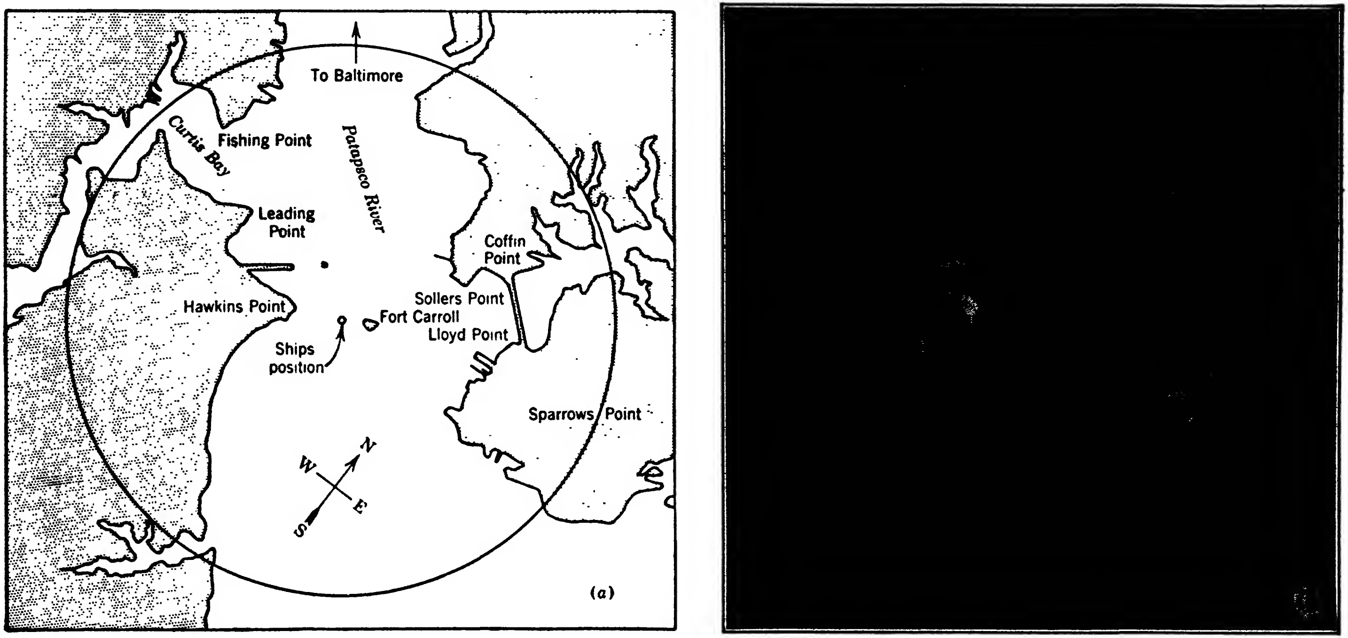


FIG. 68. RADAR P-SCOPE. The radar scope (*b*) presents a map-like picture of the ship's surroundings, with the ship itself at the center. Here an eight-mile radial range was in use as the ship steamed out of Baltimore. Comparison of scope details and the map (*a*), marked with a circle covering the region shown on the scope, reveals the fidelity of the radar picture. The signals in the water area are from other vessels, both at anchor (the upper group) and in motion, and buoys. Continuous patterns of this type permit navigation and prevent collision regardless of weather or visibility. Depending upon the navigator's interest, ranges of 2, 8, and 32 miles can be presented on the PPI scope. (Reproduced from *Westinghouse Engineer*, July, 1946, with the permission of Westinghouse Electric Corp.)

The usefulness of radar is greatly extended by the addition of a second cathode-ray tube or P-scope. The first cathode-ray tube is then called the A-scope. On this P-scope the cathode-ray pencil starts at the center of the screen and moves radially outward at an angle which represents the angular position *on a map* of the radar beam. The intensity of this cathode-ray beam is reduced to zero until an echo is picked up; then a bright spot appears on the screen. This bright spot thus represents the map position of the reflecting object. As the antenna rotates about a vertical axis, all objects within range—ships, planes, shorelines, even heavy clouds—are mapped on the P-scope screen. The luminescent material on this screen has a long-time phosphorescence, so that the spots do not fade out before the antenna and the cathode-ray beam have made a complete swing around and start over again. Thus a fairly complete map of the region appears on the screen, as seen in Fig. 68.

Radar, which was born of wartime necessity, offers wonderful opportunities for peacetime application. The navigation of inland waters, illustrated by Fig. 68, is only one of many.

The echo-ranging principles of radar are employed also for underwater detection of ships, submarines, and similar objects. Since radio waves cannot be transmitted through water, very short sound waves are employed instead. This system is called **sonar.** In other respects sonar is essentially the same as radar.

81. Radio Direction-Finding—Loran. One of the earliest as well as one of the most important applications of radio was in finding the position of a ship at sea. In the early method a direction-finding antenna is used aboard ship to locate two known stations on shore. Usually this antenna is a loop type (see Fig. 55), which is turned about a vertical axis to the position at which the signal is weakest; the axis of the coil then points towards the transmitting station. The intersection of two lines drawn on the map to represent the directions of the ship from the two stations then locates the position of the ship. This method is illustrated in Fig. 69. The angular positions of these lines can be determined to within a few degrees only, and the range is limited to a few hundred miles.

An entirely different system, called **Loran** (long-range navigation), was developed during the war. Its wartime service covered a large part of the oceans of the earth, and complete coverage may be expected for its peacetime services. As the name implies, it is far more sensitive than the older system, being reliable over distances of 750 to 1500 miles. In addition, it is far more precise, locating ships and aircraft to within one per cent or less of the distances measured. Loran utilizes the radar method for measuring distances, to determine how much farther the ship is away from one radio beacon than from another beacon sending out matched signals. This process is then repeated for another pair of beacons, and the position of the ship looked up on a special Loran map. Figure 70 shows how this works. Say the ship is 25 miles farther from beacon A_2 than from beacon A_1; this locates her somewhere on a line, aa', every point of which is 25 miles farther from A_2 than A_1. This line is called a Loran line, and a Loran map contains a whole series of such lines, each representing a different distance. If the ship is also 40 miles farther from beacon B_2 than from beacon B_1, this locates her on the line bb', the 40-mile Loran line for beacons B_1, B_2, at the intersection of line bb' with line aa'.

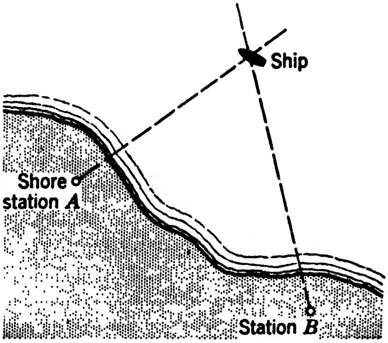


FIG. 69. POSITION FINDING WITH LOOP ANTENNA ON SHIP.

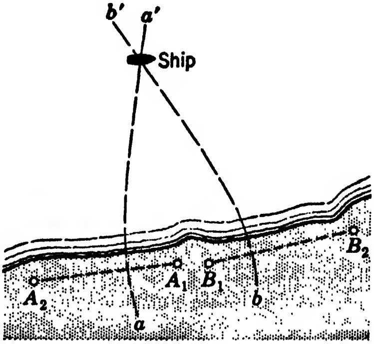


FIG. 70. POSITION FINDING WITH LORAN.

82. Radio Control. Electronic controls have been applied to many operations, from the opening of doors to the functioning of the most intricate computing machines, and radio methods extend such controls to any distance to which reliable radio communication may be carried. One of the simplest and one of the most useful of these applications is in reading instruments which are located at inaccessible places. An outstanding example of this is the radiosonde, a tiny "weather station" which is sent high into the atmosphere by means of a free balloon. A small radio set, carried along with the meteorological instruments, transmits a continuous record of the temperature, pressure, and humidity readings to the Weather Bureau receiver on the ground. Simultaneously the height and angular position of the balloon are observed and recorded by radar. Similar radio-equipped balloons are employed to study cosmic radiations at great heights above the earth.

Another example of radio control is the proximity fuze, which proved of such immense value in the war. This tiny but unbelievably rugged radio mechanism, when placed in the nose of a shell, rocket, or bomb, fires the projectile automatically when it comes within damaging distance of its target.

Another application is the guidance of moving vehicles. Full-sized airplanes have been operated without crews, their flight being directed from a distant ship or plane and followed by television. The television transmitter was located in the pilotless plane and imaged on the screen of the television receiver what would have been seen by the pilot if he had been there. The possibilities for such automatic control of vehicles are almost endless. It may seem fantastic to consider taking a motor trip with no one at the wheel, but it is now *technically* possible to equip motor cars and highways so that the cars could travel any prearranged course without human guidance, safely avoiding all traffic hazards except being run into by human-guided vehicles.

PROBLEMS

1. The coil of a frequency-meter circuit has an inductance of 15 millihenries. The condenser is variable from 30 to 250 $\mu\mu$f (micro-microfarads). Compute the range of frequencies which may be measured by this meter.

2. An oscillating electric circuit is made with a coil whose inductance is 8.0 millihenries and a condenser of 120 $\mu\mu$f capacitance. (*a*) Compute the frequency of its oscillations. (*b*) What capacitance would be required to tune this coil to the second harmonic of the original oscillations?

3. A radio station broadcasts on a frequency of 1450 kilocycles. Compute the wavelength of the radio waves.

4. The antenna of a short-wave radio transmitter consists of a horizontal metal rod 160 cm long. What frequency of radio waves will it radiate most efficiently? Explain.

5. Two antennas such as the one described in problem 4 are placed parallel to each other, the one being 160 cm vertically above the other. They are kept oscillating exactly in phase, at the frequency determined in problem 4. In what directions will their combined radiations be greatest? In what directions will they be least? Explain, using diagrams.

6. A 630-kilocycle radio transmitter is amplitude-modulated by an audiofrequency tone of 512 vibrations per second. How many radio oscillations will there be in *one* of the groups of oscillations shown in Fig. 56(*c*)?

7. Two identical microwave sources, operating *in phase* at a frequency of 3000 megacycles, are located 25 cm apart. A receiver is located so far away that the rays reaching it from both sources may be considered parallel. Find the angular positions for this receiver at which the receiver signal is a maximum, and at which it is zero. Solve graphically. (Cf. Fig. 25.)

8. A radar set operating at 400 megacycles is located 400 cm above the earth's surface. (*a*) Show that, because of interference between the direct beam and the beam reflected from the earth's surface, the resultant beam is very weak close to the horizon. (*b*) Find the angles with the horizon at which the direct beam and the reflected beam are in phase, and the angles at which they are exactly out of phase. Do this graphically. (*Note:* The reflected beam appears to come from an *image-*

source which is located according to the laws of optics and is 180 degrees out of phase with the real source. Consider the earth flat over the area involved.)

9. A radar signal returns 85 microseconds after being sent out. How far away is the object located?

10. A sonar signal locates a ship which is 3 miles away. How long does the signal travel before its return? Take the speed of sound to be 5000 ft per sec in sea water.

11. Explain why the reflector in a mattress array (see Sec. 77) must be placed one-quarter wavelength behind the dipole sheet.

12. Take two points 10 cm apart and plot a line that is everywhere 2 cm farther from one point than from the other. This is a Loran line (see Sec. 81).

REFERENCES

For Additional Reading

ARRL *Radio Amateur's Handbook.* (Issued annually.)

J. Mofenson, "Radar Echoes from the Moon," *Electronics*, Vol. 19, pp. 92–93, April, 1946.

"The Loran System," *Electronics*, Vol. 18, pp. 94–99, November, 1945.

"Proximity Fuze," *Electronics*, Vol. 18, pp. 110–111, November, 1945.

For Advanced Reading

F. E. Terman, *Radio Engineering* and *Radio Engineers' Handbook.*

W. L. Everitt (editor), *Fundamentals of Radio.*

H. H. Skilling, *Fundamentals of Electric Waves.*

D. G. Fink, *Radar Engineering.*

M.I.T. Radar School, *Principles of Radar.*

S. Ramo and J. R. Whinnery, *Fields and Waves in Modern Radio.*

S. K. Haynes and W. J. Jackson, "The Physics of Radar," *American Journal of Physics*, Vol. 14, pp. 143–164, 1946.

J. B. Fisk, H. D. Hagstrum, and P. L. Hartman, "Magnetron as a Generator of Centimeter Waves," *Bell System Technical Journal*, Vol. 25, pp. 167–348, April, 1946.

W. W. Salisbury, "The Resnatron," *Electronics*, Vol. 19, pp. 92–97, February, 1946.

D. G. Fink, *Principles of Television Engineering.*

V. K. Zworykin and G. A. Morton, *Television.*

R. J. Evans, "Echo Ranging Sonar," *Electronics*, Vol. 19, pp. 88–93, August, 1946.

L. B. Arguimbau, *Vacuum Tube Circuits.*

CHAPTER VI

THE ELECTROMAGNETIC SPECTRUM—INFRARED LIGHT, ULTRAVIOLET LIGHT, AND X-RAYS

83. The Spectrograph. If the various components of light are to be carefully studied, they must be spread out into a pure spectrum by some such means as the grating spectrograph shown in Fig. 26, or the prism spectrograph illustrated in Fig. 71. In this prism spectrograph the light to be studied enters through the narrow slit S placed at the focus of the lens L_1, which delivers it to the prism as a beam of parallel rays. This first part of the apparatus is called the collimator. After passing through the prism the light enters the camera through the lens L_2 and falls upon a ground-glass focusing screen or a photographic plate placed at the focal plane of the lens. Let us assume that the light contains only a few wavelengths, as it would if it came from a source such as a mercury-vapor lamp. Then the prism splits the beam of light from the collimator into a diverging series of overlapping parallel beams, one for each wavelength, and the camera lens focuses each beam to a different place on the photographic plate, forming a spectrum of bright lines. (Rays are drawn in Fig. 71 for but two of these wavelengths.) Each line is an image of the collimator slit, and may be made very fine by making the slit correspondingly narrow. If the source is an incandescent solid such as an electric lamp filament, an infinite number of overlapping beams will exist, and the slit images will run together to form a continuous spectrum.

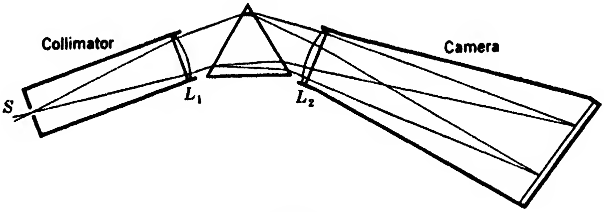


FIG. 71. PRISM SPECTROGRAPH.

84. Invisible Light. If a continuous spectrum is produced in a spectrograph and recorded upon a photographic plate by a suitable exposure, the limits of the blackening produced upon the plate will be quite different from the limits corresponding to the visible spectrum as seen on the ground glass. If the plate is one of the ordinary sort used for commercial work, it will be entirely free from any blackening at the orange and red end of the spectrum, but it will be strongly blackened for a long way beyond the violet or short-wave limit of the visible spectrum. An orthochromatic emulsion, which is the most common type for amateur camera film, will blacken farther into the long waves so as to include some orange light; panchromatic emulsions will cover most of the visible spectrum (with a curious low sensitivity in the green); and special "infrared" plates will blacken well beyond the red or long-wave limit of the visible spectrum. All these photographic materials are strongly sensitive to the short-wave range of the spectrum to as far as about 2200A, and even panchromatic emulsions must be used with special color screens to make their relative color sensitivity comparable to that of the eye.

Although these invisible parts of the spectrum which are detected by the photographic plate are not "light" in the literal sense of the word, since they are invisible, they have all the other *physical* attributes of light, and we therefore call them invisible light: ultraviolet light for the region of wavelengths shorter than the visible limit in the violet, and infrared for the region of longer-than-visible waves at the other end of the spectrum. Even with the aid of special photographic plates we cannot go far into the infrared or the ultraviolet regions of the spectrum as long as we use glass prisms and lenses, since glass, strangely enough, becomes opaque not far beyond each end of the visible spectrum. For this reason special apparatus is required for the investigation of each of these extensions of the spectrum.

85. Infrared Light. By using a prism of rock salt, and mirrors in place of lenses, the infrared region may be explored far beyond the limit set by glass optics. Below the limit of photographic sensitivity the spectrum must be "viewed" by means of thermocouples or other very sensitive thermometric devices which detect the rise in temperature of a narrow blackened strip upon which a portion of the spectrum falls. This black strip and thermometer arrangement measures the total energy of the radiation falling upon it and may be used in any part of the spectrum, even the ultraviolet. It is not very sensitive, however, and in regions where photographic plates may be used it is much inferior to them. Fortunately the infrared part of the spectrum contains much of the energy, so that this thermometric detector is useful here, where other detectors are useless. Wire-grid diffraction gratings and other interference apparatus have also been used to produce spectra in the far infrared, and infrared waves as long as ⅓ mm have been measured. (See Sec. 100.)

86. Ultraviolet Light. Spectrographs made with quartz prisms and lenses may be used for ultraviolet light down to about 1800A, at which limit quartz becomes opaque. Concave gratings may be used somewhat farther, and then the opacity of the air itself forces the use of vacuum spectrographs. In these vacuum spectrographs the spectrum is formed by a concave reflecting grating; and the grating, the light source, and the photographic plate are all placed in a large air-tight box from which the air is pumped. The short-wave limit of 2200A, which is given above for all common photographic plates, is set by the opacity, below that limit, of the gelatine in which the sensitive silver salts are imbedded. For waves shorter than this, special plates made without gelatine (called Schumann plates) are employed. As an alternative it is possible to use ordinary plates which have been sensitized by lightly coating their surfaces with an oil which will fluoresce when the very short waves fall upon it. Much of the resultant fluorescent light will be of wavelengths longer than 2200A and able to penetrate the gelatine.

87. Luminescence. The phenomena of luminescence, which have already been mentioned in connection with cathode rays (Sec. 6), provide another and very convenient means for the detection and study of ultraviolet light. If, for example, a piece of uranium glass, a common type of ornamental glass, is held in a beam of sunlight or of light from an arc lamp, it will glow with a greenish yellow light which appears to come from the glass itself, giving it a curious opaque appearance. If this piece of glass is placed in a spectrum, it will be quite transparent and free from such glow in the red, orange, yellow, and green parts of that spectrum, beginning to glow in the blue and violet parts and glowing brightest in the invisible ultraviolet part. Solutions of a great many substances, such as anthracine, esculine, quinine sulfate, and fluorescein, similarly fluoresce upon exposure to light of short enough wavelength. Some of them are excited to fluorescence by parts of the visible spectrum, as is uranium glass; practically all of them will fluoresce upon exposure to ultraviolet light. One remarkable example is chlorophyll, the green coloring matter in grass and leaves. This substance fluoresces a deep red color in the orange part of the spectrum, and again in the ultraviolet. Because of this fluorescent light, living foliage appears light when viewed through a piece of glass of a deep red color, whereas artificial foliage appears dark.

For the study of such substances it is convenient to have a source of ultraviolet light which is free from any visible light, and this is made possible by a special kind of black glass which is opaque to practically all visible light and at the same time transmits much more ultraviolet light than does ordinary glass. When this black glass is used as a window for a box containing a quartz mercury-arc lamp, or as a bulb to encase such an arc, it will be found that almost all substances luminesce to some extent when exposed to ultraviolet light, and many other substances glow very brilliantly. Each substance luminesces with a characteristic color, and brilliant stage effects have been produced with luminescent paints and dyes which reveal their colors only when the stage is darkened and irradiated by ultraviolet lamps. Some of these substances glow only during irradiation, whereas others will continue to glow for some time afterward. As has been stated in Sec. 6, the instantaneous glow is called fluorescence, and the afterglow, phosphorescence. Substances which phosphoresce are sometimes called phosphors.

Most substances which luminesce under light excitation will also luminesce under other forms of excitation, such as X-rays and cathode rays. The phenomena involved in luminescence are complex, involving first the absorption of the irradiating light, and then sooner or later the re-emission of some of the absorbed energy as light of the *longer* wavelengths representing the characteristic colors of the luminescent light. Slight traces of foreign substances will profoundly alter the color and other characteristics of the luminescence; for this reason ultraviolet lamps are of considerable use in chemical laboratories for the detection of small differences in organic substances, such as those due to adulteration or to differences in origin. They are even more useful in criminological laboratories. For example, they make it possible to distinguish between real and imitation gems, and to detect forgeries by differences in paper, ink, etc., which appear only in the ultraviolet light. Writing which has been eradicated chemically may show up clearly under ultraviolet light.

88. Infrared Phosphors. Since light emitted as luminescence is always of longer wavelength than the exciting light, it is impossible to produce visible luminescence by irradiation with infrared light alone. Never-the-less, infrared light does have an important effect upon luminescence, which may be illustrated by allowing it to fall upon a screen of phosphorescent zinc sulfide, after exciting the screen to strong phosphorescence with ultraviolet light. The portion of the screen which is exposed to infrared light will lose its phosphorescence much sooner than the rest; hence it will appear darker.

If the screen is watched closely at the instant of exposure to the infrared light, it will be observed to glow somewhat more brightly for an instant. This shows that the effect of the infrared light is to bring about a quicker release of the stored-up energy, part of it being released as visible light. When certain phosphors are cooled to very low temperatures, they are able to hold the stored-up energy almost indefinitely, the screen appearing dark soon after the ultraviolet excitation is removed. Infrared light still is able to release this energy, causing the screen to glow visibly. Recently it has been found possible to produce phosphors which possess this ability at room temperature.[1] After excitation with ultraviolet light they fluoresce briefly and then become dark. Subsequent exposure to infrared light will cause them to glow quite brightly for extended periods.

89. Applications for Infrared Light. Both infrared and ultraviolet light have been employed for secret signalling, but infrared light is considerably better, since it is less scattered by dust and fog. Infrared signalling lamps are made by covering ordinary lamps with black screens which are opaque to all visible light but transmit infrared light. The detector may be a special type of photocell (see Sec. 103) or even a screen of infrared phosphor. A gas-discharge type of lamp has been developed whose brightness may be varied, or *modulated*, at voice frequencies, so that speech may be transmitted along an infrared light beam and detected by a photocell circuit.

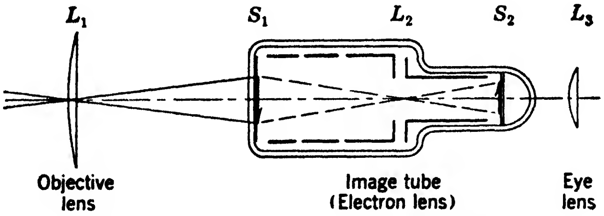


FIG. 72. SNOOPERSCOPE (schematic diagram).

If a screen of infrared phosphor (see Sec. 88) is placed in the focal plane of a lens of large aperture, a visible image of an object illuminated with infrared light will appear upon this screen; in effect, infrared *vision* is made possible. Figure 72 shows schematically a more sensitive device [2] for this purpose, which makes use of physical principles to be described in Chapter VII. The lens L_1 focuses the infrared light to form an invisible image upon the screen S_1, which is a mosaic of many thousand tiny infrared-sensitive photocells, similar to the Iconoscope screen

---

[1] See several articles in the *Journal of the Optical Society of America*, Vol. 36, June, 1946.

[2] See G. A. Morton and L. E. Flory, "An Infrared Image Tube and Its Military Applications," *RCA Review*, Vol. 7, pp. 385–413, 1946.

described in Sec. 74. Whenever light falls upon one of these cells, electrons are emitted from it into the evacuated space behind it. The strong electric fields produced between the cylindrical metal electrodes (shown in cross-section) which form the electron lens, L_2, give these electrons high speeds and focus them upon the fluorescent screen, S_2, where they produce a visible fluorescent image. The eye lens, L_3, aids the eye in viewing this image.

90. Fluorescent Lamps. The use of luminescence as a practical source of illumination is of quite recent development. In these fluorescent lamps, the source of excitation is a mercury arc whose operating conditions are so adjusted that 95 per cent of the light is emitted in the single ultraviolet wavelength, 2536A. This arc discharge takes place in a long tube 1 inch or 1½ inches in diameter, and the luminescent mineral is placed as a thin coating on the *inside wall* of this tube. The color of the light depends, of course, upon the mineral used. Artificial willemite gives the highest luminous efficiency, in a light that is green in color. Blue-white ("daylight"), yellow white, yellow, pink, and blue are obtainable with different minerals, and still other colors are produced by making the tubes of colored glass. A small amount of visible mercury light is mixed with the fluorescent light. These lamps operate at a very low temperature and with a luminous efficiency that is much higher than that of any other light source. The power required depends upon the length of the tube, amounting to about 10 watts per foot.

91. Discovery of X-Rays. In 1895, W. C. Röntgen, professor of physics at Würzberg, was engaged, as were so many of his contemporaries, in the study of electrical discharges through highly rarefied gases. He was using a well-exhausted tube of a form devised by Crookes (Fig. 73) which required a high voltage to operate it, and he observed that a fluorescent screen lying on the table near by would glow when the tube was operated, even though the tube was inclosed in a light-tight cardboard box. His hand cast a shadow upon this fluorescent screen, and the bones cast a denser shadow than the flesh. By following the shadow back to the source he quickly discovered that this glow was produced by some kind of radiations coming from the glass at the end

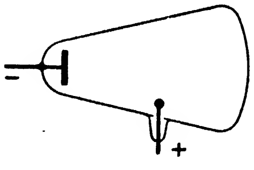


FIG. 73.

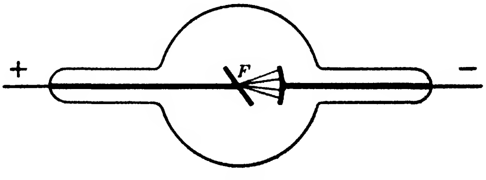


FIG. 74.

of the tube where it was being bombarded by the cathode rays, and simple tests convinced him that these were an entirely new kind of radiations, which he named X-rays. As soon as Röntgen had announced his discovery, other scientists also busied themselves with the study of these new radiations. Soon it was discovered that a dense metal plate placed inside the tube made a better target for the cathode rays than the wall of the tube, and that a concave cathode could then be used to concentrate the cathode rays to a small area of the target and thus produce nearly a point source of X-rays. Within less than a year the X-ray tube acquired the familiar form shown in Fig. 74.

92. Detection of X-Rays. As has been related above, it was by means of *luminescence* that X-rays were discovered. Most of the solid substances which luminesce under ultraviolet-light excitation or cathode-ray excitation luminesce also when exposed to X-rays, although the relative intensities of luminescence may be quite different for each of these different modes of excitation, and those which glow most strongly in X-rays may not be the ones which are most sensitive to ultraviolet light or to cathode rays. Barium platinocyanide, willemite, and zinc sulfide which has been specially prepared are the most sensitive to X-ray irradiation and the most commonly used for *fluoroscope* screens, that is, screens upon which X-ray shadows, such as those of the hand and other portions of the body, may be viewed.

A second means by which X-rays may be detected is a *photographic plate* or film. If X-ray shadows fall upon a photographic plate for a sufficiently long time a permanent photographic record is the result. The exposure may, of course, be made with the film or plate still in its light-proof wrapping. With a powerful X-ray tube the exposure may be quite short, and it may further be shortened by sandwiching the photographic film between two screens of a suitable fluorescent material. The fluorescent light from these *intensifying* screens is several times more effective than the direct action of the X-rays upon the film, the combined effect of fluorescent light and direct X-ray action shortening the exposure by a factor of ten or more. There is, however, some loss of detail when intensifying screens are used.

A third means for detecting X-rays is the *ionization* which they produce. An electroscope at a considerable distance from an X-ray tube will soon lose its charge when the tube is operated; and if air from the vicinity of the X-ray tube is blown towards the electroscope, it will discharge more quickly still, showing that the passage of X-rays through the air has made it electrically conducting, by ionizing it. Although ionization cannot of course be used for X-ray pictures, it is by far the most sensitive of these three means for the detection of X-rays, and has the further advantage that it may be used to make quantitative measurements. Very weak beams of X-rays may be measured by means of an ionization chamber such as is shown in diagram in Fig. 75. The outer metal cylinder, *A*, is charged to a suitable potential, so that when X-rays are sent through it lengthwise, as indicated, they ionize the air or other gas within the cylinder, and a small current flows to the insulated wire, *W*, and thence to the electroscope. The figure shows a gold-leaf electroscope for measuring this current. In practice a more sensitive measuring instrument is employed, such as an electrometer or a special type of electron-tube circuit, devised especially for this purpose. (See Sec. 15.)

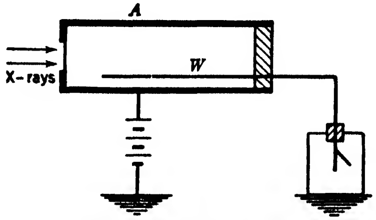


FIG. 75. IONIZATION CHAMBER.

93. X-Ray Pictures. Nothing is completely transparent to X-rays. Even air absorbs them appreciably, as is evidenced by the ionization which results from their absorption. On the other hand, nothing is entirely opaque to X-rays, the opacity being roughly proportional to physical density, quite without regard to transparency or opacity to visible light. Thus dense substances, such as bone, metal, and glass, cast deep shadows in an X-ray beam, whereas wood, flesh, hard rubber, and other substances of low density cast faint shadows. It was also early observed that, when the vacuum in the X-ray tube was so high as to require a high voltage for its operation, the rays were more penetrating than those produced by the lower voltages which could be used when the vacuum was less perfect. The fundamental nature of this difference, which is described by calling the rays "hard" and "soft" respectively, will be explained in Sec. 95. When the voltage is high, so that hard X-rays are produced, soft X-rays also will be produced, but these soft X-rays may be filtered out by interposing a sheet of aluminum in the X-ray beam.

The usefulness of X-ray pictures depends upon these penetration properties of the rays, which reveal bones within flesh, nails in shoes, pipes in walls, etc. Hard X-rays are used where thick layers of matter must be penetrated, and where shadows of bones, metal, and other dense materials are to be depicted. Soft X-rays are used for materials of low density, such as the flower parts shown in Fig. 76. As almost everyone knows, strong shadows of body cavities may be obtained by temporarily filling them with some dense but harmless substance such as barium sulfate, which is used in the study of the alimentary canal.

FIG. 76. X-RAY PHOTOGRAPH OF A COLUMBINE FLOWER. Made by Dr. H. F. Sherwood of the Eastman Kodak Company Research Laboratories, who has developed the soft X-ray equipment necessary for making such pictures.

All X-ray pictures are *shadow* pictures, and the sharpness of such shadowgraphs is dependent upon the smallness of the source of X-rays. For this reason tubes used for *radiographic* work are fine-focus tubes: that is, they are built so that the cathode rays concentrate upon a very small spot on the target.

94. X-Ray Tubes. For some time after the discovery of X-rays all X-ray tubes were of the type shown in Fig. 74. The operating voltage of such a tube is dependent upon the density of the gas left in the tube; and since this gas density changes (it usually diminishes) as the tube is used, some device must be attached to it to permit readjustment of its pressure. The modern X-ray tube, invented by Coolidge and shown in Fig. 77, is so highly evacuated that what little gas is left has no effect upon the operation of the tube. Cathode rays are obtained by thermionic emission of electrons from the cathode, *C*, which is an incandescent spiral of tungsten wire, surrounded by a metal cup which focuses the cathode rays upon the target, *A*. The electron current is flowing at the saturation value (see Sec. 11), and the intensity of the cathode-ray beam and hence of the X-rays may be controlled by regulating the heating current flowing through the filament from the battery, *B*. Hard or soft X-rays may be obtained by merely changing the voltage supplied to the tube. Either a-c. or d-c. high-voltage supply (from 30 to 100 or more kilovolts) may be used, but d-c. supply is more effective. With a-c. supply the tube is self-rectifying, operating only during the half-cycle for which the hot electrode is the cathode. A low-voltage transformer, with its secondary very highly insulated from its primary,

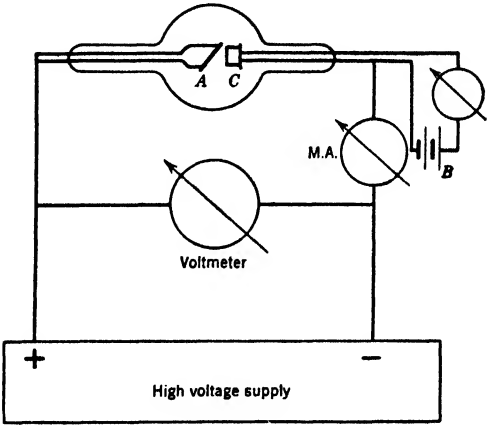


Fig. 77. Coolidge X-Ray Tube.

may be used in place of the battery to heat the filament. An all-glass tube is shown in Fig. 77. The most modern tubes are practically all metal, the only glass being the glass arms needed to insulate the electrodes.

X-ray tubes for therapeutic purposes, such as treatment of cancer, are broad-focus tubes, as distinguished from the fine-focus tubes used for radiographic work. The broad focus permits them to be operated for relatively long periods without overheating the target. Hard X-rays are used for this purpose, especially for treatment of deep-lying tissues, partly because these rays are better able to penetrate to the desired depth without overdosing the skin and other tissues through which they must first pass, and partly perhaps because they are more effective in treatment of the diseased tissue. X-ray tubes operated at more than a million volts are now being used for some radiotherapeutic work.

The efficiency of X-ray tubes is very low. Only a very small amount of the energy of the cathode-ray beam is converted into X-ray energy, the rest being wasted in producing troublesome heating of the target. In a small tube this heat energy is dissipated by radiation, the target being made large enough to allow for this without too great a rise in temperature. A large X-ray tube must be water-cooled, the target being made hollow so that the cooling water may be pumped through it. A tube of intermediate size may be water-cooled or air-cooled; if air-cooled, the target is supported by a heavy metal rod which is sealed through the wall of the tube and carries a set of thin metal cooling-vanes on the end which projects outside of the tube.

95. The Physical Nature of X-Rays. For a long time the usual tests for the nature of radiations gave only negative results when applied to X-rays. All experiments show that X-rays are never deflected by either electric or magnetic fields, no matter how strong those fields may be, and none of the early experiments revealed any reflection, refraction, or diffraction of the ordinary sorts. Whether X-rays were some kind of uncharged particles or whether they were waves of some sort remained a matter of debate for nearly twenty years after their discovery by Röntgen. If they are waves similar to light waves, then the failure to discover any reflection or refraction means that X-ray waves must be thousands of times shorter than the waves of visible light; and in 1913 a theoretical physicist, von Laue, suggested an experiment to test this out.

Friedrich and Knipping, two friends who had the necessary experimental equipment, carried out this experiment, which is illustrated in Fig. 78. The pinholes in the lead plates, *L*, limit the X-rays to a fine beam which shoots through a thin crystal plate, *C*, and thence onto a photographic plate, *P*. If the X-rays are only *particles*, then those which are scattered by the atoms of the crystal should blacken the photographic plate more or less uniformly around the point, *O*, where the direct beam falls, the density of blackening being greatest at *O* and decreasing gradually as the distance from *O* increases. That is, the deflected particles would be scattered at random, and the blackening of the photographic plate would somewhat resemble a target which had been shot at a great many times by an unskillful marksman, so many times that the individual shots have run together.

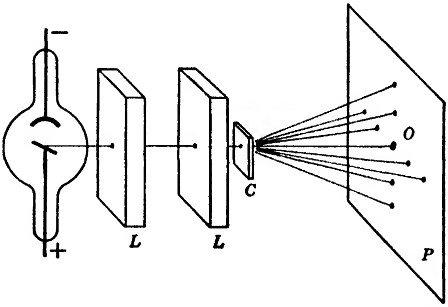


FIG. 78. DIFFRACTION OF X-RAYS: LAUE'S EXPERIMENT.

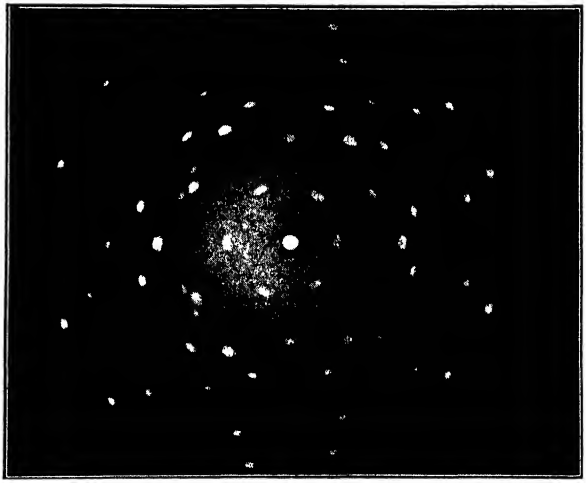


FIG. 79. LAUE PATTERN—DIAMOND CRYSTAL. Made by Dr. C. J. Ksanda of the Geophysical Laboratory of the Carnegie Institution of Washington.

If X-rays are waves, quite a different result might be expected. The plane faces and exact angles of a natural crystal show clearly that its elementary particles (which we now know to be atoms) are arranged in some regular fashion. Each atom will scatter some X-rays, and if the X-ray waves are shorter than the distances between atoms, the scattered waves will interfere so as to reinforce each other's effects in certain directions, and cancel in all other directions. In other words, the regular arrangement of atoms in the crystal will act as a *three-dimensional diffraction grating*, and an interference pattern will appear upon the photographic plate. This was the prediction of von Laue, and Friedrich and Knipping obtained such interference patterns. Figure 79 shows the interference pattern produced in this manner by a thin diamond crystal. Although the patterns could not at first be interpreted quantitatively, since the exact nature of crystal structure was still unknown, *this was convincing evidence for the existence of X-ray waves.*

96. Bragg's X-Ray Spectrometer. Only a few months after Laue's experiment had been performed, W. H. Bragg discovered a much simpler arrangement for the diffraction of X-rays, and one with which could be constructed an X-ray spectrometer of high precision. This spectrometer is shown by Fig. 80. The lead slit-plates, L, limit the X-rays which fall upon the crystal to a narrow beam, as shown. If this beam is of a single frequency, it will not in general be reflected by the crystal, C. If C is turned slowly about an axis through the point of incidence, certain angles will be found for which quite strong reflection occurs. This may be detected photographically, or by means of an ionization chamber, as is shown in the figure.

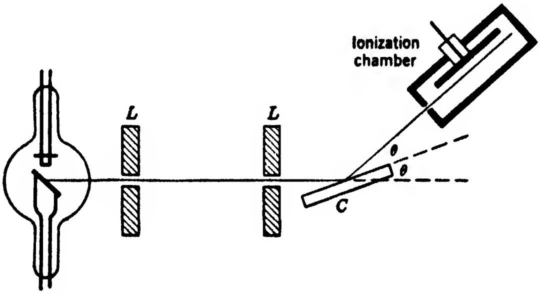


FIG. 80. BRAGG X-RAY SPECTROMETER. The crystal table and the arm which carries the ionization chamber are geared together so that the angle of diffraction is always equal to the angle of incidence.

Why this reflection occurs is explained by Fig. 81, which represents a greatly magnified section of the crystal, with a beam of X-rays incident upon it at an angle θ_2. (θ_2 is measured between the X-ray "ray" and the *surface* of the crystal, rather than the normal to that surface. This departure from usual optical practice is for the practical reason that, in this case, this is the smaller angle.) The dots represent the atom centers, which are arranged in horizontal layers called crystal planes. A few of the incident wave fronts are represented by the heavy solid lines, while the dotted lines similarly represent a few of the waves which have been reflected by the crystal planes.

The "reflected" waves are formed by constructive interference of the X-ray wavelets which are scattered by the atoms in the crystal.[3] All the wavelets scattered by *any one layer of atoms* will be in phase with each other in the direction of regular reflection, which is here shown. In other words, each crystal plane may be looked upon as a nearly transparent plane mirror for X-rays. (This is demonstrated, in Fig. 81, by the fact that the incident and the reflected waves intersect each other in the crystal planes.) In general the reflections from successive planes will not be in phase with each other, so that the crystal as a whole does not then reflect the X-ray beam. When, however, the angle of incidence has exactly the right value, strong reflection does occur. This condition is shown in Fig. 81, which illustrates the *second-order* reflection.

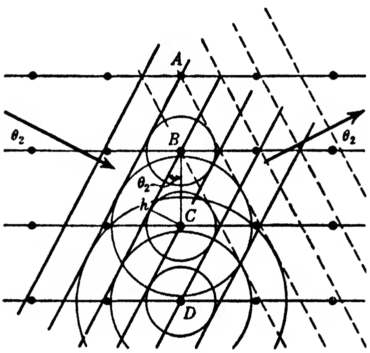


FIG. 81.

Just how this takes place may be seen by considering the wavelets scattered by the atoms whose centers are at A, B, C, and D, the wavelets which are represented by the circles drawn about these points. For example, the *incident* wave which is scattered by B in this case has traveled a whole wavelength farther than has the one scattered by A at the same instant, and the scattered wavelets from B must then travel another whole wavelength before they catch up with the reflected wavefront through A. The X-ray waves reflected by the crystal plane through B are thus *two* whole wavelengths behind those reflected by the crystal plane through A. Similarly those reflected by the plane through C are two wavelengths behind those reflected by the plane through B, etc., and hence all of the reflected waves add up to produce a strong reflected X-ray beam.

---

[3] Although X-ray scattering is caused by the electrons in an atom rather than by its nucleus, the average scattering effect of an atom has its center at the nucleus, as is here represented.

The value for the angle of incidence, θ_2, at which this takes place may be obtained from the triangle, BCh, formed by drawing Ch perpendicular to the incident wave fronts. Since Ch equals the wavelength, λ, and BC is the grating space, d, or the distance between adjacent crystal planes,

$$\sin \theta_2 = \frac{Ch}{BC} = \frac{\lambda}{d} \qquad (6\cdot 1)$$

This is the *second-order* reflection since, as has been shown above, the waves from adjacent crystal planes differ in path length by *two* wavelengths. The first-order reflection occurs when the path difference is *one* wavelength, and the distance corresponding to Ch is but $\lambda/2$, that is, when

$$\sin \theta_1 = \frac{\lambda}{2d} \qquad (6\cdot 2)$$

Generally, when strong reflection occurs, the path difference may be any number, n, of whole wavelengths, and

$$\sin \theta_n = \frac{n\lambda}{2d} \qquad (6\cdot 3)$$

This is Bragg's Law.

97. Crystal Structure. Neither Laue nor Bragg knew beforehand the structure of the crystals used. They knew that some kind of particles must be lined up to form the natural cleavage planes of crystals, but it was usually supposed that these particles were molecules. To Bragg goes the credit for solving this problem. By diffracting X-rays from

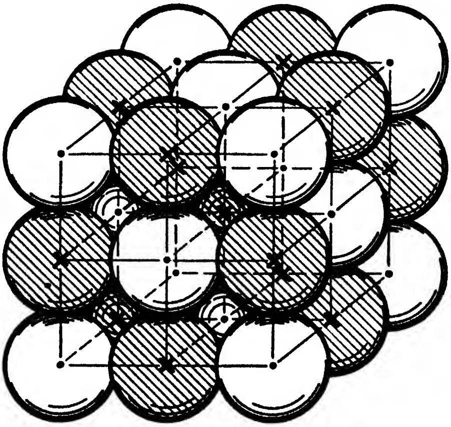


FIG. 82. CRYSTAL STRUCTURE OF ROCK SALT. • Sodium atoms. × Chlorine atoms.

different surfaces of a rock-salt crystal, surfaces obtained by cutting it in various directions in addition to the natural surfaces, he proved that atoms and not molecules line up to form crystals. His results for rock salt are represented by the crystal model shown in Fig. 82; the light balls represent the sodium atoms, and the dark ones, the chlorine atoms. The sodium and the chlorine atoms are lined up alternately in rows, in three mutually perpendicular directions. Since every sodium atom is equidistant from *six* chlorine atoms, and each chlorine atom is in like manner surrounded by six sodium atoms, there is indeed no molecule in the simple sense of the word. The molecular symbol, NaCl, means merely that there are equal numbers of both kinds of atoms in a crystal of rock salt.

Once the arrangement of the atoms in a crystal is known, it is possible to compute the distance between the atoms by means of the data found in Chapter II. For example, for rock salt,

$$\text{Specific gravity} = 2.163$$

$$\text{Molecular weight} = 23.00 + 35.46 = 58.46$$

and one gram molecule contains 6.023×10^{23} atoms of sodium and an equal number of chlorine atoms, or 12.046×10^{23} atoms of both kinds. Hence there are

$$\frac{2.163}{58.46} \times 12.046 \times 10^{23} = 44.57 \times 10^{21} \text{ atoms}$$

in one cubic centimeter of rock salt. The number of atoms along each edge of such a centimeter cube is determined by taking the cube root of this number, or

$$3.546 \times 10^{7} \text{ atoms per centimeter}$$

and the distance between centers of adjacent atoms is the reciprocal of this, or

$$2.820 \times 10^{-8} \text{ cm} = 2.820\text{A}$$

Other crystals may be analyzed in similar fashion. Indeed, one of the very important scientific uses for X-rays is the analysis of crystal structure. See Sec. 186.

98. X-Ray Spectra. Once the grating space has been computed for the crystal of a Bragg spectrometer (Iceland spar crystals are frequently used for this purpose), X-ray wavelengths may be measured and X-ray spectra studied. The results of such studies will be discussed in Chapter X. At present it will be sufficient to state that X-ray spectra consist of a continuous spectrum, analogous to white light of the visible spectrum, upon which is superimposed a simple "line" spectrum, the frequencies of these "lines" being determined by the atomic nature of the target from which they come. X-ray wavelengths range from a small fraction of an angstrom unit, for the hardest, up to several angstrom units for the softest ordinarily used.

99. Refraction of X-Rays. Later experiments have shown that X-rays are refracted very slightly by prisms, the refraction being in the direction opposite to that in which visible light is refracted; they are bent slightly towards the *vertex* of the prism. This means that they travel faster through the prism than they do in free space. For this reason they are also totally reflected at almost grazing angles with the surface, just as light is totally reflected at the inside surface of a 45-degree prism.

Because of this total reflection it is possible to measure X-ray wavelengths with an ordinary ruled diffraction grating. It is used as a reflection grating, the X-ray beams making very small angles with its surface. The wavelength values obtained in this manner are now in acceptable agreement with those obtained with crystal gratings, although this was not so at first. (The discrepancy was due principally to the erroneous value of e by means of which the crystal-grating space was then computed.)

100. The Electromagnetic Nature of Light Waves. In the introductory paragraph of Chapter V a question was raised for which we may now consider a more complete answer. What is the physical nature of light waves? This question no longer pertains to visible light alone, but to the full spectrum of light, visible and invisible, from the longest infrared waves to the shortest waves in the ultraviolet. To this we may add also X-rays at the one end and radio waves at the other end of the spectrum. All these waves, diverse as their respective specific properties may be, have certain fundamental characteristics in common. Most significantly, all are transverse waves, all travel in a vacuum, and in a vacuum all of them travel with the same speed. These common properties prove that all these waves are of essentially the same type, the physical characteristics of all being revealed most clearly by the radio waves.

This is to say that light waves, including X-rays, are electromagnetic waves identical to radio waves except for differences due to wavelength or, more exactly speaking, frequency. Both light waves and radio waves are waves of pulsating electric and magnetic force-fields at right angles to each other, as shown in Fig. 50. As long as they are in empty space, nothing moves. When they pass through matter, electric charges are set into motion, usually by the electric field. It is the electric field which we detect by the scattering experiment described in Sec. 48, and it is the direction of the electric field which is referred to as the direction of "vibration" in connection with other phenomena of polarized light. The magnetic field is for the most part an indispensable but comparatively "silent" partner.

Aside from differences in frequency, the only distinctions between these different ranges of wavelengths are those which relate to their origins: radio waves are produced by oscillatory electrical circuits, light waves by thermal or electrical excitation of atoms and molecules, and X-rays by electron bombardment of the target of a special type of tube. At one time these ranges were quite limited in extent, being separated from one another by broad gaps in the spectrum in which no waves had been detected. These gaps no longer exist, so that even this artificial distinction has disappeared. The entire electromagnetic spectrum has now been studied, and the physical identity of the waves produced by these different methods has been further verified by direct comparison in the overlapping regions.

The gap between the ultraviolet and X-rays was closed by means of ruled diffraction gratings used at near-grazing incidence (as mentioned at the end of Sec. 99) and in a vacuum. The overlap is from about 200A, for the longest X-rays, to about 20A, for the shortest ultraviolet waves measured. The gap between the infrared and radio waves was closed in 1923 by Nichols and Tear. Their radio transmitter was of the spark-gap type used by Hertz, but of microscopic size, the antenna consisting of two tungsten rods, each $\frac{1}{10}$ mm long by $\frac{1}{10}$ mm in diameter, placed end to end with a tiny spark gap between them. Its fundamental wavelength was 0.44 mm. In their detector for these waves tiny strips of platinum metal deposited upon a mica vane served as a multiple receiving antenna. The electric currents induced in these platinum strips by the radio waves falling upon them heated the surface of the mica vane, and the residual air (this receiver vane was suspended in air at very low pressure) produced a radiometer force against the mica vane which turned it by a measurable amount. A lens of paraffin was used in front of the transmitter to produce a parallel beam, and a similar lens was placed in front of the receiver to focus the beam upon the mica vane. A special form of interference apparatus, really a reflection type of diffraction grating, served to measure the wavelengths. With this apparatus Nichols and Tear measured both the fundamental wavelength of their transmitter and also its first overtone, or second harmonic, having a wavelength of 0.22 mm. Furthermore, with this same detector they measured heat-produced infrared waves of 0.324-mm wavelength.

Even earlier (1901) Nichols and Hull performed an experiment which gave other very direct evidence for the electromagnetic nature of light waves. According to the electromagnetic theory, light waves should exert a pressure against any surface upon which they fall.[4] If the surface absorbs the light, the pressure is equal to the intensity of the beam divided by the speed of light. If the surface is perfectly reflecting, the pressure is twice this; but under any conditions it is very small and very difficult to detect. Nichols and Hull were able, however, to measure this pressure, and they found it of the magnitude predicted by the electromagnetic theory. All of the evidence, including much more which may not be mentioned here, proves the electromagnetic nature of light waves as certainly as anything may be proved concerning light. On the other hand, the phenomena which will be discussed in the next chapter cannot be accounted for by any kind of light waves.

PROBLEMS

1. Compute the wavelength of X-rays which are reflected by a calcite crystal at an angle of incidence of 8 degrees (with the surface). Take the crystal grating space as 3.03A.

2. At what angles of incidence must an X-ray beam of 0.50A wavelength fall upon the face of a rock-salt crystal to be strongly reflected?

3. X-rays of 0.35A wavelength are reflected at an angle of 7.2 degrees. Compute the crystal-grating space, if this is the third-order reflection.

REFERENCES

For Additional Reading

J. A. Eldridge, *The Physical Basis of Things*, Chapter XIX.
Physics Staff, University of Pittsburgh, *Atomic Physics*, Chapters V and VII.
G. E. M. Jauncey, *Modern Physics*, Chapters X and XI.
G. F. Hull, *Modern Physics*, parts of Chapter VII.
J. B. Hoag, *Electron and Nuclear Physics*, Chapter IX.
J. D. Stranathan, *The "Particles" of Modern Physics*, Chapter 7.
E. C. Watson, "Discovery of X-Rays," *American Journal of Physics*, Vol. 13, pp. 281–290, 1945.
A. W. Hull, "Thirty Years of X-Ray Research at the General Electric Research Laboratory," *American Journal of Physics*, Vol. 14, pp. 71–79, 1946.

---

[4] This is not the cause of motion in the radiometer sometimes seen in a jeweler's window. The radiometer motion is of thermal origin, and is caused by the gas molecules which remain in the bulb in spite of its relatively high evacuation. The radiometer action is greatest against the blackened surfaces of the vanes, since these surfaces are most heated by the sun's rays, and is least against the polished surfaces; this is just the reverse of the effect produced by true light pressure.

E. F. Nichols and J. D. Tear, "Joining the Electric Wave and Heat Wave Spectra," *Annual Report of the Smithsonian Institution*, 1923, pp. 175–185.

J. E. Hoover, "Physical Science in the Crime Detection Laboratory," *Review of Scientific Instruments*, Vol. 9, p. 335, November, 1938.

G. A. Morton and L. E. Flory, "An Infrared Image Tube and Its Military Applications," *RCA Review*, Vol. 7, pp. 385–413, 1946.

For Advanced Reading

W. H. Bragg and W. L. Bragg, *X-rays and Crystal Structure*.

R. W. Wood, *Physical Optics*, Third Edition, Chapter XX ("Fluorescence and Phosphorescence").

G. L. Clarke, *Applied X-rays*.

CHAPTER VII

LIGHT PARTICLES AND ELECTRON WAVES

101. Introduction. In the preceding chapters there have been described the phenomena which verify the existence of light waves and which establish their electromagnetic character. The evidence is so complete and the conclusions are so beautifully consistent that it seems impossible that this description of the nature of light should not be capable of explaining any phenomenon involving light. Nevertheless, H. Hertz, the same man who devised the experimental proof for the existence of electromagnetic waves, discovered also a phenomenon which cannot be satisfactorily explained by light waves. This phenomenon is the photoelectric effect.

102. Photoelectric Effect. The photoelectric effect may be demonstrated very simply by means of an insulated zinc plate connected to a gold-leaf electroscope, as shown in Fig. 83. If this plate, Z, is charged *negatively*, and then exposed to light from the bare arc lamp, A, it will lose its charge, as will be shown by the falling of the leaf of the electroscope. If the arc is close to the plate the loss of charge is rapid, but if it is moved farther away the rate at which charge escapes will be less. Quantitative experiments show that the rate at which the charge escapes is directly proportional to the brightness of the illumination falling upon the plate. On the contrary, if this plate is charged *positively*, no such loss of charge takes place. Evidently light falling upon the plate enables only negative charges to escape from it; and more detailed experiments show that this negative charge is carried by electrons. If, for example, the cathode, C, of the tube shown in Fig. 4 is made of zinc, and the tube itself made of quartz (or provided with a quartz window near C) and highly exhausted, an electric discharge may be sent through the tube when this cathode is

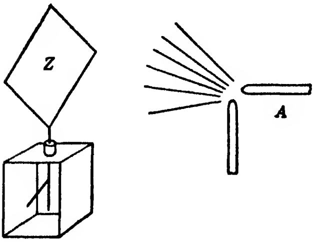


FIG. 83.

irradiated by ultraviolet light. The narrow beam of rays which will then pass through the hole in *D* will cause fluorescence at the end of the tube, it will be deflected by electric and by magnetic fields, and the magnitudes of these deflections will prove that these particles are indeed electrons.

If, when the zinc plate of Fig. 83 is charged negatively, a clear glass plate is interposed between the arc and the zinc plate, the movement of the electroscope leaf immediately stops. Evidently the photoelectric effect in zinc is due to the very short ultraviolet waves which are absorbed by clear glass, and the visible light and the small bit of ultraviolet light which do pass through the glass plate are incapable of producing any photoelectric effect.

Experiments with an aluminum plate show similar results but to a lesser degree, and quantitative experiments show that the wavelengths of light required to produce a photoelectric effect with aluminum are considerably shorter than those required for zinc. On the other hand, sodium and potassium surfaces are sensitive to even longer waves than zinc, potassium being sensitive to the shorter waves of the visible spectrum. Cesium is sensitive throughout the visible spectrum and into the infrared. It is impossible, however, to show the photoelectric effects with these metals by such a simple experiment as that used for zinc, since the surfaces of these metals become contaminated immediately upon exposure to air. Even zinc and aluminum must be freshly scrubbed with sandpaper to show good results, and experiments with the more sensitive metals are of necessity performed in a vacuum.

103. Photoelectric Cells. Such experiments are nowadays readily performed, since photoelectric cells, or evacuated bulbs containing such photosensitive surfaces, usually of cesium, are manufactured in great numbers for commercial purposes. Such a commercial photoelectric cell is shown in Fig. 84. *P* is the photosensitive surface (a copper or silver plate coated with cesium metal) which serves as the cathode of the cell, and *A* is a wire which serves as the anode to collect the electrons which escape from *P*. The construction of such a cell requires the greatest care. First the bulb is evacuated and baked until it is free from all gases which can contaminate the sensitive surface, then the surface of *P* is lightly oxidized by admitting just the right amount of oxygen, and finally cesium vapor is introduced, usually by a chemical reaction produced within the bulb itself. The cesium vapor condenses upon the oxidized surface of *P* in a layer about one atom deep. The bulb is then sealed off with the space within it evacuated as perfectly as is possible with the best of modern pumps. Similar cells are also made with sodium, potassium, or rubidium surfaces. Cells for use with ultraviolet light must be made of fused quartz instead of glass, or else have quartz windows.

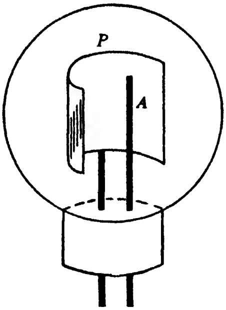


Fig. 84. Photoelectric Cell.

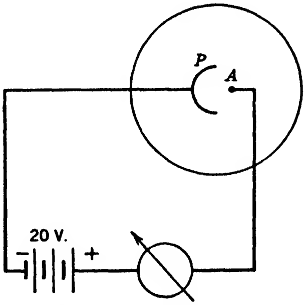


Fig. 85. Photocell Circuit.

104. Photocell Circuits. Figure 85 shows a circuit for use with such a photocell. With this circuit, it may be shown that the *current* which flows through the galvanometer *is proportional to the intensity of the light* falling upon the inside surface of P. For any particular cell the spectral sensitivity extends from some low frequency limit, which is determined by the nature of the photosensitive surface, through a maximum, and to an upper frequency limit set usually by the transmission of the walls of the bulb enclosing the cell. Cesium photocells may be made which have approximately the sensitivity range of the eye. Photoelectric currents are always very small and must either be measured by a sensitive galvanometer or else be amplified by some means such as the electron-tube amplifier shown in Fig. 86. The operation of this circuit is as follows:

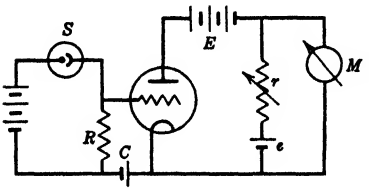


Fig. 86. Photocell-Amplifier Circuit.

When the photocell S is dark, so that no current flows through the high resistance R, the potential on the grid of the amplifier tube is negative, and equals the e.m.f. of the cell C. When light falls upon the photocell, the electrons liberated by the light will set up a current in the resistor, R, thereby changing the potential of the grid by an amount equal to the product of that current by the value of R, which may be 10^7 ohms or higher. This change in the grid potential will cause a corresponding change in the plate current, which will be measured by the meter M; and this change in current may be as much as 10,000 times as great as the original photocurrent. In order that a sensitive meter may be used for M, the current which flows through the amplifier tube when the photocell is dark is by-passed by the balancing circuit, r, e. This circuit is adjusted so that the current × resistance drop through r just balances the e.m.f., e, when the cell is dark. Any *change* in this current will, however, flow almost entirely through the low-resistance meter M.

Some commercial photocells contain a very small amount of an inert gas such as argon, which has the effect of amplifying the photocurrent. Although such cells are considerably more sensitive than the high-vacuum types, they are less stable in their operation and require quite critical adjustment of their operating voltage to take advantage of their full sensitivity.

105. "Sandwich"-Type Photoelectric Cells. Still another type of photoelectric cell consists of a very thin layer of some semi-conducting substance such as cuprous oxide, sandwiched in between two metallic conducting layers, as is shown in Fig. 87. The bond between the oxide layer and the copper base has been so formed by special chemical and physical treatment that when this surface is illuminated electrons will flow across it from the oxide to the copper. The upper metallic layer is made thin enough to let most of the light through. The sensitive surface may be formed at the top instead of at the bottom of the "sandwich," and other semi-conducting substances may be used, such as selenium, which is much more sensitive to light than cuprous oxide is.

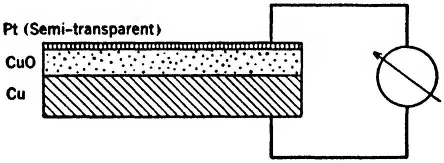


FIG. 87. "SANDWICH"-TYPE PHOTOCELL.

Fairly large currents (as compared to those obtainable from other photocells) will flow through a low-resistance galvanometer connected to such a photocell, and no battery is needed. The light itself provides the energy needed to drive the current. When the galvanometer resistance is sufficiently low, the current from such a cell will be quite nearly proportional to the intensity of the light falling upon it. But when the galvanometer resistance is high, most of the current set up by the light flows back through the semi-conducting layer itself, and the current through the galvanometer increases more and more slowly as the light intensity increases. For this reason the currents from such cells cannot readily be amplified, since most amplifying devices have high input resistances.

106. The Photoelectric Equation. Current will flow in *any* photoelectric cell under the influence of light alone, just as it does in the sandwich type; but in the vacuum-type cell the current is much smaller, and a very sensitive galvanometer is required to measure it. Furthermore, a sufficiently sensitive meter will show that in general current continues to flow even when the potential of the photosensitive surface is made positive by a considerable fraction of a volt. This means that some at least of the electrons leave the photosensitive surface with enough kinetic energy to counteract the tendency of the positive potential to prevent them from escaping, and the *maximum* kinetic energy possessed by any of the electrons can be measured by finding the potential which will *just stop* the current.

This experiment cannot be performed with commercial cells and is exceedingly difficult even with cells especially constructed for the purpose. It was first performed by Millikan, in 1916. Since light waves are electromagnetic waves, it is natural to expect that, the more intense the light, the stronger would be the electric force exerted by its waves upon the electrons in the photosensitive surface, and the greater consequently would be their kinetic energy upon leaving that surface. But this proves to be not at all the case. Millikan found that the kinetic energies of these electrons depend upon only the *frequency* of the light, regardless of whether the light is weak or strong. Furthermore, he found that the *maximum* kinetic energy produced by light of any particular frequency, f, is directly proportional to the amount by which that frequency exceeds a frequency, f_0, which is the limiting frequency below which no electron emission can occur. If the proportionality factor is represented by h, this experimental law can be written

$$\tfrac{1}{2}mv^2 = h(f - f_0) = hf - hf_0 \qquad (7\cdot1)$$

and if hf_0 is written as w_0, this may be rewritten as

$$hf = w_0 + \tfrac{1}{2}mv^2 \qquad (7\cdot2)$$

This is the photoelectric equation of Einstein. For reasons to be explained in Sec. 109, the proportionality factor, h, is called the "quantum" constant or Planck's constant. Its value, as found by this and other experiments, is

$$h = 6.62 \times 10^{-27} \text{ erg} \times \text{sec}$$

A more precise value is given in Appendix II.

107. Light Particles (Photons). As has been suggested in the previous section, this photoelectric equation cannot be explained by the wave theory of light. If, however, we knew nothing of the many phenomena which so convincingly demonstrate the existence of light *waves*, we should hail this phenomenon of the photoelectric emission of electrons as a beautiful proof for the existence of light *particles*. That is to say, if a light beam may be considered to be a stream of particles (which will hereafter be called **photons**), and if the kinetic energy of each of these photons is equal to the quantity hf, then the photoelectric equation given above, (7·2), represents the *energy* equation for a collision of one of these photons with the photosensitive surface in such a way that *all* its kinetic energy is given to *one* electron. According to this idea, w_0 is the amount of this energy which is used up (or converted into potential energy) in pulling the electron away from the surface, which attracts it, and $\frac{1}{2}mv^2$ is the kinetic energy which the electron then has left.

This theory also accounts in a very simple way for the existence of a low frequency limit for the photoelectric effect; the quantity of energy, w_0, is the least amount of energy that an electron can receive and still escape from the surface attraction. Its value depends upon the nature of the photosensitive surface, and it is called the work function for that surface; see Sec. 191. When the frequency f is less than f_0, the energy hf which an electron receives from the light is less than the amount w_0 needed for its escape; it may start away from the surface, but it will fall back again.

According to this theory, **the frequency of the light is only a measure of the amount of energy possessed by the photon.**

The intensity law, or the exact proportionality between the photocurrent and the intensity of the light, is also accounted for very nicely by this theory, which explains the intensity of light as being proportional to the *number* of photons falling upon the surface per second per unit area.

Additional and equally good evidence for this photon theory of light is given by X-rays. When an X-ray tube is operated at a constant potential, V, the frequencies of the X-rays emitted from its target are all lower than a frequency limit, f_1, and X-rays of higher frequency may be obtained only by increasing V. Duane and Hunt proved experimentally that f_1 is directly proportional to V, and, furthermore, that this proportionality could be written

$$Vc = hf_1 \qquad (7\cdot3)$$

where h is the same "quantum" constant found for the photoelectric equation. Indeed, save for omission of the factor w_0 (which, with

X-rays, is entirely negligible in comparison with the other terms), this *is* the photoelectric equation, and the explanation for this phenomenon is fundamentally the same, the process here being the *inverse* of the photoelectric process. In the photoelectric process the metal surface is bombarded by photons, and electrons are emitted. In this process the metal surface is bombarded by electrons, and photons are emitted; thus the maximum frequency of X-rays is emitted when a cathode-ray electron gives all its energy to one photon of that maximum energy. Equation (7·3) is the energy equation for this exchange, since Ve is the energy of the cathode-ray electron, and hf_1 is the kinetic energy of the emitted X-ray photon.

The direct photoelectric effect is also produced by X-rays, and the agreement with the photoelectric equation is equally good, although here the phenomenon is more complicated. It will be discussed more fully in Chapter X.

All these experiments present the best of evidence that light consists of *photons*, or particles which have kinetic energy equal to hf, and we may further speculate upon what kind of particles these photons may be. It is certain that they cannot be ordinary particles. In the first place they must all travel at the same speed, namely, the speed of light, c, and the differences in kinetic energy must be due to differences in mass. This means that, when a photon loses some of its kinetic energy, its speed does not slacken: it just *loses mass.* And if it is absorbed so that it loses all its energy, it does not come to rest: it just ceases to exist! Mathematical theory shows that for such restless particles as these the kinetic energy is mc^2, or twice as great as would be computed by the kinetic-energy formula for ordinary particles.

These properties of photons may be summarized as follows:

$$\text{Kinetic energy} = hf = mc^2 \tag{7·4}$$

$$\text{Mass} = m = \frac{hf}{c^2} \tag{7·5}$$

$$\text{Momentum} = mc = \frac{hf}{c} = \frac{h}{\lambda} \tag{7·6}$$

Only the first of these relations has been directly verified by the experiments so far considered. The others are verified by the phenomenon to be described in Sec. 108.

108. The Compton Effect. A. H. Compton has shown that the scattering of X-rays by electrons furnishes a means for measuring the momenta of X-ray photons, and that the momenta thus measured are in exact agreement with the value, h/λ, given by the photon theory of light. X-rays are scattered in all directions when a beam of X-rays falls upon a block of some material such as carbon, the scattering being entirely comparable to the scattering of visible light by colloidal particles (see Sec. 42), except that the scattering particles in the case of X-rays are mostly the electrons of the scattering substance. One would naturally expect the scattered X-rays to be of the same wavelength as that found in the original beam, and this is partly so. A considerable portion of the scattered X-rays, however, are of longer wavelength, the increase in wavelength being greatest for those scattered directly back in the direction of the source. Again this change in wavelength is a phenomenon which cannot be explained by the wave theory but, as Compton showed, can be beautifully explained by the photon theory. According to this theory, a change in wavelength occurs when the scattering is the result of a collision between an X-ray photon and a free electron (or an electron in the outer part of an atom, where it is so loosely held that it is practically free); and this collision takes place just as does any collision between two perfectly elastic bodies. *The total momentum remains unchanged.*

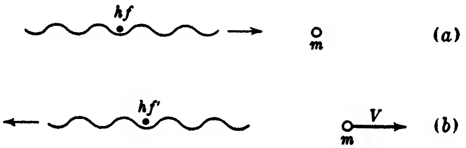


FIG. 88. COMPTON EFFECT. (a) Before collision between the photon and the electron. (b) After collision.

Consider for simplicity a head-on collision. Even an X-ray photon is much lighter than an electron, so that it rebounds with almost its original energy and with its momentum almost completely reversed—almost, but not quite. (See Fig. 88, which represents these conditions schematically.) Before collision the electron is at rest, and the energy of the photon is hf, where f is the frequency of the X-rays in the primary beam. After collision, the energy of the electron is $\frac{1}{2}mv^2$, the frequency of the scattered X-rays is f', and the energy of the rebounding photon is hf'. Since no energy is lost,

$$hf = hf' + \tfrac{1}{2}mv^2$$

or

$$h(f - f') = \tfrac{1}{2}mv^2 \tag{7·7}$$

Also, while the momentum of the photon has changed from h/λ to

$-h/\lambda'$, and that of the electron from zero to mv, the total momentum has not changed, so that

$$\frac{h}{\lambda} = -\frac{h}{\lambda'} + mv$$

or

$$mv = \frac{h}{\lambda} + \frac{h}{\lambda'} = \frac{2h}{\lambda}, \text{ approximately} \qquad (7\cdot8)$$

These two equations together will predict the value of the change of frequency $(f - f')$, to be found experimentally for X-rays scattered backwards towards the source, if photons do exist and this theory is correct in other respects. Solving them simultaneously,

$$f - f' = \frac{2h}{m\lambda^2} \qquad (7\cdot9)$$

The change in wavelength, $(\lambda' - \lambda)$, then may be obtained by dividing both sides of equation (7·9) by c, remembering that $f/c = 1/\lambda$, and making the approximation that $\lambda\lambda' = \lambda^2$. The result is

$$\lambda' - \lambda = \frac{2h}{mc} = 0.048 \times 10^{-8} \text{ cm} = 0.048\text{A} \qquad (7\cdot10)$$

For scattering in other directions a similar proof shows that the change of wavelength should be

$$\lambda' - \lambda = \frac{h}{mc}(1 - \cos\theta) = 0.024(1 - \cos\theta)\text{A} \qquad (7\cdot11)$$

where θ is the angle between the direction of the scattered beam and the direction of the primary beam. This change of wavelength is large enough to be measured with precision by an X-ray spectrometer, and the measurements agree, within the precision of measurement, with the predicted values for all angles measured.

Compton went further and designed a cloud-chamber experiment in which the directions taken by the recoiling electron and by the scattered X-ray photon both could be observed, and the measured angles checked with the angles computed for collision between two perfectly elastic bodies, the one having the mass of the electron, and the other the mass given by equation (7·5) for the mass of a photon.

The photon theory of light is the only theory that will consistently explain all these phenomena, and the evidence for the existence of photons is now just as good as the evidence for the existence of light waves. To satisfy all the evidence, *light must be both particles and waves!* Since the evidence exists for both theories, the physicist accepts both, using the one or the other theory as it best suits his need. As will be seen later, it turns out that these two theories are not so antagonistic as at first they appear.

It is interesting here to note that a single phenomenon may at the same time illustrate both the particle and the wave properties of light. Thus the process of scattering of X-rays, which has just been shown to demonstrate most conclusively the particle character of X-rays, also exhibits one of the fundamental properties of waves. The X-rays that are scattered in a direction perpendicular to the primary beam are polarized, just as visible light, scattered by colloidal particles, is polarized. (See Sec. 48.)

109. Quantum Theory (Historical). The beginnings of these ideas relating to the particle theory of light were made about 1900 by Planck, who at that time developed what is known as the "quantum" theory to explain the relatively complex phenomenon of thermal radiation from a black body. (See Sec. 200.) In order to explain the distribution of energy among the wavelengths in the spectrum of light radiated by a hot body he made the bold assumption that, for some unknown reason, energy is radiated in this case only in "quanta," that is, in discrete amounts, of sizes proportional to the frequencies of the waves radiated. The proportionality factor is again the factor h, defined in Sec. 106, and for this reason this factor is called the quantum constant, or Planck's constant.

In 1905 Einstein proposed the *photoelectric equation* as a logical theoretical consequence of Planck's theory, and the experimental verification of this equation, and the new and fundamental ideas which have followed, have already been discussed. Here, as happens relatively often in the history of scientific discovery, a new and revolutionary fundamental idea originated in connection with a relatively involved phenomenon (thermal radiation in this case), and it was only later that the simpler and more fundamental bases for it were discovered. We, who have the advantage of being able to view the development of this idea in retrospect, have been able here to begin consideration of these phenomena with those which present it most directly.

110. Electron Waves. The dilemma in which the experimental evidence leaves the theories of light is not unique. In 1924, Louis de Broglie made the apparently ridiculous suggestion that *particles of matter* could sometimes manifest themselves as *waves*, whose wavelengths would be related to the momenta of the particles in just the same way that the wavelength and the momentum of a photon are related, i.e.,

$$\text{Momentum} = \frac{h}{\lambda} \tag{7·12}$$

or

$$\lambda = \frac{h}{\text{Momentum}} = \frac{h}{mv} \tag{7·13}$$

Also, the frequency of their vibrations would be expressed by the equation[1]

$$f = \frac{mc^2}{h} \tag{7·14}$$

which is again the exact counterpart of the relation between frequency and mass for photons, given by equation (7·4).

For golf balls or marbles or even grains of sand these matter waves would be unmeasurably short. For cathode-ray electrons, however, they would be comparable with X-rays in length and might be measured with a crystal grating. For 100-volt electrons the wavelength is 1.225A, and for 10,000-volt electrons, 0.1220A.

111. Electron Diffraction. Ridiculous as did this suggestion appear to be, it has been proved true. First Davisson and Germer (in 1927) in this country, and very soon afterwards G. P. Thomson (who, interestingly enough, is the son of J. J. Thomson, the man who measured the particle properties of the electron), performed experiments in which electrons are diffracted just like X-rays. Davisson and Germer used an apparatus analogous to a Bragg X-ray spectrometer, and worked with electrons whose motions were produced by a potential difference of only a hundred volts or so. When they projected these fairly slow electrons against the face of a large single crystal of pure nickel, they found that the electrons were reflected by the crystal planes only for angles which corresponded to constructive interference for the electron waves.

Thomson worked with electrons which were set into motion by a potential difference of 10,000 to 100,000 volts or more, and used a method similar to that of the Laue experiment with X-rays. When a very fine pencil of such high-speed electrons is shot through a very thin film of metal or other crystalline substance and thence onto a fluores-

---

[1] At low and moderate speeds this frequency is constant and equal to 1.235×10^{20} per second for an electron. At high speeds both theory and experiment show that the electron mass increases (the significance of this will be explained in Sec. 212) and the frequency of pulsation increases accordingly.

cent screen or a photographic plate some distance away, a pattern such as the one shown in Fig. 89 is produced upon the plate. This diffraction pattern consists of rings about the central spot instead of the pattern of spots seen for X-rays in Fig. 79, since the metallic film contained many small crystals arranged in random order. The diffraction rings are formed by those crystals which happen to be oriented at such angles with the primary beam that their crystal planes are able to reflect the electron waves in phase to the screen. An identical pattern would be produced by X-rays of the same wavelength, passed through this same polycrystalline film.

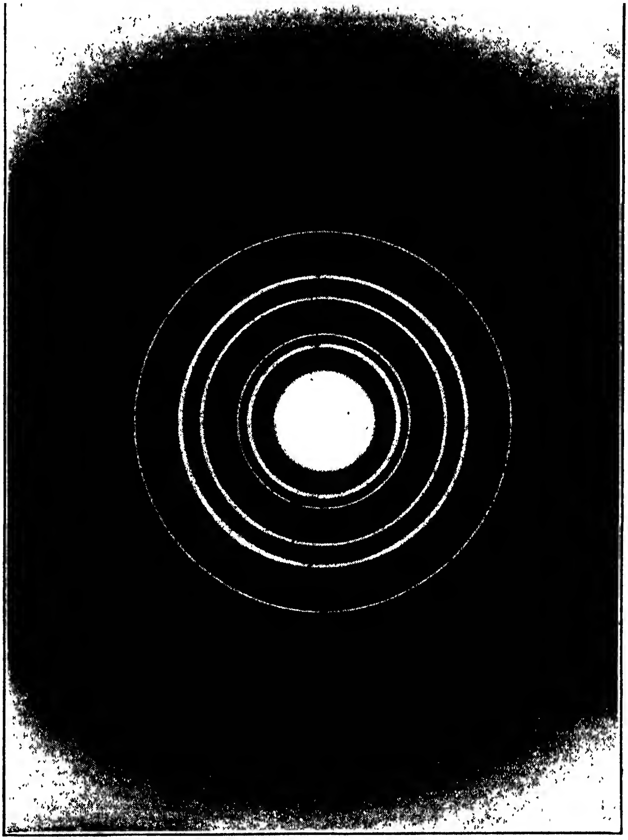


FIG. 89. DIFFRACTION OF ELECTRONS BY AN ALUMINUM FILM. Photograph by Dr. Lester H. Germer of the Bell Telephone Laboratories.

These experiments present evidence for the existence of electron waves which is just as good as any evidence that we have for the existence of X-ray waves or for visible light waves. Furthermore, at the same time that these diffracted electron beams are exhibiting wave properties, they are still electrically charged particles, since a transverse magnetic field will deflect the whole pattern up or down upon the screen or photographic plate. To satisfy all the evidence, *electrons must be both particles and waves!*

112. Waves and Particles. Thus we must not only consider that light waves are particles, but also that electrons, as well as probably all other things which we are accustomed to think of as particles, are also waves. This second dilemma, instead of increasing the mental discomfort of the physicists, actually relieved it. The scientist likes even his dilemmas to be consistent, and if *everything* in the physical world is capable of manifesting itself both as particles and as waves, perhaps he can find a rational way of making these two aspects of light and of matter compatible with each other. To a very considerable degree such an objective has been attained, although we may hope to understand the relationship between waves and particles better at a future time, and perhaps then be able to express it better. The currently accepted description of this relationship is as follows:

> **The motions of particles are guided entirely by the waves to which they belong.**

Light particles or photons, for example, follow paths determined by the light waves to which they belong. Where the waves are of zero amplitude, there the photons cannot go. Where the waves are of maximum amplitude, there photons are most likely to be found, although this does not exclude the possibility of their being found sometimes where the amplitude is smaller. Consider, for example, the fringes formed by a biprism, which are described in Sec. 31 and illustrated in Fig. 19. No photons strike the screen at the centers of the dark fringes, while the greatest numbers of photons fall upon it at the centers of the bright fringes; the probability that a photon will strike the screen at any point is proportional to the square of the amplitude of the waves (that is, to their intensity) at that point. According to this description, the energy of the light is carried by the photons, and the waves themselves have a somewhat unreal existence, serving only to guide and to direct the energy by way of the photons.

113. Group Velocity. Before considering the extension of this description to electron waves we must first consider a very peculiar characteristic of these waves: The discerning student is going to discover that, since the wavelength of a cathode-ray electron is h/mv, and its frequency is mc^2/h, the wave speed, which is the product of these two expressions or c^2/v, must be greater than the speed of light. This puzzles him. If the electron waves travel that fast, why do they not leave the electron far behind? The explanation is that the wave representing any particular electron is not a continuous wave of constant amplitude, but rather a *wave group* such as is diagrammed in Fig. 90; it is this *group* as a whole which travels with the particle speed, v. The individual waves travel through the group with the much higher speed, c^2/v, appearing at A, the rear of the group, swelling to maximum amplitude at the center of the group, and then shrinking and finally disappearing as they travel out of the group at B. Water waves behave in similar fashion, as anyone who throws a stone into a still pond may observe.

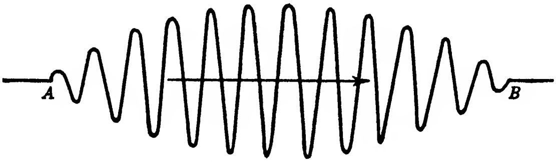


FIG. 90. WAVE GROUP.

114. The Uncertainty Principle. Such a group of waves is the equivalent of a combination of many continuous waves of slightly different wavelengths. At the center of the group these continuous waves would all be in phase and add up to a maximum effect, but at some distance in either direction from the center they would get out of phase and cancel each other's effects from there on. If now the probability of finding an electron at any place is determined by the intensity of its wave at that place, then the most probable position for the electron which travels along with the group will be at the center of the group, although it may also be anywhere else in the group where the intensity is not zero. This indefiniteness or *uncertainty* as to its exact *position* is an unescapable feature of the wave-particle theory. At the same time, since the wave group is made up of a range of continuous waves of different wavelengths, each wavelength corresponding to a slightly different value of the momentum, the *momentum* of the electron is likewise *uncertain*, in that its value has a chance of lying anywhere within this range of values. Both these ranges of uncertainty may be very small, their product being of the order of magnitude of the very small quantity h (more exactly speaking, $h/2\pi$), but neither the position nor the momentum may be made exact without making the other entirely indeterminable. Thus exact momentum means a continuous wave extending indefinitely in both directions, and the probability of finding the electron is the same anywhere along the wave. On the other hand, a wave group short enough to locate the electron within a space as small as its own diameter would require a group of waves containing an extremely wide range of wavelengths, with a correspondingly enormous uncertainty as to its momentum.

> **The range of uncertainty of the position of a particle, multiplied by the range of uncertainty of its momentum, will always be of the order of magnitude of the constant, $h/2\pi$.**

This statement is called the uncertainty principle. Even before the discovery of the wave properties of electrons, this principle was enunciated by Heisenberg (1925), who pointed out that such uncertainties in the position and the momentum of a particle could never be detected experimentally, since the very act of making the observation by any experimental means would disturb the position or the momentum of the particle by amounts as great as those given by this principle.

115. Wave Mechanics. Almost all the experimental evidence for the existence of matter waves has been obtained for electrons, since the electron is the only particle which is light enough to have waves of easily measurable length. Even the proton (the hydrogen nucleus) has waves almost too small to measure, although some observations have been made which verify the existence of proton waves. Although heavier particles are out of the question with present known methods, there is no reason for believing that this dual wave-particle aspect of matter, which is revealed by electrons and protons, is not true for all particles great or small; and Schrödinger, Heisenberg, and others have built up from the theoretical suggestions of de Broglie a complete mathematical theory to describe the wave systems corresponding to particles of matter under any circumstances.

This theory is called *quantum mechanics* or *wave mechanics*, and its essential features, in so far as they may be described qualitatively, have been given above. For large particles, such as grains of sand, the waves are so small, as compared to the sizes of the particles themselves, that the positions of the particles are quite definite and the motions predicted by the wave-mechanics theory are exactly the motions described by the simpler "ordinary" laws of mechanics. As smaller and smaller particles are considered, the ordinary laws of mechanics become only approximately correct, whereas the wave mechanics is still exact.

It must be emphasized that matter waves are not the same as light waves or other electromagnetic waves. As has been described above, our experiments always detect the electrons, etc., as particles, not as waves, and our only knowledge of the existence of the waves is *infer-*

ential. This might perhaps also be said of the shorter electromagnetic waves, from light waves on down to X-rays, but, thanks to our ability to produce long electromagnetic waves, we have very considerable knowledge of the nature of the waves themselves. Concerning matter waves we have no such information. If we compute their amplitudes, we may predict (within limits) where to find the particles which they represent, but we can say little else concerning what they themselves may be.

116. Electron Lenses. One of the important developments in applied electronics is the electron microscope, and a consideration here of electron lenses and electron microscopes not only will be interesting in itself but also will bring out several interesting points in connection with the theoretical topics of this chapter.

It will be recalled that the limitation upon the useful magnification of a microscope is set ultimately by the length of the light waves themselves. This is discussed in Sec. 41. No matter how perfect or how powerful the microscope lenses, no detail can be seen that is much smaller than a half-wavelength of the light used. If finer detail is to be seen, shorter waves must be used. X-rays have the requisite short wavelengths, but unfortunately there exists no way of focusing X-rays.

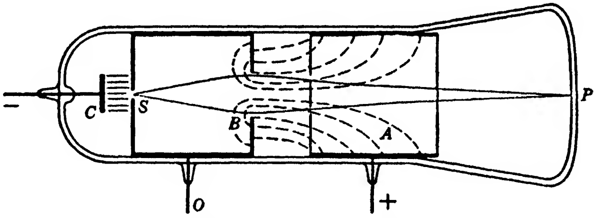


Fig. 91. Electric-Field Electron Lens.

All X-ray pictures are shadowgraphs, and their sharpness is limited by the size of the spot to which the cathode rays converge upon the target. Electrons, on the contrary, have the requisite short wavelength, and both electric and magnetic fields may be used to focus cathode-ray electrons.

An electric-field electron "lens" may be a simple circular hole in a metal plate, with a stronger electric field on one side than on the other side of the hole. Figure 91 shows such an electron lens in cross-section at *B*. Electrons from the cathode, *C*, are accelerated through the small hole, *S*, by a small voltage between *C* and *S*. Thence they go through the electron lens, *B*, which focuses an image of *S* upon the fluorescent screen at the far end of the tube. The electron paths for the outside portions of this cathode-ray beam are shown by the fine lines drawn from S to P. The electric field between the lens plate and the anode, A, is shown by the dotted lines; and the focusing action of the lens is due to those lines of force, which go through the hole and terminate on the left-hand side of the lens plate. As will be seen in the figure, these lines of force have radial components which push the electrons of the cathode-ray beam inward toward the axis of the tube, thus changing it from a divergent beam to a convergent one. The strength of this field determines the power of the lens, so that the beam may be focused by adjusting the voltage between the lens plate and the anode, A, which is a cylindrical tube seen in cross-section.

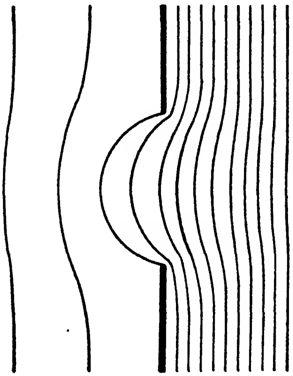


Fig. 92. Equipotential Surfaces in the Neighborhood of an Electrostatic Lens.

The foregoing explanation of the action of this lens is given in terms of the particle properties of the electron, and it is interesting to see how its action could be explained in terms of electron waves. According to the wave-mechanics theory, the electric field changes the refractive index of the region through which the cathode rays pass, and the variations of the refractive index in the neighborhood of the hole are such as to form a lens for the electron waves. A crude representation of this lens may be made by drawing the equipotential surfaces in the neighborhood of the hole, as is done in Fig. 92. To make a comparison with the optical case, one must imagine the region to the left of the hole, where the electron speed is low and the wave speed is high, to have a low value of refractive index,[2] with a higher value of refractive index in the region embraced by the curved equipotential surfaces near the hole, and a still higher value to the right of the lens plate, where the electron speed is high and the wavespeed correspondingly lower. The electron waves are refracted to a focus at P by this "lens," and the electron particles follow the guidance of the waves. Here the phenomenon is on a large enough scale to make both explanations (which may be compared to the ray and the wave explanations in an optical example) agree with each other. A cathode-ray oscillograph (see Sec.

---

[2] By analogy to the optical case, we may define the index of refraction for electron waves in an electrostatic field as $n = c/u$, where u is the wave speed (see Sec. 113) and equals c^2/v. It follows that $n = v/c$. The index of refraction for electron waves in an electrostatic field is thus directly proportional to the electron speed, v.

13) uses an electron lens, similar to this one, to focus its beam of electrons to a fine point upon the fluorescent screen.

A magnetic-field electron lens focuses the electron beam in quite a different manner, but does it just as well if not better. Most electron microscopes which have so far been built have magnetic lenses. As shown in Fig. 93, the focusing field is a magnetic field produced by electric current flowing in a circular coil of wire, *C–C*, wrapped around the outside of the tube, and shown in cross-section in the figure. An electron which moves straight down the axis of the tube will be moving parallel to the magnetic lines of force and will reach the screen without being deviated from its path. On the contrary, an electron which starts out from the cathode in a direction making an angle to the axis will have a component of its velocity perpendicular to the magnetic field and will be deflected sidewise, its path through the magnetic field appearing circular to anyone looking along the axis of the tube.

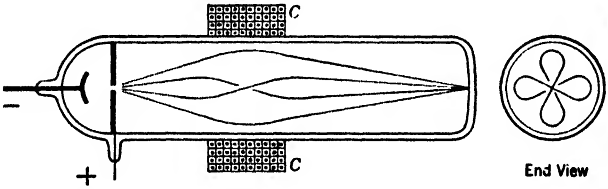


FIG. 93. MAGNETIC-FIELD ELECTRON LENS.

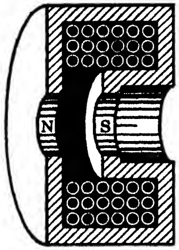


FIG. 93(*a*).

This is indicated in the end view which is shown in Fig. 93. This circular motion, combined with the original forward motion parallel to the axis, makes a helical path which, if the magnetic field is of the correct strength, brings it back to the axis again when it reaches the screen at the far end of the tube. In this manner the cathode-ray beam is focused to a point upon the fluorescent screen. An attempt has been made to draw a few of these helical paths in Fig. 93. A better picture of them is presented by the twisted filaments of colored glass sometimes seen in glass marbles.

For a more powerful lens the magnetic field of the coil is intensified and concentrated on the inner side of the coil by a soft-iron sheath which surrounds the coil, except for a slot along its inner circumference. Figure 93(*a*) shows such a magnet, cut in two across its middle; the two edges of the slot form the poles, N, S, of the magnet.

117. Electron Microscope. Electron lenses may be used to form images in much the same manner as optical lenses. Thus either an electric-field lens or a magnetic-field lens may be focused so as to project a magnified image of the cathode surface upon the fluorescent screen, and if the lens is close enough to the cathode, the magnification may be as much as one hundredfold or more. Such simple electron microscopes have been used to make some very informative studies of the process of electron emission from hot cathodes and from photoelectric surfaces. Electron microscopes may also be used to produce images of objects placed in the path of the electron beam. To make a comparison with optical microscopes, we may say that such objects are "illuminated" by the electron waves. In order that the electron beam may penetrate even the thinnest objects, high-voltage electrons must be used, and the magnetic type of electron lens is found to be more suitable. Figure 94 shows schematically an electron microscope designed for this purpose.

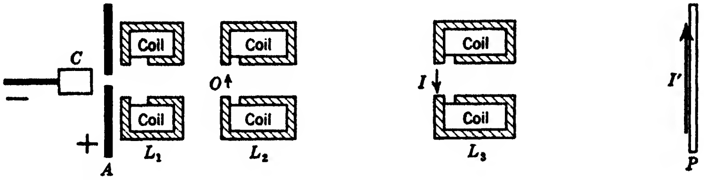


Fig. 94. Electron Microscope. Schematic diagram of the microscope shown in Fig. 95.

Electrons which are emitted by the hot cathode, C, are given high speeds by the potential difference between C and the anode A, and then are condensed upon the object O by means of the magnetic field of the coil L_1. To obtain the large magnifications desired, two focusing coils are used; the system is analogous to a compound projection microscope of the optical type. The magnetic field of the first of these coils, L_2, acts as an objective lens to form the intermediate image, I, and this image serves as object for the second "lens," the magnetic field of the coil L_3, which forms the final image, I', upon the plate P. This plate may be a fluorescent screen, or it may be a photographic plate upon which the cathode rays will leave a developable image.

One of the early commercial electron microscopes is shown in Fig. 95. Magnifications up to 25,000 diameters may be produced in this microscope, the intermediate image being magnified 100 diameters and the final image further enlarged 250 times. The final image is so sharp that useful enlargements up to 100,000 diameters may be made,

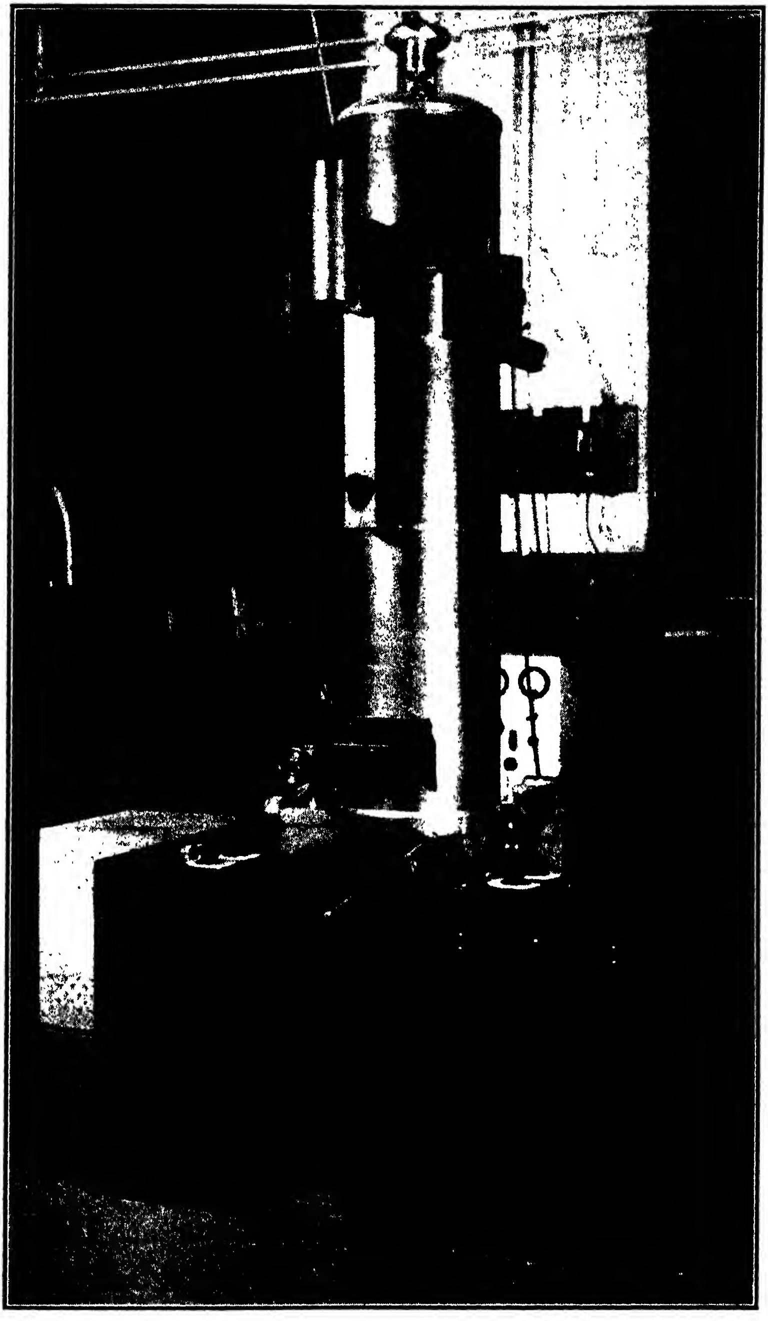


FIG. 95. ELECTRON MICROSCOPE. Designed by Dr. L. Marton, and manufactured for sale by the Radio Corporation of America.

and details as small as 50A may be resolved. Compare these figures with those given in Sec. 41 for optical microscopes! Figures 96 and 97 show specimens of the results obtainable. To attain this precision both the accelerating voltage (from 30,000 volts to 100,000 volts) and the currents through the focusing coils must be held steady to within a few hundredths of one per cent, a precision made possible only by means of very ingenious electronic control circuits.

FIG. 96. STREPTOCOCCI BACILLI, magnified 10,000 diameters.

FIG. 97. TYPHOID BACILLI, magnified 10,000 diameters. The plant-like tendrils are the bacilli. The sharp lines at the left image the curled edge of the supporting film.

The specimens to be examined are introduced into the microscope through an air lock or small antechamber, which is then closed and evacuated before the door into the microscope proper is opened and the specimens pushed into position for use. They are supported upon very thin films of some light plastic material which offers little obstruction to the electron beam, and must themselves be very thin. These requirements, as well as the necessity of being examined in a vacuum, place rather severe limitations upon the types of specimens which may be examined. In spite of these limitations, however, many applications have been found for the electron microscope, and the number continues to increase. For example, the surface of a metal may be studied by making a casting of it in a thin film of some plastic, then examining this casting in the electron microscope. Later models are more compact than the one shown in Fig. 95 and have various other improvements, but they do not greatly surpass it in resolving power.

Electron microscopes may also be made with electric-field lenses. Although the possible magnification for a magnetic-field lens is considerably greater than for an electric-field lens, a microscope made with electric-field lenses would possess the distinct advantage that no elaborate voltage stabilization would be required; it is necessary only that the accelerating voltage and the focusing voltages change proportionally. A microscope of this kind has been built in the General Electric Company laboratories. It will produce an overall magnification up to 8000 diameters, which is sufficient to serve many needs. An additional advantage possessed by this microscope is its small size, much smaller than that of other types.

PROBLEMS

1. The long-wavelength limit of a photoelectric surface is 5500A. Compute the work in ergs needed to pull electrons out of this surface.

2. (*a*) If light of 2537A wavelength falls upon the surface described in problem 1, what will be the maximum kinetic energy of the escaping electrons? (*b*) What opposing voltage will be needed to stop these electrons from escaping?

3. What is the minimum wavelength which may be excited by 30 kilovolts applied to an X-ray tube?

4. Compute the energy and the mass of X-ray photons whose wavelengths are 0.300A.

5. (*a*) Compute the momentum given to an electron (initially at rest) when it is struck squarely by an X-ray photon of 0.300A wavelength. (See Sec. 108.) (*b*) Compute the speed given to this electron by the collision.

6. Electrons have been set into motion by a fall of potential of 1200 volts. Compute: (*a*) their speed; (*b*) the wavelength of their waves; (*c*) the frequency of these waves; (*d*) the speed of these waves.

7. Compute the wavelength of a hydrogen ion which has been set into motion by 100 volts potential difference.

REFERENCES

For Additional Reading

R. A. Millikan, *Electrons, Protons, Photons, Neutrons, and Cosmic Rays*, Chapters X and XI.
Physics Staff, University of Pittsburgh, *Atomic Physics*, Chapter VIII.
J. D. Stranathan, *The "Particles" of Modern Physics*, Chapter 6, Secs. 3 and 4.
G. E. M. Jauncey, *Modern Physics*, Chapter IX.
Max Born, *The Restless Universe*, Chapter III.
J. B. Hoag, *Electron and Nuclear Physics*, Chapters III and IV.
G. F. Hull, *Modern Physics*, Chapter V.
G. Gamow, *Mr. Tompkins in Wonderland*, Third Lecture, and Second and Fourth Dreams.

A. H. COMPTON, "Scattering of X-Ray Photons," *American Journal of Physics*, Vol. 14, pp. 80–83.
L. MARTON, "A New Electron Microscope," *Physical Review*, Vol. 58, p. 57, 1940.

Many articles relating to the practical uses for electron microscopes are to be found in the *Journal of Applied Physics* from 1940 on.

FOR ADVANCED READING

HUGHES and DUBRIDGE, *Photoelectric Phenomena.*
A. H. COMPTON, *X-Rays and Electrons.*
RUARK and UREY, *Atoms, Molecules, and Quanta*, Chapter XXI.
ZWORYKIN and WILSON, *Photocells*, Second Edition.
ZWORYKIN and MORTON, *Television*, Chapters III and IV (Electron Optics).
C. H. BACHMAN and S. RAMO, "Electrostatic Electron Microscopy," *Journal of Applied Physics*, Vol. 14, pp. 8–18, 69–72, 155–160.

CHAPTER VIII

THE HYDROGEN SPECTRUM AND THE HYDROGEN ATOM

118. Atoms. When last we considered the atom (Sec. 28) we had arrived at a picture of it in which practically all its mass is concentrated in a nucleus that is not much more than 10^{-12} cm, or one ten-thousandth of an angstrom unit, in diameter, and carries a positive electric charge. Around this nucleus, in the normal state of the atom, are clustered enough electrons to neutralize this nuclear charge; and in spite of the fact that these electrons are comparable in smallness of size to the nucleus itself, they somehow manage effectively to fill up an approximately spherical volume several angstrom units in diameter. Just how are these electrons arranged, and how do they manage in effect to fill up a volume so large in comparison to their own size? This was the point at which we turned away from consideration of the atom to consider the nature of light. Now we may return to consideration of these questions. Our source of information concerning the electronic structures of atoms will now be the light which comes from the atoms when their electronic structures are disturbed. The word "light" is of course used here in its physical sense and includes the entire electromagnetic spectrum, from the infrared through X-rays.

119. Line Spectra. When atoms are crowded close together, as they are in a solid, a liquid, or even a very dense gas, the light which they emit depends primarily upon the temperature of the substance, and forms a *continuous* spectrum. It will be white light if the temperature is high enough. The physical laws which describe such emission of light will be discussed in Chapter XIV. On the contrary, if the atoms are relatively far apart, as they are in a gas or vapor of fairly low density, the light emitted by them will contain only a limited number of separate frequencies and will produce a *bright-line* spectrum. The emission of light by a gas or vapor may be the result of high temperature, as it is in the sodium flame. In stars this is the primary cause of the light emitted, but in the laboratory the highest temperatures which may be produced artificially produce comparatively low excitations; higher excitations are produced only by electrical means, such as the electric arc, the Geissler tube, and the spark discharge, the intensity of excitation increasing in the order named.

Even though the electric arc is very brilliant, the degree of excitation is comparatively low; it is partly electrical and partly due to the high temperature. The bright-line spectrum comes of course from the luminous volume of gas and vapors between the electrodes. The light emitted by the terminals of the electrodes, which are usually heated to incandescence by the discharge, forms a continuous spectrum; it is usually kept out of the spectrograph by using a condensing lens to focus an image of the arc upon the slit of the instrument. The arc spectrum shows spectra both of the gases through which the discharge occurs and of the substances which are vaporized from its terminals. If the electrodes are of metal, the arc will show a bright-line spectrum for that metal. The iron arc spectrum is usually formed in this manner. More often, however, the arc is formed between carbon rods into which axial holes have been drilled, these holes being filled with the element whose spectrum is desired, or with some chemical compound of it. Carbon is very useful for this purpose, since it has very few lines in its arc spectrum, and none in the visible spectrum. In this class fall also the vacuum arcs, of which the mercury arc is a familiar example. Here the spectrum is that of the mercury vapor alone, the vapor being formed by evaporation from the mercury terminal in a closed and evacuated bulb or tube. Another example which is becoming familiar is the sodium arc, which has had considerable use for highway lighting.

Electrical discharges through rarefied gases, such as take place in the Geissler tubes of the laboratory and in the innumerable electric discharge tubes used for luminous signs, produce more intense excitations, the spectra here being those of the gases in the tubes, uncontaminated by vapors from the terminals. The highest excitations are those which may be obtained in the electric spark. The spark, like the open arc, produces spectra both for the gas through which the spark passes and for vapors of the metals which form the spark terminals. The degree of excitation in a spark may be varied considerably by altering the discharge circuit and other conditions for the discharge. The highest spark excitations are produced between terminals placed fairly close together in a high vacuum.

These different modes of excitation do not differ in kind, but only in degree. Regardless of the method, excitation results from collision between the atom and some other particle, and the degree of excitation is determined by the energy involved in such a collision.

The frequencies of the lines appearing in a spectrum are characteristic both of the atom which emits the light and of the degree of excitation employed, and the existence of such characteristic spectra has been of immense practical value to scientists, quite without regard to their relation to atomic structure. The chemist, for example, by examining the spectrum of the light produced by a compound which he wishes to analyze, can detect the presence of elements in smaller quantities than by any other method of analysis. Indeed, by observing the intensity of the lines which identify such an element, he may sometimes even estimate the quantity. Since all the elements have many lines in their spectra (iron has thousands of lines in the visible spectrum alone) such identification usually requires exact measurements of wavelength. It requires also that the wavelengths of at least the principal lines of the elements be known quite accurately; for this reason, and for many others as well, the wavelengths of most of the lines in the near infrared, visible, and ultraviolet regions have been very carefully measured for all the elements.

120. Stellar Spectra. The astronomer has made still more spectacular use of spectra. By examining the spectra of stars he can analyze them chemically, he can estimate their approximate temperature, and he can measure their motions towards or away from us through space. Sometimes stellar spectra are bright-line spectra such as described above; more often they are absorption spectra such as described in Sec. 135. In either event spectrochemical analysis is almost as easy as in the chemical laboratory. Although this is true now, it was not true when stellar spectra first began to be studied, since stellar spectra differ from laboratory spectra in a number of particulars. Consider, for example, the spectra of stars of the same general size and brightness as our sun. These spectra are practically identical, thus indicating that all stars of this class have the same chemical composition, but they often differ from the sun's spectrum in one important respect; in some all spectrum lines may be displaced a little towards higher frequencies, in others towards lower frequencies, because of the motions of these stars along the line of sight.

This phenomenon is an example of the Doppler effect, and the explanation for it is the same in principle as that given for the rise in pitch of a sound heard as the source approaches one, and the fall in pitch when it is moving away. From the magnitude of the shift the speed of the star in its motion towards or away from us may be computed. After leaving the source (the star), all light waves travel with the same speed, *c*. Because of the motion of the star, however, each successive wave is emitted from a point a little farther along the path the star is traveling. Thus the waves are crowded a little closer together in front of the star'smotion, and correspondingly separated a little more behind it. It follows that, if the star is moving towards us, the frequency of the light will be increased; if it is moving away from us, the frequency will be lowered.

The change of frequency will equal v/c times the original frequency, where v is the speed of the star towards or away from us.

This Doppler effect was rather easily and quickly explained, but other differences between stellar and laboratory spectra remained puzzling until quite recent times. Thus certain characteristic spectrum lines which are very strong in laboratory spectra are faint or even absent in some types of stellar spectra, whereas other lines are found to be much stronger in these stellar spectra than in the spectra observed in the laboratory. As a result of better knowledge of the origin of spectra, it is now possible to interpret these differences in relative intensities of spectral lines as due to the very high temperatures of these stars, and to use them as a basis for estimates of star temperatures.

121. Origin of Line Spectra. The frequencies corresponding to the characteristic spectrum of any element must depend in some manner upon the structure of the atom of that element, and the problem which confronted the theoretical physicist was that of working backward from the spectrum to the structure from which it originates. Since light waves are electromagnetic waves, and since the electrons which form the outer part of the atom are electrically charged particles, it seemed only natural to expect that these frequencies are the fundamental and "overtone" frequencies of the natural vibrations of the electrons when they are disturbed from their equilibrium positions within the atom; but all attempts to find such a system of related frequencies proved futile. Although in the course of such investigations many relations were found between the frequencies, none of them could be explained in terms of fundamental and overtone vibrations.

122. The Hydrogen Atom. Consider, for example, the very simplest atom, the hydrogen atom, whose nucleus carries but one positive elementary charge, and which has consequently but one electron in its outer structure. Various experiments indicate that this atom has a diameter of approximately one angstrom unit (10^{-8} cm), so that the nucleus and the electron must be separated by about half this value, or 0.5×10^{-8} cm. This separation is the first thing to be accounted for by any description or "model" of the atom.

As long as they are at rest with respect to each other, there is nothing which we know concerning the nucleus or the electron, as *particles*, which will keep them apart. If, however, they are revolving about each other, this separation is easily accounted for by their inertia. Since the electron is so much lighter than the nucleus, such motion may be considered a revolution of the electron about a stationary nucleus, and a steady circular orbit is possible if the centripetal force required to keep the electron in that orbit equals the electrical attraction between the two particles. If m represents the mass of the electron, r the radius of its orbit, and v its speed, this condition is expressed by

$$\frac{mv^2}{r} = \frac{e \times e}{r^2} \qquad (8 \cdot 1)$$

For $r = 0.5 \times 10^{-8}$ cm, this formula gives a speed, v, equal to 2200 kilometers per second, and a revolution frequency equal to approximately 7×10^{15} cycles per second.

Such a planetary model for the hydrogen atom is an exceedingly attractive one, but it is unable to explain the spectrum observed for hydrogen. Instead, it would radiate any wavelength whatsoever, depending upon the radius of its orbit. Even worse, it would never *stop* radiating until it had completely collapsed. That is to say, unless the physical laws for electrically charged particles such as electrons differ from the laws which are found for ordinary electrically charged particles, such a revolving electron would constitute an electrical oscillator, which would *continually* radiate electromagnetic waves of a frequency equal to the frequency of its revolutions. And the resultant continuous loss of energy by radiation would bring the electron nearer and nearer to the nucleus until finally the atom would be entirely collapsed. The frequency increases as the electron approaches the nucleus, so that the radiation from many such atoms in many different stages of this motion would form a continuous spectrum.

As an alternate possibility we might consider the existence of some unknown repulsion force between the nucleus and the electron, effective only at short distances, such that a balance between the attraction force and this repulsion force would hold the electron stationary at a distance of approximately half an angstrom unit. This atom model would be a stable one, since it would not radiate unless set into vibration by some excitation process. Also, if it were to be excited into vibration, it would emit a series of frequencies which would form a line spectrum. So far this model does very well, but, when one tries to make the frequencies which could be emitted by such an atom model fit the spectrum actually observed for hydrogen, it proves to be impossible.

123. The Hydrogen Spectrum. The frequencies emitted by such an atom model must form an harmonic series, that is, they must be multiples of the fundamental frequency of the electron vibration. As may be seen in Table I, this is not the case. These observed lines form a series, it is true, but instead of forming an harmonic series, with the frequencies separated by equal frequency intervals, it is a series in which the frequencies get closer and closer together, approaching a limit at 822×10^{12} cycles per second.

Many attempts were made to find an explanation for this spectrum on a "classical" basis, that is, in terms of the laws of mechanics and of electricity which are found to apply to large-scale phenomena. All attempts were unsuccessful, although a high-school teacher named Balmer, by empirical (cut-and-try) means, did find a formula by which he could compute with high precision all the wavelengths in this spectrum. Balmer himself wrote this equation in terms of wavelengths, but if we translate his equation into frequencies, we obtain one which is both simpler and more fundamental. It is

TABLE I

HYDROGEN SPECTRUM IN THE VISIBLE AND NEAR ULTRAVIOLET

Wavelength, A	Frequency	Color
6563	457×10^{12}	Red
4861	617	Blue-green
4340	691	Blue
4102	731	Violet
3970	755	
3889	771	(Ultraviolet)
3835	782	

$$f = 3288 \times 10^{12}\left(\frac{1}{2^2} - \frac{1}{N^2}\right) \text{ cycles/sec} \qquad (8\cdot 2)$$

When $N = 3$, this formula gives the frequency 457×10^{12}, which is the frequency of the first line of the hydrogen spectrum, as is given in Table I. Likewise, for $N = 4, 5, 6$, etc., this formula gives the frequencies of all the other lines listed in Table I.

This is but an empirical formula, as has been explained above, and it tells us nothing concerning the structure of the hydrogen atom which will account for these frequencies; nevertheless it is an important step forward, and it laid the foundation for the work of Bohr, who did find a way to explain it.

124. Bohr's Hydrogen-Atom "Model." In 1913 Niels Bohr invented a model for the hydrogen atom which successfully accounts for all the lines of the atomic spectrum of hydrogen. Bohr started with the planetary model of the atom, consisting of an electron revolving around the nucleus in a circular orbit, which has been discussed in Sec. 122. He then met the difficulties with this model which are there discussed by the simple but bold assumption that, if experiment shows that the normal atom does not radiate, then it does not! This by itself would have been only a way of begging the question if he had stopped there. To this assumption, however, he added the two conditions or postulates stated below, and these postulates could not give agreement with the experimental facts unless they define some very fundamental properties of the hydrogen atom.

Postulate 1. An electron may revolve without radiating in any orbit for which 2π times its angular momentum is an *integral* number, *N*, times the quantum constant, *h*; i.e.,

$$2\pi \times \text{angular momentum} = 2\pi \times mv \times r = Nh \qquad (8\cdot 3)$$

Such orbits are called "stationary" orbits.

In the normal state of this atom model the electron is considered to be in the innermost orbit, for which N equals unity, and for this orbit the formula which will be derived in the next section gives a radius of 0.53×10^{-8} cm. This value is in good agreement with those estimated from various experiments, and thereby satisfies one of the requirements for a hydrogen-atom model. The other orbits which are also defined by this postulate, the ones for which N equals 2, 3, 4, etc., are orbits in which the electron would have more energy than it has in the innermost one, its energy increasing as it goes farther out. Bohr assumed that the electron could occupy such orbits only if it were driven into them from the inner one by some excitation process such as one of those mentioned in Sec. 119. It could then stay in such an orbit for a considerable time, perhaps as long as an hundred-millionth of a second. (Certain experiments indicate that this is the order of magnitude of the average time spent by an atom in an excited state.) If this seems to be a very short time, it is only because the second is so large in comparison to the time scale of atomic processes. In this apparently brief stay in such an orbit the electron could make on the average something like ten million revolutions. For comparison consider an electric motor running at the ordinary speed of 1800 r.p.m. This motor would take nearly 4 days to rotate that many times. Ultimately, however, according to Bohr's theory, the electron will spontaneously fall back to an orbit nearer to the nucleus, and the extra energy which it had in the larger orbit will be lost. The theory now borrows the "quantum" idea for energy which originated with Planck, and the second postulate states how this energy is lost.

Postulate 2. When the electron falls from an outer orbit to one closer in, the atom emits a *photon* of light whose energy equals the energy lost by the electron. That is, if W'' is the energy which the electron has in the outer orbit, and W' the energy which it has after it has fallen to the inner one,

$$hf = W'' - W' \tag{8·4}$$

Here f is the frequency of the light emitted, and hf is consequently the energy of the photon.

The force equation, (8·1), holds for each of the "stationary" orbits. Using this equation, together with the first postulate (which defined the size of each orbit), Bohr computed the energy for each "stationary" orbit. (This equation is derived in Sec. 125.) When he then used the second postulate to compute, from these energy values, the frequencies of the spectrum lines, he obtained values which were exactly those obtained by experiment. This will be explained in more detail in the next few sections. So far as the hydrogen atom is concerned, this atom "model" is a complete success! One further point in this theory should be noted: Only one frequency is emitted by one atom at any one time. The spectrum as observed is the total emission from countless atoms, in some of which the electron drops from the second orbit to the first, in some from the third to the second, in some from the third to the first, etc. All possible changes of orbits may occur, but not all are equally probable; the most intense lines correspond to changes which occur most often.

125. Derivation of Bohr's Equation. Solving simultaneously the force equation, (8·1), and equation (8·3), which represents postulate 1, we may obtain for the radius of a "stationary" orbit

$$r = \frac{N^2h^2}{4\pi^2me^2} \tag{8·5}$$

and for the speed of an electron in such an orbit

$$v = \frac{2\pi e^2}{Nh} \tag{8·6}$$

For $N = 1$, $r = 0.53 \times 10^{-8}$ cm, which, as has already been stated in Sec. 124, is in good agreement with the values for the radius of the normal atom as estimated from experimental data. The radii of the first five orbits are shown in Fig. 98.

The energy of a "stationary" orbit is partly kinetic and partly potential energy. The kinetic energy is $mv^2/2$. The potential energy is found by computing the work needed to bring the two charges to the distance r from each other, starting with them "infinitely" far away (where the potential energy is assumed to be zero). Since they attract each other, this work is *negative*, and can be shown to be

$$\text{Potential energy} = -\frac{e^2}{r} \tag{8·7}$$

The total energy is then

$$W = -\frac{e^2}{r} + \frac{mv^2}{2} \tag{8·8}$$

Putting into this equation the values of r and v obtained above,

$$W_N = -\frac{2\pi^2 me^4}{h^2} \times \frac{1}{N^2} = -\frac{2178 \times 10^{-14}}{N^2} \text{ erg} \tag{8·9}$$

126. Energy Levels and Frequencies. The energy values which may be computed from this equation are plotted in Fig. 99 as a series of horizontal lines, which we shall hereafter refer to as *energy levels*. Along the right side of the diagram are given the energy values; along the left side are given the values of N, the *quantum number*, as it now will be called. The lowest level, of quantum number equal to 1, represents the energy of the normal state of the atom, while the other levels, of higher quantum numbers, represent the energies of various excited states. The top line, labeled ∞ (infinity), represents *zero* energy, or the energy of the electron when it is entirely removed from the atom. Note that this peculiar choice of the level of zero energy gives to all the other levels

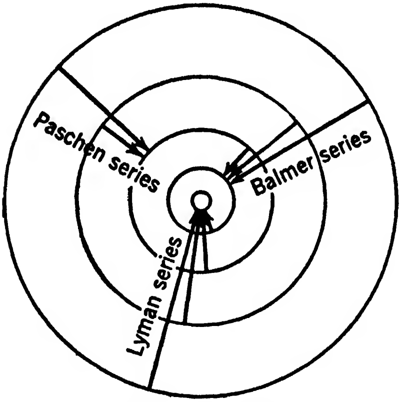


FIG. 98. BOHR ORBITS.

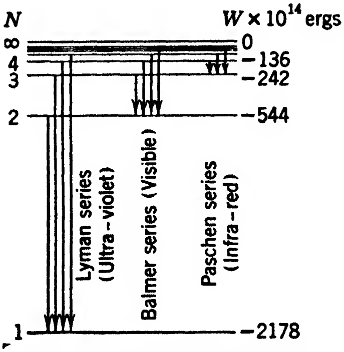


FIG. 99. ENERGY LEVELS—HYDROGEN ATOM.

negative values. It may be compared with the method of indicating levels in a mine; the lowest levels are represented by the biggest numbers.

The emission of light by the dropping of the electron from an outer level to one closer in may be represented in this diagram by a vertical line drawn between the corresponding energy levels. For example, the first line of the spectrum series given in Table I under Sec. 123 is emitted when the electron drops from the third orbit to the second one, and is represented by the line which is drawn vertically downward from the third level to the second level in this diagram. The change of energy can be gotten from the values given at the right, and thence the frequency of this line may be computed; i.e.,

$$\begin{aligned} hf &= W_3 - W_2 \\ &= (-242 \times 10^{-14}) - (-544 \times 10^{-14}) \\ &= 302 \times 10^{-14} \text{ erg} \end{aligned}$$

and

$$f = 456 \times 10^{12} \text{ cycles per sec}$$

The frequencies for the rest of the lines in Table I are obtained similarly and are represented by the series of vertical lines in Fig. 99 which are labeled *Balmer series*. All the frequency values computed by this equation check the experimentally measured values to a much greater accuracy than is indicated by these figures. Whatever weaknesses may later be found in the original Bohr theory, it gives the correct values for the energy levels of the hydrogen atom to a high degree of accuracy.

This Balmer series does not include all possible changes in position for the electron. In particular, it does not include the transitions which bring the atom back to its normal state, the ones which are indicated at the left of Fig. 99 and are labeled the Lyman series. Since all these drops are so much greater than those corresponding to Balmer's series, these frequencies are all very much higher. They agree very closely with the frequencies for a series of lines found by Lyman in the far ultraviolet, and thus provide additional demonstration of the exactness of the energy levels given by Bohr's theory. Likewise the series of lines given by transitions from higher levels down to the third level, which is labeled Paschen series, is one which has been found experimentally by Paschen in the infrared. Transitions terminating on still higher levels (corresponding to $N = 4$, 5, etc.) would produce lines of still lower frequency, or still farther into the infrared. Although measurements in this region of the infrared are very difficult to make, two additional series of lines have been found, the Brackett series and the Pfundt series.

A general formula may be derived which will give the frequencies for all the lines of the hydrogen spectrum. If, in equation (8·4), which represents the second postulate, there is inserted the values for W given by equation (8·9), the result gives

$$f = \frac{2\pi^2 me^4}{h^3}\left(\frac{1}{N^2} - \frac{1}{N'^2}\right) \tag{8·10}$$

When $N = 2$, this is the exact equivalent of Balmer's formula, and gives the frequencies for the lines of Balmer's series when $N' = 3, 4, 5$, etc. When $N = 1$, it similarly gives the frequencies for Lyman's series, while Paschen's series is obtained when $N = 3$.

127. Deuterium Spectrum. Deuterium, or heavy hydrogen (see Sec. 22), is chemically identical to ordinary hydrogen, but the mass spectrograph shows that its atomic mass is about twice as great. Its spectrum also appears to be identical to that of hydrogen unless it is examined very closely indeed, and then it is discovered that all lines are very slightly higher in frequency than the corresponding ones in the hydrogen spectrum. This difference, far from disproving Bohr's theory, gives additional proof for it. The deuterium atom consists of a nucleus bearing a single elementary positive charge, with a single electron circling about it (in quantized orbits), so that the theory as developed above would give it a spectrum identical to that of hydrogen. But in that derivation a very small approximation was made: For the sake of simplicity, it was assumed that the nucleus remained stationary, with the electron revolving about it, whereas actually the nucleus and the electron both revolve about their common center of mass. Since the nucleus is 1840 times as massive as the electron, the center of mass is so close to the nucleus as to make the approximation a very good one. In the deuterium atom the center of mass is still closer to its nucleus, and this makes a slight but observable difference in the two spectra. Although this difference is only 27 parts in 100,000, it can be observed in the spectra, and the spectrum measurements check up with the theoretical values.

128. Spectrum of Ionized Helium. Still further support for Bohr's theory is found in the spectrum of ionized helium. The helium atom contains two electrons and, as one might expect, has quite a different spectrum from that for hydrogen. But when helium gas is very highly excited additional spectrum lines appear, and these lines form *hydrogenlike* series. The helium atoms which emit these lines have been *ionized*, so as to lose one electron completely, and these spectrum series arise from changes in energy of the remaining electron. Since this helium *ion* is then exactly like a hydrogen atom, except for the larger electric charge on its nucleus, Bohr's theory must apply to it. Because its nucleus carries two elementary charges instead of one, the energy-level values are all 2^2 or 4 times those given by equation (8·9); when this is taken into account, the frequencies computed by theory agree very closely with the observed values.

It is interesting to note that, in this spectrum, the series of lines which is analogous to Balmer's series for hydrogen occurs in almost exactly the same place in the ultraviolet as does Lyman's spectrum for hydrogen, while the lines for ionized helium which appear in the visible region correspond to a far infrared series for hydrogen. This visible part of the helium spectrum was first observed by Pickering, in stellar spectra. It was ascribed initially to hydrogen because every other line (the ones corresponding to even values of N') coincides almost exactly with a line of Balmer's series.

Similar hydrogenlike spectra have been observed in the light from very highly ionized lithium vapor (due to doubly ionized lithium atoms), and in the light from several more elements whose atoms have been so highly ionized as to have but one electron remaining. Bohr's theory very precisely predicts the frequency values for all such spectra. The general formula which fits all such cases is

$$f = Z^2 \frac{2\pi^2 mc^4}{h^3} \left(\frac{1}{N^2} - \frac{1}{N'^2} \right) \tag{8·11}$$

where Z is the atomic number, that is to say, the number of positive elementary charges upon the nucleus.

129. Electron-Wave Explanation for Bohr's First Postulate. The existence of electron waves, as has been explained in Sec. 110, provides a rational explanation for Bohr's first postulate. The "stationary-state" orbits are just those for which the electron waves make *standing-wave patterns*. Consider, for example, the Nth hydrogen orbit. From postulate 1,

$$2\pi \, mv \, r = Nh \tag{8·12}$$

The circumference of this orbit is consequently

$$2\pi r = N \frac{h}{mv} \tag{8·13}$$

But, according to Sec. 110, the wavelength of the electron wave is h/mv, while N is a whole number. Hence the circumference of this orbit is equal to N wavelengths, or the waves form a standing-wave pattern such as is sketched in Fig. 100 for $N = 4$.

It will be explained later on that this is not an exact picture—that, indeed, Bohr's theory is not complete, although both Bohr's theory and this electron-wave explanation for his first postulate are adequate to explain these simple hydrogenlike spectra. And it is perhaps possible to see at this point an indication of where the incompleteness of such theory lies. The Bohr orbits and this wave picture lie in a plane—in only *two* dimensions—whereas waves under similar conditions would spread out in all directions, in *three* dimensions. When this oversight is corrected for, and the wave-mechanics theory which is briefly outlined in Sec. 115 is applied to the solution of the problem of atomic structure, the results obtainable are adequate to explain, exactly, the structure of all atoms, in a manner which agrees with the spectra emitted by them. More will be said concerning this later on, but first let us consider in Chapter IX a fairly simple spectrum to which Bohr's theory will not apply without modification.

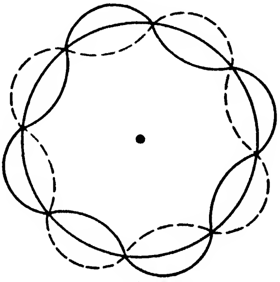


FIG. 100.

PROBLEMS

1. Compute: (*a*) the velocity of the electron in the smallest Bohr orbit; (*b*) the frequency of its revolution in this orbit.

2. Compute the frequencies and the wavelengths of the first three lines of the Lyman series, using the data given in Fig. 99.

3. Compute the wavelengths of the first few lines in the spectrum of ionized helium, for the series whose transitions end on the level for which $N = 4$. (This is the Pickering series.)

4. In the spectrum of a star, the hydrogen alpha line has a wavelength of 6563.25A instead of the value of 6562.85A which is found in the laboratory. Is this star moving towards us or away from us, and at what speed?

REFERENCES

See end of Chapter IX.

CHAPTER IX

OPTICAL SPECTRA AND ATOMIC STRUCTURE

130. Lithium Spectrum. Lithium is the third element in the periodic table, and its spectrum is a fairly simple one, as is shown by Fig. 101(*a*). At first sight, no regularities at all appear in this spectrum, but this is only apparent. By procedures which are fairly simple to the spectroscopist (although they are somewhat involved to be explained here) it is possible to separate these lines into several overlapping series, such as those shown in (*b*), (*c*), and (*d*) of Fig. 101. Each of these series resembles the series of the hydrogen spectrum in that the lines become progressively weaker and closer together as they increase in frequency, and it is also possible to derive a formula for each series (Rydberg's formula) which will give the frequencies of the lines very closely. Now, by using Bohr's second postulate and working *backward*, it is possible to figure out the *energy levels* which represent the normal and the excited states for the lithium atom. The results of such analysis are shown in Fig. 102.

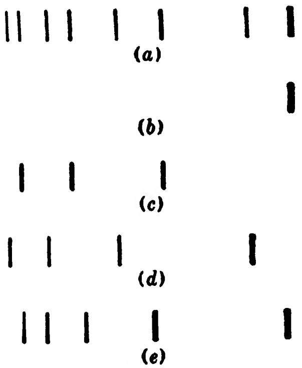


FIG. 101. LITHIUM SPECTRUM (visible range only).
(*a*) Complete spectrum.
(*b*) "Principal" series.
(*c*) "Sharp" series.
(*d*) "Diffuse" series.
(*e*) Hydrogen spectrum, for comparison.

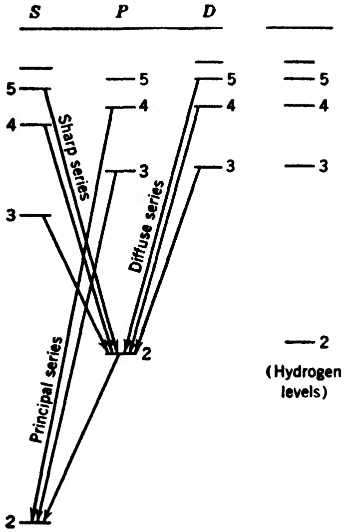


FIG. 102. ENERGY LEVELS—LITHIUM ATOM.

One difference between these energy levels and those shown in Fig. 99 for the hydrogen atom is immediately apparent. Instead of one set of energy levels there are three sets (there would be more if a more complete analysis had been made), and the transitions which give the spectrum lines *always* take place from one set to the next one to either side, *never* between levels of the same set. For example, the "principal" series corresponds to transitions from successive levels of the P set to the lowest level of the S set; the "sharp" series, to transitions from S levels to the lowest P level; etc. Let us now see how nearly Bohr's theory can come to accounting for these energy levels.

Although (as we know from evidence to be presented later on) lithium has three electrons circling about its nucleus, there is reason to believe that these energy levels belong to a single one of these electrons. First of all, there is the chemical evidence which is represented by the position of lithium directly under hydrogen in the periodic table. Lithium, like hydrogen, is univalent, which means that only one of its three electrons is easily removed; the other two appear to be much more tightly bound to the nucleus. (The exact wave-mechanics theory shows that this is indeed so, and even gives a reason for there being only two electrons in this innermost group. See Sec. 150.) Let us assume this to be so, and attempt to apply Bohr's theory to the motion of the outer or *valence* electron.

The two inner electrons surround the nucleus fairly uniformly, so that, in so far as the electric field outside of their orbits is concerned, their charges practically neutralize two of the three positive charges carried by the nucleus. That is to say, as long as the third electron stays outside the region occupied by the inner electrons, it is attracted towards the nucleus by a force which is nearly equivalent to the attraction of a single positive charge. But this is exactly the situation found in the hydrogen atom, and the energy levels as obtained by Bohr's theory should then be the same for both the hydrogen and the lithium atom. This is clearly not true, as may be seen by comparison of the lithium energy levels shown in Fig. 102 with the corresponding hydrogen energy levels, which are plotted to the right of the figure. Although the higher P and D lithium energy levels approach those for hydrogen rather closely, the lower P levels, and all the S levels, fall below the corresponding hydrogen levels; the $2S$ level is very much lower down.

131. Elliptical Orbits. Sommerfeld suggested an explanation for this. He suggested that there might be, even in the hydrogen atom, elliptical orbits in addition to the circular ones; and for answer to the obvious query, "How much elliptical?" he gave additional quantum rules. These rules are quite mathematical, and all that need be stated here are the consequences which follow. As a result of this theory each electron orbit is specified, not by a single quantum number, N, but by two quantum numbers, k and n, which Sommerfeld called the azimuthal and the total quantum numbers, respectively. When $k = n$, the orbit is circular. When k is less than n, the orbit is elliptical; and the smaller the ratio of k to n, the more elongated is the orbit. This is shown in Fig. 103, which represents the first few of these orbits for the hydrogen atom. The large figure labeling each orbit designates the value of n, while the subscript gives the k value for that orbit. Note that the 1_1, 2_2, and 3_3 orbits are circular, since for these $k = n$, but that all the others are elliptical. The nucleus does not lie at the center of the ellipse, but at a point near one end which is called its focus, and the more elongated the ellipse, the nearer this focus is to the end of the ellipse, as is seen in Fig. 103.

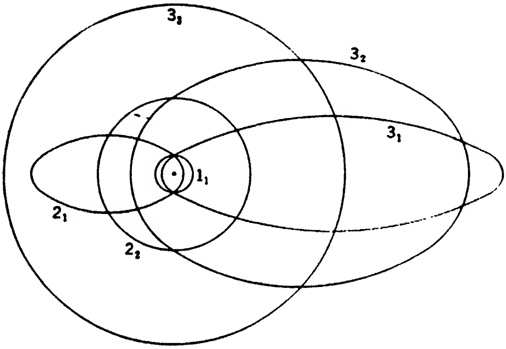


FIG. 103. ELLIPTICAL ORBITS—HYDROGEN ATOM.

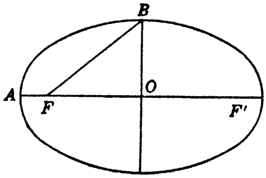


FIG. 104. FOCI OF AN ELLIPSE. Focus F is located by drawing line $BF = AO$. Focus F' is located in like manner. $OF = OF'$.

The azimuthal quantum number, k, which may have any value from 1 to n, determines the angular momentum of the electron, according to Bohr's first postulate. It then follows from equation (8·3) that

$$\text{Angular momentum} = \frac{kh}{2\pi} \tag{9·1}$$

For the *hydrogen* atom (but generally not for other atoms) the total quantum number, n, determines the energy value for the orbit, just as did the quantum number, N, in the original Bohr theory. Equation (8·9) then becomes

$$W_n = -\frac{2\pi^2 me^4}{h^2n^2} \tag{9·2}$$

That is, all orbits in the hydrogen atom which have the same total quantum number, n, have the same energy, regardless of their ellipticity;[1] so far as energy levels are concerned, there is no means of distinguishing between elliptical and circular orbits. It is for this reason that Bohr's original theory, although incomplete, was so successful with the hydrogen atom.

132. Lithium Energy Levels. Consider next the lithium atom. Referring to Fig. 102, it may now be explained that the energy levels are here sorted out into sets according to the values of the azimuthal quantum number, k. For the S set, $k = 1$; for the P set, 2; for the D set, 3; etc. For reasons stated in Sec. 130, many of the orbits shown in Fig. 103 for the hydrogen atom might also fairly well represent possible orbits for the outer electron of the lithium atom. First of all, however, the 1_1 orbit must be eliminated, since the quantum number $n = 1$ belongs to the two inner electrons. The reason for this was originally only a postulate, that is to say, an assumption which made the results of the theory agree with the experiment; but it has since been justified by the exact wave-mechanics theory. These inner electrons must not of course be thought of as circling about in this orbit, which has been drawn for the hydrogen atom, but they will remain inside a volume of roughly this diameter or less. Since this part of the atom, with its two electrons, resembles the helium atom, it is convenient to refer to it as the helium core, or simply as the core of this atom.

The 2_2, 3_2, and 3_3 orbits, and all other orbits for which k is greater than 1, lie entirely outside this core, in the region where, as has been explained above, the electric field is almost the same as that of the hydrogen nucleus; and the corresponding energy values are likewise almost the same, as may be seen for the P and D levels in Fig. 102. (The slight differences are due to imperfect screening of the nuclear charge by the two inner electrons.)

The S levels show big deviations from the hydrogen values, and this may be understood when the corresponding 2_1, 3_1, etc., orbits of Fig. 103 are observed. All of these orbits for which $k = 1$ cut through the core, and exposure of the electron for even this brief part of its path to the stronger field of the unshielded nucleus considerably modifies its energy. (Properly speaking, the path is no longer an ellipse, but may be roughly described as an ellipse that is constantly shifting its position, so that its

---

[1] There is actually a very small difference, which appears only as a "fine structure" of the hydrogen spectrum lines when they are examined with powerful interference spectrometers; this "fine structure" verifies the existence of (the wave-mechanical equivalent of) these elliptical orbits.

axis rotates about the nucleus.) The energy levels are all lowered in consequence, as is seen in Fig. 102.

This elliptical-orbit atom model explains, qualitatively at least, the energy levels of the lithium atom, and it may be used with equal success to account for the spectra of sodium, potassium, and the other alkali-metal (hydrogenlike) elements. Each of these atoms has, as has lithium, a single valence electron whose possible energy levels account for its normal spectrum, and each has a core of inner electrons whose structure resembles one of the inert gases. For lithium this core is heliumlike, as has been explained; for sodium it is neonlike; for potassium, argonlike.

Various modifications have been suggested to improve this orbital model for the atom. One of these is a change of the azimuthal quantum number, k, which in Sommerfeld's theory represents the angular momentum of the electron. It was found that many of the phenomena of spectrum emission could be explained somewhat better by assuming the existence of less angular momentum than this, and a smaller quantum number, l, equal to $k - 1$, was correspondingly introduced. According to this modified theory,

$$\text{Angular momentum} = \frac{lh}{2\pi} \qquad (9\cdot3)$$

In the original model, k had values from 1 to n. In this model, l has values from zero to $n - 1$. The most significant thing about this change is the result that the S states, which now correspond to $l = 0$, are states in which the electron has no orbital angular momentum, and this is a result that is required in order to explain the fine structure of spectra (Sec. 151), the Zeeman effect (Sec. 154), etc.

133. Electron-Wave Atom Model. At this point, if not sooner, the reader may well have become impatient with all this talk about "models." If the physicist knows how the atom is built, why does he not describe it without further ado? If he does not, why does he not admit his ignorance? In any event, why all this waste of time talking about "models"? In answer to these questions the reader must be reminded that, first of all, we can describe something new only in terms of things already familiar. This is what we mean by making a "model." In the second place, it must be remembered that our evidence for atomic structure is at the best indirect evidence, from which we must infer a structure which we can never directly see, and that this evidence is acquired little by little. In this respect it is helpful to compare our construction of successive models for the atom with the work of an artist (not a surrealist) who attempts to paint a picture of a complicated device which he sees at first from a great distance. Certain broad features of this picture will be correct; but closer approach will show him that many details exist which he did not see at first, and that some details which he at first imagined to exist were merely illusions produced by distance. From some position, not too close, the details appear remarkably sharp and clear, and the picture is unusually satisfying to the artist. Still closer approach may reveal that features which appeared sharp and definite are in reality of such peculiar and unfamiliar appearance as to defy reproduction by any medium which the artist may possess.

This is the situation with respect to our knowledge concerning the atom. At first it seemed possible that, if an atom were to be magnified sufficiently, we should indeed see it as little electrons moving in orbits about a central nucleus, just as has been represented in these earlier "models." Actually, the nature of the atom transcends our experience with things of familiar size, and any model which pictures it to us in terms of particles moving in orbits is inexact. Our information concerning the structure of the atom is now more exact than ever before, but for that very reason it is now impossible to represent it exactly by any model in which the electrons are located in orbits, or, indeed, by a model in which the electrons are located with any definiteness at all!

The most recent model, and the one which is so far entirely successful in explaining all known phenomena of spectra, is one whose structure is determined by the wave properties of electrons which have been presented in Chapter VII. In Sec. 129 this theory has already been applied, in an elementary way, to account for Bohr's first postulate. Although correct as far as it goes, that treatment is inadequate to the general problem, since an electron wave spreads out in three dimensions, and is there represented as being in but two dimensions. When the wave-mechanics theory is exactly applied to the problem of the hydrogen atom, it is found that each possible "stationary" state for the electron is represented by a separate wave pattern. Each of these wave patterns may be identified by three quantum numbers, l, m, and n. The significance of m will be discussed in Sec. 153. The theory shows that this quantum number may have integral values from $-l$ to $+l$, including zero. That is, if $l = 2$, m may have the values -2, -1, 0, 1, and 2.

The other two quantum numbers have the same significance which they have had previously in the revised orbital model. The value of l still determines the angular momentum of the electron in each stationary state, and in general it is found that the quantum numbers and quantum rules, which were assumed arbitrarily by the earlier quantum theory in order to make a theory which would fit the experimental facts, are now obtainable as logical mathematical consequences of the theory itself. Although the orbits vanish in this electron-wave or wave-mechanical-theory model for the atom, there is still a close mathematical correlation between the orbital picture and the wave properties of the atom. In consequence of this, the orbital model is still quite generally used in discussing the properties of atoms, because of the greater ease with which its properties may be visualized.

134. Spectrum Analysis. Spectra for atoms other than hydrogen and lithium possess complexities which require further revision of theory, and even the hydrogen and lithium spectra possess fine structure which is not fully explained without these revisions. Some of this theory will be given in the next chapter. Regardless of these revisions, Bohr's second postulate always holds. The emission of every spectrum line results from transition of an electron from one stationary state to a lower one, and h times the frequency of that line is always equal to the difference between the energy levels of the two states. In a few cases it is possible to compute the energy levels from theory, but usually they must be obtained experimentally by working backward from the frequencies of the spectrum lines. *Spectrum analysis*, as this process is called, has been carried out for a large share of the elements. It has required an immense amount of labor, and much of it could not have been accomplished without the aid of most complicated and ingenious automatic computing and measuring machines. The energy levels for an atom which, like lithium, has only one outer or valence electron, form a single group of S, P, D, $\cdots$ sets, such as shown in Fig. 102. These are the simplest cases. When an atom has more than one valence electron, its spectrum is more difficult to analyze. The stationary states then are determined by considering the resultant angular momentum, including the momenta of *all* the valence electrons, and the energy levels which represent them form two or more groups of S, P, D, $\cdots$ sets. The transitions which are "permitted," that is, which lead to spectrum lines, are specified by various "selection" rules which at first were empirical but since have been shown to be rational deductions from the wave-mechanics theory. Further details are given in the references listed at the end of the chapter.

135. Absorption Spectra. Under certain conditions characteristic atomic spectra may occur as *dark-line* or *absorption* spectra rather than as bright-line or emission spectra. Fraunhofer observed that the solar spectrum is crossed by a number of dark lines, and Kirchhoff and Bunsen observed that these *dark* lines correspond in their positions in the spectrum to certain of the *bright* lines in the atomic spectra of sodium, calcium, hydrogen, etc., and hence reveal the presence of these elements in the sun. Since Fraunhofer's time hundreds of such dark lines have been observed in the solar spectrum, and by means of these lines it has been proved that the sun contains most of the elements which are present in the earth. It is interesting to recall here that one of the now familiar elements was "found" in the sun before it was discovered here in the earth, a fact which is given permanent record in its name, helium. The spectra of many stars are also dark-line spectra which similarly reveal their chemical composition.

A simple experiment will explain the formation of the dark Fraunhofer lines. If the light from a carbon-arc lamp is passed through fairly dense sodium vapor, and the light then examined in a spectroscope, a dark line (really a pair of lines very close together) will appear across the continuous spectrum in the place where the familiar yellow sodium lines usually appear. The sodium vapor is capable of absorbing the same frequencies which it can emit as a bright-line spectrum, providing that the continuous spectrum light comes from a source considerably hotter than the vapor. In the solar (or stellar) spectrum the continuous background comes from the incandescent core of the sun (or star), and the dark lines result from absorption by its outer and cooler atmosphere.

For a long time this absorption was explained as a sort of resonance effect, although it was observed that this "resonance" did not occur for all the lines of a spectrum, but only for certain lines which were called "resonance" lines. Thus hydrogen gas in its normal state is entirely transparent to visible light, although it has several members of the Balmer series in its visible emission spectrum. Hydrogen atoms in their normal state will absorb only the series of lines in the ultraviolet called the Lyman series. Likewise cool sodium vapor will absorb only the series of lines called the principal series (the double yellow line at 5890, 5896A is the first member of this series), and this is true also for other elements. *A gas or vapor whose atoms are in their normal states will absorb only the lines of the series which it emits when the electron returns from an excited state to its normal state.* The spectroscopist finds this a very useful aid in sorting out spectrum series.

The quantum theory explains these absorption spectra in a very simple manner. *A photon of light passing through an atom may be absorbed if, and only if, it has just enough energy to raise the atom from its normal state to one of its excited states.* This means that the frequency of the absorbed photon is just the same as the frequency of the photon which is emitted by the return of the atom to its normal state: in other words, that the frequency of each dark line in the absorption spectrum must be the same as that of a corresponding bright line in the emission spectrum, exactly as is observed to be the case.

A very interesting experiment illustrating the absorption of light by atoms may be performed with mercury vapor. Mercury vapor, like hydrogen gas, is entirely transparent throughout the visible spectrum, although mercury has a number of visible lines in its emission spectrum. The explanation is the same as that given above for hydrogen. The first "resonance" line is in the ultraviolet, having a wavelength of 2537A, and mercury vapor absorbs light of this wavelength very strongly. When light of this wavelength is used, even the slight amount of mercury vapor above an open dish of mercury at room temperature will cast a visible shadow upon a fluorescent screen, as has been pointed out by Wood, in his *Physical Optics*.[2] The experiment may be performed with an inexpensive mercury-vapor "sterilizing" lamp, which emits almost all its radiation in this one wavelength.[3] A fluorescent screen is placed several feet from the lamp, and a beaker containing a small amount of mercury is held in front of it. If now the air above the mercury is gently blown from the beaker, it will cast a smoky shadow upon the screen.

The same experiment may be performed even more simply with sodium vapor, since the first resonance line for sodium forms practically all of its visible spectrum. A Bunsen burner is placed between a sodium-vapor-arc lamp and a white screen, so that the shadow of the Bunsen flame falls upon the screen. This shadow will be quite faint unless some sodium vapor is introduced into the flame, by holding a little moist salt on asbestos paper, or a rod of soft glass, in the base of the flame. Then the shadow becomes inky black.

136. Fluorescence of Atoms. What becomes of the energy which is absorbed by the atoms in the process described above? There is always the possibility that a collision with another atom will convert this energy into extra kinetic energy of the two colliding atoms. This is indeed possible for any excited atom. But it is also possible that the electron which has been lifted to an excited level by the absorption of light will again return to the normal level by the re-emission of the energy as light. This re-emitted light is called *fluorescent* light, and will

---

[2] R. W. Wood, *Physical Optics*, Macmillan, Third Edition, p. 590.

[3] This lamp is like the fluorescent lamps described in Sec. 90, except that the tube is made of a special glass which is quite transparent to ultraviolet light, and is not coated with fluorescent material. For best results, all but an inch or so of the tube should be covered over so as to have a fairly small source. The fluorescent screen must be one which is sensitive to the 2537A light; most X-ray fluoroscope screens are suitable.

of course be emitted in all directions from the atom. It is most easily seen by looking into the gas or vapor at right angles to the exciting beam, and is generally weak enough to require care in eliminating from view all directly reflected or scattered light. If monochromatic light is used to produce this fluorescence, the process may be followed in detail. Thus, if the frequency of the exciting light is such as to lift the electron only one energy level above its normal level, the same frequency will be emitted as fluorescent light when the electron returns to the normal level. When the exciting frequency is higher, so that the electron is lifted several energy levels above the normal level, several return paths are then possible. In some of the atoms the electron may return directly to the normal level, but in others it may first fall to one of the intermediate levels, thus returning to the normal level by two or more stages. It follows that the fluorescent spectrum will then contain not only the line corresponding to the original exciting frequency, but several other lines of *lower* frequency as well. If the exciting frequency is high enough to lift the electron to zero energy or above it, it will ionize the atom, and the fluorescent spectrum will consist of the complete spectrum of the atom.

It is a fundamental law of luminescence, called Stokes's law, that the frequency of fluorescent or phosphorescent light is always lower than, or at the most equal to, the exciting frequency. The reason for this law is easily understood from the explanation given above for fluorescence in a gas.

137. Ionization Potential. The values which are obtained spectroscopically for the energy levels of an atom may be checked by entirely independent electrical methods. It is particularly easy to do this for the normal level. All that need be done is to admit a small amount of the gas to be studied into a two-electrode electron tube of the type shown in Fig. 7, and then to study the flow of electrons through the tube as the voltage across its terminals is slowly increased. As long as the voltage is kept low enough, the presence of this gas will have little effect. If perchance one of the electrons coming from the hot cathode bumps into an atom of the gas, it bounces off *elastically* with but negligible loss of energy, and such collisions have no appreciable effect upon the current through the tube. As the voltage is increased, this current will rise gradually, as is shown by the lower part of Fig. 8, and also by Fig. 105. The current will start to rise very much more rapidly, as is shown in Fig. 105, as soon as the voltage reaches a value called the *ionization potential*, and if attempt is made to increase the voltage above this value, the current will rise sharply to its saturation value, as the figure shows. This sudden rise of current is accounted for by the positive ions which are produced when gas atoms are bombarded by electrons which have been set into motion by a voltage equal to the ionization potential. These electrons possess enough kinetic energy to knock the outermost or valence electron completely out of an atom with which it may collide, thus leaving the atom positively charged, or *ionized.*

Although these positive ions are responsible for the big increase of current which takes place, they do not of themselves carry very much of the current, because the number of positive ions produced is quite small as compared to the number of electrons flying between the electrodes. But, because they are heavy and slow-moving, it takes only a comparatively few new ones per second to keep the space between the electrodes pretty well filled up with ions, full enough so that everywhere except close to the cathode their positive charges may just equal the negative charges carried by the faster-flying electrons. From this condition it results that the electric field is concentrated close to the cathode. In effect, the positive ions bring the surface of the anode close up to the cathode. The electric field there is then very intense, and the current goes to its saturation value, as has been described above, if attempt is made to raise the voltage higher. The value for the ionization potential is given by the voltage at the point c, Fig. 105, for which the current just begins to increase at a greater rate than it would in a vacuum. (The dotted line, a, indicates the current values corresponding to vacuum conditions.)

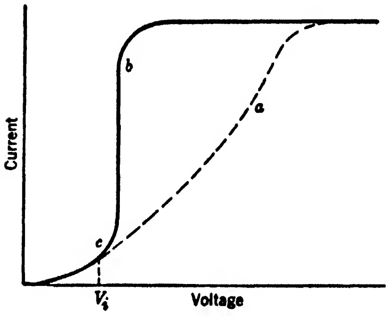


FIG. 105.

If the value of the ionization potential, V_i, is multiplied by the charge carried by the electron, the product is

$$V_ie = W_0 \tag{9·4}$$

where $+W_0$ equals the work required to lift the valence electron entirely out of the atom from the normal level, and hence $-W_0$ is the value for the *energy of the valence electron in its normal energy level.*

If special electron tubes are used, tubes which contain one or more wire-gauze electrodes or grids in addition to the cathode and the collector plate, much greater precision is possible in the measurement of the ionization potential; it is possible also to perform experiments which measure values for other energy levels as well. In all such experiments the agreement between the electrically and the spectroscopically determined energy level values is very satisfactory.

138. Excitation by Electron Bombardment. A cross check upon all these data may be obtained by observing the light emitted by a gas as it is bombarded in such tubes by electrons of known energy. The gas begins to "glow" as soon as the voltage across the tube has passed the *resonance* potential, that is, the potential which measures the energy needed to lift the valence electron to the first excited level. This "glow," which may be in the ultraviolet and hence invisible, will consist spectroscopically of just one line, the *resonance* line referred to in Sec. 135. The explanation is evidently the same as that given in Sec. 136. If the voltage is still higher, the atom will be excited to higher levels, and more lines may appear. Again the explanation is identical to the one given in Sec. 136 for the corresponding case where the excitation was by means of photons. Only when the potential exceeds the ionization potential will the complete normal spectrum be emitted. With still higher voltages the *enhanced* spectrum, corresponding to excitation of the ionized atom, may be obtained.

The light which is emitted from the atom in these experiments may be measured optically, as described above, or it may be measured photoelectrically, by means of a photosensitive surface placed inside the tube and properly screened from the discharge. Olmstead and K. T. Compton[4] used this method to observe the energy levels for the hydrogen atom. Since hydrogen gas at ordinary temperatures is diatomic, they first had to heat the hydrogen to a temperature of 2500° C to dissociate the molecules into atoms. They then determined electrically the potentials at which emission of new spectrum lines just began. The values which they obtained in this manner for the first seven energy levels agreed with the spectroscopic values to better than 0.1 per cent. For the lowest energy level (the ionization potential) they obtained 13.54 volts, as compared with the spectroscopic value of 13.539 volts.

Whatever changes may have occurred or may yet occur in theories for atomic structure, the existence of energy levels has been established experimentally beyond any doubt. The values of these energy levels for the various atoms represent our most certain and valuable information concerning them.

139. Commercial "Gas-Filled" Electron Tubes. It is extremely interesting to observe how often a scientific discovery which is made with no practical purpose in mind turns out to be of great practical value. This has proved to be so for the phenomena described in Sec. 137. If, for example, a very small amount of a suitable gas is admitted to a tube of the type shown in Fig. 7, there is thereby produced a much more

---

[4] *Physical Review*, Vol. 22, p. 559, 1923.

efficient rectifier tube for circuits such as those shown in Figs. 9 and 10. Mercury vapor is commonly used, and also argon and neon gases. This tube will serve just as well as does the vacuum-type rectifier tube to prevent the passage of current through the circuit when its plate is negative, but when its plate is positive, it will pass as large a current as may be required (up to the safe limit of the tube) without the voltage across the tube ever rising appreciably above the ionization potential (about 11 volts for mercury vapor). The explanation for the action of this tube has been given in Sec. 137. Since, when the tube is conducting, the effective anode is brought close to the cathode by the presence of positive ions, the efficiency of the tube may be further increased by partially inclosing the hot terminal so that less power is required to heat it. There is one limitation to the use of such tubes. When the voltage falls to zero, the electron flow stops, but the positive ions require an appreciable time to become neutralized by drifting to the electrodes. One one-thousandth of a second is allowed for this in the design of commercial tubes for 60-cycle operation. If the reverse voltage builds up to too high a value before these ions are neutralized, the unneutralized ions make possible a reverse current which will ruin the tube and may damage other parts of the apparatus.

Small amounts of gas may also be used in three-electrode tubes equivalent to the one shown in Fig. 13, and here the presence of gas in the tube entirely changes its properties. To be sure, if the grid potential is sufficiently negative *before* any potential is applied to the plate, no current can start, whether there be gas present or not—see Sec. 14. Once the current starts, however, the grid has no further control over it. The gas *ions* which are produced by the electron current then surround the grid wires and neutralize, by their positive charges, the effect upon the current of any negative charges on the grid, even when the grid has a high negative potential. These tubes, which are called thyratrons, are made in all sizes from the tiniest sub-miniature tubes to those capable of handling many amperes of current and working at thousands of volts. The smaller ones are used for hundreds of special purposes in laboratory work, while the larger tubes of this type have many important commercial uses, including the conversion of d-c. power into a-c. power at high voltages.

140. Atom Building. It is now possible to consider, tentatively at least, the changes which take place in the electronic structure of atoms as we go from element to element in the periodic table. Let us start with the hypothesis that the number of electrons in a normal atom equals Z, the atomic number of the atom (its nucleus consequently carrying an electric charge equal to $+Ze$), and see what support can be found for this hypothesis in the evidence given by atomic spectra. First of all, the spectra of hydrogen, of ionized helium, and of doubly ionized lithium show that the nuclei of these atoms carry respectively one, two, and three positive elementary charges. For these three atoms the spectra are of the type accurately described by Bohr's theory, and the size of the nuclear charge is directly deduced from that theory, as is explained in Sec. 128. So far, the data agree perfectly with our hypothesis.

It is difficult to obtain such hydrogenlike spectra (which require the stripping off from the atom of all but one of its electrons), for elements much beyond lithium, but the same principle can be applied to the interpretation of other spectral evidence. Thus singly ionized lithium, doubly ionized beryllium, etc., have heliumlike spectra, indicating that each of these ions has but two electrons remaining. Similarly, singly ionized beryllium, doubly ionized boron, etc., have lithiumlike spectra. As long as we limit our consideration to the first twenty elements in the periodic table (see Table I, Appendix III), we shall find that the spectrum of a singly ionized atom is of the same spectral type as that of the normal atom just below it in the periodic table; the spectrum of the doubly ionized atom, of the type of the normal atom second below; etc. Such evidence may be confidently interpreted as meaning that the number of electrons in an atom increases one by one as the atomic number increases, or is numerically equal to the atomic number, Z. This is also in accord with the experiments of Rutherford (Sec. 27) on the scattering of alpha particles.

141. Electron Shells. This evidence furthermore indicates that we might imagine each atom as being built from the one just below it by increasing the nuclear charge by one elementary charge and adding another electron, leaving the *arrangement* of the other electrons unchanged. (They will of course be crowded closer together by the increased attraction of the larger nuclear charge.) The evidence also indicates qualitatively what this arrangement may be. As has been explained in Sec. 131, the spectra of hydrogen, lithium, sodium, and potassium are all spectra which may be accounted for by a single electron; and these spectra also show that the total quantum number, n, which represents the normal undisturbed state of this single electron, equals 1 for hydrogen, 2 for lithium, 3 for sodium, and 4 for potassium. In each of these atoms, then, the other electrons must be much more tightly bound, which means that they must be identified by lower values of n. In lithium, for example, the other two electrons must be in states for which $n = 1$. The complexity of the helium spectrum, as compared to the spectra of hydrogen and lithium, indicates that the two electrons in the helium atom both occupy normal states for which $n = 1$, and this is confirmed by detailed analysis of this spectrum. We shall now postulate, subject to later experimental confirmation, that the first two electrons in *any* atom are in normal states for which $n = 1$, and have equal energies. This group of electrons, which lies closest of all to the nucleus, is commonly called the ***K* shell.**

Since the spectra of the elements between lithium and sodium likewise increase in complexity as the number of electrons increases (130 series of lines have been identified in the neon spectrum), we may conclude that the additional electrons in the atoms of these elements must possess energies comparable to that of the third one, and likewise be identified by $n = 2$. They form a group of electrons called the ***L* shell,** which reaches a total of eight in neon. Similarly, we may account for the building of an ***M* shell** between sodium and potassium, identified by $n = 3$, and reaching a total of eight electrons in argon. Then, in potassium, the valence electron begins the building of an ***N* shell.**

Beyond potassium the electron structure is not built up so regularly, and analysis is more difficult. The evidence of X-rays (Sec. 147) has verified the existence of such electron shells and provides our best evidence for the number of electrons in all but the outermost shells. By this means it has been shown that the *K* shell is complete with two electrons, and the *L* shell, with eight; but the *M* shell, which was apparently complete with eight electrons, later on builds up to eighteen (in copper). All the inert gases beyond helium have outer shells of eight electrons, and there are very few atoms with more electrons than this in their outermost shell. The electronic structures for all the elements are given in Appendix III.

PROBLEMS

1. Using the data given in Fig. 99, compute the ionization potential of hydrogen.

2. Compute the "resonance" potential for the mercury atom, if the "resonance" line has a wavelength of 2537A.

3. The lithium atom will be just ionized by the absorption of light of 2300A wavelength. Compute the ionization potential for lithium.

4. Compute the velocity with which the valence electron is thrown out of a lithium atom, by the absorption of light of 1849A wavelength (cf. problem 3).

5. Describe the electronic structure of carbon, nitrogen, and oxygen.

REFERENCES

For Additional Reading

Physics Staff, University of Pittsburgh, *Atomic Physics*, Chapters VI and IX.
J. A. Eldridge, *The Physical Basis of Things*, Chapters XIV to XVII.
J. B. Hoag, *Electron and Nuclear Physics*, Chapter VII.
G. F. Hull, *Modern Physics*, Chapters VI, VII, and X.
Foote and Mohler, *Origin of Spectra.*
Ruark and Urey, *Atoms, Molecules, and Quanta*, Chapter XIII.
J. D. Stranathan, *The "Particles" of Modern Physics*, Chapter 6, Sec. 2.

For Advanced Reading

F. K. Richtmyer, *Introduction to Modern Physics.*
R. W. Wood, *Physical Optics.*
H. E. White, *Introduction to Atomic Spectra.*
Ruark and Urey, *Atoms, Molecules and Quanta*, Chapters V and VII.
L. B. Loeb, *Atomic Structure*, Chapter XII.
Harnwell and Livingood, *Experimental Atomic Physics*, Chapters VII and VIII.

CHAPTER X

X-RAY SPECTRA AND MORE ABOUT ATOMS

142. Characteristic X-Ray Spectra. As has been mentioned in Sec. 98, the X-ray spectrum consists of a spectrum of narrow "bright" lines superimposed upon a continuous spectrum. We are at present concerned only with the line spectrum, which is characteristic of the element from which the target of the X-ray tube is made; and the most astonishing thing about this line spectrum is its extreme simplicity. In the optical range of wavelengths we have become accustomed to seeing spectra which for the most part are very complex and which differ widely one from the next as we go from element to element in the periodic table—spectra in which series are discovered and sorted out only as the result of much expert and painstaking labor. Yet in the characteristic X-ray spectra of any element we find only a comparatively small number of lines, making up several widely separated series which in this respect much resemble the hydrogen spectrum.

143. Moseley's Law. These series are called the *K*, *L*, *M*, etc., series, respectively, in order of increasing wavelength. Figure 106 which shows the *K* lines for a few of the elements, reveals an even more

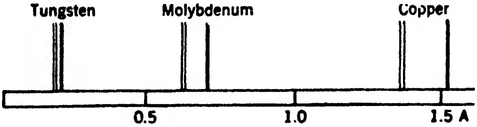


FIG. 106. *K* X-RAY SPECTRA.

astonishing fact: By stretching the spectra for the heavier elements, or by compressing those for the lighter elements, all these spectra could almost be made to coincide. Except for the *scale* of their wavelengths, these *K* series spectra are almost exactly alike![1] This similarity in

---

[1] The strongest line, which is also the line of lowest frequency, is called the *K-alpha* line, and the others are designated as the *K-beta*, *K-gamma*, · · · lines. The *K-alpha* line is actually a pair of fairly closely spaced lines, and the others should also be double, although this is rarely evident, even for the *K-beta* line.

X-ray spectra was discovered by Moseley when, in 1914, he made a systematic study of such spectra for a large number of elements. Furthermore, he found that, for the *K-alpha* line, the change in scale can be expressed by a very simple law, namely:

> **The frequency of the *K-alpha* line of any element is proportional to $(Z - 1)^2$, where Z is the atomic number of the element. This is Moseley's law.**

144. Origin of X-Rays. When a cathode-ray electron strikes the target of an X-ray tube, it is most likely that it will lose its energy by processes which may be described as collisions with the electrons in the atoms of the target. Practically all this lost energy ultimately appears as heating of the target. There is also the possibility that, when it comes close to an atomic nucleus, it will lose some or all of its energy quite abruptly, by a process which converts the lost energy into an X-ray photon. These X-rays form the continuous X-ray spectrum, whose limit is set by those electrons which lose their energy all at once by this process. The characteristic, or "bright-line," spectrum of X-rays comes from the readjustments which must be made in the electron configurations of the atoms after high-energy cathode-ray electrons have passed through them.

Let us consider first the origin of K X-rays. If the voltage across the X-ray tube is high enough, an occasional cathode-ray electron will, in its passage through an atom of the target, drive one of the K electrons entirely out of that atom. This leaves the atom with one electron gone from its K shell, but perhaps otherwise relatively undisturbed. It is now in an *excited state* for X-rays. In a short time the vacant place in the K shell will be filled by an electron falling from one of the shells farther out, and a K X-ray photon will be emitted. The energy lost by this electron as it drops to this position closer to the nucleus will determine the frequency of this X-ray photon in accord with Bohr's second postulate. If an L electron fills this vacancy, the frequency is that of the *K-alpha* line; if an M electron, that of the *K-beta* line; etc. Since the K shell is refilled from the shells next farther out, these shells must in turn be refilled before the atom regains its normal state. This results in the emission of L, M, etc., X-rays. (L X-rays are emitted when a vacancy in the L shell is filled; M X-rays, when the vacancy filled is in the M shell; etc.) Thus the emission of K X-rays is always accompanied by the emission of all the softer (longer-wavelength) X-rays. On the other hand, if the cathode-ray electrons do not have energy enough to dislodge a K electron, only the X-rays of longer wavelengths will be emitted, as has been explained in Sec. 107.

Proof for this explanation for the origin of characteristic X-rays is given by the agreement found between the observed X-ray frequencies and the theory given in the next section. It should be noted in passing that the notation of K, L, M, etc., as found here was first used to designate the different series in X-ray spectra, and was applied to the electron shells after the origin of these spectra had been discovered.

145. Derivation of Moseley's Law. Consider first an hypothetical atom with a nucleus having a charge equal to $+(Z - 1)e$, and having but one electron circling about it. Its energy levels could be computed by Bohr's theory to be $(Z - 1)^2$ times those found in Sec. 125 for the hydrogen atom. Hence they are given by

$$W = -\frac{2\pi^2 me^4(Z - 1)^2}{h^2}\frac{1}{n^2} \tag{10·1}$$

$$= -21.8 \times 10^{-12}(Z - 1)^2 \frac{1}{n^2} \text{ erg}$$

At first sight the conditions inside an atom in the target of an X-ray tube appear to be just the opposite of this hypothetical case. The atom in the target has all its inside shells completely filled until the cathode-ray electron comes along and knocks out one of them. But if it is a K electron that is knocked out, and if in addition we postulate that in all atoms *there are but two K electrons,* then the part of the atom inside the L shell very much resembles the hypothetical situation described above. Between the K shell and the L shell the electric field is that produced by the charge, $+Ze$, on the nucleus, and the charge, $-e$, on the remaining K electron, which may be considered in effect to surround the nucleus. This *screening* effect of the remaining K electron may be described either as a result of its rapid orbital motion about the nucleus, or by saying that the wave representing this electron is symmetrical in all directions about the nucleus.

However we describe it, the field between the K and L shells is essentially that due to an effective nuclear charge equal to $(Z - 1)e$, or just the field which is found in the corresponding region of the hypothetical atom first considered. When an electron falls from the L shell into the vacancy in the K shell, its loss of energy is then the same as the loss of energy in the hypothetical atom when its electron falls from the energy level for which $n = 2$ to the level for which $n = 1$, namely,

$$hf = 21.8 \times 10^{-12}(Z - 1)^2\left(\frac{1}{1^2} - \frac{1}{2^2}\right)$$

$$= 16.3 \times 10^{-12}(Z - 1)^2 \text{ erg} \tag{10·2}$$

And

$$f = 24.7 \times 10^{14}(Z - 1)^2 \text{ cycles/sec} \qquad (10 \cdot 3)$$

is then the frequency predicted by this theory for the *K-alpha* X-ray line.

The theory thus predicts that the frequency of the *K-alpha* line should be proportional to $(Z - 1)^2$, and this is the law which Moseley found to be true experimentally. Furthermore, the numerical agreement is good. For example, 193×10^{16} is the frequency value computed from the formula for the *K-alpha* line for copper ($Z = 29$), while the measured value is 194×10^{16}.

For the other lines in the K series, as well as for the lines of the other X-ray series, the theory is not so simply developed, although the principles involved are the same.

146. Atomic Structure as Revealed by X-Rays. This agreement which is found between Moseley's experimental law and the theory derived above on the assumption that there are *two* electrons in the K shell is proof that the *K shells in all atoms* whose X-ray spectra have been studied *contain* two and *only two electrons.* Thus a rule which was shown by optical spectra (Sec. 141) to hold for all the lighter elements (above hydrogen) is now verified, by X-rays, for all the rest of the elements.

In similar fashion the L X-ray spectra are interpreted to show that, beginning with neon, all atoms contain *eight* electrons in their L shells. The softer X-rays do similar service for the M, N, $\cdots$ shells, so that the electron groupings for all the elements are now known with certainty. Thus this X-ray evidence explains the amazing similarity of all the rare earths, elements 57 to 71 in the periodic table. All these elements have the same outer structure of two P electrons and either eight or nine O electrons, in consequence of which their chemical properties are almost the same, and their optical spectra very similar. The only difference in these atoms is in the number of electrons in their N shells; this is revealed by their X-ray spectra, which show clearly that these N electrons increase from 18 in lanthanum to 32 in lutecium, as is shown in Appendix III.

147. X-Ray Absorption Spectra. The coefficient for the absorption of X-rays by any substance is inversely proportional to the thickness of a layer of that substance which will reduce the intensity of the X-ray beam to a specified fraction (1/2.72) of its original value.[2] The mass-

---

[2] The law for the absorption of X-rays may be written

$$\log_e \frac{I_0}{I} = \mu x \quad \text{or} \quad \log_{10} \frac{I_0}{I} = 2.30\mu x \qquad (10 \cdot 4)$$

where I_0 = intensity of the original beam, I = intensity of the beam after passing through a layer of thickness = x, and μ is the absorption coefficient.

absorption coefficient is defined as the result obtained by dividing this absorption coefficient by the density of the substance. In Sec. 93 it has been stated that the absorption of X-rays is roughly proportional to the physical density of the substance considered, which is equivalent to saying that the mass-absorption coefficient is the same for all substances

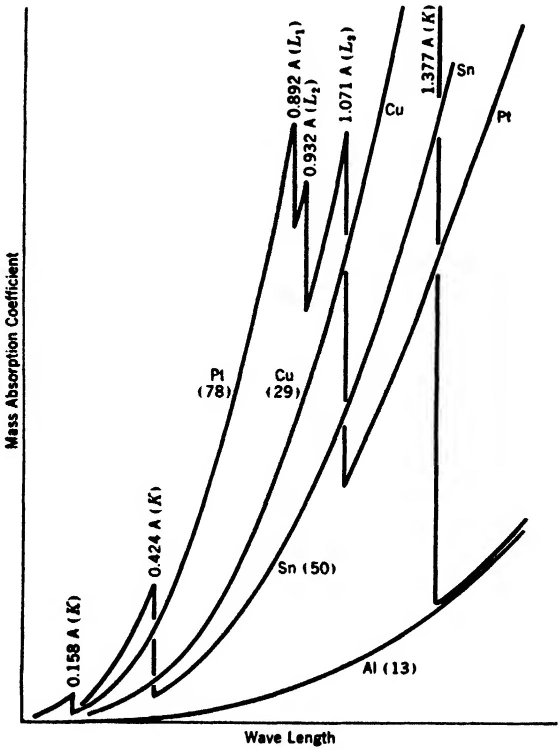


FIG. 107. X-RAY ABSORPTION SPECTRA.

for the same X-ray beam. How rough an approximation this is, when monochromatic X-rays are considered, is shown by Fig. 107. Consider, for example, the values given for platinum. As the wavelength is increased from very low values, the mass-absorption coefficient at first increases rather rapidly, but when a wavelength of 0.158A is reached the platinum becomes suddenly very much more transparent. As the wavelength is further increased, the absorption again increases rapidly, but again drops abruptly at 0.892A, which is the first of three such breaks in the plot which occur close together, the other two being at 0.932A and 1.071A. These breaks are called **absorption limits.** Copper and tin both show one such absorption limit in the range of wavelengths here represented, and aluminum also would show one if the range of wavelengths were extended. In general, a series of such absorption limits will be found for any element, the number of them increasing with the atomic number of the element. For reasons that will soon be apparent, the first of these limits for any element is called the K limit; the next three, the L limits; the next (five, close together), the M limits; etc.

These absorption limits can be directly interpreted in terms of the theory given in Sec. 144. A portion of the apparent absorption of X-rays represents losses due to scattering by elastic collisions with the outer electrons of the atom. This is discussed in detail in Sec. 108. The rest of it, and the part with which these absorption limits is concerned, may be explained as a kind of photoelectric effect. If an X-ray photon possesses enough energy to lift an electron entirely out of the atom from its normal state in one of the inner shells, then this event will sometimes occur, the photon being "absorbed" or vanishing in the process. It may also liberate an electron when it has more than enough energy to do so (it has been shown that the excess energy goes off as kinetic energy of the liberated electron), but the decrease of the absorption for frequency values higher than the absorption limit indicates that the X-ray photons are most likely to be caught when they have just enough energy to liberate an electron.

Consider now the absorption limit which is called the K limit because photons of this frequency are just able to liberate a K electron. The electron must be lifted *entirely out of the atom,* since there is no other place for it to go. All the other shells (with the exception of the outermost shells, whose energy levels are practically zero as compared to the energy-level values for the inner shells) are already filled up. This K absorption limit thus measures *directly* the *energy* of a K electron within the atom.

Similarly the L absorption limit measures the energy of the L electrons within the atom. Any explanation for these *three* L absorption limits requires that there be *three kinds* of electrons in the L shell, a requirement which had already been found necessary from other considerations, particularly to explain the existence of two *K-alpha* lines. The three kinds of L electrons will be accounted for in Sec. 151. The M, N, $\cdots$ absorption limits are explained in similar fashion. The X-ray emission spectra and absorption spectra are entirely in agreement with each other, the absorption spectra serving further to confirm the evidence concerning the innermost electronic structure of the atom which is given by the emission spectrum.

148. X-Ray Fluorescence (Secondary X-Rays). Since the absorption of X-rays by an element leaves its atoms in excited states, a substance which is irradiated by X-rays will emit its characteristic spectrum as secondary X-rays, or fluorescent X-rays. Compare this with the corresponding optical case described in Sec. 136. If the complete characteristic spectrum of an element is to be emitted, then of course the exciting X-rays must be at least as high in frequency as the *K* absorption limit for that element. For all these phenomena both theory and experiment are in agreement.

Both the absorption spectrum and the fluorescent spectrum of X-rays give means for studying the inner electronic structure of atoms without the necessity of building the elements into an X-ray tube, as targets. They also make possible the study of X-ray spectra for elements whose physical nature might not permit their use in an X-ray-tube vacuum. Data obtained in all these different ways have been utilized in the analysis of atomic structure, which is reviewed in Sec. 146.

In closing this consideration of X-ray spectra, the reader is urged to compare carefully the production and properties of X-ray spectra with the production and properties of optical spectra. In addition to the comparisons noted in this section and elsewhere, he should note the comparison between the excitation of X-rays in an X-ray tube with the excitation of optical spectra, which is described in Sec. 138. He should note also that, while an atom will not emit its optical spectrum unless the element is in the form of a gas or vapor of low density, X-rays are emitted from elements in the solid state as well as from any other state. This is understood when it is remembered that X-rays originate in the depths of the atom, where the presence of close atomic neighbors has little or no effect, and where the energies involved are vastly greater than the energies of molecule or crystal formation. (Very slight shifts of the positions of the X-ray absorption limits of an element, due to differences in the chemical composition of the substance containing the element, have been observed in a few cases.)

149. Pauli's Exclusion Principle. It is now possible to show how beautifully the electron-wave atom model accounts for these electronic structures. As was explained in Sec. 133, the permitted stationary states for an electron within an atom are those for which the electron waves form standing-wave systems. And since, in a very exact sense, the wave system *is* the electron, each electron must have a wave system which is distinct from that of any other electron. This fundamental principle may be stated more explicitly in terms of quantum numbers. Since each stationary state, that is, each standing-wave pattern, is represented by a set of quantum numbers,

No two electrons can have the same set of quantum numbers.

This rule, known as **Pauli's exclusion principle,** is one of the most important laws of wave mechanics.

Let us see how this works out for the quantum numbers, l, m, and n. The rules for the values of these quantum numbers are given in Sec. 133. Thus l may have all integral values from 0 to $(n - 1)$, inclusive, while m may have integral values from $-l$ to $+l$, inclusive.

When $n = 1$, $l = 0$, and $m = 0$. This allows only *one* electron to have the quantum number $n = 1$, or to occupy the K shell, whereas the experiments find *two* electrons there.

When $n = 2$, four distinct sets of quantum numbers are possible: For $l = 0$, $m = 0$; for $l = 1$, m may equal -1, 0, and $+1$. This allows only four electrons to have the quantum number $n = 2$, and hence to occupy the L shell. Again the value is but *half* of that found by experiment. A similar check for $n = 3, 4, 5, \cdots$ gives each time half as many electrons for each shell as the experiments show to be possible. So far it looks bad for the theory.

150. Electron Spin. This difficulty might be explained by saying that we have not used enough quantum numbers, although it would hardly do to invent a new one just to help out the theory. Fortunately there is need for an additional quantum number, for other reasons. There exists a lot of experimental evidence (some of which will be reviewed below) which indicates that the electron itself is *spinning* about an axis through its center of mass, and an additional quantum number is required to specify the angular momentum of its spin. All the evidence indicates that the electron spin is constant in magnitude, giving to the electron an intrinsic angular momentum which is one-half[3] as great as the angular momentum of the orbital motion of an electron in an orbit for which $l = 1$. Hence this new quantum number, which is called the *spin* quantum number and is designated by s, has either the value of $+\frac{1}{2}$ or $-\frac{1}{2}$, and no other values.

This additional quantum number is just what is needed by Pauli's principle, since it allows two electrons for *each* combination of values for l, m, and n. This doubles the number of electrons possible in each shell and brings the theory into agreement with the experimental evidence.

151. "Fine Structure" in Spectra. The first evidence for the existence of electron spin appeared as a *fine structure* of the lines in optical

---

[3] This is on the basis of the orbital model. According to the wave-mechanics model, this ratio is 0.615 instead of $\frac{1}{2}$. This difference is not important unless one is actually making numerical computations.

spectra. For example, the well-known yellow line in the sodium spectrum is a double line, as any good spectroscope will show, the wavelengths being 5890, 5896A, and all the other lines of the sodium spectrum are likewise double.[4] The spectra of all of the alkali metals, lithium, sodium, $\cdots$ cesium, are of this same type, although in the lithium spectrum the separation of the pairs of lines is so small as to be very hard to observe, whereas the distance between the two lines which form the first pair in the principal series of cesium is more than one-tenth of the length of the whole visible spectrum, and the name fine structure is hardly appropriate. This fine structure of spectrum lines may of course be attributed to a corresponding fine structure of the energy levels, and the existence of *electron spin* makes possible a theoretical explanation of this.

According to this theory, for any combination of the quantum numbers, l, m, and n, there may exist two stationary states, one for which $s = +\frac{1}{2}$, and the other for which $s = -\frac{1}{2}$. In the older theory the energy of a stationary state was determined by the values of n and l. (The quantum number m has nothing to do with the energy unless the atom is in a strong magnetic field. See Sec. 154.) In this modified theory, the energy is determined by n, l, and s.[5] When $l = 0$, the energy will be the same whether s equals $+\frac{1}{2}$ or $-\frac{1}{2}$. In all other cases, there is a pair of energy levels for each value of n and l. Thus all the P, D, . . . energy levels of Fig. 102 are double levels. For the lighter elements the computed effect of electron spin upon the energy is small; that is, the doublet separation is small. But the effect increases with the atomic number of the atom, and all the energy-level values computed from this theory are in excellent agreement with the values obtained from the spectra. According to this theory, also, hydrogen should have doublet levels, and this is revealed when its spectrum lines are examined by very powerful spectrometric apparatus of the interference type.

The three L absorption limits found in X-ray absorption spectra may also be explained as a type of fine structure in the energy levels of the L shell. As has been pointed out in Sec. 150, there are eight L electrons, which represent four different combinations of l and s, namely, $l = 0$, $s = +\frac{1}{2}$; $l = 0$, $s = -\frac{1}{2}$; $l = 1$, $s = +\frac{1}{2}$; and $l = 1$, $s = -\frac{1}{2}$. These, however, represent but three possible energy values, since, as has been explained above, when $l = 0$ the energy is the same

---

[4] Transitions between the P, D, F, $\cdots$ levels also show a faint third line.

[5] The quantum numbers usually specified are n, l, and j, where $j = l + s$ and is always positive. Thus, for all S levels ($l = 0$), $j = \frac{1}{2}$; for each pair of P levels ($l = 1$), j equals $\frac{1}{2}$ and $\frac{3}{2}$; for each pair of D levels, j equals $\frac{3}{2}$ and $\frac{5}{2}$; etc.

for both values of s. And these are the three energy values corresponding to the three L absorption limits. The M, N, $\cdots$ shells have respectively 5, 7, $\cdots$ separate energy levels which are similarly revealed by the X-ray spectra, and which may be explained in similar fashion by this theory.

152. Atomic Magnets. According to the original Bohr atom model, with its circular orbits, the hydrogen atom in its normal state is a little magnet. The electron revolving in its circular orbit is equivalent to a current flowing in a tiny circuit, and the equivalent magnet has a magnetic moment,

$$M_0 = \frac{eh}{4\pi mc} \tag{10·5}$$

This may be shown as follows: In electrical theory it is proved that the magnetic moment of such a circuit equals iA/c, where i is the current, and A is the area of the circuit. In this case, if v is the speed of the electron in its orbit, and r is the radius of the orbit,

$$M_0 = \frac{iA}{c} = \frac{1}{c}\cdot\frac{ev}{2\pi r}\pi r^2 = \frac{evr}{2c}$$

From Sec. 125, $v = 2\pi^2 e/h$, and $r = h^2/4\pi^2 me^2$. Hence

$$M_0 = \frac{e}{2c}\cdot\frac{2\pi e^2}{h}\cdot\frac{h^2}{4\pi^2 me^2} = \frac{eh}{4\pi mc} \tag{10·5}$$

This magnetic moment value is called the **Bohr magneton.**

In 1921 Stern and Gerlach devised an experiment by which they could measure the magnetic moments of single atoms. This was essentially an experiment in which a moving beam of the atoms was deflected by a magnetic field, the deflections being proportional to their magnetic moments. A magnetic field which was very much stronger on one side than on the other was required, since, as is well known, a uniform magnetic field will not move a tiny magnet, for example a tiny iron filing, one way or the other, but will only tend to line it up with the field. For this reason a pointed magnet is used to pick up small articles, and the same principle was applied in the Stern-Gerlach experiment. One pole of the very powerful magnet which they used was made *wedge shaped,* with a sharp edge, while the other pole was made broad.

Figure 108 shows schematically the arrangement of their apparatus. It was inclosed in an air-tight vessel which was highly evacuated. Silver atoms were used in their first experiments. Silver metal was heated in the furnace, F, from which silver vapor then escaped through very fine slits which formed it into a fine ribbonlike beam that passed parallel to the sharp edge of the wedge-shaped pole piece, N, and close to it, as shown. Just beyond the magnet poles this beam condensed upon a cold glass plate. When the poles were not magnetized, this beam condensed in a short fine line upon this plate. But when the poles were strongly magnetized, the deposit of silver upon the plate appeared as shown in the figure at *D*, although very much smaller. As is indicated by the dotted lines, the middle portion of the beam of atoms, the part which came close to the sharp edge of N, was split by the magnetic field into two beams of equal intensity, one being deflected towards the pole N, and the other *away* from it by an equal amount. This must mean that half of the atoms lined up with their north poles pointing in the direction of the magnetic field and were consequently pulled towards N, while the other half lined up in the opposite direction, or with their north poles pointing towards the north pole of the magnet, and hence were repelled. (This peculiar behavior is explained in the next section.) The deflection was only about 0.15 mm, but the beam was sharp enough to make possible a good computation of the magnetic moment of the silver atoms. The value of this magnetic moment was found to be, within the limits of accuracy of the experiment, one Bohr magneton.

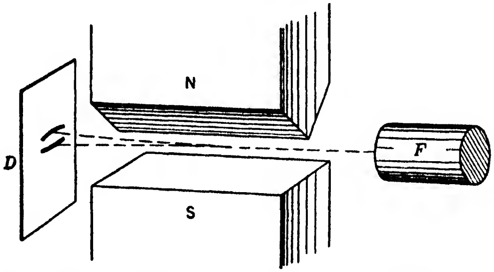


FIG. 108. THE STERN-GERLACH EXPERIMENT.

The same experiment has since been performed with atoms of other elements, by Stern and Gerlach and by various other experimenters. For hydrogen, sodium, and potassium the value found was likewise one Bohr magneton.

The result for hydrogen is in agreement with Bohr's original theory as outlined above. Bohr's theory, however, has been found inadequate to account for all the data provided by atomic spectra, and the modified orbital model and the wave-mechanics model, both of which agree much better with the spectrum data, both predict that the angular momentum of the hydrogen atom, and hence its magnetic moment, must be zero for its normal state. According to these models, the Stern-Gerlach experiment should show the magnetic moment of the hydrogen atom to be zero instead of one Bohr magneton.

Again the spinning electron saves the theory. Since the electron is an electrically charged body, its spinning gives it the properties of a little electromagnet. This then can account for the magnetic moment found for the hydrogen atom by the Stern-Gerlach experiment, which thus shows that the intrinsic magnetic moment of the spinning electron is one Bohr magneton.

Since one Bohr magneton is the value found for each of the elements sodium, potassium, and silver, it follows that the magnetic moment for each of these atoms is also due to one spinning electron. This result is of important significance to atom-building theory. It shows that, in these atoms, the directions of the spins and the orientations of the orbits for all the rest of the electrons are such as to make their angular momenta and their magnetic moments cancel. For sodium, the rest of the electrons form closed K and L shells; for potassium, they form closed K and L shells, plus a group of 8 electrons in the M shell; while in silver, they form a core consisting of a closed K shell (2), a closed L shell (8), a closed M shell (18), and a group of 18 electrons in the N shell. Both this evidence and the evidence given by spectra show that the angular momentum and the magnetic moment of a closed shell are both zero, and that this is also true for an incomplete shell of 2 or 8 or 18 · · · electrons. It is for this reason that all the alkali metal spectra can be assigned to *one* electron. It explains also why boron, aluminum, gallium, indium, and tellurium have one electron spectra (shown by the doublet character of their energy levels). Aluminum, for example, has 2 K, 8 L, and 3 M electrons. As explained above, the K, L, and the first two of the M electrons have zero resultant angular momenta, so that the normal spectrum of aluminum may be assigned to the third M electron alone.

153. "Magnetic" Quantum Number. The peculiar behavior of the atomic magnets in the Stern-Gerlach experiment, in the manner in which they lined up either with the magnetic field or against it, needs further consideration. Suppose that a large number of *spinning* magnets are placed in a uniform magnetic field. Our first expectation might be that they should all line up with their north poles pointing in the direction of the field. When, however, we remember that these magnets are spinning, we will realize that they might also remain at many different angles to the field, precessing around the lines of the field in the same manner that a top precesses about the vertical lines of gravity force. And just as the precession of the top produces reaction forces which prevent it from falling, so the precession of a spinning magnet holds it at an angle to the magnetic field.

According to the quantum theory, somewhat the same thing will happen if a large number of atoms are placed in a magnetic field, but with a most remarkable difference. To make the situation as simple as possible, let us at first neglect all the *spin* properties of the electron, just as we did at first for all the atom models. An atom in an S state ($l = 0$) will then have zero angular momentum and zero magnetic moment, and will be unaffected by a magnetic field. Let us consider it first in a P state ($l = 1$) in which it will have angular momentum equal to $h/2\pi$, and a magnetic moment equal to one Bohr magneton ($eh/4\pi mc$). Then, according to the quantum theory, it is possible for the atoms to occupy only the three positions indicated schematically in Fig. 109. They may be lined up so that their equivalent magnets are (a) in the same direction as the magnetic field, or (b) perpendicular to the magnetic field, or (c) pointing in exactly the opposite direction to the magnetic field.

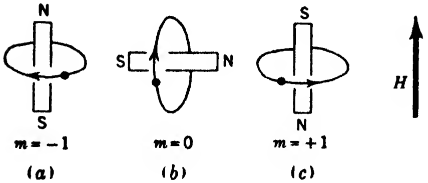


Fig. 109.

The quantum number m, which first appeared in Sec. 133 in direct consequence of the electron wave theory of the atom, but for which we have hitherto found no use, now serves to designate these different possible positions for atomic magnets in a magnetic field. The values of m corresponding to each position are indicated under each diagram in Fig. 109, and Fig. 110(a) illustrates the scheme whereby these values are determined.

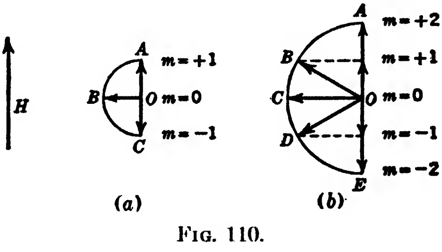


Fig. 110.

> **If, for each permissible position of the atomic magnet, there is drawn a vector line of length equal to l, the component of this vector in the direction of the magnetic field will have an integral value which will be the value of m for that position.**

Figure 110(*a*) shows this construction for $l = 1$. *OA* represents the position (*c*) of Fig. 109; *OB*, position (*b*); and *OC*, position (*a*). This is better seen when $l = 2$, as is illustrated in Fig. 110(*b*). In this case the electron orbits may occupy five different positions. Three of these, namely *OA*, *OC*, and *OE*, correspond to those possible when $l = 1$, but *OB* and *OD*, the other two, are additional ones. In general, there are $2l + 1$ possible positions, corresponding to values of m from $+l$ to $-l$, and including 0; and these values of m are just the ones allowed by the wave theory.

None of these cases corresponds to the state of the silver atoms used in the Stern-Gerlach experiments, since a normal silver atom is actually in an S state, and both its angular momentum and its magnetic moment are due to the spin of the electron, and to that alone. When this is so, the theory (as explained below) shows that there are only two possible positions for the atom, and that these positions are just the ones found by the experiment.

Whenever the electron spin (of one electron) must be considered, the *space quantization* is determined by the quantum number, $j = l + s = l + \frac{1}{2}$. (See Sec. 151, footnote 5.) The number, j, instead of l, is represented by a vector line, and the possible positions for the atomic magnets are those for which the components of j in the direction of the magnetic field are equal to m. When j is an integer plus a half, m must likewise have half-integral values. Thus, for an atom in an S state ($l = 0$), $j = \frac{1}{2}$, and m can have only the values $+\frac{1}{2}$ and $-\frac{1}{2}$, corresponding to the two positions actually observed in the Stern-Gerlach experiments for silver, hydrogen, etc.

When l is greater than zero, other possibilities exist. For example, if the Stern-Gerlach experiment were performed with atoms in states for which $j = \frac{3}{2}$, four separate beams would be formed by the magnetic field, corresponding to the four values of m: $-\frac{3}{2}$, $-\frac{1}{2}$, $+\frac{1}{2}$, $+\frac{3}{2}$. In the most general case, j (now written J) is determined by the l and s values for *all* the electrons in an atom; frequently, however, it may be ascribed only to the outermost one or two electrons, for reasons discussed at the end of the last section. When two electrons must be considered, J is an *integral* number (whichever way the spins are directed, their algebraic sum is a whole number), and m is an integral number again. Further consideration of these more complex atoms must be left to more advanced books.

154. The Zeeman Effect. Further information concerning these atomic magnets may be obtained from the spectra of atoms which are placed in a strong magnetic field. When an electric discharge tube, or other light source emitting a bright-line spectrum, is placed between the poles of a powerful electromagnet, a very powerful spectroscope (such as a concave diffraction grating of long focal length) will show that each spectrum line has been split by the action of the magnetic field into three or more close-spaced but separate lines. Further experiments will show that the amount of splitting, or the separation of the lines, is directly proportional to the strength of the magnetic field. This phenomenon is called the Zeeman effect, after its discoverer, and the group of lines into which each original line has been split is called a Zeeman pattern. Figure 111 shows several such Zeeman patterns. Sometimes this Zeeman pattern consists of only three lines, one in the position of the original line, and the other two equally spaced to either side of it. This is called a "normal" Zeeman pattern, although it occurs less frequently than do the more complex ones. (Quite a large number of the more complex patterns will appear to be normal Zeeman patterns when the spectroscope is not powerful enough.) When the source of light is viewed in a direction at right angles to the magnetic field, it will be found also that the light which produces the lines of the Zeeman pattern is plane polarized; for some of them it is polarized in a direction parallel to the magnetic field, and for others, at right angles to it. In the normal Zeeman pattern, the (electric) vibrations in the central line are parallel to the magnetic field, while those in the other two lines are at right angles to the field.

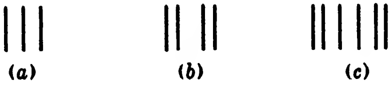


FIG. 111. ZEEMAN PATTERNS. (*a*) "Normal" Zeeman pattern. For the *Hg* singlet line at 5790A, and H = 10,000 oersteds, the lines are separated by 0.156A. (*b*), (*c*) Two examples of "abnormal" patterns.

The Zeeman effect depends for its explanation upon the orientation of electron orbits, which has been discussed in Sec. 153, and may indeed be taken as additional evidence for the existence of such space quantization of electron orbits. (It is actually the first of such evidence to be obtained.) When an atom is placed in a strong magnetic field, it will set itself into one of the possible quantized positions discussed in Sec. 153; its energy state then will be changed slightly, by an amount of magnetic energy which will depend upon its resultant magnetic moment, upon the strength of the magnetic field, and upon its position in the magnetic field. And since (with the exception of an S state for an atom in which the electron spins cancel) there are several such possible quantized positions corresponding to each original energy state (same values of n, l, and s), each original energy level is split up into several close-spaced levels, according to the allowable values of the magnetic quantum number, m. The additional transitions which these new energy levels make possible then account for the observed Zeeman patterns. The Zeeman effect is of great importance to spectroscopy in a practical way, since the same Zeeman pattern is found for all the lines belonging to the same spectral series, and may therefore be used as a means for sorting out lines belonging to different series.

If we limit our consideration to the normal Zeeman effect, we may now analyze this phenomenon in more detail. The normal Zeeman pattern is produced in an atom in which the electron spins cancel. (Such a condition is found in an atom of an element such as helium or mercury, in which there are two valence electrons, when those two electrons have their spin axes pointing in opposite directions.) To make the treatment still simpler, we shall consider a principal series line, corresponding to a transition from a P level to an S level.

The conditions will then be exactly those described in Sec. 153. The S state will be unaffected by the magnetic field, since S orbits have no magnetic moment. For the P state, however, three positions for the atom are possible, as has been explained in Sec. 153. Originally the atom in this P state had the same amount of energy in all these positions. In the magnetic field, however, the orbit whose position is shown by (c) in Fig. 109 gains magnetic energy equal to the product of its magnetic moment by the strength of the magnetic field, or

$$w = \frac{eh}{4\pi mc}H \tag{10·6}$$

The one shown by (a) loses an equal amount of energy, while the one shown by (b) remains unchanged. In this way the original P energy level is split into three equally spaced energy levels which are, however, very close together, even when the magnetic field is quite strong. The three Zeeman lines which now appear are obviously lines which represent transitions from each of the three P levels to the single S level, and when their separation in the spectrum is measured it is found to be just the amount here predicted by the theory. Equally good agreement between theory and experiment is found for other Zeeman patterns.

PROBLEMS

1. If the *K-alpha* line of nickel is 1.655A, what is the wavelength of the *K-alpha* line of silver? Use Moseley's law.

2. Plot a curve with ($Z - 1$) as abscissae (x-values), and square roots of the frequency of the *K-alpha* lines as ordinates (y-values), for the following elements (consult Appendix III for Z values):

Element:	K	Cu	Mo	Sn	W
Wavelength:	3.73	1.54	0.71	0.487	0.209

Explain how this plot verifies Moseley's law.

3. (*a*) Compute the numerical value of the Bohr magneton. (*b*) Compute the energy of a Bohr magneton when placed in a magnetic field of 10,000 gauss, for each of the three positions shown in Fig. 109.

4. Show that Pauli's principle allows a maximum of 18 electrons in the M shell, and 32 in the N shell.

REFERENCES

For Additional Reading

MAX BORN, *The Restless Universe*, Chapter IV.
J. A. ELDRIDGE, *The Physical Basis of Things*, Chapters XVIII to XXII.
PHYSICS STAFF, UNIVERSITY OF PITTSBURGH, *Atomic Physics*, Chapter X.
G. E. M. JAUNCEY, *Modern Physics*, Chapter X.
R. A. MILLIKAN, *Electrons, Protons, Photons, Neutrons, and Cosmic Rays*, Chapter XII.

For Advanced Reading

F. K. RICHTMYER, *Introduction to Modern Physics*.
L. B. LOEB, *Atomic Structure*, Chapters VI to XI.
RUARK and UREY, *Atoms, Molecules, and Quanta*, Chapters VIII and IX.
MAX BORN, *Atomic Physics*, Chapters V and VI.
HARNWELL and LIVINGOOD, *Experimental Atomic Physics*, Chapter IX.

CHAPTER XI

MOLECULAR MOTIONS

155. Molecules. The philosophical concept of "atoms" as the indivisible "ultimate" particles of matter (roughly equivalent to what we now call molecules) is a concept which goes back at least as far as the ancient Greek philosophers, and is probably as old as philosophy itself. The concept of atoms as the ultimate entity of a chemical element, and of molecules as combinations of such atoms, is one of much more recent origin, which was still being used somewhat tentatively by chemists as late as the beginning of the nineteenth century. And the recognition of molecules as *physical* entities was not firmly established until the kinetic theory of gases had been developed, or about the middle of the nineteenth century.

All this is ancient history as compared to the developments in physical science which have so far been considered in this book, and it may appear as if we were progressing backwards in now considering the kinetic theory of matter. It is necessary, however, to review the kinetic theory here in order to set up a proper background for the understanding of the latest developments, and of course the molecule as we now know it is not the simple mechanical particle which it at first appeared to be.

156. Solids, Liquids, Gases. It is rarely that we may deal with particles of atomic dimensions as individuals. Individual helium nuclei, shot out of a radioactive element, or ions accelerated by Van der Graaf generators or by cyclotrons, may produce separately visible flashes of light upon a luminescent screen or separate tracks in a cloud chamber. With few other exceptions, we must deal with atoms and molecules in countless numbers, in the form of solids, liquids, or gases. Even a single spot on a mass spectrogram, although it may represent only one kind of atomic or molecular ion, is the result produced by a stream of many, many ions of that kind which have been sorted out as to speed and mass by the magnetic and electric fields of the mass spectrograph. All of our "practical" dealings with matter have to do with solids, liquids, and gases of various kinds.

The kinetic theory is significant to modern physics as the first scientific attempt to envision the invisible: to determine experimentally the physical nature of things which, because of their infinitesimal sizes, could never be seen, and to do this by experiments which could be performed only upon aggregates containing vast quantities of such particles. The evidence obtained in this way is all indirect—circumstantial evidence, if you will; and entirely new methods of interpreting such evidence—statistical methods—had to be developed. The first success for these methods was obtained in explaining the physical nature of a gas. To the ancient Greek philosophers the solid state represented the simplest state of matter, and it seriously taxed their ingenuity to account for the infinite expansibility of the gaseous state.

This is now quite reversed; in the light of present-day knowledge of the component parts of matter, the gaseous state appears in many ways to be the simplest state.

157. Gas Laws. First of all, this is evidenced by the simplicity of the physical laws which specify the compressibility and the thermal expansion of a gas. For solids and liquids, these quantities vary quite irregularly from one substance to the next. But *all* gases, if sufficiently low in density,[1] have the same compressibility, as stated by Boyle's law:

For a fixed mass of gas at constant temperature, the volume is inversely proportional to the pressure.

Likewise (again subject to the same restriction as regards density), it is found that

All gases have the same coefficient of thermal expansion.

The value for this coefficient is 1/273 for one centigrade degree rise in temperature, when the initial volume is measured at 0° C and the pressure is held constant. This is Gay-Lussac's law.[2]

It is always possible to state a law expressing change with temperature of some physical quantity (volume in this case) more simply by revising the temperature scale so that the *revised zero point* is the point at which the law predicts zero value for that quantity. In this case, $-273°$ C is the temperature indicated by the value of the coefficient of expansion as the temperature at which the volume of any gas will

---

[1] The gas density must be very much smaller than that of its liquid or solid phase. This same restriction applies throughout this chapter where gases are concerned.

[2] Although this law [or more often, equation (11·1), which is deduced from it] is often called Charles's law, historical evidence shows that the credit for it belongs to Gay-Lussac. See an article by W. J. Lyons, *American Physics Teacher*, Vol. 6, p. 256, October, 1938.

become zero. If then we take our new zero point 273 centigrade degrees below the melting point of ice, we have a **gas scale** of temperatures such that

$$\frac{V}{V_0} = \frac{T}{T_0} \quad \textit{when the pressure is held constant} \tag{11·1}$$

T, T_0 are temperatures measured on the revised temperature scale, and V, V_0 are the corresponding volumes.

When this law is combined with Boyle's law, the general gas law[3] results, namely,

$$PV = RT \tag{11·2}$$

158. The Gas-Law Constants.[4] The value of R, which is equal to P_0V_0/T_0, will depend upon the mass of gas used, and the kind of gas. When, however, the mass of gas taken is numerically equal to the molecular weight (this mass of substance is called a *gram molecule*, or a *mole*) then it is found that the constant R has a value R_0, which is the same for all gases, and equal to 8.31×10^7 ergs/centigrade degree. Since a *mole* of any substance contains 6.02×10^{23} molecules, this is equivalent to saying that R_0/N_0 is a constant for all gases. This constant is called **Boltzmann's constant,** and its numerical value is

$$k = \frac{8.31 \times 10^7}{6.02 \times 10^{23}} = 1.38 \times 10^{-16} \text{ erg/centigrade degree}$$

---

[3] *Proof*: This proof is generally given in elementary physics books, but will be repeated here for convenience. Starting with the algebraic expression for Gay-Lussac's law,

$$V_1 = V_0\left(1 + \frac{1}{273}t\right)$$

$$= V_0\left(\frac{273 + t}{273}\right) \text{ (pressure constant at the value } P_0\text{)} \tag{11·3}$$

If now the gas scale of temperatures is used, $273 = T_0$, $273 + t = T$, and

$$V_1 = V_0\frac{T}{T_0}, \text{ when the pressure } = P_0 \tag{11·4}$$

Next, if V_1 is changed to V by changing the pressure from P_0, the value which it has so far held, to P, then (in accord with Boyle's law)

$$PV = P_0V_1 = \frac{P_0V_0T}{T_0}$$

or

$$PV = RT \tag{11·5}$$

[4] More precise values for these constants are given in Appendix II.

The gas law may now be written

$$PV = NkT \tag{11·6}$$

159. Avogadro's Law. The volume, V_0, of a mole under standard conditions of pressure and temperature will likewise be the same for all gases. Its value is 22.4 liters at 0° C and one atmosphere. Since a mole of any substance contains the same number of molecules, this is equivalent to saying that, in general:

> **At the same temperature and pressure, equal volumes of all gases contain the same number of molecules.**

This is Avogadro's law. It was not accepted by the chemists within Avogadro's lifetime, but is now taken as a primary basis for determining molecular weights, after the ratios of the elements in a compound have been determined.

160. Kinetic Theory. How can these gas laws be accounted for in terms of molecules? If all the molecules are in sufficiently rapid motion, it is fairly easy to do this. We may make a "model" of a gas by imagining a large room containing a great many tennis balls. To be specific, say a room 10 yards square by 5 yards high, containing 4000 balls. If all these balls are at rest, they will occupy about 1 cubic yard of volume, or cover about one-fifth of the floor area. But if, in some manner, they are kept in motion at high speeds, then they will fly about in all directions, bumping into each other; and as a result of such collisions they will be, *on the average*, about equally distributed throughout the room. The average "density" of this distribution will then be 8 balls per cubic yard.

Many of the characteristics of such motions may be demonstrated with the two-dimensional gas model shown in Fig. 112. In this model a dozen or more ping-pong balls, in a shallow glass-walled box, will simulate the molecules of a gas if they are kept in rapid motion. Since the ping-pong balls are not perfectly elastic, they must be supplied with energy continuously, by means of a paddle at the bottom of the box. Figure 113 shows a series of "slow-motion" movie frames of the motion of these balls, taken at the rate of 128 pictures per second. It will be noted first of all that the distribution of these balls was far from uniform and varied widely from time to time, that is, from picture to picture, although we have said that the *average* distribution should remain the same. This will always be so when a small sample is taken; averages must always include a very large number of particles. In this model it is observable also that the weight of the balls causes the density of their distribution in general to be greater towards the bottom.

This will be true also for gas molecules, although the difference may be negligible unless the container is very high, since the gas molecules are so very light. The distribution may be considered uniform if the average kinetic energy of the balls is very much greater than the difference

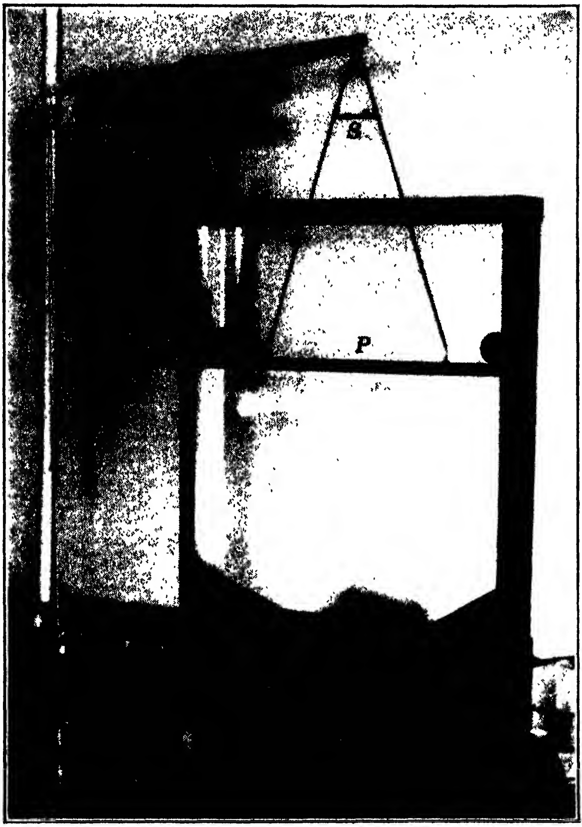


FIG. 112. TWO-DIMENSIONAL GAS MODEL WITH PING-PONG BALL "MOLECULES." The weight of the "piston" P is balanced with the counterweight B, so that the pressure against P may be measured with weights placed on the scale-pan S. (For a fuller description of this apparatus, see T. B. Brown, *American Journal of Physics*, Vol. 9, 1941, pp. 168-172.)

between the potential energy of a ball at the top of the container and that of one at the bottom.

Furthermore, if the balls are kept in motion in some manner so as to keep the *total kinetic energy* constant, the average speed of all the balls will remain the same. The individual speeds will differ widely; this may be seen in Fig. 113, where some of the balls are practically at rest,

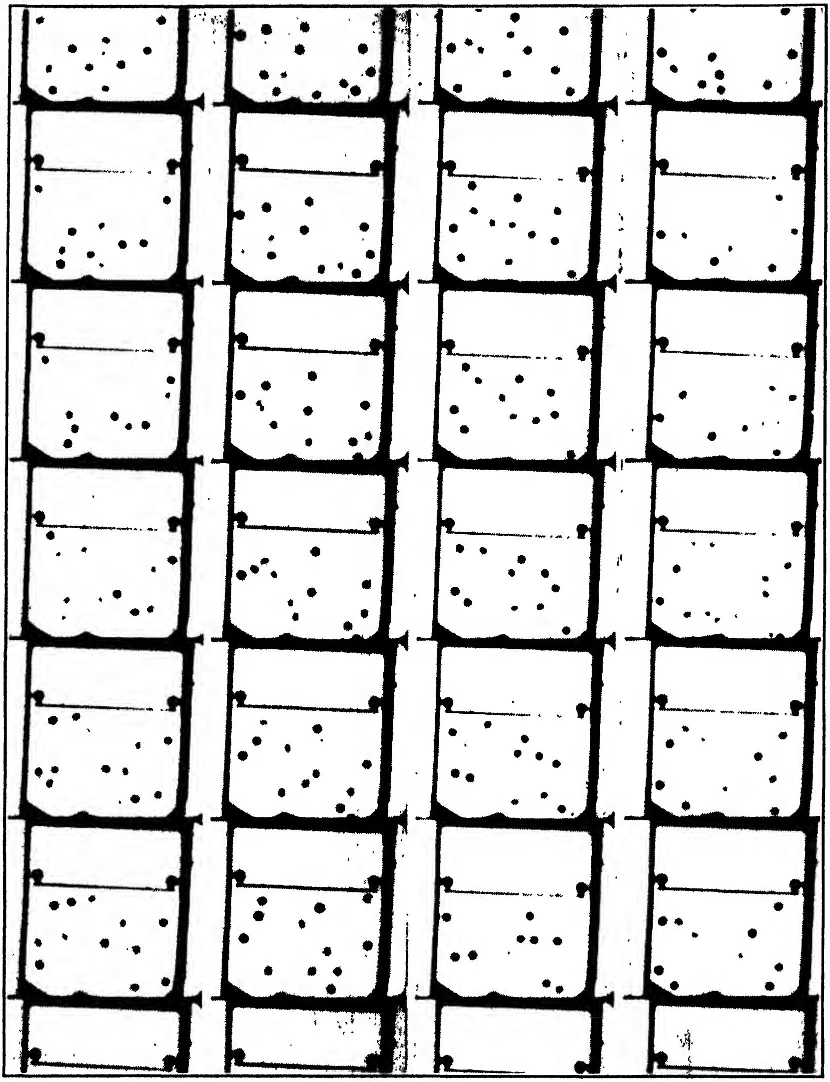


FIG. 113. "MOVIE" FRAMES OF PING-PONG BALL GAS. Taken against a bright background. Read strips from the top down. Strips are not consecutive. Fast balls produce blurred images; very fast balls may not show at all.

while others are moving so fast as to have badly blurred images. Each ball will change its speed at every collision. Sometimes it may be moving very fast, and sometimes it may be nearly at rest; but its average speed over a long enough time will remain constant. Also, as a result of collisions, the average speed in any one direction will remain constant, and the average value will be the same for all directions. The pressure exerted by a gas against the walls of its container may now be explained, in the tennis ball model, by the bumping of the balls against the confining walls. Each ball of itself exerts a sharp blow of short duration, but the average effect of many blows per second is a nearly steady push distributed evenly over the whole wall.

The derivation of a formula by which to compute the value of this push is greatly simplified by means of a mathematical subterfuge first used by Joule: This is to assume that all the balls have the same speed,

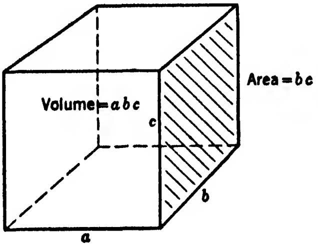


FIG. 114.

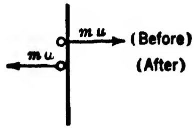


FIG. 115.

and that one-third of them are moving east and west, one-third north and south, and one-third up and down, and without any collisions except with the walls. Although the individual balls are actually moving in entirely random directions, with speeds which vary both above and below the average speed, the *average* effect of *all* of them is mathematically the same as that produced by the highly artificial and physically impossible arrangement just described. Let us also assume that the balls and the walls are perfectly elastic, so that we do not have to account otherwise for the energy which keeps them going. To make the solution of general use, call the dimensions of the room a, b, c (see Fig. 114); the number of balls, N; and their common speed, U.

Consider the balls which are moving in the east-west direction. When one of these balls hits the east wall, it rebounds with the same speed, but in the reverse direction. Since its velocity has been just reversed, the *change of momentum* is $2mU$ for each collision (see Fig. 115); and, as it travels across the room and back again between collisions, it makes $U/2a$ such collisions every second. Newton's second law of motion now tells us that the total force exerted by the east wall against all the balls striking against it in one second is the total change of momentum given by it to all the balls in one second. That is,

F = number of balls × number of bumps per ball per second × change of momentum per bump

$$= \frac{1}{3}N \times \frac{U}{2a} \times 2mU = \frac{NmU^2}{3a} \qquad (11\cdot7)$$

This same result will represent the actual case if the average speed, U, assumed for this proof, is not the *numerical* average, but the speed determined by the *average kinetic energy* of the balls. (This is sometimes called the square-root-mean-square value, for reasons that will be recognized by mathematically minded students.)

The pressure against the walls will then be

$$P = \frac{F}{\text{Area of wall}} = \frac{F}{bc} = \frac{NmU^2}{3abc} \qquad (11\cdot8)$$

Since $abc = V$ = the volume of the room,

$$PV = \tfrac{1}{3}NmU^2, \text{ and is a } \textit{constant} \qquad (11\cdot9)$$

But this is the same relation between pressure and volume which is expressed by Boyle's law. Our tennis-ball model gas "obeys" Boyle's law! Indeed, if we make the same assumptions for the molecules of a real gas at constant temperature which we have here made for the room full of tennis balls, obviously we shall arrive at the same result.

161. Maxwell's Law. If this tennis-ball model for a gas is to be in equally good agreement with the rest of the general gas law, a further assumption must be made, such that

$$\tfrac{1}{3}NmU^2 = RT \qquad (11\cdot10)$$

The nature of this assumption will be better seen if we rewrite this as

$$\tfrac{2}{3}N(\tfrac{1}{2}mU^2) = RT = NkT \quad \text{(see Sec. 158)}$$

or

$$\tfrac{1}{2}mU^2 = \tfrac{3}{2}kT \qquad (11\cdot11)$$

If this equality is to hold true, then the **average kinetic energy of the balls (or molecules) must be proportional to the gas-scale temperature, *T*.** This is **Maxwell's law.** Numerically,

$$\tfrac{1}{2}mU^2 = \tfrac{3}{2} \times 1.38 \times 10^{-16}T = 2.07 \times 10^{-16}T \text{ erg} \qquad (11\cdot12)$$

(Note that this law applies to molecular motions of *translation* only. Molecular motions of rotation and vibration will be considered in Chapter XII.)

A "working model" of a gas may thus be made by assuming that the molecules are: (*a*) very numerous and all alike, (*b*) very small as compared to their average distance apart, (*c*) hard, perfectly elastic spheres, (*d*) moving about at high speeds, in entirely random directions. It must also be assumed (*e*) that there are no forces acting between molecules (except at the instant of collision), and (*f*) that their average kinetic energy is proportional to the gas-scale temperature, *T*.

162. Transpiration. Direct and independent experimental evidence for Maxwell's law is given by the passage of a gas through a porous wall. If, for example, oxygen is inside and hydrogen gas outside (air and illuminating gas will also serve to demonstrate this) the unglazed porcelain cup shown in Fig. 116, then the gases will mix by passage through the pores of the cup. When the experiment is carried out, the manometer shows that the pressure inside the cup increases; although initially both gases were at the same pressure, hydrogen is passing into the cup more rapidly than oxygen is passing out. The experiment may be modified so as to have gas on one side of such a porous wall and a vacuum on the other. If the measurements are then made with different gases, it is found that the number of molecules escaping through the pores per second is directly proportional to the square root of the gas-scale temperature and inversely proportional to the square root of the molecular weight of the gas. Thus, in the example first cited, hydrogen goes through the pores of the cup four times as fast as oxygen does.

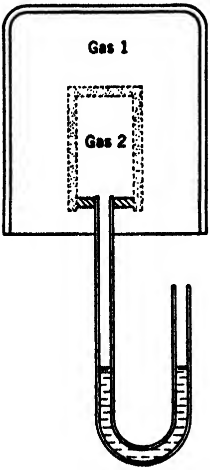


FIG. 116. TRANSPIRATION EXPERIMENT.

The photographs of the ping-pong-ball model, shown in Fig. 117, illustrate what is going on in this experiment. The pores of the cup are so small that the gas cannot stream through, but must shoot through one molecule at a time, just as the balls in this model must go through the hole in the partition which represents one pore of the cup. A ball will shoot through this hole only when it happens to be headed directly towards it, and the *chance* that this will occur is directly proportional to the average number of times per second that all the balls hit the partition. It is easy to see that this, in turn, is proportional to the average number of balls on that side of the partition, and to their average speed.

For molecules, the average speed is given by Maxwell's law (equation 11·11) as

$$U = \sqrt{\frac{3kT}{m}} \tag{11·13}$$

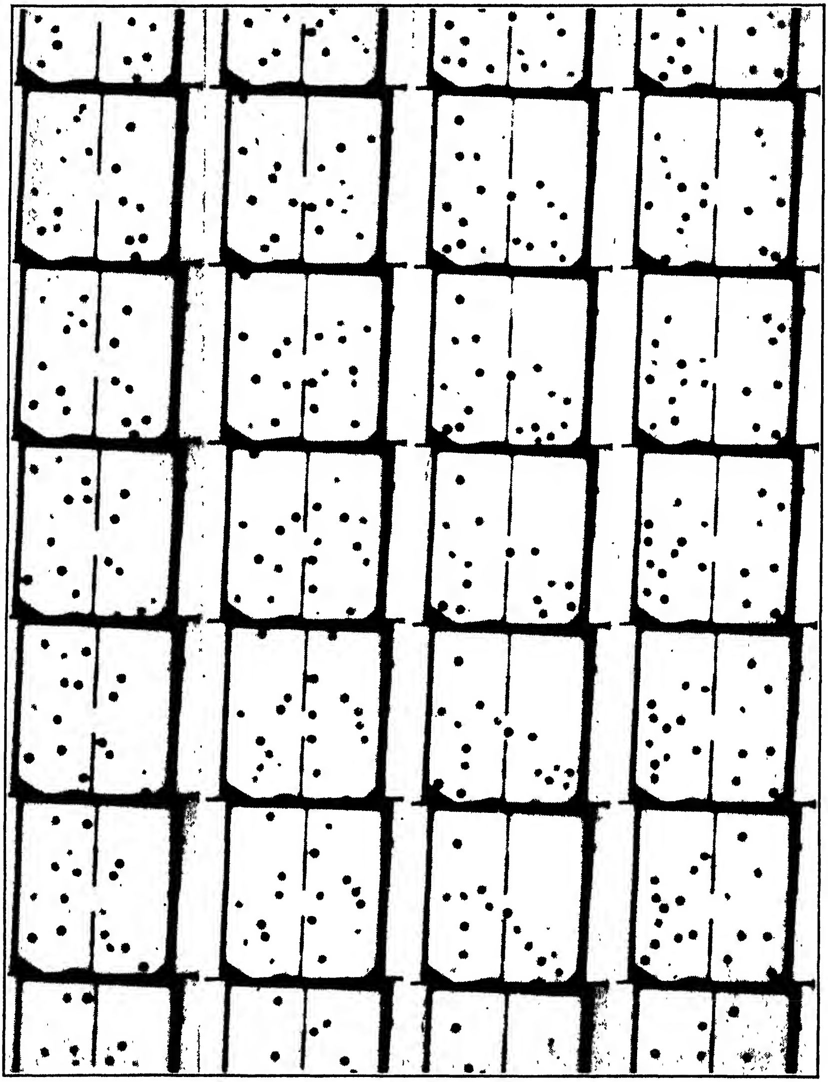


FIG. 117. TRANSPIRATION FOR A PING-PONG BALL GAS. Read each strip from the top down. Strips are not consecutive.

Hence, for two gases at the same temperature and pressure, the average speeds are inversely proportional to the square roots of their molecular masses, or of their molecular weights. Since the number of molecules per unit volume is the same under these conditions, the rate of transpiration, or the number of molecules passing through the pores of the cup per second, will then be inversely proportional to the square roots of their molecular weights, just as was discovered by experiment. The kinetic theory of gases thus passes another test for its validity.

If a mixture of gases is put into such a porous container, the lighter component will escape the faster, leaving a higher concentration of the heavier component behind. The phenomenon is comparable to the partial separation of two substances of different volatility (e.g., water and alcohol) by distillation. Successful use of this method for the partial separation of isotopes was made first with chlorine, and then with mercury vapor.[5]

163. Absolute Temperature. Maxwell's law gives a new and more tangible meaning to the physical concept of temperature. In terms of this law, **temperature (of a gas) may be defined as a quantity proportional to the average kinetic energy of the translational motions of its molecules.** The zero point of this scale is evidently that temperature at which the average kinetic energy is zero, and hence the temperature at which all the molecules are at rest. It is quite obviously the lowest possible temperature; so that in this sense it may be called **absolute zero,** and the temperature scale, the absolute scale.[6]

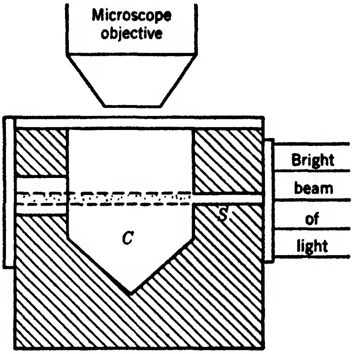


FIG. 118. BROWNIAN MOVEMENT APPARATUS. Light passing through the slit, *S*, illuminates a thin layer of smoke particles in the chamber, *C*. In the field of the microscope each smoke particle appears as a tiny point of light, in minute, erratic motion.

164. Brownian Movements. Although it is impossible to verify the motions of gas molecules by direct observation, it is possible to come surprisingly close to doing this. If the tiny particles of a smoke cloud (produced by a burning match or cigarette) are viewed through a microscope, they will be seen to be in continuous, erratic motion. Figure 118 shows an arrangement for viewing these motions, which are called

[5] An interesting discussion of the various methods employed for separation of isotopes is given by D. E. Wooldridge in the *American Physics Teacher*, Vol. 6, pp. 171–181, August, 1938.

[6] A more exact definition of absolute temperature is made in the theory of thermodynamics, but it is quite exactly equivalent to this one if the gas density is kept low enough. Even at absolute zero the gas molecules must not condense, but must be imagined as coming to rest in free space.

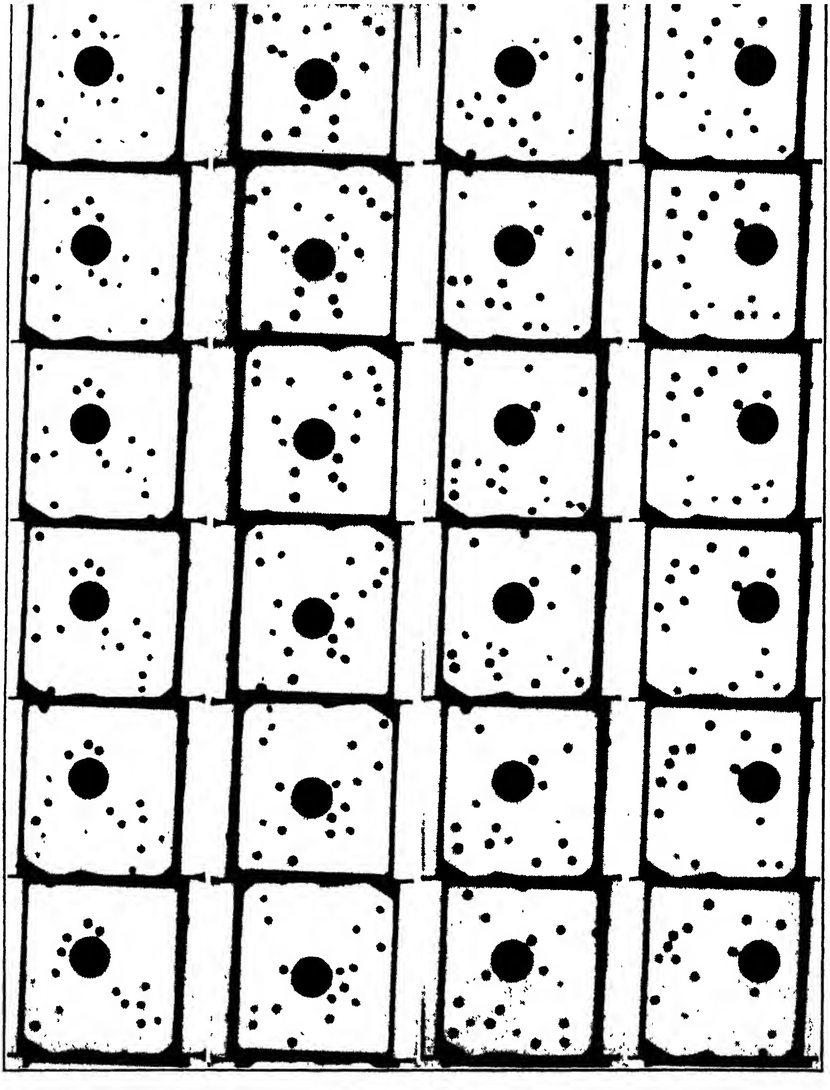


FIG. 119. BROWNIAN MOTIONS IN A PING-PONG BALL GAS. The large circular shadow is a wood disk suspended by a wire. Its "Brownian motions" consist of slow horizontal displacements which show up only by comparison of the several strips taken at rather long intervals apart.

Brownian movements, after the name of the botanist who first discovered them (1827). Erratic though these motions are, Perrin found means of making experimental measurements upon motions of similar particles, and he showed that *the average kinetic energy of these particles* *is the same as that given by Maxwell's law for the average gas molecule.* Although these smoke particles are immensely larger than the molecules, the molecules bumping against all sides of these smoke particles share their energy with them.

These Brownian motions also may be demonstrated with the model of Fig. 112, as shown in Fig. 119. Here the wood disk, which represents a Brownian particle, moves so sluggishly that frames several seconds apart must be compared to detect its motion. These pictures show very clearly how the irregular distribution of the balls about the disk accounts for its erratic motions.

165. Distribution of Speeds. Another important fact concerning Brownian movements is the random distribution of the speeds of these particles. Although the average speed remains constant (unless the temperature changes), the speeds of individual particles differ very greatly from one to the next, and are continuously changing. More are to be found with speeds near the average than far different from it; but a few may have nearly zero speeds, and a few, very high speeds. The same distribution in speeds applies also to molecules. Perrin's data may be interpreted to show that the distribution in speed of the molecules is just that distribution which is predicted by the laws of probability or chance. A variety of other experiments similarly verify this "law of chance" distribution of molecular speeds, which is shown in Fig. 120 for nitrogen molecules at 25° C. The abscissa values represent the speeds, and the height of each block is proportional to the number of molecules having speeds within the short range of speeds represented by the width of the block, divided by the width of the block. If the blocks are made very narrow, their tops will approximate to the smooth curve.

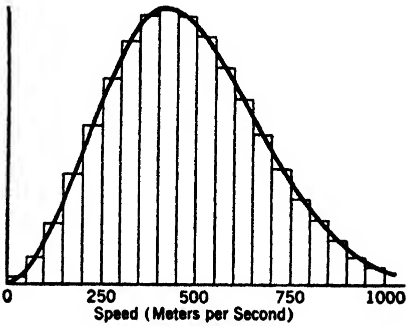
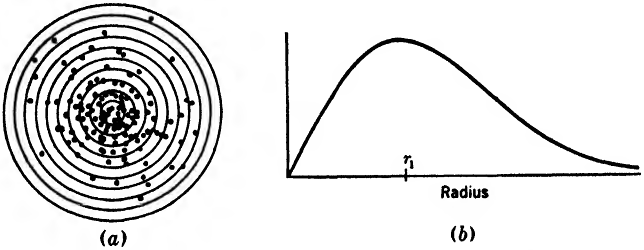


Fig. 120. Distribution in Speed of Nitrogen Molecules at 25° C.

166. Law of Chance. The distribution of shots in a target may look somewhat like Fig. 121(a). If we use targets containing a large number of rings of equal width, as shown, and count the number of shots in each ring for a very large number of targets made by marksmen of equal ability, we shall find that the counts may be represented by a plot such as shown in Fig. 121(b). Although the density of shots is greatest at the bull's-eye, the rings increase in area as they get larger, so that the maximum number of shots occurs in the ring of radius r_1.

With a less skillful group of marksmen the "scatter" will be greater, and the radius of this ring will be larger. The average distance of each shot from the bull's-eye will be approximately equal to r_1, but not exactly so.

In a two-dimensional gas (e.g., marbles rolling about over a frictionless surface) the distribution of velocities will be mathematically the same as the distribution of shots on a target. If a plot is made of these velocities, each point in the plot representing the x- and y-components of the velocity of a molecule of the gas, the plot will exactly resemble the target shown in Fig. 121(*a*); and Fig. 121(*b*) will similarly represent the number of molecules having speeds in the ranges of speed represented by the successive rings of the target. The "scatter" of the points would increase with the temperature of the gas; the radius, r_1, would be proportional to the square root of the absolute temperature. Since a real gas is three-dimensional, the curve of Fig. 120 corresponds to an equivalent "target" which is likewise three-dimensional, in which the rings of the flat target are replaced by spherical shells.



(*a*) (*b*)

FIG. 121. DISTRIBUTION OF SHOTS IN A TARGET.

Mathematical development of this idea leads to an equation for this curve, which is

$$y = \frac{A}{\sqrt{T}} \frac{a}{e^a} \tag{11·14}$$

In this equation, A is a constant, $e = 2.72$ (the base for natural logarithms), and

$$a = \frac{mu^2}{2kT} \tag{11·15}$$

where u is the speed.

Mathematically minded students will be interested to note that both the arithmetic average, U_1, and the speed, U, corresponding to the average kinetic energy, are somewhat higher in value than U', the value of the speed which corresponds to the peak of the curve (the "most probable" value). All three of these average values are proportional, however, and change together as the temperature changes; i.e.,

$$U' : U_1 : U = 1.00 : 1.13 : 1.22 \tag{11·16}$$

167. Mean Free Path. Under standard conditions of pressure and temperature there are 26.9×10^{18} molecules in each cubic centimeter of any gas. If the molecules are assumed to be arranged in regular rows, the average distance between molecules may then be computed as

$$\frac{1}{\sqrt[3]{26.9 \times 10^{18}}} = 33 \times 10^{-8} \text{ cm} = 33\text{A} \qquad (11 \cdot 17)$$

How far on the average can one molecule move before striking another?

Let us view this problem in terms of a shooting gallery again. Figure 122 represents a single layer of molecules (still assumed to be regularly spaced, with their centers 33A apart) with a square target drawn about each one. The total area of each target is then $(33\text{A})^2$. The bull's-eye at the center of each target is drawn so that a bull's-eye shot will represent a collision with the molecule at its center. Two molecules will collide if the distance between their centers equals their diameter (see Fig. 122, where *a* and *b* represent two such molecules), so that the radius of the bull's-eye is made equal to the *diameter* of a molecule. A fairly representative value for the diameter of a gas molecule is 3A. Using this value, the area of the bull's-eye is $\pi(3\text{A})^2$. Since the molecules are moving entirely at random (shooting blind), the *chance* of hitting the nearest molecule is measured by the ratio of the area of the bull's-eye to the whole area of the target. Here this chance is $\pi(3\text{A})^2/(33\text{A})^2$ or about 1/40. On the *average* then each molecule will pass 40 molecules before hitting one. The average distance between collisions is then $40 \times 33\text{A} = 1300\text{A}$, or 1.3×10^{-5} cm. This distance is called the *mean* (average) *free path* of the molecule.

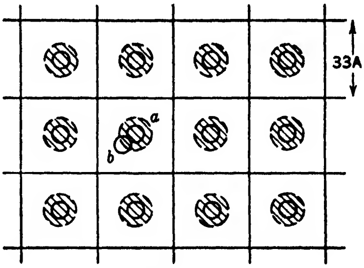


FIG. 122.

This of course is entirely too simple a proof; in actuality the molecules are not regularly arranged, nor are they standing still, waiting to be hit. The irregularity of spacing averages out; but the motions of the molecules, curiously enough, bring them somewhat more often into each other's way, so that the mean free path computed by the more exact theory is somewhat less than that computed above, although the order of magnitude is the same. Our simple theory gives the mean free path as

$$L = \frac{1}{\pi N_1 d^2} \qquad (11 \cdot 18)$$

while the more exact theory gives this value divided by $\sqrt{2}$. In this formula, L is the mean free path, N_1 is the number of molecules per unit volume, and d is the diameter of the molecule. For the example cited above, $L = 900\text{A}$ or 0.9×10^{-5} cm by the more exact formula.

The mean free path, as is indicated by equation (11·19), is inversely proportional to the density of the gas, so that in what are termed "high vacua" the mean free path may be quite large. For example, at a pressure measured by 10^{-6} mm of mercury, or about one-billionth of an atmosphere, the mean free path in nitrogen is about 100 meters, although there are still left about 27 million molecules of gas to each cubic millimeter. And there is spectral evidence to show that, in those vast glowing regions in the heavens which are called nebulae, the mean free path may be measured in miles.

168. Heat Conduction in Gases. In the process of heat conduction through a gas, heat is carried from the hot side to the cold side of a layer of gas by collisions between molecules, and the farther the molecules can travel between collisions, the faster the heat will be transferred. It follows that the rate of heat transfer through a gas is directly proportional to the mean free path, L. A detailed application of the kinetic theory to this phenomenon shows that the thermal conductivity, K (the heat conducted through unit cross-section area, per second, when the temperature gradient is one degree per unit thickness), is, except for a possible constant correction factor,

$$K = \tfrac{1}{3}\rho U_1 L C_v \qquad (11\cdot 19)$$

In this formula U_1 is the average speed of the molecules [see equation (11·16)], ρ is the density of the gas, and C_v is the specific heat of the gas under conditions of constant volume.

The phenomena of mixing of gases by diffusion, and of viscosity in a gas, are also easily interpreted by the kinetic theory, and their coefficients are likewise directly proportional to the mean free path.

As has been shown in Sec. 167, the mean free path for the molecules in any gas depends upon their diameter, and this diameter is not easy to determine, although a rough value may be estimated by the methods described in Sec. 23. More often the computation is made the other way around: the mean free path is determined experimentally from some phenomenon such as thermal conductivity, and the diameter of the molecules then computed from the mean free path. This method for obtaining molecular diameters has been discussed earlier, in Sec. 24.

169. Vacuum Pumps. The evacuation of a vessel by means of a vacuum pump is not a process of pulling the gas molecules out of the vessel, but rather one of trapping them as they come out in consequence of their own natural motions. When the gas pressure is still large, the gas comes out as a stream, the portions nearer the outlet being driven out by the molecular bombardments of those behind. But when the pressure has been lowered until the mean free path is comparable to the dimensions of the outlet and of the tube leading to it, they come out one by one, whenever their random motions carry them towards and through the outlet. Consider, for example, the pump diagrammed in Fig. 123, which is one form of a type of pump much used to produce high vacua. An ordinary vacuum pump, connected to the side outlet, *S*, reduces the pressure in the whole system to a few hundredths of a millimeter. The mercury at *B* boils easily at this pressure, and a stream of mercury vapor flows up the tube, *T*, is deflected downward by the mushroom cap, *C*, and is finally condensed along the walls of the outside pump tube, *P*, which are kept cool by the water jacket, *J*. All gas molecules in the pump tube below the cap are carried along downward by this stream of mercury vapor and then carried away by the fore pump. Gas molecules from the vessel which is sealed to the top of the pump tube are carried off in similar fashion: whenever their random motions carry them through the space between the cap and the wall of the pump tube, they are caught up by the vapor stream and carried by it to the fore pump. The only limit to the vacuum possible is set by the pressure of the mercury vapor that escapes condensation on the pump-tube walls. In recent years various organic oils, of high stability and very low vapor pressure, have often been substituted for mercury in such pumps.

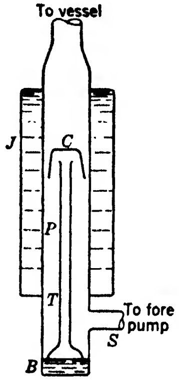


FIG. 123. DIFFUSION PUMP.

170. Molecular Motions in Liquids; Osmosis. There is every evidence for believing that the average kinetic energy of the molecules of a liquid is the same as that of the molecules of a gas at the same temperature. As a matter of fact, Brownian movements were originally observed for tiny colloidal particles in a liquid "suspension," and Perrin made his measurements, which are referred to above, on such liquid suspensions of colloidal particles. It was not until comparatively recent times that quantitative observations were made on the Brownian movements in a gas. Numerous other phenomena, such as diffusion and osmosis, provide further evidence as to the applicability of Maxwell's law to liquids.

Osmosis is a particularly good example. Consider a strong sugar solution separated from pure water by a semipermeable membrane (one through which water molecules will pass freely, but through which, for some reason, the sugar molecules may not pass). At the start the pressure on both sides of the membrane is the same. Water molecules will go through the membrane in both directions; but since there are more water molecules per unit volume on the pure water side of the membrane, more molecules will go through from that side than from the side of the solution. This will continue unless sufficient pressure is exerted upon the solution to compress it to the point where the number of water molecules per unit volume is the same on both sides of the membrane. *The water pressures will then balance, and the osmotic pressure (which is measured by the external pressure applied to the solution) is that pressure which is due to the bumping of the sugar molecules against the membrane.* Unless the concentration is too high, this pressure should be approximately that given by the gas formula

$$P = \tfrac{1}{3}N_s m_s U_s^2 = N_s kT \tag{11·20}$$

and the experiments verify this. (N_s is the number of sugar molecules per unit volume of solution, while m_s and U_s are respectively the mass and the root-mean-square speed of the sugar molecules.)

The evidence for molecular motions in solids is more difficult to secure, and the best is that given in Chapter XIII. For the most part such motions are confined to oscillations of the atoms about fixed positions, although in the surface layers, and at elevated temperatures throughout the solid, some migration of atoms, or diffusion, does take place. The amount of this diffusion is so minute that it is only since the discovery of induced radioactivity that means for its systematic study have been available.

PROBLEMS

1. What volume will be occupied by a gram of CO_2 at 100° C and 2 atmospheres pressure?

2. Compute the density of nitrogen gas (N_2) at 300° C and a pressure measured by 96 cm of mercury.

3. A certain gas contains 6 parts of carbon to 1 part of hydrogen by weight. Its density under standard conditions is 0.00125 gram/cm^3. Find its chemical formula.

4. The highest commercial vacuum is measured by about a millionth of a millimeter of mercury. How many air molecules per cubic millimeter are left?

5. A machine gun discharges one hundred 50-gram bullets per minute with a speed of 400 meters per sec. Compute the average force of recoil.

6. Repeat the kinetic-theory proof for the pressure of a gas, using numerical values, as follows: $m = 27 \times 10^{-24}$ gram, $U = 400$ meters/sec, $N = 21 \times 10^{20}$

molecules, in a box 30 by 40 by 50 mm. Reduce the pressure which you find to atmospheres.

7. Compute the average speed of a molecule of oxygen at 600° C.

8. Compute the average Brownian-movement speed of a water droplet 1/10,000 cm in diameter, floating in air at 25° C.

9. Count the "hits" in each ring of the target shown in Fig. 121(*a*), and plot a distribution curve for them.

10. Assuming the mean free path in air at atmospheric pressure and room temperature to be about 10^{-5} cm, compute the mean free path in air at the pressure given in problem 4, and the same temperature.

11. Compute the mean free path for the tennis-ball gas described in Sec. 160.

12. Prove equation (11·18) by repeating the procedure followed in Sec. 167 with letter values.

13. A sugar solution contains 100 grams of sugar per 100 cm^3 of solution. Compute the osmotic pressure of this solution, at 27° C. *Note*: First compute the number of sugar molecules per cubic centimeter. Then assume that they behave like a perfect gas. Molecular weight of sugar = 342.

REFERENCES

For Additional Reading

SIR WILLIAM BRAGG, *Concerning the Nature of Things*, Chapters I and II.
MAX BORN, *The Restless Universe*, Chapter I.
HENRY CREW, *Rise of Modern Physics*, Chapter IX.
PHYSICS STAFF, UNIVERSITY OF PITTSBURGH, *Atomic Physics*, Chapter II.
R. A. MILLIKAN, *Electrons, Protons, Photons, Neutrons, and Cosmic Rays*, Chapter VII.
J. A. ELDRIDGE, *The Physical Basis of Things*, Chapters III and V, and parts of Chapter IV.
G. F. HULL, *Modern Physics*, Chapter I.
G. E. M. JAUNCEY, *Modern Physics*, Chapter VII.

For Advanced Reading

J. PERRIN, *Brownian Movements and Molecular Reality.*
MAX BORN, *Atomic Physics*, Chapter I.
L. B. LOEB, *Atomic Structure*, Chapter XIX.
L. B. LOEB, *Kinetic Theory of Gases.*

CHAPTER XII

MOLECULAR ENERGY

171. Internal Energy. Any attempts to picture, or to make a "model," of the atomic or molecular structure of material substance (gas, liquid, or solid) must take into account the phenomena of heat, since, as has been known for a long time, the heat energy of a substance consists of the kinetic and potential energies of its atomic and molecular components. Since the term "heat energy" is sometimes used with several meanings, the more exact term "internal energy" will be used hereafter for whatever potential and kinetic energy the molecules and atoms of a gas possess as a result of their positions with respect to each other, and of their random motions (as distinguished from any motion they may have in common, as for example the motion of a stream of gas, or of a bullet). The internal energy of our "tennis-ball" gas would be all kinetic energy of translation, which, as is shown in Sec. 161, equation (11·12), is

$$w_i = 2.070 \times 10^{-16}\, T \text{ erg for each molecule} \qquad (12\cdot 1)$$

For a mole (gram molecule) which contains 6.023×10^{23} molecules, it is

$$W_i = 6.023 \times 10^{23}(2.070 \times 10^{-16}\, T) \text{ ergs}$$
$$= 12.47 \times 10^{7}\, T \text{ ergs} \qquad (12\cdot 2)$$

172. Molar Heat. We cannot measure the total internal energy of a real gas, for obvious reasons. But we can measure the *increase* in internal energy for a 1-degree rise in temperature; and this quantity, when computed for one mole, is called the molecular heat, or *molar heat.* From equation (12·2) it is evident that, for the ideal gas, the molar heat is

$$12.47 \times 10^{7} \text{ ergs} \quad \text{or} \quad 2.98 \text{ calories}^{1} \qquad (12\cdot 3)$$

For any real gas the molar heat may be obtained by confining the gram molecule of gas at constant volume and measuring the heat quan-

---

[1] The proper dimensions for molar heat are "calories per degree centigrade per mole," but this is awkward to use, and no great ambiguity is involved by the use of the shorter unit, "calories."

tity required to produce the 1-degree rise in temperature. It is equivalent to the specific heat at constant volume (C_v), multiplied by the molecular weight.

Let us try this for argon, whose molecular weight is 40, and for which C_v is 0.074 calorie/gram-degree.

$$\text{Molar heat} = 0.074 \times 40 = 2.96 \text{ calories} \qquad (12\cdot4)$$

The agreement is excellent. Evidently in this respect, also, the tennis-ball model is a good picture of the argon molecule. And the same excellent agreement is found for other *monatomic* gases such as helium, neon, and mercury vapor.

Oxygen gives quite a different value; for oxygen, $C_v = 0.155$, molecular weight = 32, and hence

$$\text{Molar heat} = 0.155 \times 32 = 4.96 \text{ calories} \qquad (12\cdot5)$$

Many other *diatomic* gases, such as nitrogen and hydrochloric acid, have very close to this value of the molar heat, which is bigger in the ratio of 5:3 than the value of 2.98 calories predicted by the simple tennis-ball model.

Evidently the tennis-ball model is too simple and will need to be modified. Indeed, the modification is fairly obvious. Each molecule must be imagined to be a little rigid dumb-bell (the two atomic nuclei rigidly bound together at a fixed distance apart) which is not only flying about with the random motions already discussed, but is also spinning end over end as well. If the temperature of such a gas is to be raised one degree, not only must the total *translational* kinetic energy be increased (by about 3 calories, as already seen), but there must likewise be an increase of the total kinetic energy of this *rotational* motion. And it may be that this will account for the increase in molar heat (about 2 calories) observed for oxygen and similar gases.

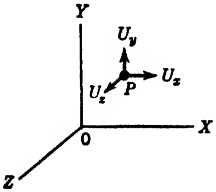


FIG. 124. DEGREES OF FREEDOM. OX, OY, and OZ are three mutually perpendicular axes.

173. Degrees of Freedom. But why, for these diatomic molecules, should the increase in rotational energy be just two-thirds as much as the increase in translation energy? Let us consider the translation case again. Any particle, P (Fig. 124), has three independent *degrees of freedom:* That is, its motion parallel to any one of the axes, say OX, is quite *independent* of its motion parallel to either one of the other axes. And, since the motions of the molecules of a gas are entirely random motions, one-third of the kinetic energy of these molecular motions can be attributed to the components of their motions in each of three mutually perpendicular directions. (This is the significance, in the simplified proof of Sec. 160, of assuming that one-third of the molecules are traveling in each of three such directions. It may be proved more generally, as is done in more advanced treatises.) In this case, then, there is **equipartition of the energy** among the three degrees of freedom. It follows that the molar heat of translation is likewise equally divided among the degrees of freedom, giving

$$\tfrac{1}{3} \times 2.98 = 0.993 \text{ calorie for each degree of freedom} \qquad (12 \cdot 6)$$

(This is, by a very curious but useful coincidence, very close to 1 calorie per degree of freedom.)

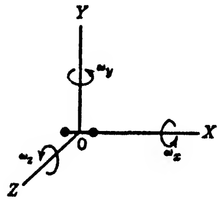


FIG. 125. ROTATIONAL DEGREES OF FREEDOM. ω_x, ω_y, and ω_z are angular velocities about the x, y, and z axes, respectively.

If, now, a molecule can spin as well as fly through space, it has additional degrees of freedom due to rotation. In general, there will be three such degrees of freedom, corresponding to rotations about each of the three axes (see Fig. 125). If we *postulate* that the principle of *equipartition of energy* extends to these degrees of freedom also—that, as a result of collisions between molecules, each of these *rotational* degrees of freedom acquires as much kinetic energy as is possessed by each of the *translational* degrees of freedom—then the total kinetic energy of our gas will be doubled, and the molar heat will also be doubled. Since there are now *six* degrees of freedom, with 0.993 calorie for each one,

$$\text{Molar heat} = 6 \times 0.993 = 5.96 \text{ calories} \qquad (12 \cdot 7)$$

This still does not explain the molar heat of oxygen. It is now too *big*. But if, *for some reason*, a dumb-bell molecule *cannot* rotate about the line connecting the two atomic nuclei (the axis OX, Fig. 125) then it will have only two rotational degrees of freedom, or only five in all, and

$$\text{Molar heat} = 5 \times 0.993 = 4.97 \text{ calories} \qquad (12 \cdot 8)$$

which agrees as nicely as could be asked with the experimental value for oxygen and similar diatomic gases.

But what is the reason that rotation is impossible about this one axis? Why, indeed, cannot a spherical molecule, such as an argon molecule, also rotate and have kinetic energy of rotation? In the past, some very ingenious explanations have been given in answer to these questions, but the simplest and most satisfactory is provided by the quantum theory. A dumb-bell molecule cannot rotate[2] about the line joining the atomic nuclei because its *moment of inertia* about that axis is *too small!* This apparently ridiculous statement is given a rational explanation by the quantum theory.

174. The Quantum Rule for Spinning Bodies. Whenever the angular momentum of a spinning body changes, it changes by a jump of just one quantum of angular momentum,[3] or an amount equal to $h/2\pi$.

This is the rule which, as we have seen in Chapter VIII, applies to every change in motion of the electron in an atom, and it appears also to apply to the motion of any spinning body whatsoever. These jumps could never be detected for a spinning top, because its moment of inertia is proportionally very large. For a molecule the situation is quite different. An oxygen molecule, turning end over end (about an axis such as OZ, Fig. 125), has a moment of inertia about this axis of only 19×10^{-40} gm cm^2, and must spin at the rate of 88 billion r.p.m. to have even one quantum of angular momentum. At room temperature the average oxygen molecule is spinning with about 12 times this speed (i.e., with about 12 quanta of angular momentum). This means that, at ordinary temperatures, the average kinetic energy per degree of freedom is sufficient to excite the molecule into its rotational states of motion (for the *two* degrees of freedom represented by the OY and OZ axes, Fig. 125), so that these *two* degrees of freedom may take up their shares of the total energy. But about the third axis (the axis OX, running through the two atomic nuclei) the moment of inertia is practically zero. Indeed, the moment of inertia about this axis is principally that due to the electrons themselves; it is so small that the energy required to give the molecule even one quantum of angular momentum about this axis is far more than may be acquired from collisions between molecules at ordinary temperatures. This explains why only two rotational degrees of freedom may be excited, and why consequently the molecular heat is but 5 calories.

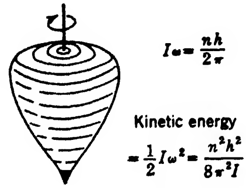


FIG. 126. QUANTUM CONDITIONS FOR A SPINNING BODY.

---

[2] More exactly speaking, its rotation about this axis cannot *change*. It may have permanent intrinsic rotations due to various causes, but these do not affect the matters under discussion here.

[3] Here and in the following sections we are going to use the simpler quantum theory, and neglect the wave-mechanics refinements to that theory. The results, except for minor details, are essentially the same.

The hydrogen molecule is perhaps the most interesting example. Because its atoms are lighter and closer together, its end-over-end moment of inertia is only one-fortieth of that of the oxygen molecule, so that at room temperature its rotation on the average represents only two quanta of angular motion. At room temperature its molar heat is about 5 calories, in agreement with that for oxygen. At lower temperatures, however, it begins to diminish, until at the temperature of liquid air it is only about 3 calories. At liquid-air temperature, the hydrogen molecule cannot be excited into rotation, and behaves like a monatomic (tennis-ball) gas.

Hydrogen gas exhibits other properties which are quite as interesting. Experiments and theory both show that there are actually two kinds of hydrogen molecules, which are designated as para-hydrogen and ortho-hydrogen. Ordinary hydrogen gas contains both kinds, in the ratio of 1 to 3 respectively. At very low temperatures the molar heat for both kinds is 3 calories, as stated above; and both kinds have a molar heat of 5 calories at sufficiently elevated temperatures. But the change from 3 calories to 5 calories at intermediate temperatures follows quite a different law for para-hydrogen from the one it follows for ortho-hydrogen. The two kinds of molecules may also be differentiated by distinctive differences in their band spectra (see Sec. 177).

The difference between these two kinds of hydrogen molecules may be described in terms of the orientations in them of the hydrogen nuclei: The hydrogen nucleus possesses a "spin" analogous to the spin of the electron; in ortho-hydrogen these spins for both atoms are in the same directions about parallel axes, whereas in para-hydrogen the axes are parallel, but the spins are in opposite directions.

The relative proportions of para-hydrogen and ortho-hydrogen in hydrogen gas remain the same at ordinary temperatures. At very low temperatures, however, there is a slight tendency for ortho-hydrogen to change over to the para- form, since this form has the lower energy. This tendency is accelerated by the presence of charcoal (which acts like a chemical catalyst), and by this means nearly pure para-hydrogen has been prepared and its properties studied.

Other molecules consisting of two like atoms show similar ortho- and para- forms.

175. Vibrating Molecules. Whenever a diatomic gas is found to have a molar heat higher than 5 calories, its molecules must have a motion of vibration, in addition to their other random motions. Vibrations of such molecules consist of motions of the atoms towards and away from each other along their line of centers, as indicated in Fig. 127, with a frequency determined by their masses and by the forces which hold them together to form the molecule; unless the amplitude is large, this motion will be a simple harmonic motion. Since the vibrations take place in but one direction (along the line of centers), this adds but *one* degree of freedom to the total motion of the molecule. But (if all the molecules are vibrating to some extent), it adds two calories to the molar heat: one for the kinetic energy of vibration, and one for the potential energy which is, on the average, equal to the kinetic energy.

FIG. 127.

This makes, for such a gas containing flying, spinning, vibrating molecules, a total molar heat equal to $3 + 2 + 2 = 7$ calories. For molecules composed of heavy atoms, as, for example, those of iodine, the experimental measurements show the molar heat to be close to 7 calories, in accord with this simple theory. This is likewise true for most diatomic gases at high enough temperatures; but at lower temperatures the molar heat may lie somewhere between 5 and 7, indicating that for some reason only part of the molecules are vibrating. We must again turn to the quantum theory to explain why this is so, as well as to explain why, at still lower temperatures, no vibrations at all may take place.

Quantization of momentum is found for vibrations (and, indeed, for all *periodic* motions), just as for the rotations which we have already considered. The rule for the quantization of vibrational momentum is not so simple as the one for angular momentum, and it will be sufficient for our purposes here to state the equation for the energy values which this rule allows:

Because of this quantization of all vibrational motions, simple harmonic vibrations can occur only with such amplitudes that the energy[4] is

$$W_v = n_v h f_v \tag{12·9}$$

Here f_v is the frequency of the vibration, and n_v is a whole number, the *vibrational quantum number.*

When the collisions, on the average, are able to excite the molecules into one or more quanta of vibration, the molar heat will be found to be 7, as explained above. If the frequency of molecular vibration is low, this may occur at moderate temperatures, as it does for iodine vapor. But if this frequency is high, then, unless the temperature is also high, the vibrations may not be excited at all (molar heat equal to 5)

---

[4] The more exact theory of wave mechanics indicates that there is always a minimum amount of vibrational energy equal to $hf/2$ (even at absolute zero), so that the equation should be written: $W_i = (n_v + \frac{1}{2})\, hf_v$. Since we are concerned only with additions of energy, the simpler equation is just as satisfactory here.

or only occasionally, when the collisions involve considerably more than the average energy. This explains molar heats lying between 5 and 7.

Molecules made up of more than two atoms may have still more degrees of freedom, and still higher molar heats. An interesting example of frequent occurrence is the molecule having three atoms in a straight line, as illustrated for carbon dioxide in Fig. 128. Since the atoms are in line, this molecule has but two degrees of freedom for rotation, the same as the diatomic molecule has. However, it has three possible degrees of freedom for vibration, corresponding to the three normal modes of vibration illustrated by (*a*), (*b*), and (*c*). The solid arrows indicate one phase of these vibrations, and the dotted arrows, the opposite phase. The bending type, (*a*), is the lowest in frequency and the one most easily excited. At moderate temperatures carbon dioxide has a molar heat of about 7 calories, indicating that but this one mode is being excited. Both the symmetrical and the antisymmetrical modes, (*b*) and (*c*), of to-and-fro vibration are of considerably higher frequency, and consequently may occur only at considerably higher temperatures. If all modes were fully excited, the molar heat would be 11 calories. When the atoms in a molecule are not in line, three degrees of freedom for rotation are possible of course.

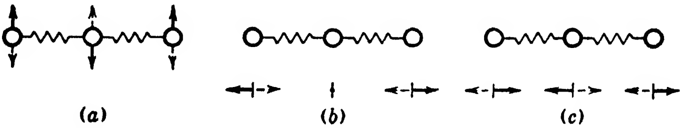


FIG. 128.

176. Infrared Spectra. The quantum rules for the rotations and vibrations of molecules obtain very pretty verification from the spectra emitted (or absorbed) by such molecules. Away down in the far infrared absorption spectrum of a gas is found a series of equally spaced "lines," the spacing being equal to the frequency corresponding to one quantum of angular momentum for the molecule concerned. They represent light which is absorbed when the *rotation* of the molecule is changed by just one quantum. They are very difficult to observe, and not many such "rotation" spectra have been studied.

Farther up on the spectrum, but still far into the infrared, are found groups of absorption lines representing the absorption of light by *vibrating* molecules. Let us consider for simplicity a diatomic molecule, which has but one vibrational frequency. If the rotational motion of the molecule did not change, there would appear a vibrational spectrum consisting of fairly regularly spaced lines, the spacing being approximately equal to the vibration frequency of the molecule. (If the vibration of the molecule were truly simple harmonic motion, these lines should be exactly equally spaced, but the frequency decreases as the amplitude of vibration increases.) However, these vibration lines do not appear. Instead, in the place in the spectrum where each such line should appear, there is found a group of lines such as is shown in Fig. 129, the spacing of the lines (the change in frequency from one line to the next) being determined by the fundamental frequency of *rotation* for the molecule.

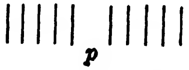


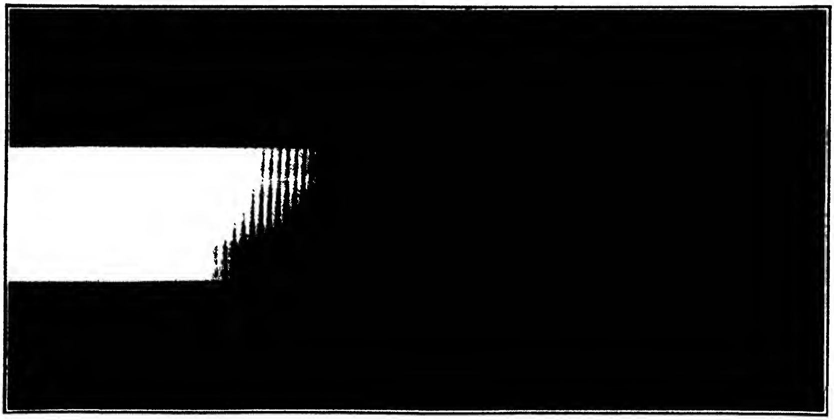
FIG. 129. VIBRATION-ROTATION SPECTRUM FOR A MOLECULE. The point, *p*, represents the change of vibrational energy.

The line representing the change of vibrational energy does not even appear as one of the lines in this group. Instead, a gap appears at this point in the spectrum, in the middle of this group of rotation-vibration lines, and the appearance of this gap shows that the motion of vibration cannot change without the motion of rotation changing at the same time. This rule has already been encountered in atomic spectra, where we saw that the quantum number, *l*, which represents the angular momentum of the electron in its orbit, must change if any change at all takes place. In both that case and this one the angular momentum must increase or decrease by just one quantum.

177. Band Spectra. If a gas or vapor is highly excited, the molecules will be broken apart, or dissociated, and the spectra appearing will be those of the atomic components of the molecules. But if the excitation is not too intense, a very complicated-appearing molecular spectrum will be produced. When photographed with a spectograph of moderate power, these spectra appear to consist of groups of broadened lines, or bands, occurring at fairly regular and closely spaced intervals throughout each group, these lines or bands usually being sharp at one side and spread out diffusely at the other. This may be seen in the spectrum photographs reproduced in Fig. 130. If these spectra are photographed with a very powerful spectrograph (such as a concave grating of 21-foot focal length), each band is found to consist of many very fine lines. These band spectra, as they are called, represent changes of *electronic* energy (similar to those which take place in an atom, only the molecular electrons are involved here), plus changes of *vibrational* energy (which determine the spacing of the bands in each group), plus changes of *rotational* energy (which account for the many fine lines in each band).

Although these spectra appear to be very complex, they agree with theory so nicely that it is often possible to determine from them such interesting and important things as the arrangement of the atoms

(*a*)



(*b*)

FIG. 130. BAND SPECTRA. (*a*) Aluminum oxide (emission spectrum). (*b*) Iodine vapor (absorption spectrum).

within the molecule (including their distances apart), and the various possible frequencies with which the molecule may vibrate. Further, the changes in the electronic energy which appear in these spectra show that, as is suggested above, when two or more atoms combine to form a molecule, a part of the electrons from both atoms (at least the valence electrons) rearrange themselves so as to form an electron "shell" that belongs to the molecule as a whole rather than to the separate atoms. Thus the two electrons belonging to the hydrogen atom form a sort of heliumlike *K* shell for the hydrogen molecule; the four outermost electrons from each oxygen atom form a complete *L* shell of eight electrons for the oxygen molecule, etc. These spectra also furnish data for the determination of the binding energies of molecules, but further discussion of this subject is beyond the scope of this book.

There are many other optical phenomena, including the absorption effects which account for the color of dyes and pigments, which are explainable in terms of molecular structure, but, with the exception of the Raman effect, they are too complicated to consider here.

178. Raman Spectra. In discussing the scattering of light (Secs. 42 and 43) it was mentioned that even molecules can scatter light to some extent, and this may be demonstrated for the molecules of a liquid by means of the apparatus diagrammed in Fig. 131. The liquid (purified, both physically and chemically, with the greatest of care) is sealed into the glass (or quartz) tube shown, and the sealed end is bent and blackened so as to form a perfect black background for the scattered light, which is viewed from the opposite end. (This technique was invented by Wood.) A powerful mercury lamp, provided with suitable light filters, F, illuminates the tube from the side with very intense light of (preferably) a single wavelength. The light scattered by the liquid is photographed by the spectrograph. The scattered light is very faint, and most of it is of the original wavelength. If an exposure of many hours is made, however, it will be found that, in addition to the line on the plate which represents the scattered light of the original wavelength (considerably overexposed), there are other, much fainter lines to either side of this strong one.

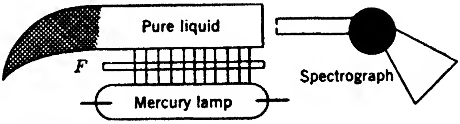


FIG. 131. APPARATUS TO PHOTOGRAPH RAMAN SPECTRA.

These additional lines, which were first observed and explained by Raman, also represent scattered light. Those of longer wavelength represent scattered light whose photons, upon collision with the scattering molecule, have excited that molecule to a higher vibrational state, and thereby have lost that much energy. Since, for a photon, loss of energy means loss of frequency, these waves are longer; and the frequency of vibration of the molecule may be determined by measurement of the displacement of this line from the strong one. The scattered light of higher frequency is similarly explained, except that in this case the molecule *loses* energy in the collision, and the photon *gains* it. (This is crudely comparable to a baseball gaining speed by collision with a bat.) This phenomenon provides still another means for the study of the structure of molecules which, in spite of the difficulties involved (tedious purification processes, and long exposures) has proved very valuable.

179. Interatomic and Intermolecular Forces. The "chemical" forces which hold atoms together to form molecules are for the most part of electrical origin, although they cannot very often be described in an elementary manner. The simplest cases are those where the separate atoms may be considered to be oppositely charged. In HCl, for example, the binding force may be considered to be a simple electrical attraction between the positively charged H ion and the negatively charged Cl ion. Such forces are called *ionic* forces. But this explanation fails to account for the force which holds an extra electron onto the neutral Cl atom to make of it a negative ion. At this point the quantum theory supplies a good explanation. Since neutral Cl has but 7 electrons in its outer (M) shell, there is room for one more electron in this shell. (The remaining 11 M electrons cannot be added unless 2 N electrons are present. See Appendix III.) And an electron may go (if the two atoms are brought close enough to each other) from the K shell of a neutral H atom to this vacant place in the M shell of the Cl atom because, in spite of giving to the Cl ion a negative charge, the net effect is a loss of energy by the hydrogen-chlorine system (HCl molecule). The lost energy is in fact the *binding energy* of the HCl molecule.

The energy form of argument just used is a valuable one for dealing with forces not easily visualized. The most stable state of any physical system is the one in which the total energy is *lowest*, and if possible the system will change to that lowest energy state.

When a molecule consists of two identical atoms, ions obviously cannot be formed, and a different type of force must act. Consider, for example, the hydrogen molecule; in this molecule the two H atoms may be considered to be held together by *sharing* their two electrons, the two electrons belonging to the molecule and not to the separate atoms, as has been mentioned in Sec. 177. Again the energy argument is applicable. The hydrogen molecule sticks together because this arrangement for the two H nuclei and the two electrons has *less energy* than is possessed by the separate neutral atoms. The wave-mechanical theory develops this idea in a more adequate and exact form which is, however, not capable of satisfactory visualization. Because of the mathematical form of this theory such forces as these are called *exchange* forces.

In still other cases the forces are even less direct in character, being frequently much like the forces which hold molecules together to form liquids and solids. Such forces are often called *van der Waals* forces, for reasons that will appear below.

Some of the phenomena which demonstrate such forces between *molecules* will be considered in the remaining sections of this chapter.

180. Van der Waals' Equation. In developing the kinetic theory, in Sec. 160, it was explicitly assumed that molecules have negligible size, that whatever forces they may exert upon one another are negligible except during collisions, and that the time occupied by collisions is negligible in comparison to the time of flights between collisions. As long as the density is kept low enough, these assumptions are quite valid, but allowances must be made for all these factors when the density of the gas is not exceedingly small as compared to that of the liquid or solid phase of the same substance.

The effect of molecular size is to reduce the space in which the molecules may move about between collisions; the collisions are then more frequent, and the pressure within the gas is consequently higher than it would be for an ideal gas at the same temperature. The effective volume is then $(V - b)$, where b represents the reduction of volume due to the finite size of the molecules; and the pressure, P', due to the motions of the molecules, is now found by putting $(V - b)$ into the gas-law equation in place of V, that is, by the equation

$$P'(V - b) = RT \qquad (12\cdot10)$$

If the molecules of a gas attract one another, then the pressure which is actually measured in the gas will be less than P' whenever the density of the gas becomes appreciable. This is easily understood if we imagine an infinitely thin plane diaphragm (shown edgewise by S, Fig. 132) to be placed in the gas, and then compute the force, per unit area, that the gas to one side of S exerts upon the gas to the other side of S. Against each side of S the gas molecules exert a pressure, P', due to their thermal agitations, whose value is given by equation (12·10). At the same time there will be an attraction force acting across this boundary between the gas molecules lying to either side of it.

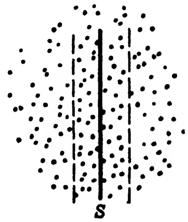


Fig. 132.

These attractions are limited to those molecules which lie close to S, and this is indicated by the dotted lines parallel to S, which represent approximately the limiting range for the intermolecular forces. Since within this range each molecule to the left of S is attracted by each molecule to the right of S, the magnitude of the attraction force, per unit area, is approximately proportional to the product of the concentration of molecules to the left of S and the concentration of molecules to the right, or to the *square* of the density of the gas. Since the volume, V, is inversely proportional to the density, this force may be represented by $-a/V^2$, where a is a constant of proportionality, and the negative sign indicates that the force is one of attraction. The net pressure across this boundary is then

$$P = P' - \frac{a}{V^2} \tag{12·11}$$

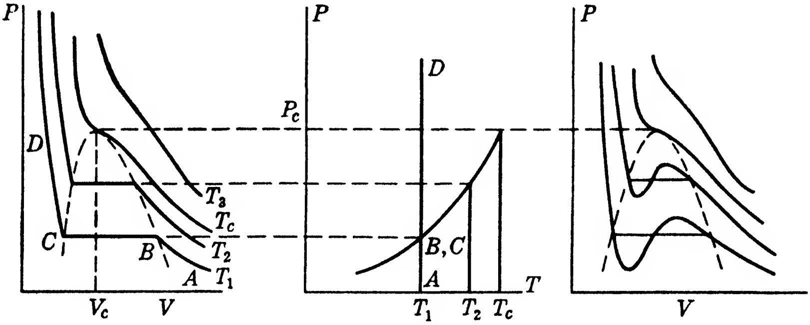
and the gas law becomes

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT \tag{12·12}$$

This theory was first developed by van der Waals, and the equation is known by his name. Although the assumptions are somewhat too simple to make it an exact equation, it fits surprisingly well in many cases, even when the gas becomes a saturated vapor and then condenses into a liquid.

181. Condensable Gases. Figure 133(a) shows a series of isothermal plots of pressure and volume for a condensable gas. The plot for T_1 may be interpreted as follows: At the pressure and volume represented by A the substance is a gas (or vapor). As its volume is decreased, its pressure rises until the conditions represented by B are reached. At this point it is *saturated*, and any further decrease in volume causes condensation, with no rise in pressure. Point C represents complete condensation, and points from C to D represent the liquid phase, the pressure rising very rapidly with very small further decrease in volume. At the higher temperature, T_2, the same changes are noted, with, however, a shorter range of condensation. The start and finish of condensation for other temperatures are indicated by the dash line passing through B and C. The temperature T_c is the boundary value. Above this value condensation never occurs. Thus at T_3 the substance has all the characteristics of a gas throughout its range of values. The temperature represented by T_c is called the **critical temperature,** and the corresponding pressure and volume, P_c and V_c, are called the **critical pressure** and the **critical volume.**

Another representation of these conditions for a gas may be made by a plot of pressure and temperature such as is shown by Fig. 133(b). The corresponding isothermals are indicated here as vertical lines, and corresponding pressure values are indicated by the horizontal dotted lines connecting (*a*) with (*b*). Indeed, Figs. 133(*a*) and (*b*) may be considered to be two different views of a three-dimensional diagram or model, such that, in (*b*), the viewpoint is directly in line with the line *B–C*, and all of the curve *ABCD* in (*a*) is represented by the vertical line *ABCD* in (*b*). The curve plotted in (*b*) is called the saturation-vapor-pressure curve, since, for each temperature, it represents the pressure at which condensation (or vaporization) occurs. At its upper end it terminates at the critical point. At its lower end it terminates at a point called the triple point, which represents the temperature and pressure at which all three phases, solid, liquid, and vapor, can exist together in equilibrium. Further details concerning this and many other related phenomena will be omitted here, since they may be found in so many other books and are not directly related to the other matters here being considered.



(*a*) (*b*) (*c*)

FIG. 133.

Figure 133(*c*) shows plots of van der Waals' equation for the temperatures represented in (*a*). The curves for T_c and T_3 appear very similar in both figures, and can indeed be made to agree quite well quantitatively by choosing *b* equal to $\frac{1}{3}V_c$, and *a* so as to make the curves match at the critical point. At lower temperatures the van der Waals plot has a curious dip and hump in place of the horizontal line which represents condensation in the experimental plot. Otherwise it resembles the corresponding experimental plot, and is in fair agreement with it. If a horizontal line is drawn as shown in Fig. 133(*c*), so as to make the area below it, in the dip, equal to the area above it, in the hump, this line will correspond quite well to the condensation line. The parts of the van der Waals plot which make the dip and hump then may be considered to represent conditions of instability which are averaged out by the horizontal line.

This is much more than a subterfuge to make theory fit experiment. If exceeding care is taken, it is possible to reduce the pressure upon a very pure liquid (in a closed vessel) to values considerably below that at which vaporization normally occurs, without any vapor being formed: this process is represented by the left-hand side of the dip of the van der Waals curve. Likewise it is possible, with suitable precautions, to supersaturate a vapor, or to carry it along an isothermal to values represented by the van der Waals extension of *AB* above the condensation line.

TABLE II

Critical-Point Data

Substance	Absolute Temperature, °C	Pressure, atmospheres	Volume, cc/mole
Hydrogen (H_2)	33.3	12.8	65
Helium (He)	5.3	2.26	58
Nitrogen (N_2)	126.1	33.5	90
Oxygen (O_2)	154.4	49.7	74
Neon (Ne)	44.5	26.9	41
Carbon dioxide (CO_2)	304.3	73.0	96
Ammonia (NH_3)	405.6	111.5	72
Ethyl ether ($C_4H_{10}O$)	377.0	35.5	283

182. Critical-Point Data. Table II gives values of critical data for a number of substances. Ethyl ether is liquid at ordinary room temperatures, and ammonia (NH_3), which is normally a gas, can exist as a liquid at room temperature if under high enough pressure. For these substances the critical temperature is higher than room temperature. With the exception of carbon dioxide, which will be considered in detail later, the other substances listed in Table II cannot be condensed to the liquid phase at room temperature, no matter how high the pressure exerted upon them. For these substances the critical temperatures are lower than room temperature; and, as Fig. 133 clearly shows, no substance can exist as a liquid, no matter how great the pressure exerted, unless its temperature is lower than its critical temperature. Then it may be condensed by raising its pressure to the value given for that temperature by the saturation-vapor-pressure curve [Fig. 133(*b*)] for the substance. If the temperature is lowered sufficiently below the critical temperature, the substance may usually be obtained as a liquid at atmospheric pressure. This is the state in which liquid air and similar low-temperature liquids are obtained. For carbon dioxide, however, the triple point pressure is above atmospheric pressure, so that carbon dioxide, at atmospheric pressure, is a solid, the familiar "dry ice" of commerce.

Carbon dioxide is of interest because its critical temperature comes so close to ordinary room temperature, and because it was the substance with which Andrews demonstrated the existence of the critical point. In the steel cylinders in which it is sold commercially carbon dioxide is usually liquid in the lower part, with saturated vapor above. On a hot summer day, however, the temperature may exceed the critical temperature, and then the boundary between liquid and vapor disappears, the cylinder then being filled with a homogeneous substance whose properties are essentially those of a gas.

The same transition through the critical point is sometimes illustrated with ethyl ether sealed into a stout glass tube. If the initial proportions of liquid and vapor are correct, the boundary surface or interface between the two phases will remain practically fixed in position as the temperature is raised, but will disappear completely as the critical temperature is passed. Although the volume of the liquid phase appears unchanged, evaporation across the interface is continually going on, continually reducing the density of the liquid and correspondingly increasing the density of the vapor as the temperature rises, until at the critical temperature these two densities become equal and the two phases become one. Upon subsequent cooling through the critical temperature the liquid phase is re-formed as droplets throughout the tube; violent precipitation then takes place until the interface is again formed. It is interesting to speculate upon what may have taken place when the surface of the earth cooled through the critical temperature for water vapor and the first rainstorms occurred, the raindrops having a temperature of 374° C, the critical temperature for water!

183. Joule-Thomson Porous-Plug Experiment. The most direct evidence for molecular forces between the molecules of a gas is given by the porous-plug experiment of Joule and Thomson. In this experiment the gas is allowed to diffuse through the pores of a "plug" of unglazed porcelain, the gas density at the one side of the plug being quite high, on the other side, quite low. Section 162 describes the process by which the molecules pass through the pores of the plug. The pressures on both sides of the plug are kept very steady, so that no compressions or expansions of the gas may occur. Under these conditions, if the molecules have no attractions for one another, they will have the same average speed after passing through the plug as they had before doing so.

If, however, they attract one another, they will lose speed upon going through the plug, since in so doing they are pulling away from their closely crowded neighbors in the high-density region. Such a loss of speed would be made evident by a lowering of the temperature, that is, by a difference between the readings of thermometers placed to either side of the plug. The test is a very sensitive one. For most gases the effect is a cooling; and the lower the initial temperature, on the high-density side, the greater proportionally is the cooling. By making this process regenerative—that is, by utilizing the first cooling to lower the temperature of the incoming gas—an efficient process was developed for the production of low temperatures. This is the process ordinarily employed for the production of liquid air and other liquid gases.

Hydrogen is interesting in that, when the initial temperature is room temperature, the porous-plug experiment shows heating rather than cooling of the gas. After hydrogen has been cooled by liquid air, however, it cools as do other gases, and may then be liquefied by the same process.

184. Evaporation. As is evident from the discussion of Sec. 181, evaporation is a process in which the potential energies of the molecules are increased as they pull away from the liquid surface. *The heat of vaporization is a measure of this potential energy per gram of molecules.*[5] When evaporation takes place without the supply of any external energy (heat), this gain of potential energy by the molecules occurs at the expense of their kinetic energies, and the process is therefore a cooling one, as is well known. Both the vapor and the liquid are cooled. The vapor is cooled, since the escaping molecules all lose speed as they pull away from the liquid surface. The liquid is cooled, since those of the molecules in the liquid which have the greater speeds are more able to escape, and do escape in greater numbers. This lowers the average kinetic energy of those remaining, and hence lowers the temperature of the liquid.

In a closed container, an equilibrium condition is reached in which evaporation appears to have ceased. No further change takes place in the liquid volume, and no cooling occurs. Actually, evaporation is proceeding just as rapidly as ever, the rate of evaporation being determined solely by the temperature. At the same time, however, condensation is likewise occurring, and the equilibrium is a dynamic

---

[5] The heat of vaporization includes also the work done against external pressure in expanding from the volume occupied by one gram of the liquid to the volume of the saturated vapor at the same temperature. For one gram of water at its normal boiling point (100° C) the increase in internal energy is 498 calories, while the external work is 40.5 calories.

state in which the number of molecules escaping the surface by evaporation just equals the number returning to the liquid (condensing).

PROBLEMS

1. The molecular heat of a gas is found to be 8 calories. Discuss its possible modes of motion.

2. Compute the molecular heat of water and of water vapor. ($C_v = 0.36$ for water vapor.)

3. Compute the number of quanta of angular momentum possessed by a top that is spinning 300 r.p.m. Take the top to be a circular disk 21 mm in diameter and weighing 5 grams.

4. Compute the moment of inertia of an iodine molecule about the axis perpendicular to the line joining the two iodine atoms, if the atoms are 2.66A apart.

5. At what temperature would the iodine molecule have just 15 quanta of angular momentum, on the average?

6. A bead of mass equal to 0.80 gram is supported by a stiff spring, so that it vibrates with a frequency of 20 per second. Neglecting the mass of the spring, compute: (*a*) the force when it is displaced 0.20 mm; (*b*) the energy when it is vibrating with this amplitude; (*c*) the number of vibrational quanta corresponding to this motion.

7. When light of 3330A wavelengths shines through a liquid, a line at 3680A appears in the scattered light. Explain the origin of this line, and compute the frequency which accounts for it.

8. A molecule may vibrate with a frequency of 3×10^{13} sec^{-1}. (*a*) Compute the wavelength corresponding to this frequency. Will this wavelength appear in the spectrum? Explain. (*b*) If light of 4000A wavelength is scattered by such molecules, what will be the wavelength for the nearest Raman line?

9. Assume that the molecules of a certain gas are of such a size that b in the van der Waals equation equals 45 cm^3 per mol, but that they do not attract each other. Compute the pressure required to compress this gas to 1/50 of its volume at atmospheric pressure, without change of temperature.

REFERENCES

For Additional Reading

Sir William Bragg, *Concerning the Nature of Things*, Chapter III.
M. Mott-Smith, *Heat and Its Workings*.
Edwin Edser, *Heat*.
Physics Staff, University of Pittsburgh, *Atomic Physics*, Chapter XI.
J. A. Eldridge, *The Physical Basis of Things*, Chapter XXIII.
J. M. Cork, *Heat*, Chapters VII, VIII.

For Advanced Reading

Ruark and Urey, *Atoms, Molecules, and Quanta*, Chapter XII.
R. W. Wood, *Physical Optics*, Third Edition, Chapter XIV.
L. B. Loeb, *Atomic Structure*, Chapters XVI–XVIII.
L. B. Loeb, *Kinetic Theory of Gases*.
M. Born, *Atomic Physics*, Chapters 7 and 8.

CHAPTER XIII

CRYSTALS

185. Liquids and Solids. In the gaseous state the molecule is a distinct entity having independent properties which are quite unaffected by the presence of its nearest neighbors. In liquids, and in some solids, the molecule still has definite identity, although the close crowding of its neighbors and the strong forces which they exert upon one another seriously modify most of its physical properties, as well as some of its properties which might almost be called chemical. The intensity of intermolecular forces in solids does not need comment, but they may not be so apparent in a liquid such as water. Since water has no rigidity, it appears very "weak." But if one tries to pull it apart, under conditions such that there is no free surface at which a break may start, an astonishing strength is discovered. Experiments have shown that, if water is painstakingly purified, with especial care for the removal of all dissolved gas, and is then completely inclosed, it will withstand more than 2000 $\mathrm{lb/in.}^2$ of *tension!* Theoretical considerations such as are discussed in Sec. 180 indicate that its tensile strength may be many times greater.

186. Crystals. In liquids and solids, then, the "physical" forces which hold the molecules together are approaching comparability with the "chemical" forces binding the molecule together. In a crystalline solid they have become one and the same, and the identity of the molecule is completely lost. The crystal is the chemical as well as the physical unit, and the forces holding it together are principally electrical forces. Consider, for example, a crystal of rock salt, for which the arrangement of atoms is shown in Fig. 82 and described in Sec. 97. As far as crystal salt is concerned, the chemical formula, NaCl, means merely that there are equal numbers of Na and Cl atoms in the crystal. In the interior of the crystal each Na atom is surrounded on all six sides by Cl atoms, while each Cl atom is similarly surrounded by Na atoms; and the forces binding each Na atom to each Cl atom are the same in every direction.

NaCl is an example of an *ionic crystal.* In the formation of the crystal each Cl atom robs a Na atom of its valence electron, using this electron to build around itself a closed *M* shell of eight electrons;

this gives it an excess negative charge. The Na atoms are left with an excess positive charge and a closed *L* shell of eight electrons. Each atom is then spherically symmetrical, and the NaCl crystal is the most compact arrangement of spherical atoms that is possible, when adjacent atoms have opposite charges. Another example of an ionic crystal is $CaCO_3$ (Iceland spar). In this crystal the Ca atoms are positively charged, and the CO_3 radical groups are negatively charged. These positive and negative ions alternate in the crystal just as do the Na and Cl ions in the NaCl crystal; but since the CO_3 group is not spherical, the crystal is not cubical but has the rhomboid form which is shown in Fig. 35 and described in Sec. 50.

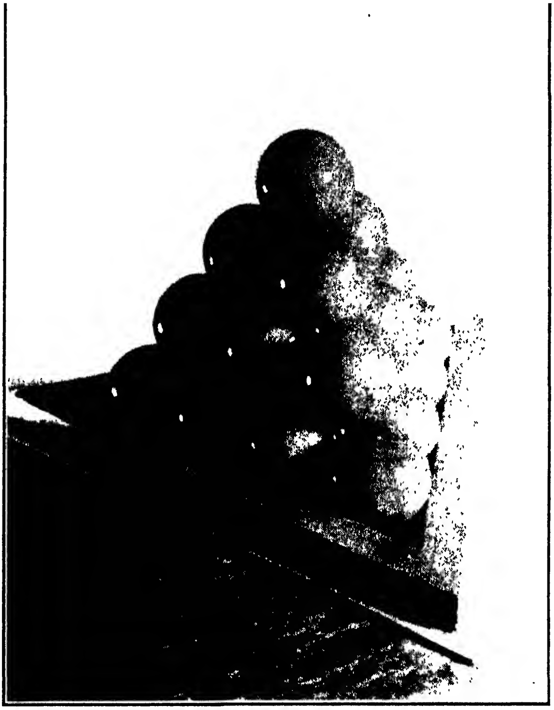
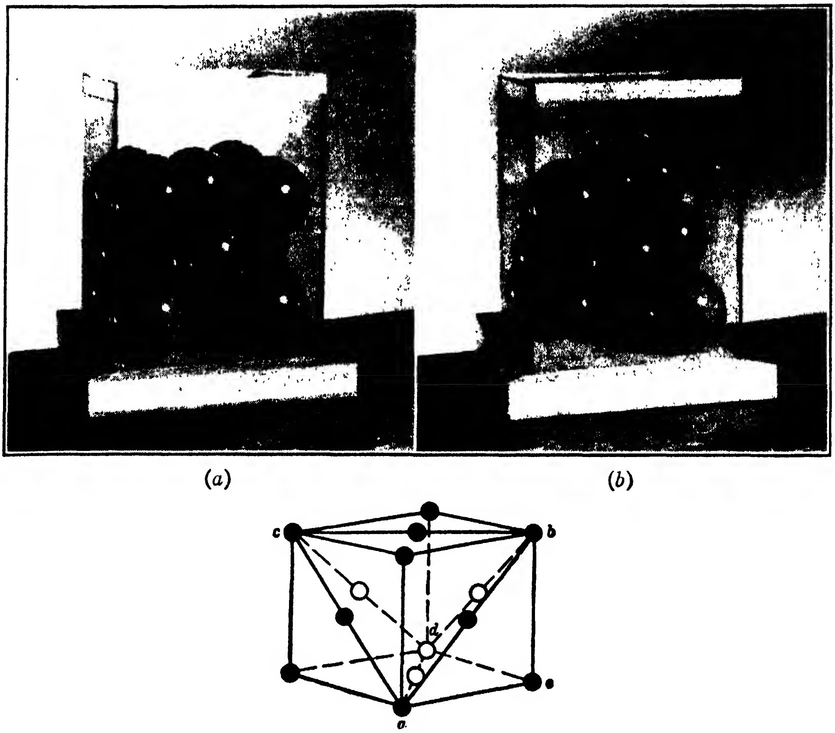


FIG. 134.

If, on the contrary, all the atoms are spherically symmetrical and are not charged, a closer packing can be made, of the kind that is represented by the familiar shot-pile arrangement shown in Fig. 134. The forces holding the atoms together are still electrical, but are accounted for as interactions between electrons and nuclei, rather than directly between atoms. Such a pile is really a regular tetrahedron, a figure



(a) (b)

(c)

FIG. 135. FACE-CENTERED CUBIC ARRANGEMENT. (b) is the same arrangement as (a), with "atoms" removed from alternate corners to show the tetrahedral structure corresponding to Fig. 134. In (c), which shows only the centers of the atoms, the diagonals drawn across the faces of the cube form the edges of the tetrahedron.

having four faces, each of them an equilateral triangle. Out of the inside of such a pile could be taken a *cubical* arrangement of atoms, having atoms at each corner of the cube, and one atom in the center of each of its faces. This *face-centered cubic* pattern is shown in Fig. 135(*a*); and the equivalence to the shot-pile pattern is shown by Fig. 135(*b*), in which the tetrahedron is revealed by removing four of the balls representing the atoms from the four alternate corners of the cube. A large number of metallic crystals are of this form. Aluminum, silver, gold, and platinum are good examples. Such metals are usually soft, and ductile rather than brittle. The reason is evident when one observes how readily the shot-pile arrangement re-forms in a box of marbles.

Magnesium, zinc, cadmium, and mercury form crystals in which the arrangement of atoms is very near to the shot-pile arrangement, and just as compact. The difference may be explained with the aid of Fig.

136, which shows one layer of close-packed spheres, and a portion of a second layer. There are now two possible ways of adding the third layer, and these two ways correspond respectively to the two possible close-packed crystal structures. If the spheres of the third layer are placed directly above those in the first layer (the ones which are shaded in the diagram), the arrangement known as *hexagonal close-packed* is produced. Each layer in the pile then repeats the arrangement of the next layer but one beneath it. This is the arrangement of atoms to be found in zinc and in the other metals named above, all of which tend to crystallize into hexagonal prisms. The reason for this crystalline form is indicated by the little holes or channels, *a*, *a′*, etc., which run straight through all the layers of atoms. The crystal prisms form with their axes in the direction of these channels, and rows of such channels are exposed on the crystal faces. The heavy dashed lines in Fig. 136 indicate where two such crystal faces might have been formed.

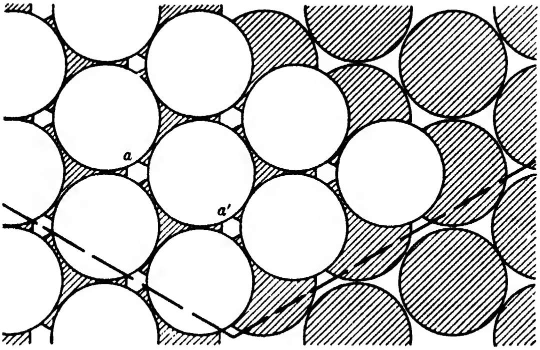


FIG. 136.

The alternate manner of piling the spheres is to place the third layer above the points, *a*, *a′*, etc., so as to close up these channels. This forms the shot-pile arrangement of Fig. 134; the corresponding crystal structure is the face-centered cubic crystal, as has been explained earlier.

Figure 137 shows still a third important class of cyrstal, called the *body-centered cubic* crystal; (*a*) is the usual representation of this arrangement, and (*b*) indicates how the whole crystal may be considered a double system of interlocking cubes. Because of this interlocking action, such crystals are very strong, and all the toughest metals—iron, tungsten, chromium, etc.—have this crystal structure. This is not a close-packed arrangement, and the atoms which form this type of crystal are *not* spherically symmetrical, but exhibit definite poles of attraction for one another.

Definite poles of attraction are even better shown by the carbon atom, which has four poles of attraction, equally spaced over its surface (a tetrahedral symmetry). It forms two kinds of crystals, both of them a type of hexagonal structure. But one of them (diamond) is the hardest crystal known, while the other (graphite) is so soft that it is used as a lubricant. This structure, as well as others which may not be mentioned here, is best seen in crystal models.

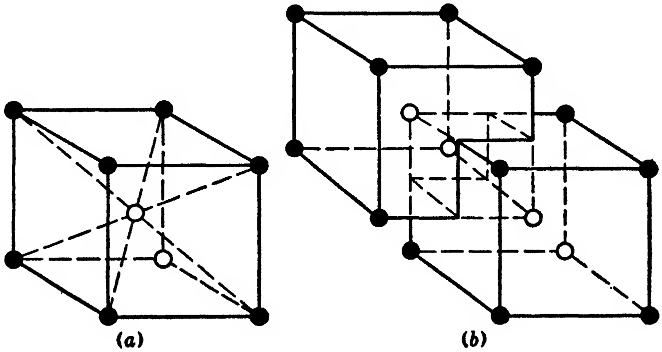


FIG. 137. BODY-CENTERED CUBIC ARRANGEMENT. (*a*) Usual representation. (*b*) Interlocking cubes. The center atom of (*a*) may be considered as the corner of a second cube, interlocking with the first one.

187. Atomic Heats. What can be learned about the motions of the atoms within a crystal? We have found that much may be learned concerning molecules in gases by study of their specific heats. Likewise, specific heats may be able to supply us with an answer to our question concerning the motions of atoms in crystals. Here we shall find it convenient to use the *atomic heat*, or the specific heat for a *gram atom* of the crystalline substance.

Before considering the experimental evidence, let us speculate a little about what the atomic heat of a crystalline substance might be expected to be. In the first place, the average position of each atom is fixed: The only motions the atom may have are *vibrations* of small amplitude about that fixed point, and in general each atom will have *three* degrees of freedom, with both kinetic energy and potential energy (in equal average amounts) for each degree of freedom. If then we assume that every atom is vibrating independently, with the same average kinetic energy as that possessed by a gas molecule at the same temperature, the atomic heat should be about 1 calorie of kinetic energy plus 1 calorie of potential energy for each degree of freedom, or 6 (more exactly 5.95) calories. (See Sec. 175.)

This prediction, which was made in this way many years ago, is known as Dulong and Petit's law. Many substances have atomic heats close to this value, and all substances approach it at high temperatures (and sometimes slightly exceed it). The atomic heats of other substances, however, fall far short of it at room temperature (diamond is the most notable exception to this law, with an atomic heat of 1.45), and the atomic heats of all substances approach zero as the temperature approaches absolute zero.

Again the quantum theory provides the explanation of these discrepancies, and one that is by this time fairly familiar to the reader. At the lower temperatures, many of the atoms cannot vibrate because they cannot receive a quantum of vibrational energy. If the atom vibrates independently at all, it must receive at least one quantum of vibrational energy—an amount of energy which, as has been stated in Sec. 175, is equal to *h* times the vibrational frequency. If then the vibration frequency is low, it may be possible for all the atoms, on the average, to be excited into vibration, and the atomic heat will be about 6 calories, as predicted by the law of Dulong and Petit. In the diamond crystal, however, the carbon atoms are light, and the intensity of the forces binding them together is testified to by the hardness of the crystal. The possible frequencies of independent vibration for the carbon atom are consequently so high that a single quantum of vibrational energy greatly exceeds the average energy of a gas molecule at ordinary temperatures; only a small fraction of the carbon atoms will then be excited into independent vibrations at room temperature. The very low atomic heat is thus easily explained. At correspondingly lower temperatures similar low values will be obtained for other substances, with the same explanation.

This was Einstein's theory. It explained the experimental results fairly well, although it predicted that atomic heats would diminish somewhat more rapidly with decrease in temperature than was found experimentally. Debye explained this discrepancy. He pointed out that not only is it possible for atoms to vibrate independently; it is also possible for them to vibrate in groups, in a sort of standing wave. Even when the chance is negligibly small that an atom will be excited into vibration by itself, such *groups* of atoms might continue to vibrate, since the greater the number of atoms moving in phase, the less their frequency of vibration, and the less correspondingly the excitation energy required. Using this idea (expressed in exact mathematical form), Debye derived a formula for atomic heats which fits the experimental results as exactly as might be wished.

188. "Free" Electrons in a Metal. What becomes of the electrons in a crystal? If the crystal is an electric insulator, such as sodium chloride or quartz, this is somewhat of an academic question, although of very considerable importance to the theoretical physicist. But if the crystal is an electric conductor, as practically all metallic crystals are, it becomes a very practical question as well, since its properties as a conductor of electricity depend upon the existence of free electrons, forming a sort of electron gas in the spaces between the atomic cores (their nuclei and their undisturbed inner shells of electrons). The steady drift of these free electrons, superimposed upon whatever *random* motions they may have, constitute the electric current. Likewise, these free electrons provide what appears to be an easy way to explain the fact that metals are much better conductors of heat than are any other substances. If, as was at first assumed by this theory, the average kinetic energy of a free electron in this electron gas is the same as that of a molecule of an ordinary gas (at the same temperature as that of the metal), then the random motions of these electrons will serve to transfer heat from the regions of high temperature to those of lower temperature; the process would then be the same as the conduction of heat by an ordinary gas. The much greater effectiveness of the electron gas could be accounted for by the large density of the electron gas and the high speeds of the electrons. (Since the mass of an electron is but 1/1840 part of the mass of a hydrogen atom, its speed will be $\sqrt{1840}$ or 43 times the speed of a hydrogen atom having the same kinetic energy.)

The heating of a metal by an electric current might similarly be explained by this theory. The steady flow of electrons which constitutes the electric current is only an average effect; the motions of individual electrons are continually being interrupted by collisions with the atoms of the metal, and such collisions increase the average kinetic energy of the atoms—that is, heat the metal.

Neither of these explanations is correct.

As long as these free-electron ideas are used only to explain these phenomena of thermal conduction and of electric conduction, they work beautifully. But when all the implications of this theory are considered, serious difficulties develop. The first arises in connection with the atomic heats of metals. If, as is reasonable to suppose, there is at least one free electron for each atom in a metal,[1] and if, as has

---

[1] In molecules it proved satisfactory to assume that, although the deep-lying electrons belonging to each atom continued to belong to the component atoms after the molecule was formed (although the energy levels of these electrons may be somewhat changed in value by the nearness of the other atoms), the outermost or valence electrons rearranged themselves to form a group of electrons belonging to the molecule as a whole. Likewise, in a crystal the deep-lying electrons retain their identity as atomic electrons, and it is the valence electrons (one or two per atom) which become the "free" electrons in a metal.

been assumed above, these electrons have the same average kinetic energy as do the atoms of the metal, then they should contribute materially to the value of the atomic heat of the metal. Since each free electron has three degrees of freedom of translation, they should contribute approximately *3* calories (see Sec. 173) to the atomic heat. But they do nothing of the kind. At ordinary temperatures, the observed atomic heat, as has been described in the preceding section, can be almost entirely accounted for by the motions of the atoms themselves, just as if there were no free electrons present. (When the atomic heat exceeds 6 calories, the electrons must make some small contribution.) This difficulty could be dodged by assuming that, after all, only a few *free* electrons may exist, but then another difficulty arises: If the free electrons are so few in number that their contribution to the atomic heat of a metal may be neglected, then they are too few to account for the conduction of heat and of electricity in the manner which has been assumed!

189. Fermi-Dirac Statistics. This dilemma may be avoided only by considering that the motions of the electrons in the interior of a metal are entirely different from the motions of the molecules of an ordinary gas. At low densities this is not so. Very excellent experiments show that the average kinetic energy of the electrons emitted by the hot cathode of an electron tube is the same as that of the atoms in the hot-cathode surface, and that the velocity distribution is the same among the electrons as among the molecules of a gas at this temperature. But the density of the electron gas in the interior of a metal, even if it contains but one electron per atom, is billions of times greater than the density at which these experiments were made; and when electrons are crowded as closely together as this, the distribution of velocities among them is entirely changed.

The difference is explained by the statistical theory developed by Fermi and Dirac. In an electron gas, according to this theory, the momentum of every electron is "quantized" and must be different from that of every other one. (Since the wavelength of an electron is determined by its momentum, this is equivalent to saying that every electron must have a different electron wave.) The magnitude of this difference will depend upon the closeness of the electrons, and at low densities may be so very small that the electrons will have speeds distributed in exactly the same fashion as for molecules at the same temperature. This is in accord with the experiments mentioned above. But, when the electrons are as closely crowded together as they are in the interstices of a metallic crystal, the required difference in momentum becomes so large that, even at absolute zero, some of the electrons have energies as high as those possessed by the average gas atom at temperatures in the neighborhood of 80,000° C. Employing a terminology made familiar by its use in connection with electrons in atoms, we may say that, in spite of their tremendous speeds, these electrons cannot go any slower *because all the lower energy levels* (slower speeds) *are already filled up.* Hence, even should they collide with an atom, they *cannot give any energy to the atom.*

This theory gives us then this strange picture of a crystal at absolute zero; all the atoms are at rest (except for the half-quantum of zero-point energy mentioned in Sec. 175), and the "free" electrons are moving about between them with speeds varying from zero up to speeds which, in the older theories, would represent upwards of 80,000° C. At room temperature (or even at temperatures close to the melting point of the crystal) the picture, as far as the electrons are concerned, is not much different. The atoms are vibrating, of course, and a very few of the highest-speed electrons may, as a result of collisions with the atoms, have a little *extra* speed, but this change is small in comparison with the very high speeds which they already have. Another consequence of this theory (which comes out as a result of the wave character of the electrons) is that the mean free paths may be many times longer than could be accounted for by the other theories. At absolute zero the electrons may move the full length of the crystal without collision, and at room temperature the mean free path is still of the order of magnitude of hundreds of times the distance between atoms.

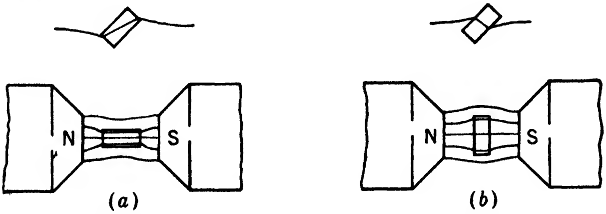
190. Conduction of Heat and of Electric Current. Heat conduction is now explained by these highest-speed electrons, which gain some energy from the more violently vibrating atoms in the hotter portions of the crystal and transfer it to the less agitated atoms in cooler portions. The comparatively small number of electrons participating in this process is compensated for by their high speeds and long free paths. Similarly, if an electric field is set up in the crystals of a wire (as by connecting the wire to a battery), this electric field will increase the velocities in the direction of the field of these highest-speed electrons, and the resultant drift of the electron gas will constitute the electric current. When these electrons collide with atoms, they lose more or less of this extra speed, the lost energy accounting for the heating effect of the current.

191. Work Function. Additional information concerning the electrons inside a metal is provided by the *work function* of its surface, or the amount of work, w_0, required to remove an electron from that surface. The value of w_0 depends upon the nature of the metal and the condition of its surface; it is measured by electrical means, such as the photoelectric effect (see Sec. 107). The existence of this work function may be explained by saying that the interior of the metal is at a positive potential, V_0, with respect to the space outside of it, and the older electron-gas theory for a metal indicated that V_0 should be approximately equal to w_0 divided by the elementary charge e; values computed in this manner ranged from below 2 volts to above 6 volts.

The existence of this inner potential has been confirmed by the *refraction* of electron beams as they pass through the surface of a metal. This phenomenon was observed by Davisson and Germer in connection with their electron diffraction experiments (Sec. 111), but the values of V_0 obtained in this manner are always much higher, sometimes by as much as 10 volts. On the basis of the older theory this was a hopeless discrepancy. According to theory of Fermi and Dirac, however, this difference is accounted for by the very high kinetic energies already possessed by the electrons in the uppermost energy levels of the electron gas. This energy is so high that it takes only the small amount of additional energy represented by w_0 to permit them to escape through the surface.

This work function also plays an essential part in explaining the emission of electrons by a metal surface when it is raised to a high temperature, and the Fermi-Dirac theory is again in good agreement with the experimental evidence.

192. Paramagnetism and Diamagnetism. Although very few substances show strongly magnetic properties, almost all solids, and some liquids and gases, exhibit some induced magnetism in strong magnetic fields. The magnetic properties of a very weakly magnetic substance may be demonstrated qualitatively by placing a small bar of it in a strong magnetic field, supporting the bar so that it may turn freely. For a *paramagnetic* substance the bar turns so as to line up in the direction of the lines of the field, as is shown by Fig. 138(a), the induced



(a) (b)

FIG. 138. (a) PARAMAGNETISM. (b) DIAMAGNETISM. The lines of magnetic induction find easier paths through the paramagnetic bar than outside of it. For the diamagnetic bar, the reverse is true. The upper figures show lines of induction through bars obliquely inclined to the magnetic field, and explain why the bars turn as they do.

magnetism being in the same direction as it would be in a similar bar of iron, but vastly weaker. Similar bars of other substances turn so as to set themselves at right angles to the lines of the magnetic field, as shown by Fig. 138(*b*), showing that the induced magnetism in these bars is in such a direction as to *oppose* the field. Such substances are called *diamagnetic* substances. Diamagnetism is always much weaker than the weakest paramagnetism.

The modern theories of atomic structure have been fairly successful in explaining these magnetic properties of different substances. The magnetic properties of individual atoms have already been described in Chapter X, Sec. 152. Each electron, because of its intrinsic spin, is a tiny permanent magnet; and each electron orbit (except for the S orbits) is equivalent to an electric circuit which produces a magnetic field, and is hence also equivalent to a tiny electromagnet. The magnetic moment of the spinning electron is one Bohr magneton (Sec. 152), and that of each orbit is equal to l Bohr magnetons, where l is the azimuthal quantum number (Sec. 131) for that orbit. The magnetic moment of the atom as a whole will be the resultant of all such individual magnets within it, taking due regard to the orientation of such magnets with respect to each other. In some atoms, as, for example, the inert gas atoms in their normal (unexcited) states, the resultant magnetic moment is zero. For others, such as the hydrogen and alkali metal atoms, it is due to the spin of one electron alone. The magnetic moment of no atom is more than a few Bohr magnetons.

As has been demonstrated by the Stern-Gerlach experiment (see Secs. 152 and 153), free atoms which possess magnetic moments orient themselves in *quantized* positions in a magnetic field, with just as many of them pointing against the field as pointing in the direction of it. This applies to streams of free atoms which are entirely independent of one another. In a gas or a liquid in which atoms or molecules may collide with one another, there is some tendency for them all to line up in the direction of the field, rather than against it, since that represents the position of minimum energy. This tendency is at all times opposed by the converse tendency toward entirely random positions which is produced by collisions. A rough comparison of these two tendencies may be made by comparing the magnetic energy of an atomic magnet, when it is lined up in the direction of the field, with the average kinetic energy of its thermal agitation. For example, one Bohr magneton has a magnetic moment, M_0, equal to $eh/4\pi mc$, and its energy when lined up with the magnetic field, H, equals $-M_0H$. (The negative sign indicates that this is the position of least energy.) If $H = 10{,}000$ oersteds, which represents quite a strong field, then $M_0H = 9.2 \times 10^{-17}$ erg. On the other hand, the average kinetic energy of a gas molecule at 20° C is equal to $(3/2)kT = 6 \times 10^{-14}$ erg, which is 650 times the value computed for the magnetic energy. The result is that in gases, in liquids, and in most solids the effective lining up of the atomic magnets is very small, and the induced magnetism is consequently very weak. Such substances are *paramagnetic* substances.

Diamagnetism is exhibited only by substances whose atoms possess no permanent magnetic moments, and the very weak induced magnetism which appears in these substances may be explained as due to a sort of *induced electric current*, produced by the action of the magnetic field upon the electrons in their orbits. According to Lenz's law, this "induced current" must produce a magnetic field which is in opposition to the external field—that is, the induced magnetism must be in the opposite direction to the external magnetic field—just as it is found to be experimentally. According to this explanation for diamagnetism, the same effect must be produced in *any* atom. In those atoms which possess permanent magnetic moments, this very weak diamagnetic effect is entirely masked by the paramagnetic phenomena.

193. Ferromagnetism. It has been far more difficult to obtain a good explanation for *ferromagnetism*, that is, for the intense magnetic properties of the very few substances (iron, nickel, cobalt, and a few alloys and compounds) for which iron is representative. In a metallic crystal, the magnetic properties are due to the atomic electrons, and not to the valence electrons which form the free electron gas described earlier in this chapter. For example, iron atoms have completed K and L shells, 14 electrons in their M shells, and 2 N electrons which become free electrons; and the magnetic moment of the iron atom is due to the combined magnetic moments produced by the spins of 4 of its M electrons. This alone would make it no stronger than many of the atomic magnets to be found in paramagnetic substances. To be sure, owing to interactions between adjacent atoms in a crystal, there is a tendency for the atomic magnets to hold each other lined up; but this tendency is equally great in many paramagnetic crystals, and is entirely too small to account for the very intense magnetism observed in iron. It appears that, in iron and in the few other substances which exhibit comparable magnetic properties, there also exist very much more powerful electrical forces which act concurrently with the magnetic forces to hold the atomic magnets in line. In consequence of these electrostatic forces, there exists a strange regimentation of the iron atoms such that, throughout large regions or "domains" within the crystal, all the atomic magnets always line up together to make the whole "domain" a permanent magnet. Such "domains" may contain 10^{14} to 10^{15} atoms. On the scale of ordinary things they are of course still very small, about 10^5 of them existing on the average in each crystal of polycrystalline iron, or about 10^6 of them per cubic millimeter. In spite of this small size, however, their magnetic fields have been mapped in some cases by the familiar "iron-filings" method, colloidal iron oxide particles being used in place of iron filings, and viewed through a microscope.

It is to be emphasized that each domain is *always* permanently magnetized to saturation, all its atomic magnets being lined up in the same direction. In an apparently unmagnetized piece of iron they are all magnetized in different directions, as is indicated in Fig. 139(*a*). (In iron these possible directions are all parallel to the crystal axes, as the figure indicates.) When a piece of iron is placed in a magnetic field, the first effect of that field is to line up all the domains in parallel directions, as is indicated in Fig. 139(*b*). As the magnetic field is slowly increased from zero value, the domains line up one by one, all the atomic magnets in each domain abruptly swiveling around together. This part of the magnetization process corresponds to the lower part of the magnetization curve shown in Fig. 140, and in the small circle it is indicated how this part of the curve would appear as an irregular broken curve if the data could be taken minutely enough and plotted to a large enough scale. By suitable amplifier circuits these sudden changes may be revealed on an oscillograph screen, or may be "heard" as audible rustling noises or minute clicks in a loud speaker. In the final process of magnetization the magnetism in the domains is turned from the direction of easy magnetization which is nearest to parallel with the field [the con-

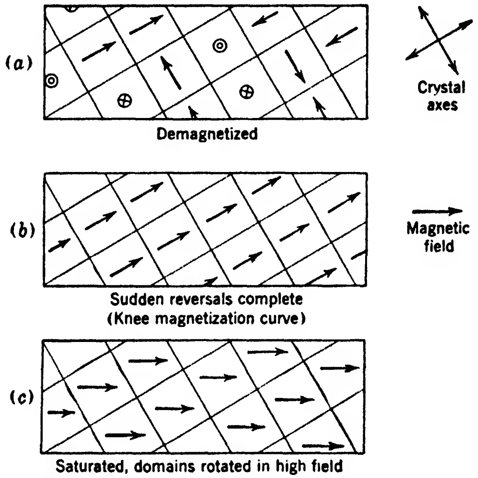


FIG. 139. DOMAINS IN A SINGLE CRYSTAL OF IRON.

dition shown in Fig. 139(*b*)] to the condition of full saturation which is shown in Fig. 139(*c*). This part of the process is represented by the upper part of the magnetization curve in Fig. 140.

Figure 139 represents highly idealized conditions within a single crystal. It is doubtful that the domains are cubical, as they are there represented, although this does not materially affect the conclusions drawn. In polycrystalline iron, moreover, the domains in different crystals will be oriented in all possible directions in space, and for this reason the magnetic properties of ordinary iron will appear to be the same for every direction of the magnetizing field. But in a single crystal of iron (and it is now possible to make single crystals large enough for such experiments), the magnetization curve obtained will depend upon the direction of the magnetic field with respect to the crystal axes, and in these specimens the directions of easy magnetization are found to be those shown in Fig. 139(*a*) for iron. (They differ for different crystals.)

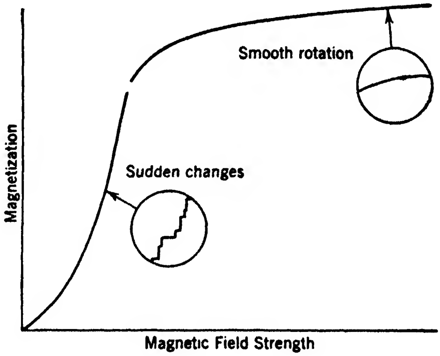


FIG. 140. A TYPICAL MAGNETIZATION CURVE.

A more complete discussion of this subject of ferromagnetism is to be found in the article by Bozorth which is listed below, and from which much of the material of this section, including Figs. 139 and 140, has been taken, with the permission of its author.

PROBLEMS

1. (*a*) Compute the length of one side of the cube shown in Fig. 135(*c*) if it is made by piling up balls 5 cm in diameter. (*b*) This cube contains the equivalent of 4 balls—½ of each of 6 balls, and ⅛ of each of 8 balls. What part of the volume of the cube is occupied by the balls, and what part is empty?

2. (*a*) Compute the length of one side of the cube shown in Fig. 137(*a*) if it is made by piling balls 5 cm in diameter. This cube contains the equivalent of 2 balls. Explain. (*b*) Compute the relative part of this cube occupied by the balls.

3. The KCl crystal structure is the same as that for NaCl. Its density is 1.99 grams/cm^3. How far apart are the atoms in the KCl crystal?

4. Calcium and copper have the same crystal form (face-centered cubic), and the "diameter" of the copper atom is 2.56A. Compute the diameter of the calcium atom, assuming that the atoms just touch in a crystal. The densities are: calcium, 1.54; copper, 8.95. Atomic weights are: calcium, 40; copper, 63.5.

5. Compute the atomic heats for copper, gold, iodine, lead, and zinc. Specific heats are: copper, 0.094; gold, 0.031; iodine, 0.054; lead, 0.030; zinc, 0.092.

REFERENCES

FOR ADDITIONAL READING

SIR WILLIAM BRAGG, *Concerning the Nature of Things*, Chapters IV, V, and VI.

G. F. HULL, *Modern Physics*, parts of Chapter IX.

R. M. BOZORTH, "Recent Advances in the Theory of Magnetism," *Scientific Monthly*, April, 1938, pp. 366–371.

FOR ADVANCED READING

SIR WILLIAM BRAGG, *X-Rays and Crystal Structure*.

L. A. DUBBRIDGE, "Electron Theory of Solids," *American Physics Teacher*, Vol. 7, pp. 357–366, December, 1939.

HARNWELL and LIVINGOOD, *Experimental Atomic Physics*, Chapter VI.

L. B. LOEB, *Atomic Structure*, Chapters XX and XXI.

R. M. BOZORTH, "Magnetism," *Reviews of Modern Physics*, Vol. 19, pp. 29–86, January, 1947.

CHAPTER XIV

BLACK-BODY RADIATION

194. Radiation from Hot Bodies. When the temperature of a gas or vapor is high enough, collisions between molecules will raise their energies to excited levels (this is the commonest type of excitation process), and light will subsequently be emitted. If the density of the gas or vapor is low enough, the emitted light will consist of bright-line spectra—principally molecular spectra at low temperatures, but predominantly atomic spectra at temperatures high enough to dissociate molecules into atoms. The details of these phenomena have been considered in Chapters VIII to X. When the gases are moderately dense, the lines are broadened (this shows up even at atmospheric pressure), and the fine lines of the molecular spectra tend to run together into a continuous band. This broadening results from the crowding together of atoms and molecules, so that their mutual interactions alter the values of the energy levels at the time of light emission. It is a curious fact that this effect is much greater when the atoms are crowded by other atoms of the same kind: a trace of one gas mixed in with another quite dense gas may give fine sharp lines even when the lines for the principal constituent of the gas mixture are quite broadened. When, however, the atoms are as closely crowded together as they are in solids and liquids, or even in very dense gases, their characteristic radiations are completely obliterated: The radiations from such a body constitute a continuous spectrum extending over a wide range of frequencies.

195. Black Bodies. The total radiation (per unit area of surface), as well as the distribution of energy according to the frequency of the radiations, depends first of all upon the temperature of the body, but it is also greatly modified by the nature of the *surface* of the body, as well as by the opacity of the body as a whole. A surface which reflects strongly the radiation which falls upon it from outside the body will equally strongly reflect back into the body the radiation which is trying to get out through that surface: a body may be a poor radiator for this reason. The best radiator should have a perfectly transparent rough surface, one that reflects none of the radiation which falls upon it. On the other hand, a body may be quite transparent throughout, absorbing very little radiation for this reason; experiments show that such substances are likewise very poor radiators. Fused quartz is a very good example of this for the visible range of wavelengths. Roughly speaking, a poor absorber is a poor radiator; a good absorber, a good radiator.

The best absorber is an opaque body with a perfectly transparent surface: Such a body will appear perfectly black in any light, and such a body, at an elevated temperature, will also be the most effective radiator of "light." ("Light," in the sense in which it will be used throughout this chapter, includes all electromagnetic radiations, both visible and invisible.) No substance completely fills this specification, but the practical equivalent of a perfectly black body may be created synthetically by the scheme shown in Fig. 141(*b*). If a cavity is made in an opaque body, with rough walls and a small opening, the opening will appear perfectly black, even when the walls of the cavity are quite highly reflecting. Figure 141(*b*) shows a cross-section of such a cavity. Whatever light enters the opening is absorbed by repeated reflections from wall to wall, and only a negligible fraction is ever reflected outward again. The blackness of the opening in a doghouse is an excellent illustration of this. The intense blackness of black velvet may be explained similarly. See Fig. 141(*d*). Likewise the surface formed by the stacked edges of a pack of safety razor blades [Fig. 141(*c*)] will appear nearly

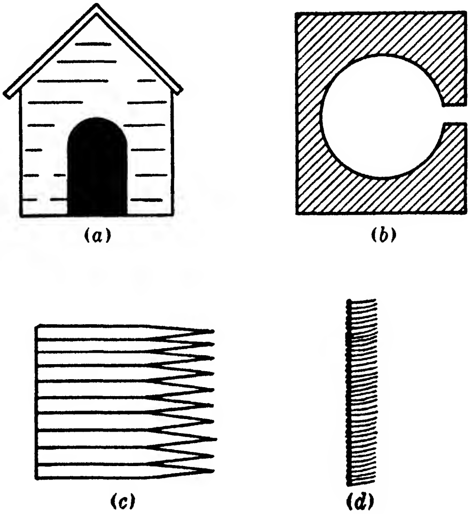

(*a*) (*b*)
(*c*) (*d*)

FIG. 141. EXAMPLES OF BLACK BODIES.

as black as velvet, although they are highly polished, because their polished faces are so inclined that they reflect the light deeper and deeper into the pack, never out again. The same principle is utilized in the Raman-effect tube (Fig. 131) to obtain a perfectly black background against which to view the very faint scattered light.

196. Kirchhoff's Law. When chinaware is heated to incandescence in a furnace to harden it—the process known as "firing"—the patterns painted upon its surface appear reversed; the lines which will appear dark in the finished ware glow more brightly than does the pure white background. If the body in which a cavity has been made [Fig. 141(*b*)] is similarly heated to incandescence, the opening to the cavity will glow brighter than will anything else at that temperature. Furthermore, *the radiation from this cavity* (per unit area), both in quality and quantity, *will depend upon nothing but the temperature of the body.* It is called **black-body radiation** and may be taken as a standard with which to compare the radiation from any other body.

If, for example, we compare the radiation from the outside surface of this body with the black-body radiation from its cavity, we may thereby compute the *emissivity* coefficient of the surface. Since practically always this coefficient depends upon the frequency of the light considered, these measurements must be made at a definite frequency, (more exactly speaking, over a narrow band of frequencies in the neighborhood of f). The **emissivity,** E_f, is defined by

$$E_f = \frac{W'_f}{W_f} \tag{14·1}$$

Here W_f is the intensity of the light emitted by the cavity (black-body radiation), or the power radiated by it per unit area per unit frequency interval; and W'_f is the intensity of the light emitted from the outside surface of the body, *at the same temperature.*

The **absorption coefficient,** A_f, for light of the same frequency range is defined as

$$A_f = \frac{\text{Light absorbed by the surface}}{\text{Light incident upon the surface}} \tag{14·2}$$

If, now, E_f and A_f are measured for the same surface, it will be found that

$$E_f = A_f \tag{14·3}$$

This is **Kirchhoff's law.** It states accurately the principle loosely stated earlier, that a good radiator is a good absorber, etc.

197. Cavity Radiation. To understand better the significance of Kirchhoff's law, let us now consider a rather large cavity in a body which is maintained at a constant temperature, T. (This may be the interior of a furnace which is entirely inclosed and very uniformly heated.) We shall also assume that the inner walls of this furnace cavity are perfectly black, although it will develop later that this is not necessary. These walls will then be continuously *radiating* and *absorbing* light,[1] the rates of emission and of absorption being determined by the temperature of the cavity, and by that alone. The rate of emission corresponding to each frequency varies in the manner shown in Fig. 142; but we are at present concerned only with the fact that, at each temperature, the walls are absorbing energy at the same rate as that at which they are emitting it. For the frequency f this rate may be represented by W_f. (W_f is the power radiated per unit area per unit frequency interval.) We may indeed think of the furnace cavity as being filled to "saturation" with radiation (light) which is in equilibrium with the black-body surface in the same sense that a saturated vapor is in equilibrium with its liquid phase at the same temperature. (See Sec. 184.) And in this sense we may also speak of the *temperature of the light*, or radiation, as being equal to the temperature of the walls.

Let us now introduce into this furnace cavity a small body whose temperature is the same as that of the furnace, but whose surface is *not* black. Since the radiation within the cavity is uniform throughout, the radiation falling upon each unit area of the surface of this body is the same as that falling upon unit area of the walls. Consider the frequency range in the neighborhood of the frequency f, for which its absorption coefficient equals A_f. At this frequency the body then absorbs light at the rate $A_f W_f$ per unit area. The same area of this body is at the same time emitting light at the rate W'_f, which is, according to equation (14·1), equal to $E_f W_f$. Hence, if Kirchhoff's law holds true, so that E_f equals A_f, at this frequency this body is emitting just as much light as it is absorbing. If then Kirchhoff's law holds true at all frequencies, the net loss or gain of energy by the body must be zero. When its temperature is that of the furnace, *this body is in thermal equilibrium* with the furnace, and *with the radiation within the furnace cavity.*

Furthermore, the energy which is not absorbed by the surface of the body is reflected by that surface, so that the surface appears just

---

[1] It must be emphasized throughout that these are two distinct and separate processes. This was first recognized by Prevost; it is sometimes called Prevost's law of exchanges.

as bright as if it were a black body! [That is, at the frequency f, the reflected light is $(1 - A_f)W_f$, and the emitted light plus the reflected light equals $A_fW_f + (1 - A_f)W_f = W_f$.] If one were to look into the furnace through a small peephole, the body could not be seen, because it would be uniformly bright, and just as bright as the walls. For this same reason the walls of the furnace itself need not be black, as was stated earlier: multiple reflections from the walls of a closed furnace fill the furnace cavity with black-body radiation, and give the walls a brightness equal to that of a perfect black body at the same temperature. This explains the effective black-body radiation emitted through a small hole in the wall of such a furnace, as illustrated by Fig. 141(*b*). In determining the emissivity of a surface, care must be taken to shield it from the radiation from the walls of the furnace, or the reflected light will make it appear brighter than it should.

198. Proof of Kirchhoff's Law. It may now be shown that the failure of Kirchhoff's law would lead to conditions contrary to all scientific experience. For, if at any frequency the absorption coefficient of a body were greater than its emissivity coefficient, then at this temperature the body, when placed in a furnace cavity, would gain energy at a greater rate than it lost it until its temperature was higher than that of the furnace. It could thus remain indefinitely at a higher temperature than that of the furnace. If, on the contrary, it were able to emit light more effectively than it could absorb it, it could maintain itself indefinitely at a temperature lower than that of its surroundings. But scientific experience has never encountered circumstances where thermal equilibrium could be maintained without all bodies being at the same temperature. Hence Kirchhoff's law must always hold.

Kirchhoff's law enables us to compute the radiation from any surface for which the absorption coefficients (for different wavelengths) are known, provided that the laws for radiation by a black body are known. When the absorption coefficient is the same at all frequencies, the body may be called a gray body. In a more restricted sense, a gray body is one for which the absorption rate is constant over the frequency range of visible light, and colored bodies are those which exhibit greater absorption in some parts of this range than in other parts—that is, bodies for which the absorption is selective. Practically all bodies show selective absorption when the entire range of invisible light as well as visible light is considered.

The remainder of this chapter is primarily concerned with the laws for radiation by an ideal *black* body.

199. Stefan-Boltzmann Law. As has already been emphasized, the radiations from an ideal *black* body, or from the "synthetic" black body formed by a furnace cavity, are determined by its *temperature* alone. And **the total energy radiated, per unit area, is proportional to the fourth power of the absolute temperature.** If W represents the total energy radiated per unit area, and T, the absolute temperature, this law may be expressed by

$$W = sT^4 = 5.67 \times 10^{-12} T^4 \text{ watts/cm}^2 \qquad (14\cdot4)$$

in which the proportionality factor s has the numerical value given. At the temperature of melting ice, this radiation amounts to 0.032 watt/cm^2, while at the surface temperature of the sun (6000° C) it is about 8.8 kilowatts/cm^2.

This law was discovered experimentally by Stefan, and the same law was derived by "classical" thermodynamic theory by Boltzmann. It applies at any temperature above absolute zero, and represents a *thermal property of the radiation itself.*

200. Planck's Law. When the "light" emitted by a black body is spread out into a continuous spectrum by means of a prism or grating, values for the intensity, W_f (power radiated per unit area per unit frequency interval), may be measured throughout the available frequency range; the results of such measurements may be represented by a curve such as is shown in Fig. 142.[2] At very low and at very high frequencies

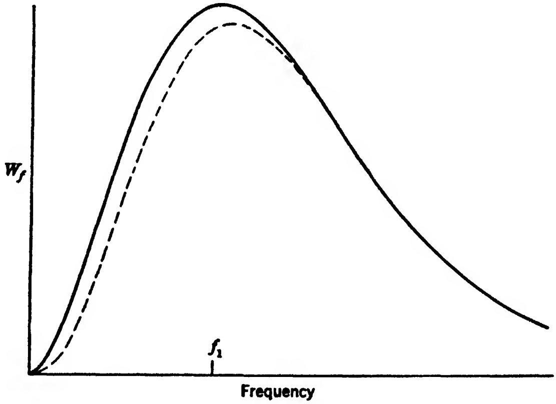


FIG. 142. BLACK-BODY RADIATION.

---

[2] This curve does not directly represent the experimental data. The thermocouple arrangement used (see Sec. 85) measures the energy in a narrow strip of the spectrum of constant width. If a prism spectrum is used, this strip will represent a band of frequencies which is broad in the infrared and narrows as the frequency increases. If a grating spectrum is used, the width of the frequency band will change in just the other direction. In either case the experimental data must be corrected for this variable frequency range, as well as for absorption in the optical parts, etc.

the intensity approaches zero, with a maximum value appearing at an intermediate frequency, f_1. As long as only the wave properties of light were known, when it was still believed that the atomic sources of light were sources of steady radiation, these experimental results could not be accounted for theoretically. This represented one of the outstanding failures of prequantum physics.

In 1901 Max Planck took the very bold step necessary to the solution of this problem: the step that laid the foundation stone for all of modern *quantum* theory which has dominated so many of the earlier chapters of this book. Planck made the then fantastic assumption that **energy is radiated,** not as a *continuous* flow of waves, but **in discontinuous amounts, or *quanta,* the size of a quantum being proportional to the frequency of the waves radiated.** (The concept of a light particle, or *photon,* was yet to be invented, by Einstein.) If w_f is a quantum of radiant (light) energy of frequency f, then

$$w_f = hf \tag{14·5}$$

where h is the now familiar proportionality factor. This is **Planck's law.** The daring which was required then to advance this tradition-breaking idea cannot now be appreciated. Even though it was successful in producing an equation that fitted the experimental data very well indeed, most physicists at first felt that this must be something in the nature of a mathematical coincidence, and that the true physical explanation might be quite different. Time has shown how very right Planck was, as we of course now know so well. With this quantum principle Planck was able to derive a formula which fitted the experimental results as closely as the values could be determined experimentally.

201. Planck's Law and the Photon Gas. The quantum idea was thus discovered in connection with a relatively complex phenomenon. The completion of this idea by the concept of light particles or photons came later on, in connection with the phenomena of photoelectricity and X-rays; these phenomena, in contrast to the phenomena of black-body radiation, give simple, direct proofs for the quantum concepts, as has been explained in Chapter VII. The more complete quantum concept now provides a much simpler description of black-body radiation, one which was first made by Bose.

In Sec. 197 the radiation in a furnace cavity was described as constituting a photon gas, and the resemblance between the curve in Fig. 142 and the curve (Fig. 120) obtained for the distribution in velocity of gas molecules seems in good accord with this description. Is it possible that the distribution in frequency of this cavity radiation may be computed from the laws of chance (statistical laws) by which the distribution in velocity for gas molecules was determined? The dotted line in Fig. 142 shows the distribution in frequency which may be computed in this manner. At high frequencies the computed values agree quite well with the experimental results, but at the lower frequencies the experimental values are all considerably higher than the computed ones.

It is evident that the analogy between a photon gas and an ordinary gas cannot be carried too far; in some respects there must be distinctive differences between photons and ordinary molecules. Bose specified three such differences: (1) the number of photons is not a constant, as it is for an ordinary gas; (2) the photons cannot be distinguished one from another; and (3) the uncertainty principle of Heisenberg (Sec. 114) must be taken into account. The significance of these differences cannot be made apparent without going through the statistical theory involved; it must suffice here to state that the statistical theory, when revised in these particulars, gives exactly Planck's formula, and hence is in exact agreement with the experimental measurements. Planck's formula is

$$\begin{aligned} W_f &= \frac{2\pi h}{c^2}\,\frac{f^3}{e^a - 1} \\ &= 4.62 \times 10^{-47}\,\frac{f^3}{e^a - 1} \qquad (14\cdot6) \end{aligned}$$

where a equals hf/kT, and e, equal to 2.72, is the "base" for natural logarithms.

The formula for the dotted line plot is the same, except that e^a replaces $(e^a - 1)$ in the denominator. It is known as **Wien's formula.**

202. Bose-Einstein and Fermi-Dirac Statistics. There seems to be no good reason why the second and the third of the restrictions specified by Bose for a photon gas should not apply to any gas whatsoever; and Einstein showed that, as a matter of fact, they probably do apply to all gases, but that they are of negligible consequence for molecular gases except at temperatures so low that other effects, of greater magnitude, are encountered. These restrictions seem indeed to be closely related to the dualistic wave-particle properties of all things physical, which is first mentioned in Sec. 112. When the wavelengths of the gas particles ($\lambda = h/mv$) are considerably shorter than the average distance between the particles, the Bose restrictions are of negligible consequence, and the ordinary laws of chance are entirely satisfactory. This is the case for molecular gases. Even the photon gas illustrates this at high frequencies (short wavelengths), as may be seen in Fig. 142: at the high frequencies the solid line, which represents the experimental values, coincides with the dotted line which was computed from the ordinary statistical laws. But when the wavelength becomes comparable with the average distance between photons, the statistical laws are quite different, as is shown by the lower frequency part of the curve in Fig. 142.

Still another example of this is given by the free electron gas in a metal. Here again the Bose restrictions are of great consequence, both because of the lightness of the particles and because of their close crowding. In this case, however, entirely different statistical laws are required to describe their behavior: those of Fermi and Dirac, as has been explained in Sec. 189.

The photon gas and the free electron gas in a metal represent two general possibilities, and it has been shown that, whenever the gas particles are so small or so close together that the ordinary laws of chance are no longer applicable, their behavior may deviate either towards that of a photon gas, or towards that of a free electron gas. In the one event their motions are described by Bose-Einstein statistics; in the other, by Fermi-Dirac statistics. Which alternative is followed by any particular gas is determined by the *spin* properties of its particles. If the angular momentum of spin is an integral number of quanta (an integral number times $h/2\pi$), Bose-Einstein statistics describes their motions; but if it is a half-integral number ($\frac{1}{2}$, $\frac{3}{2}$, $\cdots$) times $h/2\pi$, Fermi-Dirac statistics must be used.

203. Wien's Displacement Law. Figure 143 shows the intensity distribution curves for several different temperatures. As the temperature is raised, not only does the total energy of black-body radiation increase (as described by Stefan's law), but the quality changes also; the intensity increases at every frequency, but the increase is greater at the higher frequencies than at the lower ones. Thus, when a black body which is being heated first becomes luminous (at about 800° C) it appears a dull red;[3] then, as the temperature rises, it grows brighter and yellower, and approaches white when it becomes very bright. The curves are all of the same shape; the height of each one is proportional to the cube of the absolute temperature, while its width is directly proportional to the absolute temperature. And its area, which represents the total radiation, is thus proportional to the fourth power of the absolute temperature, as is stated by the Stefan-Boltzmann law. The change of color is, of course, due to shift of the peak of this radia-

---

[3] Because of a physiological phenomenon known as the Purkinje effect, the very first light seen in a very well-darkened room will be colorless (gray).

tion curve towards the higher frequencies as the temperature rises; from what has been said above it follows that:

> **The frequency at which the peak occurs for black-body radiation is directly proportional to the absolute temperature.**

This statement is **Wien's displacement law.**[4] It may be verified experimentally, or it may be derived mathematically (by the calculus method) from Planck's law. It may be stated in numerical terms as

$$f_1 = 5.88 \times 10^{10} T \text{ sec}^{-1} \tag{14·7}$$

For all ordinary terrestrial sources of light (with the exception of very brief flashes, such as those produced by the flash bulbs used for photography) this peak lies in the infrared, far below the visible range of frequencies, as is shown in Fig. 143. This is true even for the electric arc, which operates normally at a temperature of about 4000° C. At

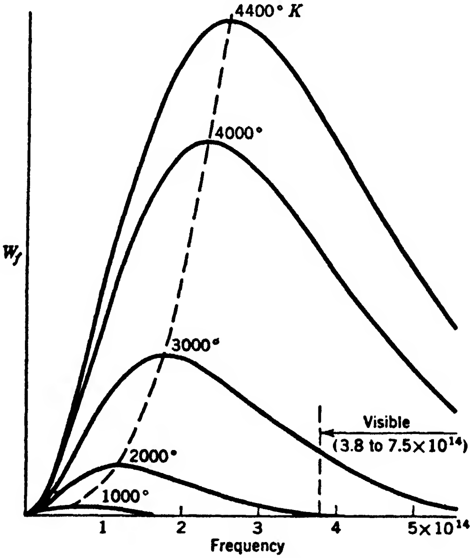


FIG. 143. BLACK-BODY RADIATION AT VARIOUS TEMPERATURES.

---

[4] It is also possible to plot a curve with intensity of light per unit difference in wavelength (this is not the same as W_f, or even proportional to it) as ordinates and wavelength as abscissas; this curve also shows a peak value, which occurs at a wavelength which is inversely proportional to T. This inverse proportionality is the original statement of Wien's displacement law. Both expressions represent the same physical principle, although the two peaks do not occur at the same place in the spectrum!

the temperature of the surface of the sun (about 6000° C) the peak is close to the lower or red limit of the visible spectrum (when the data are computed and plotted in terms of wavelength, as explained in footnote 4, the peak is found in the blue part of the spectrum). Under experimental conditions this temperature has been approached rather closely by the electric arc. When the arc is operated under a pressure of 22 atmospheres, its crater has been observed to reach a temperature of 5620° C. The surface temperatures of many stars are much higher than that of the sun, such that these stars appear blue or bluish. Practically all of the ultraviolet light reaching us from the sun and stars is cut off by the earth's atmosphere. If the radiations from these blue stars could be measured above the atmosphere, the peaks would be found in the ultraviolet. At the temperatures believed to exist in the interior of stars (see Sec. 258) the peak radiation should occur in the X-ray range of frequencies.

204. Light Sources and the Radiation Laws. Wien's displacement law is very significant in explaining the relative efficiency of artificial light sources, such as incandescent lamps. A small increase in the operating temperature of such a lamp may make a large increase in its overall brightness (visible light emitted) as well as a nearer approach to daylight whiteness. The improvements which have been made from time to time in incandescent lamps have been accomplished by discovering ways to increase the operating temperature without decreasing the useful life (nominally 1000 hours). Thus tungsten, in spite of its lower emissivity, makes a better lamp filament than carbon, because it can be operated at a higher temperature without excessive disintegration. A carbon-filament lamp can be operated at higher than its normal temperature so as to obtain as high an efficiency as that obtained from a tungsten lamp, but it will then blacken the bulb and burn out very quickly.

Tungsten in an atmosphere of nitrogen, argon, or other heavy gas which does not react chemically with it can be operated efficiently at a still higher temperature, because the gas tends to discourage the evaporation of tungsten vapor from the filament. There are some disadvantages in using gas in a lamp, principally those due to the heat carried away from the filament by the gas, but these are more than outweighed by the increased temperature attainable. They may be minimized by coiling the filament wire into as tight a coil as possible. Modern incandescent lamps are "gas-filled" tungsten-filament lamps made in this manner.

By operating a tungsten lamp at a still higher temperature a very much more powerful light source for photographic work is obtained, at the sacrifice of useful life. Lamps of this type, having a nominal life of 8 or 10 hours, are now in common use.

Flash bulbs are quite a different sort of light source. The light emitted by them results from a very brief but violent chemical reaction between the metal (aluminum) foil and oxygen contained in the bulb, the reaction being touched off by a little incandescent filament. The light consists of a continuous spectrum whose maximum corresponds to a temperature in the neighborhood of that of the sun. One of the interesting phenomena exhibited by these bulbs is the firing of one bulb by the *light* from another one. If two bulbs are placed close together, they will both flash simultaneously when the filament of only one of them is heated. The light responsible for this ignition must be of very short wavelength, since a flash bulb is unaffected by the near presence of a very powerful incandescent lamp.

205. High-Temperature Measurement. Low and medium temperatures may be measured by numerous means: The familiar mercury-in-glass thermometers are of limited range, but most convenient for use within that range; the electrical resistance thermometer has a wider range, being especially useful for low temperatures; and thermocouples made from various combinations of metals are useful over the entire range, up close to the melting points of the metals which may be used for this purpose. The most reliable thermocouples for high-temperature measurement are made with one wire of pure platinum, the other of an alloy consisting of 90 per cent platinum and 10 per cent rhodium; this is the thermocouple used in defining temperatures from the melting point of antimony to that of gold.[5]

High temperatures (beyond 1063° C, the melting point of gold) are measured by means of the radiation emitted by a black body at the temperatures in question. The total radiation has been used, but ordinarily it is more convenient to measure the radiation contained within a narrow band of frequencies (in the visible range) by means of an optical pyrometer such as the one shown schematically in Fig. 144. The lenses, L_1 and L_2, form a telescope through which the hot body is viewed, and the filament of the incandescent lamp, A, is in the position

[5] The international temperature scale defines temperatures from the boiling point of liquid oxygen to the melting point of antimony in terms of a platinum resistance thermometer; from this point to the melting point of gold, in terms of the platinum, platinum-rhodium thermocouple described above; and above that, in terms of the radiation laws, as explained in this section. The standard temperatures for calibrating the resistance thermometer and the thermocouple are: boiling points of liquid oxygen ($-182.97°$ C), of water (100° C), and of sulfur (444.60° C); melting points of ice (0° C), silver (960.7° C), and gold (1063.0° C). (All points are measured under equilibrium conditions, at normal atmospheric pressure.)

usually occupied by the cross-hairs in a laboratory telescope. The real image, I, which is formed by the objective lens, L_1, coincides with the lamp filament, so that the observer, looking through L_2, sees this image superimposed upon the filament. The appearance is somewhat as is shown in Fig. 144(*a*) and (*b*). A piece of colored glass, G, serves as a light filter to limit the light to a narrow band of frequencies; usually this is a piece of red glass which transmits a band in the neighborhood of 6650A wavelength. The brightness of the filament is controlled by means of the rheostat, R, which is adjusted until the filament is just as bright as the image. This can be done with considerable precision, since, when the brightnesses are equal, the filament disappears. The screen D is a plate of dark ("smoked") glass which reduces the brightness of the image, so that very high temperatures (bright sources) may be measured without having to heat the filament beyond a safe operating temperature.

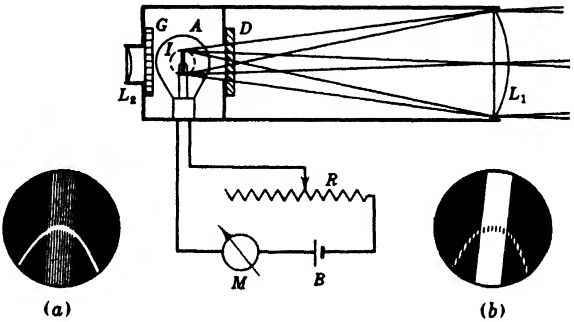


FIG. 144. OPTICAL PYROMETER. (*a*) Filament too hot. (*b*) Filament too cool.

The filament brightness is measured by means of the ammeter M, the temperature values corresponding to each ammeter reading having been determined by calibration in comparison with a standard optical pyrometer. In commercial instruments the meter may have a scale marked off directly in degrees. The primary calibration of the *standard* pyrometer must of course be made by photometric measurements, using Planck's law.

When the body whose temperature is to be measured may be provided with a cavity having a small opening upon which the pyrometer may be sighted, the pyrometer reading will give the true temperature of the body. If the emissivity of the substance is not too low, this cavity may be quite a shallow one, as for example a small hole drilled part way through it.

When, however, it is necessary to sight upon the outside surface of the body, as in measurement of the temperature of a lamp filament, the reading of the pyrometer then gives, not the true temperature, but the *apparent* temperature; and the true temperature cannot be found unless the emissivity of the surface is known. For this reason Worthing and others have made accurate measurements of emissivity for various filament materials, such as platinum, tungsten, and tantalum. The experimental procedure is essentially that described in Sec. 196. In Worthing's method the specimens consist of small tubes of the material, heated in a vacuum by passing an electrical current through them; the black-body radiation is obtained from a tiny hole drilled in the tube wall. The results of a set of such measurements are most conveniently expressed as a table of apparent temperatures and of corresponding true temperatures.

206. Lambert's Law. The emission of light from a surface takes place in all directions from that surface, and the energy values given by the formulas representing the various radiation laws are obtained by adding up the radiations in all directions from unit area of the surface. In many experiments, on the other hand, one is interested in the light emitted in one direction only, and though this quantity will be proportional to the sum total for all directions, the radiation is rarely uniform in all directions. The proportionality factor then depends upon the angle of emission.

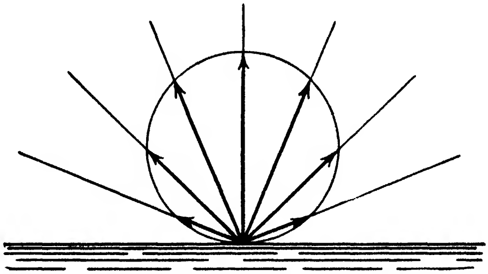


FIG. 145. LAMBERT'S LAW.

For a rough, black surface, and for most other rough surfaces, the emission is greatest in a direction normal to the surface; as the angle with the normal increases, it decreases in the manner shown by the vector lines in Fig. 145. **The intensity of the light emitted in a direction making an angle θ with the normal is proportional to $\cos \theta$.** This law is known as **Lambert's law.** If we receive this light upon a flat photometer plate placed at right angles to the light rays and at a distance D

from the emitting surface, then the intensity of illumination at the photometer plate will be

$$I = \frac{B \cos \theta}{D^2} \tag{14·8}$$

B is the brightness of the surface, and equals W, the light emitted in all directions by the surface, divided by π.

$$B = \frac{W}{\pi} \tag{14·9}$$

(This may be proved by summing up the illumination over a hemisphere of radius equal to D, and equating the sum to W. It is a calculus problem.)

As a direct consequence of this law a rough incandescent surface appears equally bright from any angle of view. At first thought this statement seems to be self-contradictory. It must be remembered, however, that when a surface is viewed obliquely it appears foreshortened or narrower, the apparent area being equal to $\cos \theta$ times the true area. Hence the emission of light by each unit of the foreshortened area is the same for any angle. It is for this reason that the incandescent-lamp filament in the optical pyrometer appears equally bright all across its diameter, so that it may be exactly matched to the uniformly bright image of the furnace opening.

Lambert's law also describes the light which is diffusely reflected from a *matte* surface, that is, from a surface which is smooth to the touch but is so rough optically that no regular reflection may occur. A flat surface "smoked" with magnesium oxide from burning magnesium metal is almost ideal in this respect.

PROBLEMS

1. If the coiled filament of a tungsten-filament lamp is examined, it will be seen that the inside of the coils is considerably brighter than the outside. Explain this.

2. The absorption coefficient for a certain kind of surface is 60 per cent. If light is reflected from three such surfaces in succession, what percentage of the incident light is reflected by the last surface? What then is the effective absorption coefficient for the three surfaces together?

3. Window glass is transparent to most of sunlight, but opaque to longer waves. Explain why a glass cover for the gardener's "cold-frame" in which seeds are planted keeps the ground warm. Does this violate Kirchhoff's law in any way?

4. Discuss the Biblical story of Shadrach, Meshach, and Abednego (Daniel iii) from the point of view of the radiation laws.

5. Assuming black-body radiation, compute the horsepower radiated from 1 cm^2 of the crater of a carbon arc operated at 22 atmospheres pressure (see Sec. 203). This power must be supplied to the arc as electrical power.

6. An incandescent lamp filament is 20 cm long and 0.030 mm in diameter. Compute the power in watts required to operate it at 3000° K. Assume the emissivity of tungsten to be 0.21.

7. Compute the temperature at which the peak of the radiation curve (Fig. 143) falls at 5000A.

REFERENCES

FOR ADDITIONAL READING

EDWIN EDSER, *Heat*.
HARDY and PERRIN, *Principles of Optics*, Chapters VIII and IX.
PHYSICS STAFF, UNIVERSITY OF PITTSBURGH, *Atomic Physics*, Chapter IV.
J. M. CORK, *Heat*, Chapter VI.

FOR ADVANCED READING

HARNWELL and LIVINGOOD, *Experimental Atomic Physics*, Chapter II.
MAX BORN, *Atomic Physics*, Chapter 7.
MAX PLANCK, *The Theory of Heat Radiation*.

CHAPTER XV

RADIOACTIVITY

207. Alchemy. Whenever we have considered the structure of the atom in earlier chapters, we have been concerned with its outer structure only. We have been content to regard the nucleus as only a tiny particle possessing mass and an electric charge. This does not mean, however, that this nucleus is unimportant. Its mass represents practically all the mass of the atom, and its electric charge is the one single factor which determines all the chemical properties of the atom, and most of its physical properties. In a very real sense the nucleus *is* the atom, with the electron structure only a superficial appendage.

If then we can alter the electric charge on the nucleus, we can achieve the end sought vainly by the ancient alchemists; we can *transmute* one element into another. As everyone now knows, this is possible, and the next few chapters tell the story of how this has been accomplished.

208. Discovery of Radioactivity. The first transmutations to be discovered are those which take place spontaneously, giving rise to the phenomena of radioactivity. The discovery of radioactivity is another of the many discoveries which make the last decade of the nineteenth century so remarkable. In 1896 Becquerel discovered that certain salts of uranium were continuously emitting radiations capable of blackening a photographic plate, even when that plate was protected by black paper wrappings. Shortly thereafter it was discovered that a gold leaf electroscope would detect in a minute or less a quantity of radioactivity so small as to require days of exposure to produce a noticeable effect upon a photographic plate.

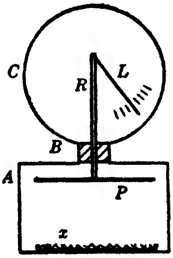


FIG. 146. ELECTROSCOPE AND IONIZATION CHAMBER TO MEASURE RADIOACTIVITY.

Figure 146 shows in cross-section an electroscope and ionization chamber suitable for the detection of radioactivity. The ionization chamber is a cylindrical metal box, A, with the electroscope case, C, mounted directly above it. The metal rod R, which passes through the insulating bushing of amber *B*, connects the circular collector plate *P* to the gold leaf *L*. The material under test is placed in a shallow metal tray on the floor of the ionization chamber (at *x*), and the collector plate *P* is charged until the gold leaf stands at a convenient angle. If the material is radioactive, the gold leaf will start to fall, and the rate at which it falls will measure the intensity of the radioactivity. The radiations from the radioactive material ionize some of the atoms of the gas within the chamber, producing equal numbers of both positively and negatively charged ions; and the flow of these ions to the plate *P* and to the walls and floor of *A* (the one half of them flowing one way and the oppositely charged half the other) constitutes an electrical current which gradually neutralizes the charge upon *P* and *L*. Greater sensitivity may be obtained by observing the gold leaf through a low-power microscope.

Using this sensitive means for the detection of radioactivity, Becquerel and the others who immediately took up investigation of this new discovery found that, whether the uranium was in the pure state or whether in chemical combination with other elements, the same mass of uranium had the same activity; and that no physical influences, such as change of temperature or exposure to light or to X-rays, could in any way alter its activity. That is, radioactivity is a property of the uranium atom itself (as we now realize, of its nucleus).

Much more powerful sources of radioactivity were soon discovered. Marie (Sklodowska) Curie who, with the aid of her husband, Pierre, had devoted herself to the study of these new phenomena from the time their discovery was first announced, found reason to suspect the existence in uranium ores of other radioactive elements. Inside of two years of vast labor, during which nearly a ton of uranium ore (from which the uranium itself had been extracted) was worked over chemically, she and her husband discovered two new chemical elements: polonium (named for Mme. Curie's native country, Poland) and radium; and both of these elements were found to be thousands of times more radioactive than is uranium.

209. Nature of Radioactivity. Discovery of these powerfully radioactive elements made possible the study of the nature of the radiations themselves, and it was soon discovered that they are of several kinds. The first and simplest evidence for this is given by their absorption by matter. Even air absorbs them, as may be demonstrated by the following experiment. An ionization chamber is provided with a wire-gauze bottom, as shown in Fig. 147, the distance between the collector plate and this gauze being only a few millimeters. If a rather broad flat surface covered with a layer of fairly powerful radioactive material is placed beneath this ionization chamber, the radiations will pass through the gauze bottom, and the fall of the leaf will measure the ionization which they produce inside the chamber. As this source of radioactivity is lowered, the ionization will not greatly change until the source is several centimeters below the bottom of the chamber, and then it will suddenly drop nearly to zero for only a slight further lowering of the source.

This abrupt termination of most of the ionization indicates complete absorption in this thickness of air of one component of the radioactive radiations, and this component is designated as the *alpha* component or as alpha rays. Their range in air, or the thickness of air which just absorbs them, depends upon the radioactive element which is their source. For radium the range is about 34 mm, for polonium it is about 39 mm, and the greatest range known is less than 90 mm.

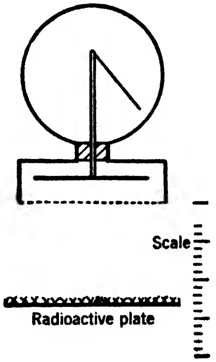


FIG. 147. APPARATUS TO MEASURE THE RANGE OF ALPHA RAYS.

If the source is polonium, it will be found that the ionization has ceased entirely when the alpha rays have been stopped; but for most radioactive substances there is still some ionization remaining, and if the experiment is continued, it will be found that the remaining radiations are much less easily absorbed. Even fairly thick metal foils will not completely absorb them, although they will reduce their intensity. The rate of absorption may be measured by interposing one foil on top of another between the source and the ionization chamber. When one or two millimeters of metal foils have been interposed in this manner, it will be found that only a very small effect is produced upon the electrometer.

A more sensitive detector, such as the Geiger-Müller counter described in Sec. 264, may now be required to show that any radiation at all is left. But this very small amount of radiation which does remain unabsorbed is found to be much more penetrating than either of the other two components, being able to pass through several centimeters of lead without being entirely absorbed. Thus it is revealed to be still a third component of the original radioactive radiations.

These three components are arbitrarily labeled with the first three letters of the Greek alphabet, in the order of their absorption. The most easily absorbed radiations are called *alpha* rays, as has already been stated; those of intermediate absorption rates are called *beta* rays; and the most penetrating are called *gamma* rays.

210. Alpha Rays. As has been stated above, the alpha rays are the most easily absorbed of all the radioactive radiations, and they are also the most powerful in their ionization effects. Indeed, these two properties go together, since the ionization produced is a direct result of the absorption. Their energy is so great that the effects of individual particles may easily be seen in the luminescence which they produce upon a screen of phosphorescent zinc sulfide. Polonium is a convenient source of alpha rays for this and other alpha-ray experiments; it is obtainable commercially in the form of a thin deposit over the surface of a copper plate. If such a polonium plate is held near a zinc sulfide screen and the screen is then viewed through a low-power microscope, something very much resembling a field of fireflies will be seen. Each individual alpha particle produces an individual flash of light. The experiment must of course be carried out in a very dark room, and the observer should remain in the dark for several minutes before attempting the experiment, to adapt his eyes to faint light.

The experiment may be performed still more simply by examining the figures on a luminous watch dial through a pocket magnifier, provided that the watch has been kept in total darkness for several hours previously to rid it of the phosphorescence which is induced in the zinc sulfide by exposure to light. Such luminous figures are painted with phosphorescent zinc sulfide with which is mixed a very small amount of radioactive material, and their luminosity is due to continuous bombardment of the zinc sulfide by the alpha-ray particles. The luminosity diminishes with time, partly because short-life radioactive materials may be used, but largely because of the destructive effect of the alpha-ray bombardment upon the zinc sulfide.

The individual effects of single alpha-ray particles are also seen in the Wilson cloud chamber described in Sec. 25. Each of the alpha-ray tracks seen in such a cloud chamber is a trail of fog which marks the passage of an individual alpha particle. This experiment also illustrates very beautifully the limited range of alpha particles. In a large cloud chamber the alpha-ray tracks spread out like the spokes of a fan, all of them terminating at practically the same distance from the source.

Alpha rays are deflected by very strong magnetic fields (it requires about 6600 gauss to deflect polonium alpha rays into a circle 1 meter in diameter), and the direction of the deflection shows that they are positively charged. Indeed, the charge on them has been experimentally determined by Geiger and Thomson, who measured the charge carried by an indirectly "counted" number of alpha-ray particles, and found this charge to be two positive elementary electric charges. The mass of the alpha particle then may be measured by the method of combined electric and magnetic deflections, just as masses were determined for electrons and for other electrically charged particles; it is found to be 6.6×10^{-24} gram, or 4.0 on the scale of atomic weights.

The evidence indicates that this alpha particle can be nothing else but the nucleus of a helium atom, and a very ingenious experiment was devised to demonstrate this directly. A small, very thin-walled glass tube containing a powerful source of alpha rays was sealed into an evacuated glass vessel, this outer vessel being provided with metal electrodes sealed through its walls so that an electric discharge might be passed into it. At the start of the experiment no discharge would pass, since the vacuum was very nearly perfect. The walls of the inner tube were thin enough, however, to be penetrated by alpha rays; after a sufficient number of alpha particles had passed through (a day or so later) an electric discharge could pass between the electrodes, and the spectrum of that discharge was the spectrum of helium gas. Although these alpha particles came from the interior of very heavy atoms, each one was a particle with two elementary positive charges and was therefore a helium nucleus. When an alpha particle has collected an "atmosphere" of two electrons and has become thereby a neutral atom, it is a helium atom.

As an alpha particle flies through space its positive charge pulls electrons off from the gas atoms through which it passes, leaving those atoms positively ionized. The electrons thus detached attach themselves to neutral atoms, forming negative ions. One hundred thousand or so of such pairs of ions are made by each alpha particle, and these are the ions which discharge the electroscope of the ionization chamber. These also are the ions upon which water vapor condenses in the Wilson cloud chamber apparatus described in Sec. 25. Each atom through which it passes slows the alpha particle down somewhat, and when all its kinetic energy is used up in ionizing atoms in this manner, it comes to rest. Hence, since all the alpha particles coming from the same element have practically the same range in air, they must *all have the same initial speed.*

211. Beta Rays. The direction in which beta rays are deflected by a strong magnetic field shows them to be negatively charged particles; and when both electric and magnetic deflecting fields are used together in the usual fashion, they are found to be very light particles traveling at very high speeds, some of them with speeds approaching that of light. Their masses are found in this manner to be much less than that of the lightest atom, but appreciably greater than that of the electron, those having the highest speeds having the largest masses. They are, as a matter of fact, electrons, and the amount by which their masses exceed that of the electron may be explained as due to their high speeds. This is not so fantastic an explanation as it may at first appear to be, since even for cathode-ray electrons the mass is found to increase with speed. Thus 5000-volt cathode-ray electrons have masses which are about 1 per cent higher than the value found for low-speed electrons, while 100,000-volt cathode-ray electrons are increased in mass by nearly 20 per cent. All such data for both high-speed cathode rays and beta rays are found to be represented quite exactly by the following formula:

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}} \tag{15·1}$$

in which m is the mass of the electron at the speed v, and m_0 is its mass at very low speeds, or its *rest-mass.*

212. Mass of Energy. This increase in mass may further be described as being just equal to the *kinetic energy* of the electron divided by c^2. For moderate values of the speed, values for which v^2 is very small as compared to c^2, this is easily demonstrated by rewriting equation (15·1) as shown below and then expanding by the binomial theorem of algebra

$$\begin{aligned} m &= m_0\left(1 - \frac{v^2}{c^2}\right)^{-\frac{1}{2}} \\ &= m_0\left(1 + \frac{v^2}{2c^2} + \frac{3v^4}{8c^4} + \cdots\right) \end{aligned} \tag{15·2}$$

If now the speed of the electron is not too great, all terms beyond the first two are negligible, and may be dropped. Then

$$m = m_0 + \frac{mv^2}{2c^2}$$

$$m = m_0 + \frac{\text{Kinetic energy}}{c^2} \tag{15·3}$$

At higher speeds this proof is not adequate, but at higher speeds the proper equation for computing kinetic energy is not $mv^2/2$, but is equation (15·4), which is given below, since the mass does not remain constant while the particle is gaining speed. It can be proved that equation (15·3) is exact for all speeds, and therefore the kinetic energy at any speed is

$$\text{Kinetic energy} = (m - m_0)c^2 \tag{15·4}$$

Since, according to equation (15 · 1), the mass of an electron approaches infinite mass as its speed approaches that of light, its speed can never become greater than the speed of light. At very high energies, its speed becomes almost equal to that of light; and any further increase of its energy will result in only an *increase in mass*, without any appreciable increase in speed. This is most strikingly illustrated by Professor Crane's very remarkable photograph, which is reproduced in Fig. 148.

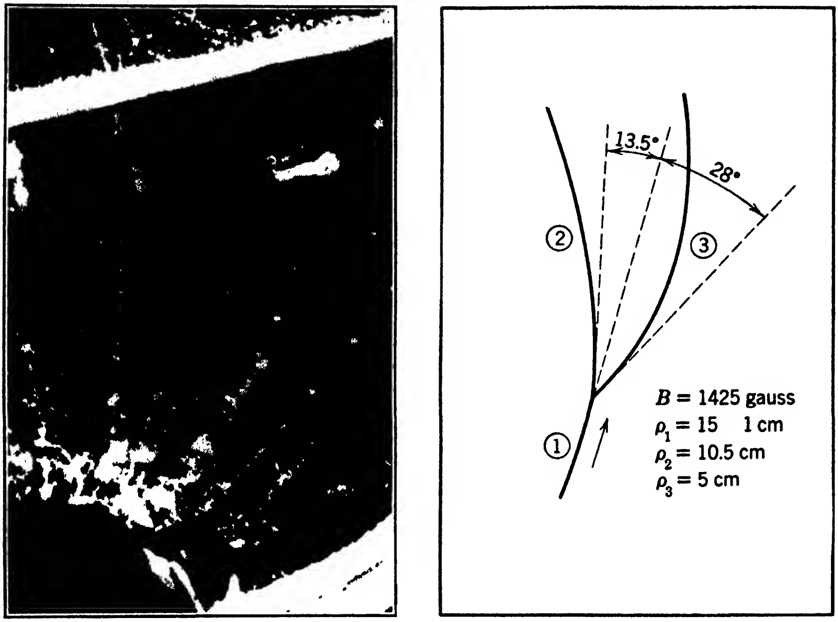


FIG. 148. COLLISION OF A HIGH-ENERGY ELECTRON WITH AN ELECTRON AT REST. Cloud-chamber photograph by Professor H. R. Crane of the University of Michigan. (Reproduced with the permission of Dr. Crane from *American Physics Teacher*, Vol. 6, p. 105, April, 1938.)

This photograph shows a chance collision between a high-energy electron and a stationary one, the collision[1] occurring in a Wilson cloud chamber in which the tracks of the electrons are made visible in the manner described in Sec. 25 for alpha-ray tracks. The electron tracks, which may be seen in the cloud chamber as white lines, have been reproduced by Professor Crane in the line drawing to the right of the photograph. Path (1) is that of the original high-energy electron, traveling in the direction of the arrow, while (2) and (3) are the paths of the two electrons after the collision.

---

[1] Only one photograph in over 10,000 showed such a collision.

The collision is an elastic one, in which no kinetic energy is lost. Had the colliding electrons been moving at ordinary speeds, their paths after collision should have been at right angles to each other, and the smallness of the angle (41.5 degrees) between paths (2) and (3) would seem to indicate that this collision was not an elastic one. The smallness of this angle may here be accounted for by the large changes in mass which result from collision between very high-energy electrons. As will presently be explained, the magnitude of the curvature which appears in path (1) shows that the speed of the primary electron was 99.69 per cent of the speed of light, and its mass was 12.7 times its rest-mass.

After the collision, both electrons were moving with very close to the speed of light, and their masses [as determined by the curvatures of paths (2) and (3)] were respectively 9.0 and 4.3 times their rest-masseses; i.e., the total mass, before the collision, was $12.7 + 1 = 13.7$ rest-masses; and after collision, $9.0 + 4.3 = 13.3$ rest-masses. The difference between these two values is 3 per cent, which is smaller than the errors involved in measuring the radii of the paths. Within the limits of experimental error, the total mass has remained unchanged; and this, according to equation (15·4), means that the total energy has not been changed, or that the collision is an elastic one.

The curvatures which appear in the electron tracks of this photograph were produced by placing the cloud chamber in a strong magnetic field (the lines of force being perpendicular to the face of the chamber) which deflected the electrons into curved paths. As is explained in Sec. 239, the momenta of these electrons may then be computed from the radii of curvature for their respective paths, using equation (17·15). Thus, for path (1) the momentum is

$$mv = \frac{Be\rho}{c} = \frac{1425 \times 4.80 \times 10^{-10} \times 15}{3 \times 10^{10}}$$

$$= 3.42 \times 10^{-16} \text{ gm} \times \text{cm per sec}$$

If we now assume, as a first approximation, that $v = c$,

$$m = \frac{3.42 \times 10^{-16}}{3 \times 10^{10}} = 1.14 \times 10^{-26} \text{ gm}$$

And since the rest-mass equals 9×10^{-28} gm,

$$m = 12.7 \text{ rest-masses}$$

If this mass value is now substituted into equation (15·1), it is found that $v = 99.69$ per cent of c, so that the approximation made in computing m is a justifiable one.

Strange as this idea of the mass of energy appears to be when it is first encountered, it is nevertheless an experimental fact which must be accepted as such. (It is also one of the mathematical consequences of the theory of relativity. This is explained in Appendix V.) We have already encountered this idea in Sec. 107 in connection with photons, and both for photons as there discussed, and for electrons as discussed above, it is found that

$$\text{Mass of energy} = \frac{\text{Energy}}{c^2} \qquad (15\cdot 5)$$

We shall see many other examples of this later on; the evidence is all to the effect that *all energy has mass.* The only reason this fact was not recognized earlier is that the mass of energy involved in all ordinary processes, even the most violent chemical processes, is too small to be detected in comparison with the other mass involved.

213. Gamma Rays. Gamma rays are not at all deflected by either magnetic or electric fields, no matter how intense the fields may be. They are very penetrating, even more so than X-rays. They ionize the air through which they pass to a degree comparable to the ionization produced by X-rays. And their wavelengths, some of which have been measured by crystal diffraction methods and some by other indirect methods, are shorter than those of X-rays. All of this evidence shows that gamma rays are electromagnetic waves (photons) of extremely high frequency.

214. Decay of Radioactivity. Even though the radioactivity of some materials, such as radium, appears to be quite permanent, all radioactivity becomes weaker as time goes on. This is quite apparent for polonium, as is shown by the curve plotted in Fig. 149. At the end of 136 days it has lost half of its activity, but at the end of another 136 days it has not lost the rest of its activity, but only half of what it started with at the beginning of the second period of 136 days. This gives the law of decay or loss of activity for polonium: *At the end of any period of* 136 *days the activity has dropped to one-half of its value at the beginning of that period.* Since infinite time would be required for it to lose all of its activity in this manner, its rate of decay is usually specified by its so-called "half-life," or this time of 136 days which is required for loss of half of its activity. The same law of decay is followed by the

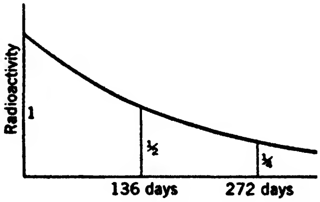


Fig. 149. Decay of Polonium Activity.

radioactivity of other substances, with different rates for each one. For some substances the rate of decay is so rapid as to be represented by a half-life of less than a second, while for others it is exceedingly slow. The half-life for radium is 1690 years, and for uranium it is 4.67 billion years.

215. Radon. Another example of radioactive decay is given by the radioactive gas, radon, which is found in radium compounds. It is an inert gas, like helium or neon, so that it always occurs in its free state. Radon loses its activity quite rapidly, its half-life being only 3.8 days. Since its half-life is so short, why has it not all disappeared long ago? The answer to this question may be found by placing some radium in a closed container connected to a vacuum pump and pumping out all of the radon. The radium will then be found to have a reduced alpha-ray activity, because part of its original measured activity was due to the radon which has been removed. In the course of time the radium will regain its original activity (in 3.8 days it will be half regained), and then more radon may be pumped off. *Evidently radon grows from radium!*

Very practical use is made of this fact, in that thin-walled glass or metal tubes ("seeds") filled with radon are used in place of radium for most therapeutic work. These "seeds" are inclosed in hollow steel needles which may be bandaged onto the diseased part or even inserted into the diseased tissue. The short life of the radon is safeguard against overexposure, and there is no danger of loss of valuable radium. The radium itself is kept in a sealed container, and the radon pumped off at regular intervals. Radon is just as good as radium for therapeutic uses, since the radiations which are of therapeutic value are the gamma rays emitted by the short-lived disintegration products following radon in the radium "family" of elements. (See Sec. 216.)

216. Disintegration Theory. It is not hard to see what happens when radon grows from radium. Radium is the 88th element in the periodic table, and therefore has a nuclear charge of 88 positive elementary electric charges. Since the alpha particle which is emitted by radium must come from the radium nucleus, the residue of this nucleus is left with a charge of only 86. It is now the nucleus of an atom of radon! Transmutation of one element into another, once only a dream of the alchemist, is here a reality; radium has been *transmuted* into radon. This change may be represented by the following equation:

$${}_{88}\mathrm{Ra}^{226} = {}_{2}\alpha^{4} + {}_{86}\mathrm{Rn}^{222} \qquad (15\cdot 6)$$

The subscript written before the symbol for each element represents its atomic number (and nuclear charge); the superscript is its atomic weight to the nearest whole number. Observe that the sum of the superscripts is the same on either side of this equation, and the same holds for the subscripts.

Radon itself is transmuted, by the escape of another alpha particle, into still another radioactive element, radium A, which is an isotope of polonium. These and subsequent transformations are shown in Table III. After radon all the elements in this series are solids, being radioactive isotopes of lead, bismuth, and polonium.

As is shown in this table, the emission of beta rays is an entirely separate process from the emission of alpha rays and occurs from different elements. Beta-ray emission also changes the nuclear charge and hence transmutes the element, *although it does not change its mass.* Beta-ray emission *increases* the atomic number by 1, as may be seen in Table I, since the beta-ray electron carries a unit negative charge.

TABLE III

THE RADIUM "FAMILY" OF RADIOACTIVE ELEMENTS

Element	Atomic Weight	Atomic Number	Emission	Chemical Name	Half-Life
Radium	226	88	Alpha	Radium	1690 years
Radon	222	86	Alpha	Radon	3.85 days
Radium A	218	84	Alpha	Polonium	3.0 min
Radium B	214	82	Beta	Lead	26.8 min
Radium C	214	83	Beta *	Bismuth	19.5 min
Radium C′	214	84	Alpha	Polonium	10^{-6} sec
Radium D	210	82	Beta	Lead	16.5 years
Radium E	210	83	Beta	Bismuth	5.0 days
Radium F	210	84	Alpha	Polonium	136 days
Radium G	206	82	None	Lead	....

* A very small percentage of the Ra-C atoms disintegrate differently, giving out alpha particles, and then get back to Ra-D by beta emission.

Gamma-ray emissions are not mentioned in this table, since the experiments show that gamma rays are not concerned with the transmutation processes but rather with the settling down of the new nucleus from the highly excited state in which it first finds itself to the normal state. Sometimes, as in the disintegration of polonium (radium F), the new atom is left in a normal state, and gamma rays are not emitted. This emission of gamma rays by the nucleus may be compared to the emission of light and of X-rays by the outer, electronic structure of the atom when this part of the atom settles down from an excited to a more stable state. The gamma-ray spectrum is found to consist at least in part of what may be called "bright lines" by comparison to optical spectra, and from the frequencies of these lines may be worked out sets of energy levels for the particles which constitute the nucleus, the procedure being similar to that used with the optical spectra to find the energy levels for the electronic structure of the atom.

In any specimen of radium there will thus be present traces of all the elements which are the products of subsequent disintegrations, and this explains the emission of beta and gamma rays, as well as alpha rays, by such a specimen of radium. Once this explanation has been worked out, the phenomena are quite easy to comprehend; but it was quite a different matter to start with a specimen of radium emitting all three radiations and to work out this explanation. The story of how this analysis was made is one of the stirring chapters in the history of scientific discovery.

Several other radioactive "families" have been discovered. Thorium and actinium each gives rise to such a sequence of radioactive elements, and the radium family is part of a longer sequence which begins with uranium. The actinium family is also derived from uranium, through an alternate branching in that sequence.

217. Isotopes. All these radioactive sequences terminate in inactive lead. For the radium sequence this lead has a mass number of 206; for the actinium sequence the mass number is 207; and for the thorium sequence, 208. In addition to these stable isotopes of lead there are a number of radioactive lead isotopes, as well as quite a number of radioactive isotopes of bismuth, polonium, radon, etc. Nowadays we have ample evidence (see Sec. 23) for the existence of isotopes, and we know that every element has several isotopes. But all this was unknown at the time of the discovery of radioactivity, and one of the outstanding achievements of the explorations of the radioactive families by Curie, Rutherford, Soddy, and others was the discovery of the existence of isotopes. For some time before that the chemists had been puzzled by the discordant results obtained for the atomic weights of lead samples coming from different localities. The discovery by Rutherford and Soddy of the existence of these lead isotopes entirely explained these discrepancies. The different samples of lead were of different radioactive origins!

218. The Age of the Earth. The discovery of radioactivity and its explanation in terms of nuclear disintegrations have provided what is at present the most reliable time-scale for geologic time. The age (since its solidification) of any uranium-bearing rock can be now determined with some certainty from its chemical analysis. As we find it today, any specimen of uranium ore contains not only uranium, but also a considerable amount of lead. If no lead was present in the ore at the time when it first solidified, then this lead which is found in it today must be the result of disintegrations of uranium nuclei, and the quantity of lead now present must be proportional to the age of the ore, from the time at which it solidified up to the present time. Since the rate at which uranium disintegrates is known from experimental measurements, this age may then be computed from the ratio of lead to uranium in the ore. Only the rate of disintegration for uranium need be considered, since the subsequent disintegrations take place at much greater rates, and the sequence of changes between uranium and lead is complete in a time that is short as compared to the age of the ore. In this manner it is possible to compute, for rocks in which uranium is found, ages which vary from about one hundred million years to nearly two billion years.

More precise use of this method requires consideration of the lead that is formed by disintegration of the thorium which may also be present in the ore. Furthermore, there is always the possibility that a small amount of lead was present in the ore at the time that it solidified. The mass spectrograph provides the most satisfactory means for meeting both of these difficulties, since it measures separately the amounts of the three lead isotopes. And since, as has been explained in Sec. 217, each of these isotopes represents a different sequence of radioactive disintegrations, three values for the age of the mineral may then be computed: one from the lead (isotope 208) to thorium ratio, another from the lead (206) to uranium ratio, and a third from the lead (207) to uranium ratio. Inconsistencies in these three values may be interpreted as due to the presence of some lead in the ore at the time of its solidification, and the amount of such primitive lead may then be computed so as to make the ratios agree with one another.

Still another means for computing the age of uranium-bearing rocks is provided by the helium which may be collected when a specimen of the rock is melted in a vacuum furnace. This helium represents the alpha particles which have been ejected in the various radioactive disintegrations over millions of years, and which, after becoming neutral helium atoms, have remained trapped in the rock until it is melted. The quantity of this helium is, of course, proportional to the age of the rock, which may then be computed from the helium to uranium ratio.

All these methods are in agreement as to the *order of magnitude* of the values which they give for the geologic ages represented by the uranium ores examined, but there is still much more work to be done before it can be said with any certainty how much more accurate than this the values are. From the present results it may be estimated that the age of the earth, since its crust was formed, is in the neighborhood of two billion years.

PROBLEMS

1. Compute the number of alpha particles escaping from 1 mg of radium in 1 sec if radium disintegrates at the rate of 0.044 per cent per year.

2. An alpha particle is ejected from radium with an initial speed of 1.6×10^9 cm/sec, and it travels 4.1 cm through air before coming to rest. Compute the average retarding force acting upon it. What do you think is the cause of this force?

3. Compute the kinetic energy of an electron traveling with a speed of 6×10^9 cm/sec. Compute the mass of this electron when moving at this speed.

4. Compute the mass in grams of the kinetic energy of a 10-ton air liner traveling at 250 miles per hour.

5. How fast must an electron move to triple its mass?

6. Prove that the momentum of a charged particle is proportional to the radius of the curved path which it follows in a magnetic field.

7. Compute the masses and the speeds of the electrons whose paths appear in Fig. 148.

8. Write the symbolic equation for the disintegration of (*a*) radon, (*b*) radium B.

9. A radioactive atom ejects an alpha particle. What are the atomic weight and atomic number of the new atom, if the original atom had atomic weight 218 and atomic number 84?

REFERENCES

For Additional Reading

PHYSICS STAFF, UNIVERSITY OF PITTSBURGH, *Atomic Physics*, Chapter XII.
J. A. ELDRIDGE, *The Physical Basis of Things*, Chapter XXIV.
J. B. HOAG, *Electron and Nuclear Physics*, Chapters XI and XII.
G. F. HULL, *Modern Physics*, Chapter XI.
G. E. M. JAUNCEY, *Modern Physics*, Chapter XVI.
G. E. M. JAUNCEY, "Early Years of Radioactivity," *American Journal of Physics*, Vol. 14, 1946, pp. 226–241.
J. D. STRANATHAN, *The "Particles" of Modern Physics*, Chapter 8.

For Advanced Reading

SIR ERNEST RUTHERFORD, *Radioactive Transformations*.
E. RUTHERFORD, C. D. ELLIS, and J. CHADWICK, *Radiations from Radioactive Substances*.
HARNWELL and LIVINGOOD, *Experimental Atomic Physics*, Chapter X.
A. HOLMES, *The Age of the Earth*.

CHAPTER XVI

NUCLEAR DISCOVERY

219. Nuclear Structure. Once it had been found that the heavy, radioactive nuclei are complex structures, capable of breaking apart to form new nuclei, it seemed reasonable to expect that *all* nuclei, except that of hydrogen, also are built up of still more elementary particles.

These ideas are new ones only in that they relate to the structure of *nuclei* rather than *atoms.* In the early days of chemical theory the hypothesis was advanced by Prout that all atoms might be made up of a single kind of building block, differing among themselves only in the number of such blocks in each atom. The chief support for this hypothesis was given by the values of atomic weights as they were then known. At that time it appeared possible that all atomic weights were integral numbers, and Prout assumed that the atomic weight for each atom represented the number of blocks contained in that atom. When better measurements showed that the chemical atomic weights for many atoms are far from being integers, this hypothesis was apparently discredited. Two discoveries of modern physics have inspired its revival. One is the discovery of radioactive transformations, as mentioned above. The other is the discovery that every element has several isotopes, and that all isotopes have atomic weights very close to integers. Now, of course, it is the *nucleus* whose composition is to be explained, and it is necessary to account for the electric charge on this nucleus, as well as its mass. To satisfy both of these conditions now requires two elementary particles.

220. The Proton-Electron Nucleus. One of these elementary particles must be the nucleus of the hydrogen atom, which we now call the **proton.** At first the electron was assumed to be the other particle. According to this hypothesis, any nucleus was built up by first packing together enough protons to give it the required mass; then, since this would give it too great an electric charge, enough electrons were tucked in to bring the nuclear charge down to the correct value. For example, the fluorine nucleus was thus considered to be made up of 19 protons and 10 electrons, to give it an atomic weight of 19 and a nuclear charge of $+9e$. In general, a nucleus of atomic weight A and atomic number

Z (nuclear charge, $+Ze$) was considered to contain A protons and $(A - Z)$ electrons.

This assumes that A is an integer. When the atomic weight is not exactly an integer, the nearest integer is taken as A, and is called the *mass number*. Since the atomic weights of isotopes do differ from integers by small but measurable amounts, the difficulty which condemned Prout's hypothesis still exists, although it is greatly reduced. The deviations from integers are now small, but it still is necessary to account for them. This can be done, as will be explained later on in Sec. 246, and this proton-electron "model" for the nucleus was given tentative acceptance for a considerable number of years. During this time the only other objections were certain theoretical ones, chiefly with respect to confinement in the nucleus of so light a particle as an electron. At first these appeared trivial. Later, however, they proved to be of vital importance and ultimately caused this model to be abandoned in favor of the far more satisfactory proton-neutron model described in Sec. 228.

221. Rutherford's Experiments. The first direct evidence for the existence of protons in the nuclei of atoms other than hydrogen was obtained by Rutherford in 1919. While carrying out his scattering experiments (see Sec. 27), he had discovered that, when alpha particles are shot into paraffin or other substances which contain many hydrogen atoms, they knocked out of such substances particles which sometimes had over four times the range of the original alpha particles. These particles were indeed hydrogen nuclei, or protons, knocked out of the hydrogenous material by direct hits upon them of the alpha particles, and their greater range is a consequence of their smaller mass and lesser charge. Then in 1919 he found similar long-range particles coming from nitrogen gas which was being bombarded by alpha particles. Since in this case only heavy nitrogen atoms were present, he concluded that these long-range particles must be *protons which had been knocked out of the nitrogen nuclei.*

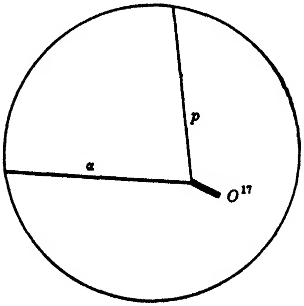


FIG. 150.

Six years later, confirming evidence was produced by Blackett, who found (out of thousands of pictures taken) eight cloud-chamber photographs which showed collisions between alpha particles and nitrogen atoms. Figure 150 shows diagrammatically such a cloud-chamber photograph. The long, thinly ionized track shows both by its length and its thinness that it is the trail left by the proton, while the short, thickly ionized track is that left by the new nucleus. After collision, the alpha particle has vanished, having been "captured" and made a part of the new nucleus. The original nucleus had a charge of +7 (elementary charges). The ejected proton carries away one of these charges, but the captured alpha particle brings in two more, so that the net charge of the new nucleus is +8. The new nucleus is that of an oxygen atom! This artificial transmutation may be represented by the following equation:

$$_2\mathrm{He}^4 + {_7\mathrm{N}^{14}} = {_8\mathrm{O}^{17}} + {_1\mathrm{H}^1} \qquad (16\cdot 1)$$

The subscripts represent the charges carried by each particle; the superscripts give the mass numbers (atomic weights in whole numbers). The new nucleus has a mass number of 17 and is hence a rare isotope of oxygen.

Rutherford found that alpha particles could in this way knock protons out of nearly all the first twenty elements of the periodic table. Beyond the twentieth element, even the most energetic of alpha particles, those from thorium C′, were ineffective. The repulsion of alpha particles by nuclear charges of greater than twenty elementary charges prevents even the thorium C′ alpha particles from approaching the nucleus closely enough to be captured by it. It seemed evident that more energetic particles were required, if further elements were to be transmuted. This need for more energetic particles, as well as the need for greater quantities of particles, spurred on many attempts to devise electrical means of producing such highly energetic particles.

222. Nuclear Energies. A better idea of the energies possessed by alpha-ray particles, and hence of the energies required to effect these nuclear transmutations, may be obtained by computing the potential differences which would be required to give the alpha particles these energies by electrical means. Since their masses and speeds are known from deflection experiments, this is easily done by means of the familiar relation

$$Vq = \tfrac{1}{2}mv^2 \qquad (16\cdot 2)$$

Here m is the mass of the alpha particle, v is its speed, and q is its charge, equal to two elementary charges. The required potential difference, V, is found from such computations to be several millions of volts for all cases, being as high as 4.38 million volts for the most energetic alpha particles, those from thorium C′.

Nuclear energies are most often expressed in millions of electron-volts. One **electron-volt** is defined as equal to the work done upon one elementary electric charge by a potential difference of one volt. It is usually abbreviated to ev. Its value in ergs is

$$1 \text{ ev} = \frac{1}{300} \times 4.80 \times 10^{-10} = 1.60 \times 10^{-12} \text{ erg} \qquad (16 \cdot 3)$$

One million electron-volts (abbreviated variously as mev, Mev, MEV, etc.) is evidently

$$1 \text{ Mev} = 1.60 \times 10^{-6} \text{ erg} \qquad (16.3a)$$

Since the alpha-ray particle carries two elementary electric charges, the energy of an alpha particle, in million electron-volts, is numerically equal to twice the potential difference required to give it that energy. A thorium C′ alpha particle, for example, has an initial energy of 2 × 4.38 or 8.76 Mev.

It is interesting to compare these alpha-ray energies with the energies involved in the production of X-rays. The hardest characteristic X-rays are the *K* X-rays of uranium, and these require for their production a potential difference of about 120,000 volts; in other words, the binding energy of the *K* electrons in the uranium atom is about 0.12 Mev. This is the largest amount of energy to be found in any atom outside of its nucleus, and it is an almost negligible amount in comparison with the quantities of energy involved in most nuclear processes.

223. The Positron. Up to the year 1932 no evidence had been found for the existence of more than two fundamental particles of matter, the proton and the electron. Of these, the heavy one, the proton, carries the elementary positive charge, and the light one, the electron, the elementary negative charge. It seemed conceivable that all aspects and properties of matter could be explained in terms of these two fundamental particles. This simple and comfortable situation was rudely destroyed in 1932, by the discovery of two new elementary particles.

The positron is one of these particles. While investigating cosmic rays with the aid of a large cloud chamber placed in a strong magnetic field, Anderson, in August, 1932, obtained several photographs of a most remarkable kind. The first one of these is reproduced in Fig. 151. The direction of curvature of the trail seen in this photograph indicated that the particle producing it was either a negatively charged particle entering the cloud chamber from above, or a positively charged one coming from below. Which of these two directions was taken by the particle was made evident by means of the thick lead plate which is placed edgewise across the cloud chamber for just this purpose. (This plate is seen as a horizontal bar across the middle of Fig. 151.) The trail is less curved below the lead plate than above it; this makes it quite evident that the particle entered the cloud chamber from below, losing momentum as it passed through the lead plate. It was then positively charged.

Everything else indicates that the mass of the particle was approximately the same as that of an electron having the same speed. Thus a heavier particle, with the momentum indicated by the curvature of this path, would have less energy and consequently be slowed up in the lead plate much more than this particle has been. Much more evidence has been accumulated since, and all of it shows that the **positron,** as Anderson named it, is a particle having essentially the same mass as the electron, but carrying a positive elementary charge.

FIG. 151. A POSITRON TRAIL IN A CLOUD CHAMBER. Cloud-chamber photograph by Dr. Carl D. Anderson of the California Institute of Technology. This is the first positron trail ever discovered.

Positrons are found to be quite as numerous as electrons in cosmic rays; hence it is most extraordinary that they never are observed under ordinary conditions. A clue to this strange difference is given by the occasional appearance in cloud chambers of pairs of trails such as shown in Fig. 152. Both trails start from the same point, in the same direction; furthermore, they have the same curvature, but in opposite directions. One is caused by an electron, the other by a positron of equal energy. Similar electron-positron pairs may be produced by shooting gamma rays of sufficient energy into a plate of lead or other dense material. It is now believed that both the electron and the positron are *simultaneously created*, at the point of common origin, *from the energy of the gamma ray or cosmic ray photon* which is absorbed at that point.

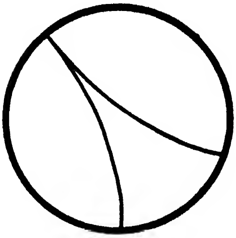


FIG. 152. ELECTRON-POSITRON PAIR.

224. Electron Creation. The possibility of the creation of electrons and of positrons from a sufficient amount of energy of some kind is now well established by many experiments. The minimum energy required may be computed from the now familiar mass-energy relation given in Sec. 212:

$$\text{Energy required} = \text{Mass} \times c^2$$

$$= 9.1 \times 10^{-28} \times 9 \times 10^{20} = 8.2 \times 10^{-7} \text{ erg} \qquad (16\cdot 4)$$

The production of an electron-positron pair by means of gamma rays, as mentioned in the preceding section, is an excellent example of the evidence for the creation of electrons and positrons from radiant energy. If both particles are created in a single process, a total of a little over one million electron-volts of energy is required. And it is found that gamma rays of less than this energy are indeed unable to produce such electron-positron pairs. When the gamma-ray energy exceeds the minimum requirement, this theory requires that the excess energy be added to the mass of the particles as kinetic energy. Again, the experiments are in good agreement with theory; the total energy of the created particles is found equal to the energy of the gamma-ray photons. Since much momentum is lost when a gamma-ray photon is transformed into an electron-positron pair, this process can take place only in the close neighborhood of a heavy atomic nucleus to which the extra momentum may be given.

The reverse phenomenon occurs when a positron vanishes. Gamma rays have been detected as emerging from a substance in which positrons are being absorbed, and the frequency of these gamma rays corresponds to about half a million electron-volts of energy, or just the energy equivalent of the positron mass! The positron disappears literally in a flash of light! The reason why the positron was not discovered sooner is explained by the fact that it appears only under the most extraordinary circumstances, and even then has only a brief existence.

Again it is interesting to note that these experimental results were predicted by theory. About 1930 Dirac developed a mathematical theory for the electron which at the same time predicted the existence of a particle such as the positron. This theory has since been shown (by Fermi and Uhlenbeck) to predict the creation of electron-positron pairs, and to account also for the brief existence of the positron by predicting a sort of "suicide pact" in which an electron and a positron might vanish together, their charges cancelling and their total energy (including their masses) being transformed into a pair of gamma-ray photons, emitted in opposite directions. When the annihilation process occurs in dense matter, this theory shows that the energy may also be emitted as a single gamma-ray photon of energy equal to the total energy of both particles, and this also has been discovered experimentally.

225. The Neutron. The other of the two particles to be discovered in 1932 is known as the neutron, and credit for its discovery is shared by five different people in three different countries. In 1930 two German physicists, Bothe and Becker, reported the discovery of a very penetrating radiation (capable of penetrating several centimeters of lead) which was emitted by beryllium when bombarded by alpha rays. They believed this to be some kind of gamma radiation. In January, 1932, two French physicists, Frederic Joliot and his wife, Irene Curie (who, it is interesting to note, is a daughter of the Curies who discovered radium) discovered that a layer of some material rich in hydrogen, such as paraffin, placed between the beryllium source of these rays and an ionization chamber, greatly increased the ionization produced therein. They then proved that this extra ionization was due to high-speed protons coming out of the paraffin, and this cast grave doubt upon the gamma-ray explanation of Bothe and Becker. Gamma-ray photons possessing as much as 55 Mev of energy would be required to account for the protons observed by the Joliots,[1] and this seemed to be much

---

[1] In a head-on collision between a moving particle and a stationary particle of equal mass, the moving particle gives most of its energy to the stationary one. (If the speeds are not too close to the speed of light, the stationary particle will take practically all of the energy. At the other extreme, when the speeds are practically those of light, the energy is divided fairly equally between the two. See Fig. 148.) But if the initially moving particle is much the lighter one, as when a photon bombards a proton, it rebounds without losing very much of its kinetic energy. Thus, to give the same energy to a proton, a photon must have initially much more energy than need be possessed by a neutron.

more energy than the other evidence indicated it was possible for them to have.

In the next month Chadwick, in England, gave the correct explanation. These rays, said Chadwick, consist of particles *of about the same mass as that of protons, but neutral particles, possessing no electric charge.* He named them **neutrons.** If neutrons have the same mass as have protons, then of course they need have no more energy than that possessed by the fastest moving of the ejected protons, in order to explain all the observed phenomena. Chadwick further substantiated this explanation by experiments in which nitrogen nuclei were similarly set into motion by neutron bombardment: if the speeds given to these nitrogen nuclei were to be explained by gamma rays, the gamma-ray photons would need to possess still higher energies than were required of them to explain the proton collisions. The required neutron energies, however, figured out to be the same both for protons and for nitrogen nuclei.

Neutrons result from numerous other nuclear reactions, but the beryllium-alpha-ray reaction is still one of the most effective of the known methods for producing them. A neutron source may be made by sealing beryllium powder into a glass tube with some radium emanation, and inclosing this tube in a lead box having walls heavy enough to stop the escape of any but the neutron radiations. The reaction which produces neutrons in this source may be represented by

$$_{4}\mathrm{Be}^{9} + {}_{2}\alpha^{4} = {}_{6}\mathrm{C}^{12} + {}_{0}n^{1} \qquad (16 \cdot 5)$$

A more powerful source of neutrons is obtained by bombarding deuterium compounds with deuterons, or deuterium nuclei, which have been given very high speeds by an ion accelerator such as a cyclotron. (See Chapter XVII.) The reaction in this case is

$$_{1}\mathrm{D}^{2} + {}_{1}\mathrm{D}^{2} = {}_{2}\mathrm{He}^{3} + {}_{0}n^{1} \qquad (16 \cdot 6)$$

226. Detection of Neutrons. Neutrons of themselves produce no ionization, since they carry no electric charge with which to pull electrons away from the gas atoms. The only evidence, indeed, for the existence of neutrons is supplied by the collisions of neutrons with atomic nuclei. This is exemplified by the protons which are knocked out of paraffin in the experiment of the Joliots which has been described. When the paraffin was not present, the electroscope should have shown no ionization at all taking place, if it had not been for the few similar secondary effects produced by collisions of neutrons with nuclei in the walls of the ionization chamber or, very rarely, in the gas of the chamber.

How infrequently this may occur is indicated by stating here that the mean free path of neutrons in lead is over five centimeters long!

Occasionally, such a rare collision between a neutron and the nucleus of a gas atom may occur within a cloud chamber, and the evidence for such a collision will then appear as random tracks such as those seen in Fig. 153. The neutron itself leaves no evidence of its passage, since it carries no charge with which to ionize the gas through which it passes. But the atomic nucleus, which is stripped of its electrons and set into rapid motion by the collision, leaves a visible trail such as those shown in Fig. 153.

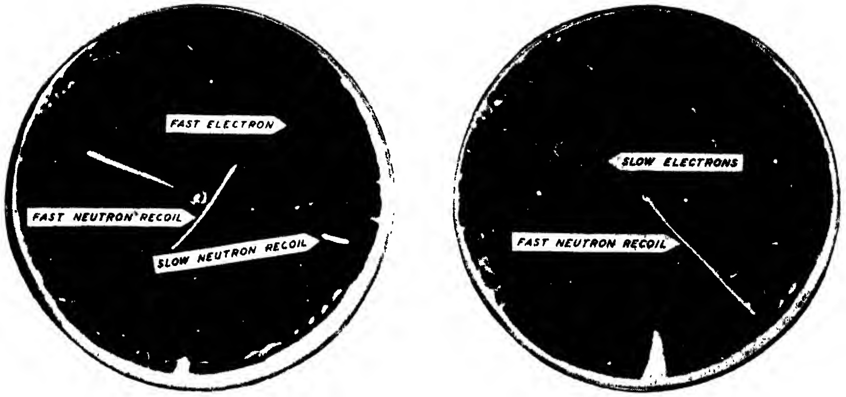


FIG. 153. NEUTRON RECOILS. Cloud-chamber photographs made by the Department of Terrestrial Magnetism of the Carnegie Institution of Washington, D. C.

A Geiger-Müller counter also may be used to detect neutrons, as explained in Sec. 264, by making its cylindrical electrode of some metal whose atoms emit electrons as a result of neutron capture.

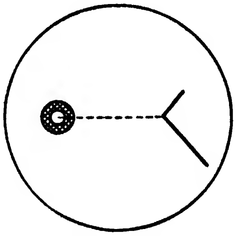


FIG. 154.

227. Neutron Transmutations. Sometimes, in a cloud chamber through which neutrons are passing, there appears a branched track such as is diagrammed in Fig. 154. If a dotted line is drawn, as shown, to represent the invisible path of the neutron from its source to the point of branching, it is seen that these tracks represent a neutron collision from which two charged particles emerge. Such a branched track was first observed by Feather, in 1932. The gas in his cloud chamber was nitrogen, and he proved that the collision had resulted in a nuclear transformation which may be represented by the following equation:

$$ {}_0n^1 + {}_7\mathrm{N}^{14} = {}_5\mathrm{B}^{11} + {}_2\mathrm{He}^4 \qquad (16\cdot7) $$

The neutron had collided with a nitrogen nucleus and had been captured by it; the resultant particle then broke down into a boron nucleus and an alpha particle. The boron nucleus left a short, thick trail of ionization, while the alpha particle left a longer, thinner trail. A good many similar nuclear transformations have since been observed. The key to the analysis of these reactions is given by the angles between the track lines, since from these angles the relative momenta of the two resultant particles may be computed, and their masses thereby verified. From these angles also the relative momentum of the neutron may be computed, and thence a fair value obtained for its mass. The best values for the neutron mass are obtained, however, by comparing the total mass, for all of the reacting particles, with the net gain (or loss) of kinetic energy by the particles, and doing this for a large number of nuclear reactions involving neutrons. In this manner a value of 1.00893 has been obtained for the "atomic weight" of the neutron.

Nuclear reactions which may be produced by neutrons are very numerous, and the usefulness of neutrons for this purpose is especially realized with the heavier elements, the ones which cannot be transmuted by means of alpha particles or other charged particles. The strong electric forces of repulsion, which limit the usefulness of charged particles for producing nuclear transformations, are not encountered by neutrons: neutrons are thus able to penetrate to the nuclei of the heaviest elements as well as of the lightest.

It is most interesting to find that slow-speed neutrons are frequently more effective than high-speed ones in effecting nuclear transmutations. Since they are not repelled by nuclei, they do not need any kinetic energy to penetrate into them; if they have too much kinetic energy, they are likely to bounce off again rather than be captured. Slow-speed neutrons may be obtained by surrounding the neutron source with a thick layer of hydrogen-rich substance, such as paraffin or water. Since neutrons and protons have essentially the same mass, each collision of a neutron with an hydrogen nucleus in the paraffin (or water) will on the average transfer about half of the energy of the neutron to the hydrogen nucleus. Thus, after twenty or thirty collisions the neutrons will have about the same average speed as have the atoms in the paraffin; that is, they will be approximately in thermal equilibrium with those atoms.

Ultimately, of course, each neutron will be captured by some nucleus with which it comes into collision. And because slow neutrons are more easily captured than fast ones, paraffin or water screens several feet in thickness make the most practical shields against unwanted neutrons. Large quantities of neutrons are produced by the operation of cyclotrons and Van de Graaf generators, so that the other apparatus in the laboratory, as well as the operators, must be protected against such neutrons by "neutron dunkers," as they have been familiarly called, which consist of thick tanks of water surrounding the parts of the laboratory in which the neutrons originate.

228. The Proton-Neutron Nucleus. Although discovery of the neutron and the positron upset the simplicity of the proton-electron concepts for the structure of matter, these discoveries have made possible the development of far more satisfactory concepts. Now it is possible to make a "model" for the nucleus, with protons and neutrons alone, which meets all the objections raised against earlier models. In this model the total number of nuclear particles of both kinds equals A, the mass number of the nucleus. The number of protons equals Z, the atomic number, and the remaining $(A-Z)$ particles are neutrons. This is the nuclear model accepted at the present time. As a first approximation, and probably a very good one for the larger nuclei, we may imagine these particles as being packed closely together so as to resemble closely a droplet of liquid. This makes the volume of a nucleus proportional to the number of nuclear particles contained in it, or the diameter proportional to the cube root of this number, and this is in good agreement with experiment. Thus the diameter of a single proton or neutron is about 3×10^{-13} cm, and that of the lead nucleus, which contains over 200 nuclear particles, is only 5 or 6 times larger. It follows that the average density is about the same for all nuclei, and has the incredibly great value of about one-quarter *billion tons* per cubic centimeter!

The nature of the forces holding these nuclear particles together is as yet unknown, although various hypotheses (see Sec. 273) have been advanced to account for them. They are so intense that the electrical forces of repulsion between protons are almost negligible in comparison, but they are very short-range forces which are effective only when the particles are close together. In this respect also nuclear forces resemble the forces which hold a liquid droplet together, even to the extent of being most apparent towards the surface of the nucleus, in a manner quite analogous to surface tension in a liquid droplet. It may be that most of the neutrons and protons are grouped together into alpha particles, as seems quite possible, but this is not at all certain.

In other respects this proton-neutron "model" does not differ greatly from the proton-electron "model" discussed in Sec. 220, and discrepancies in mass may be explained in the same manner for both models. This will be done in Sec. 246.

229. "Induced" Radioactivity. In January of 1934 the Joliots (the same Joliots who assisted in the discovery of the neutron) discovered that aluminum foil, after exposure for ten minutes or so to the alpha rays from a powerful polonium source, emitted radiations after the polonium had been removed. *The aluminum foil had been rendered radioactive* by exposure to the alpha rays, with a "half-life" of about three minutes. By ingenious chemical and physical tests they showed that the alpha rays had produced the following nuclear reaction:

$$ {}_{13}\mathrm{Al}^{27} + {}_{2}\mathrm{He}^{4} = {}_{15}\mathrm{P}^{30} + {}_{0}n^{1} \qquad (16\cdot 8)$$

The radioactivity came from the phosphorus nucleus which, after a time, emitted a positron and became a stable silicon nucleus of mass number 30. Continuing these investigations, they discovered that alpha rays would similarly produce a considerable number of other radioactive elements. Very shortly after this, Cockcroft and Walton, with their pioneer high-voltage apparatus, discovered cases of induced radioactivity produced by proton bombardment; and before the middle of the year 1934 Fermi, in Italy, had discovered many new radioactive elements which could be produced by neutron capture.

It has now been demonstrated that every known element may be represented by one or more radioactive isotopes! Various means are effective in the production of such radioactive isotopes, as well as of many new stable isotopes: bombardment by alpha rays emitted by naturally radioactive substances; bombardment by electrically energized alpha particles, protons, and deuterons; capture of neutrons; and even irradiation by very hard gamma rays. Moreover, it is often possible to produce the same isotope in several ways.

Only the very heavy, naturally radioactive elements emit alpha rays.[2] For the artificially produced radioactive elements the emissions are either electrons or positrons, the nature of the emission being determined by the direction in which the change in nuclear charge need occur in order to arrive at a stable isotope. For example, nitrogen possesses four isotopes, with mass numbers 13, 14, 15, and 16; and of these four the lightest (13) and the heaviest (16) are radioactive. Isotope 13 emits a positron and thereby becomes a stable isotope of carbon, the next lower element in the periodic table, while isotope 16 becomes a stable isotope of oxygen, the element next higher in the periodic table, by emitting an electron. This is fairly typical of the lighter elements.

---

[2] An apparent exception to this rule is the radioactive isotope of lithium of mass number 8. Lithium 8 emits an electron, and should then become beryllium 8. Instead, two alpha particles appear. The same thing happened in the famous experiment of Cockcroft and Walton, described in Sec. 234, when a proton was captured by lithium 7. In each case it appears that, for some reason, beryllium 8 cannot exist but breaks down immediately into two alpha particles. See Sec. 248.

For the heavier elements more radioactive isotopes exist, and several electron or positron emissions may occur, one following the other, before a stable isotope results.

More recently it has been discovered that an unstable nucleus may sometimes depart from the usual procedure of emitting a positron and capture a K electron instead. Evidence for this is given by the emission of the K X-rays of the stable isotope which results from this process.

230. Emission Theory. A successful nuclear model must be able to account for all the emissions which come out of it, either radioactively or as the result of nuclear reaction; these now include protons, neutrons, alpha particles, electrons, positrons, and gamma rays. It is of course quite obvious that protons and neutrons may be emitted from the proton-neutron nucleus. Likewise it is quite easy, qualitatively, to account for the emission of an alpha particle; either the component protons and neutrons may be combined to form an alpha particle at the instant of emission, or, as seems more probable, some of the nuclear protons and neutrons may exist together at all times as alpha particles within the nucleus. A quantitative explanation for the radioactive emission of alpha particles is given in Appendix VI.

The emission of gamma rays also is easily understood. Since gamma rays are photons, they may be *created* out of energy lost by the nucleus, as its component particles readjust themselves from less stable (or "excited") configurations to more stable ones. The discrete line spectra formed by gamma rays from any nucleus show that the energy of the nucleus is "quantized" into energy levels in a manner comparable with those for the electron structure of an atom or molecule.

The emission of electrons and positrons, however, is not so simple a process, although it follows somewhat closely the pattern set by the gamma rays. Since neither electrons nor positrons exist within the nucleus, the radioactive emission of these particles must involve their *creation*, within the nucleus, at the time of emission. It is not so long ago that such an idea would have been considered preposterous; now we know, from experiments such as those described in Sec. 224, that this is entirely possible. The *masses* of these emitted particles are obtained by conversion of energy into mass, just as for gamma-ray photons. These particles differ from gamma-ray photons, however, in that they carry *electric charges*; when they are created they must somehow be supplied with these charges as well. It now seems well established that these charges are taken from one or the other of the nuclear particles. Thus a positron obtains its charge from a proton, *which then becomes a neutron.* Likewise, when an electron is emitted, *a neutron becomes a proton.* According to this theory, indeed, the proton and the neutron are essentially two phases of the same particle, to which the name **nucleon** may be applied.

231. The Neutrino. Still another elementary particle seems needed to explain a very serious discrepancy between theory and experiment in the emission of beta-ray electrons and positrons. Theory indicates that beta rays for any one element should all have the same speed, whereas experiment shows that they have all possible speeds between the theoretical value and zero. The question is, what has become of the rest of the energy when a beta ray is emitted with a speed lower than the maximum? The most likely explanation makes a scapegoat of this additional elementary particle, to which the name **neutrino** is given. To this particle theory assigns a very small mass, and no charge.

To prove this theory by finding the neutrino experimentally is another matter entirely. When it is considered that even the neutron, thousands of times more massive than the neutrino, can slip through several inches of dense matter before making its presence known, it is not unreasonable to expect that so light a particle as the neutrino may never be detectable. Nevertheless it is now possible to detect it by experiments which measure, directly or indirectly, the momenta of all charged particles involved in a nuclear process which is suspected of involving a neutrino. If the vectors representing these momenta do not add up to zero, then the difference represents the momentum of the neutrino. Several such experiments now have been carried out.

One of the first of these was by Crane and Halpern,[3] who studied the radioactive disintegration of chlorine 38, which becomes argon 38 upon the emission of an electron. From cloud-chamber photographs it was possible to determine the momentum of the emitted electron, and to obtain a rough estimate of the momentum of the recoil of the argon nucleus. Whenever the electron momentum was small, these data indicated quite definitely that the recoil momentum was greater than that of the electron. The difference must have been carried away by some uncharged particle, presumably a neutrino. Jacobson, in Denmark, has obtained evidence to show that the neutrino and the electron are ejected together, in the same direction.

A reaction which produces still more positive results is the spontaneous change of a beryllium-7 nucleus into a lithium-7 nucleus by capture of one of its own extra-nuclear *K* electrons, without the emission of any observable particle, not even a gamma ray. In this process also the nucleus is set into motion, motion which can be explained only by the emission of a neutrino. Allen[4] has measured this recoil momen-

---

[3] Crane and Halpern, *Physical Review*, Vol. 53, p. 789, 1938.

[4] Allen, *Physical Review*, Vol. 61, p. 692, 1942.

tum and finds results which agree with theory, to within the limits of error in the experiment. The measurements are not precise enough to determine the rest-mass of the neutrino, except to show that it is very small, probably much smaller than that of the electron. Some theories assume it to be zero.

A third reaction which may give even better data is the disintegration of lithium 8. According to a theory developed by Gamow and Teller, when lithium 8 breaks down into two alpha particles, it emits also an electron and a neutrino. This will explain experimental observations that the alpha particles have energies varying from a maximum value down to zero; when their energies are less than the maximum, the difference is carried away by the electron and the neutrino. Rumbaugh, Roberts, and Hafstad,[5] who measured this distribution of energy, find it in good agreement with the theory of Gamow and Teller.

The existence of the neutrino seems well established by these and other similar experiments, although more and better data are needed to determine its rest-mass.

232. The Meson. In Chapter XIX will be described the discovery in cosmic rays of still other elementary particles, called sometimes **mesons** and sometimes **mesotrons.** Mesons are lighter than protons but heavier than electrons; the ones which are observed carry an elementary charge of either sign. Uncharged mesons may exist also. Mesons, like positrons, are very short-lived particles. Their other properties will be discussed in Chapter XIX.

This completes the list of known elementary particles: protons, neutrons, electrons, positrons, mesons, and neutrinos.

PROBLEMS

1. Compute the wavelength of a 2-Mev gamma ray.

2. The gamma ray specified in problem 1 creates an electron-positron pair. (*a*) Compute the kinetic energy of each of these particles. (*b*) Compute the mass of each.

3. When a moving particle strikes a stationary one head-on in a perfectly elastic collision, the speed given to the second particle is $2m_1/(m_1 + m_2)$ times the original speed of the first one (unless the speeds are close to the speed of light). (*a*) Compute the speed given to a proton by a neutron under these conditions, if the original speed of the neutron is 10^4 kilometers per second. (*b*) Compute the speed given to a nitrogen nucleus by a neutron having the same initial speed as in (*a*). (*c*) Show by computation with your numerical values that both energy and momentum are conserved, for both parts (*a*) and (*b*).

4. Using equation (7·8), compute the wavelength and hence the energy of a photon which will give (*a*) a proton the energy computed for it in problem 3(*a*); (*b*) a

---

[5] Rumbaugh, Roberts, and Hafstad, *Physical Review*, Vol. 54, p. 657, 1938.

nitrogen nucleus the energy found in problem 3(*b*). Show that the approximation made in equation (7·8) is justified with gamma rays of these energies.

5. Determine the numbers of protons and neutrons in each of the following nuclei: (*a*) He^3, (*b*) Na^{23}, (*c*) Cd^{112}, (*d*) U^{238}. (See Appendix III.)

Write the symbolic equations for each of the following nuclear reactions, and identify the new nucleus in each case. (See Appendix III.)

6. Fluorine 19 bombarded by alpha particles. Proton emitted.

7. Sodium 23 bombarded by alpha particles. Neutron emitted.

8. Chlorine 35 bombarded by neutrons. Alpha rays emitted.

9. Calcium 42 bombarded by neutrons. Protons emitted.

10. The nitrogen isotope of mass number 16 is radioactive. What does it emit, and what stable nucleus results from the emission?

11. Sodium 22 is radioactive. What does it emit, and what is the name of the new stable nucleus?

12. When silver 109 captures a neutron, it becomes radioactive. Name the stable isotope which results, and describe the nuclear reactions involved.

REFERENCES

See the list of references at the end of Chapter XVII.

CHAPTER XVII

NUCLEAR RESEARCH

233. Nuclear Research. In the preceding chapter we have considered all the basic discoveries of modern nuclear physics, together with the theories for nuclear structure and nuclear forces which may be deduced from these discoveries. It is most remarkable that all these discoveries were made by means of only *natural* sources for high-energy particles, such as natural radioactivity and cosmic rays, and that the theories deduced from such data differ only in detail from those held today. Nevertheless, the use of natural sources alone imposes severe limitations upon nuclear investigations. The particles available from radioactive sources are relatively meager in number, and limited in energy to a few million electron-volts. Cosmic-ray particles may possess far greater energies, but they are even fewer in number and far more difficult to utilize. Another difficulty with particles from natural sources is the uncertainty of their energies. Precise measurements are possible only when the energies of the bombarding particles may be determined precisely and controlled to any desired value.

For these several reasons the scientists engaged in nuclear research have developed various kinds of powerful electrical machines for accelerating electrically charged particles to very high energies. As a matter of fact, these developments were being carried on concurrently with many of the early discoveries, but their consideration is postponed to this chapter, since their principal importance is to nuclear research.

234. Transmutation by Electrical Means. The first success in the use of electrically accelerated ions to produce nuclear disintegrations was obtained in 1932. Cockcroft and Walton, working at Cambridge University, England, used a potential difference of 500,000 volts to shoot hydrogen ions, or protons, against a lithium target, and the result was the emission from the target of alpha particles nearly as energetic as those emitted by thorium C′. Later they found that even lower voltages would accomplish the same result. They interpreted this result as meaning that the capture of a proton by a lithium nucleus produced an unstable nucleus which immediately broke down with explosive violence into two helium nuclei, or alpha particles. This nuclear reaction is represented by the following equation:

$$ {}_1\mathrm{H}^1 + {}_3\mathrm{Li}^7 = {}_2\alpha^4 + {}_2\alpha^4 \qquad (17\cdot 1) $$

The correctness of this assumption was completely verified by cloud-chamber photographs which showed two alpha particles shooting out in almost exactly opposite directions, with equal ranges.

235. The Van de Graaf Electrostatic Machine. The high-voltage source employed by Cockcroft and Walton was an ingenious combination of transformer, condensers, and vacuum-tube rectifiers, operating in essentially the same manner as the circuit of Fig. 9. The practical limit for such devices is not much over one million volts. To attain higher voltages some other means must be employed, and the first electrical machine to succeed in this was invented by Van de Graaf, working at the Massachusetts Institute of Technology.

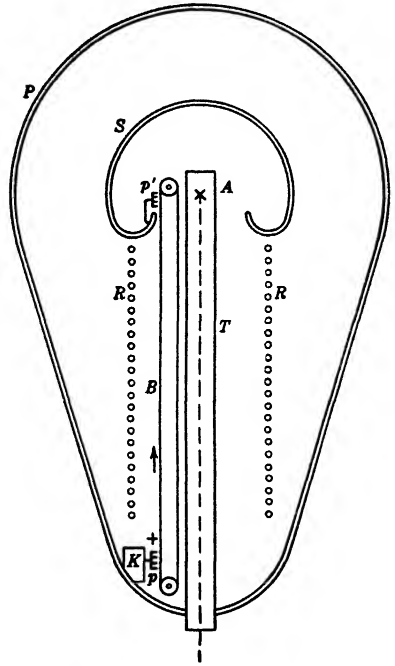


FIG. 155. VAN DE GRAAF ELECTRIC MACHINE. Schematic diagram, not to scale, of the machine shown in the frontispiece.

The Van de Graaf machine is essentially a modified form of one of the oldest types of electrical machine —the Toepler-Holtz "influence" machine. In this modern form a rapidly moving belt of paper or cloth is used in place of the rotating glass plate, and the operation is made more reliable by supplying the primary excitation from a transformer and rectifier unit of around 30,000 volts. The frontispiece shows a photograph of a working model of the big machine of this type which is now in operation at the Magnetic Laboratory of the Carnegie Institution at Washington, D. C., and Fig. 155 is a schematic diagram of the essential parts of this machine. The high-potential terminal is the so-called *corona cap*, *S*, a spherical metal shell, 19 feet in diameter, open at the bottom and with the edges of the hole tucked in as shown. This cap is supported by three stout ceramic insulating pillars which are not shown. The pear-shaped tank, *P*, 38 feet in diameter by 55 feet high, incloses and shields the whole apparatus. In the model this tank is made of a transparent plastic material; in the actual machine it is made of steel plates, and is so strongly made that a gas pressure of 50 pounds per square inch may be maintained in it when this is desirable. The manhole by which this tank may be entered shows near the ground level in the photograph.

The charging belt, B, receives its charge from the transformer-rectifier source, K, the charge being "sprayed" onto the belt by a corona (glow) discharge from the sharp points, p, which nearly touch it. This charge is carried up into the cap by the motion of the belt, which is 40 inches wide, and travels at a speed of 50 miles per hour. In the cap a second set of points, p', transfers the charge from the belt to the cap. In this way, the cap may be charged to a high potential, *this potential being produced by the mechanical work of carrying the charge on the belt up to the cap, against the electrostatic repulsion of the charge already there.* The upper limit to this potential is set by the leakage, as in all electrostatic machines. The metal rings, R, which surround the supporting columns, the belt, and the vacuum tube, T, serve as electrostatic shielding and to maintain a uniform potential gradient from the cap to the ground. In the photograph these rings conceal these other parts from view.

The vacuum tube, T, is about a foot in diameter, and extends from inside the cap to the basement room below the tank. It is kept highly evacuated by a system of powerful pumps. At the top of this vacuum tube is placed the ion source, A. This consists of a special discharge tube in which an auxiliary electric discharge through a suitable gas creates a copious supply of ions. The power plant for this ion source is housed in the cap itself. Some of these ions are injected through a small hole into the vacuum tube proper and are thence accelerated downward in a finely focused beam to the bottom of the tube, where they escape through a thin window into whatever apparatus is placed in the basement laboratory to receive them. When operated with dry air at atmospheric pressure in the tank, this machine supplies 1.8 to 2 million volts to a stream of ions which carries a current of 10 to 20 milliamperes.

At this time there are thirteen Van de Graaf generators in operation, with five more under construction or projected. The highest voltage produced by any of them is around 4 million volts.

236. The Cyclotron. About 1931 E. O. Lawrence, at the University of California, invented the *cyclotron*, an even more powerful ion accelerator which gives very high energies to ions by applying the same

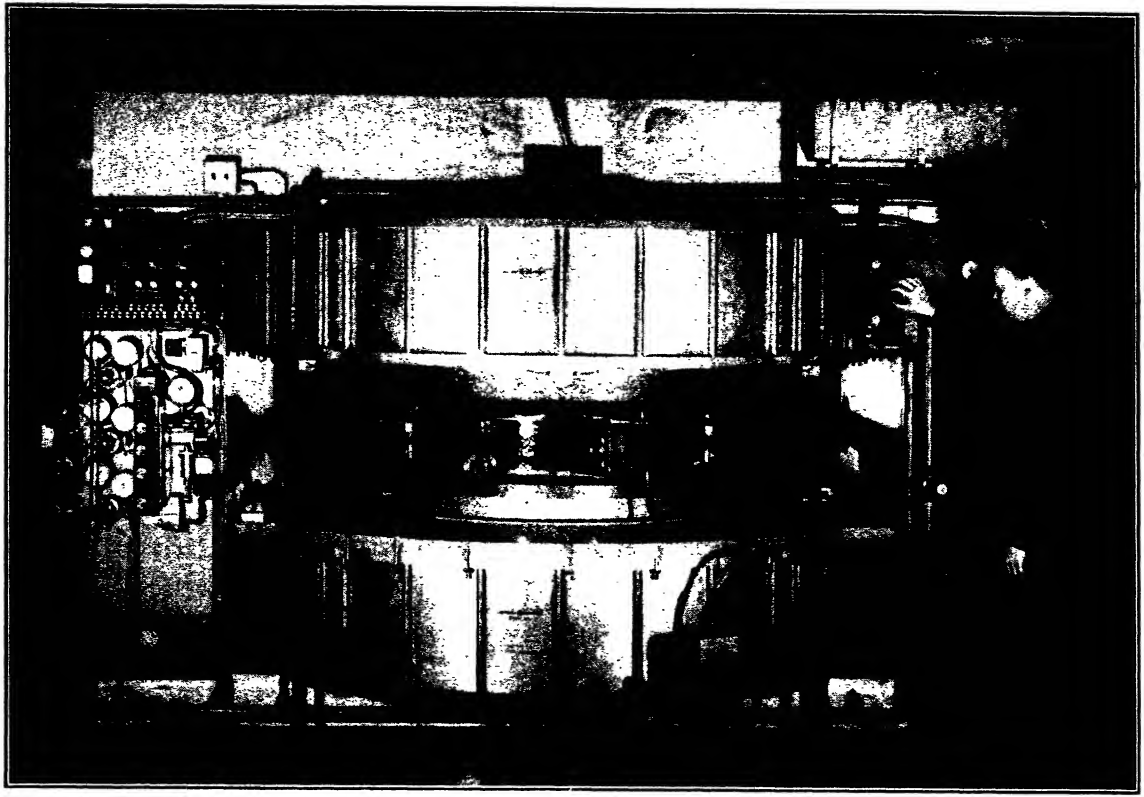


FIG. 156. THE HARVARD CYCLOTRON. The operator is Professor Roger Hickman. (Photograph supplied through the courtesy of Dr. K. T. Bainbridge.)

electric field to them many times in succession. Figure 156 shows a photograph of the Harvard cyclotron, a typical example of this type of machine. The device which tricks the ions into coming back time and time again to the electric field is the magnet, which is shown in side elevation in Fig. 157. Its cylindrical pole pieces, *N* and *S*, are supported by the massive iron frame, *A*–*B*, which serves also as a double yoke to complete the magnetic circuit. (The lines of magnetic flux are indicated by the dotted lines.) The magnetizing field is produced by an electric current flowing in circular coils placed around each pole piece; these coils are shown in cross-section in Fig. 157. The coils are covered and are water cooled. These details are also easily identified in Fig. 156. In the Harvard cyclotron the faces of the magnet poles are 42 inches in diameter, and the magnet contains 70 tons of iron and 7 tons of copper.

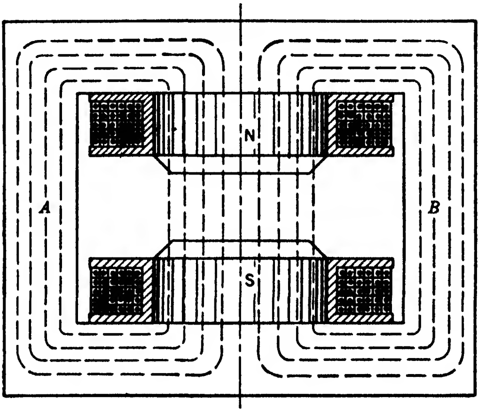


FIG. 157. THE CYCLOTRON MAGNET. The coils are shown in cross-section.

Figure 158 shows the heart of the cyclotron, the ion accelerator. *E* and *F* are the *dees*, two hollow D-shaped electrodes such as might be made by cutting in half a shallow covered circular box; the open edges (where the cut was made) face each other across a narrow gap, as is perhaps seen best in the vertical cross-section shown at (*b*). These dees, which are supported within the evacuated metal case, *B*, by means of the insulators, *M* and *N*, are connected to the terminals of the coil, *L*, with which they form a high-frequency electrical oscillator circuit. (*L* supplies the inductance for this circuit, and the dees themselves provide the capacitance.) When this circuit is maintained in electrical oscillation by means of the electron-tube driving circuit indicated schematically by *D*, a strong rapidly alternating electric field is produced across the gap between the dees; this is the field which gives to the ions the successive impulses by which they ultimately attain very high energies.

The ions themselves are supplied by an ion source, I, which is supported at ×, at the center of the evacuated case, by supports (not shown in the figures) which extend into the case from one side over the outside of one of the dees. As in the Van de Graaf machine, this ion source consists of some sort of low-voltage electric discharge through a gas which will produce the desired kind of ions. The complete assembly just fits into the space between the poles of the magnet, so that the powerful magnetic field is perpendicular to the flat sides of the dees—that is, perpendicular to the plane of the apparatus as seen in Fig. 158(*a*).

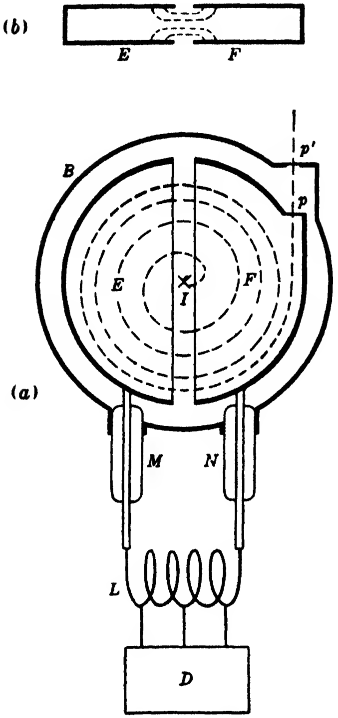


Fig. 158. Ion Accelerator for a Cyclotron. (*a*) Plan view: The spiral line represents the path of an ion. (*b*) Cross-section of the "dees."

When the electric field across the gap between E and F is directed towards F, a positive ion issuing from the ion source, I, will be accelerated into F by this electric field. Inside of F the electric field is zero, and the magnetic field will bend it into a circular path which will bring it back to the gap again. If now the magnetic field and the frequency of the electric oscillations are properly proportioned,[1] the electric field will have been reversed by the time the ion again reaches the gap, and the ion will again be accelerated, this time towards E. Thus it flies back and forth across the gap, moving faster and faster and in ever-widening circles through the interiors of the dees until at last it shoots out through the exit port, p. The target may be placed inside the vacuum case, at p', or a thin window at p' may let the ions pass out into the air, to targets, etc., placed on the outside. The stronger the magnetic field and the bigger the poles, the more trips the ion will make before escaping, and the greater will be its

---

[1] The condition is that $2\pi f = Bq/mc$, where B is the magnetic flux-density, and q/m is the ratio of charge to mass for the ions.

final energy. In order to escape, its energy must have the value (in electron-volts) given by

$$W = \frac{150e}{m}\left(\frac{BRN}{c}\right)^2 \tag{17·2}$$

where m is the mass of the ion, N is the number of elementary electric charges carried by it, R is the radius of the poles, and B is the flux-density of the magnetic field.

Since the flux-density is limited by saturation of the iron poles, the power of a cyclotron may be increased only by increasing the diameter of these poles. Lawrence's first cyclotron had poles 27.5 inches in diameter. His latest has poles 184 inches in diameter, and will produce 200-Mev deuterons, or 400-Mev alpha particles.

237. The Betatron. The cyclotron is limited in use to protons, deuterons, alpha particles, and other such heavy ions whose speeds remain well below the speed of light, even at the highest energies attainable. This restriction is necessary to maintain the *resonance* of the electrical oscillations with the angular motion of the ions. The frequency of this angular motion, which is given by

$$f = \frac{Bq}{2\pi mc} \tag{17·3}$$

remains constant only as long as the mass, m, remains constant; that is, as long as the speed remains low in comparison with c, the speed of light. This makes the cyclotron useless for acceleration of electrons, since they double their mass at less than ½ Mev of energy.

This restriction does not apply to the *betatron*, which is designed especially for accelerating electrons to very high energies. It was invented by D. W. Kerst, of the University of Illinois physics department. In general appearance the betatron somewhat resembles the cyclotron, as may be seen in Fig. 159, which is a photograph of the 100-Mev betatron of the General Electric Company. Like the cyclotron, it gives high energy to electrons by driving them around and around a circular track, keeping them in that track by means of a magnetic field. Unlike the cyclotron, however, this track is of constant radius, and there are no electrical fields acting other than the one which is produced by increasing the magnetic field rather slowly from zero to peak value!

The vacuum tube around which the electrons circulate resembles an inflated inner tube, with the electron path running along its circular

FIG. 159. 100-MEV BETATRON OF THE GENERAL ELECTRIC COMPANY.

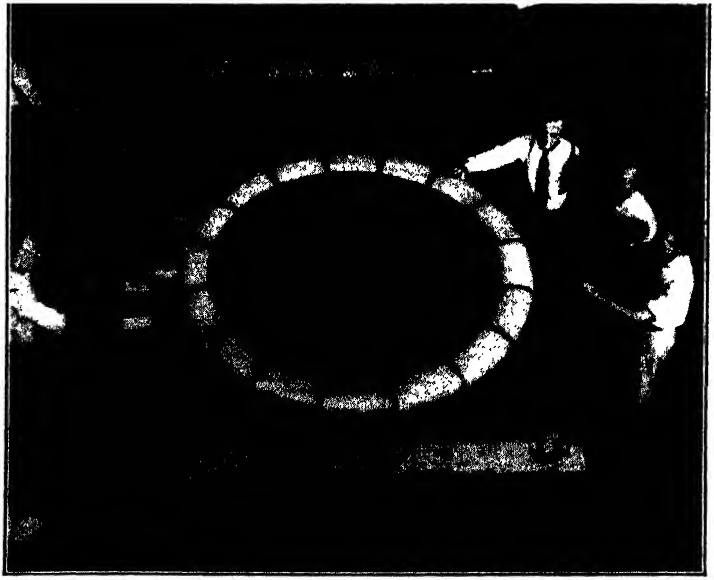


FIG. 160. 100-MEV BETATRON WITH THE TOP POLE REMOVED, SHOWING VACUUM TUBE.

axis. It is made of glass or porcelain, and need be only a few feet in diameter. It is placed between the poles of a large electromagnet, which is excited by *alternating* current. (Since this is an a-c. magnet, the cores must be laminated, like the core of a transformer.) Details of this assembly are shown in Fig. 161. The frequency of the alternating current may be the commercial 60 cycles, or it may be higher, up to several hundred cycles per second.

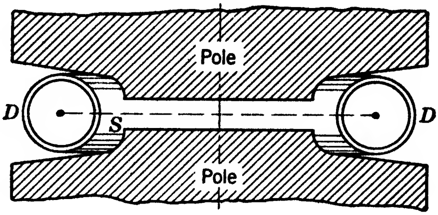


FIG. 161. BETATRON—CROSS-SECTION THROUGH POLES AND VACUUM TUBE. *DD*—doughnut-shaped vacuum tube. *S*—electron path seen edgewise. Observe that the poles are closer near the middle, so as to make the average flux-density twice that along the electron-path.

When the electrons are first injected, their speed is small and the magnetic field almost zero. Then, as their momentum increases, the magnetic field is increased proportionally *so as to keep the radius of the path from changing.* Equation (1·8) shows how this is accomplished. From equation (1·8) we may obtain

$$\frac{mv}{B} = \frac{Re}{c} \qquad (17\cdot4)$$

which shows that, if the ratio of the momentum, *mv*, to the flux-density, *B*, is kept constant, the radius, *R*, remains constant.

The momentum is caused to increase by *increasing the magnetic flux through the area bounded by the electron path.* It is well known that a changing magnetic flux will induce an electromotive-force into a closed circuit which links with it; it is not so often recognized that a changing magnetic flux will set up an *electric field* along any path which links with the flux, as in this case, even though no conductor be present. If *E* represents the strength of this electric field, then

$$2\pi RE = \text{electromotive-force around the path}$$

$$= -\frac{1}{c} \times \text{rate of change of flux through the area bounded by the path} \qquad (17\cdot5)$$

This field, *E*, is the electric field which accelerates the electron up to its maximum momentum. It can now be shown[2] that, if the magnetic flux is denser near the center of the poles than towards the edges, so as to make its average value within the path twice as strong as its value along the path, and if the flux everywhere increases at the same rate, then the electrons will continue to circle the same path from the start to the finish one quarter-cycle later, when the flux-density B has reached its maximum value, B_m. The momentum of the electrons then is given by

$$mv = \frac{B_m eR}{c} \tag{17·10}$$

The speed, v, is then very close to the speed of light, c; indeed, it approached close to c not long after the start. Putting v equal to c in equation (17·10), we obtain, as a very close approximation,

$$mc^2 = B_m eR \tag{17·11}$$

From this it follows, from equation (15·4), that the kinetic energy of the electrons is

$$\begin{aligned} W = mc^2 - m_0c^2 &= B_m eR - 8.2 \times 10^{-7} \text{ erg} \\ &= 3 \times 10^{-4} B_m R - 0.51 \text{ Mev} \end{aligned} \tag{17·12}$$

This equation shows, for example, that a betatron only 12 inches in diameter, with a value of 4000 gauss for B_m, would produce electrons having energy equal to 17.5 Mev.

A burst of electrons is accelerated in this manner for every cycle of the alternating current. When they reach maximum energy, they are deflected out of their circular path against a target. The only product

---

[2] The force on the electron is Ee, which is equal to $\frac{e}{2\pi Rc}$ times the rate of change of the flux ϕ. But, from Newton's second law, this force equals also the rate of change of the momentum, mv, of the electron. Hence, if both ϕ and mv start from zero, the value of mv at any time is proportional to ϕ; i.e.,

$$mv = \frac{\phi e}{2\pi Rc} \tag{17·6}$$

But also, from (17·4),

$$mv = \frac{BeR}{c} \tag{17·7}$$

Hence

$$\phi = 2\pi R^2 B \tag{17·8}$$

and the *average* flux-density over the area inside the path is

$$B_{av} = \frac{\phi}{\pi R^2} = 2B \tag{17·9}$$

from this target so far studied is X-rays, of gamma-ray wavelengths and intensity. These in themselves are very interesting and useful, but even more important results may be expected from the larger betatrons. One point of great interest concerning these X-rays is their distribution about the target. Ordinary X-rays spread out in all directions from the target, but these super X-rays proceed for the most part straight forward in the direction taken by the electrons as they struck the target, travelling right through the target and coming out the far side as a narrow cone of X-rays. The effective width of the X-ray beam from the 100-Mev betatron is only about 2 degrees.

238. Other Ion Accelerators. Several additional types of ion accelerators have been developed, or are being developed. All involve new combinations of one or more of the principles already described in connection with the cyclotron and the betatron. For example, the linear ion accelerator drives electrically charged particles down a straight vacuum tube by means of what amounts to a high-frequency *wave* of electric field pulses which follows the particles so as to keep in step with them and always drive them forwards. This produces the same effect as the resonance principle found in the cyclotron, so that the actual voltage produced by the high-frequency source is multiplied many times in its effect upon the moving particles. This form of ion accelerator was tried out even before the cyclotron was invented, but the early attempts were not very successful. Many of the difficulties which discouraged the early work now have been overcome by utilizing the high-frequency methods and equipment developed for radar, and recently L. W. Alvarez, of the University of California, has completed an ion accelerator of this type which gives 32 Mev of energy to protons. The protons are given an initial energy of one or two million electron-volts by means of a Van de Graaf machine, then are accelerated to 32 Mev while travelling from one end to the other of an evacuated tube 40 ft long and 4 ft in diameter. The accelerating pulses are provided by a bank of 25 or 26 radar magnetrons ranged along the sides of this tube.

The synchrotron is another device which employs the resonance principle of the cyclotron. Indeed, in its simplest form the synchrotron differs from the cyclotron only in its method of operation. It will be recalled that the cyclotron operates successfully only when the mass of the ions remains constant, that is, so long as the speed of the ions is negligible as compared with the speed of light.[3] When the mass begins to increase, the ions fall out of step with the electric field and gain no

---

[3] Compensation for small increases in mass may be made by shaping the pole pieces so that B increases towards the rim by the correct amount, but this cannot be carried very far without defocusing the ion beam.

more energy from it. In the synchrotron these restrictions are circumvented by *periodically* increasing the flux-density, B, or decreasing the frequency of oscillation, f, or by altering both B and f simultaneously. For example, the high energies produced in the 184-inch cyclotron are made possible by periodically varying the frequency of the electric field between the dees. The beam of ions then emerges as a series of pulses, one for each decrease in the frequency of oscillation.

When the synchrotron principle is employed for electrons, the conditions become quite simple to describe. Within a short time from the start the speed of the electrons becomes practically the speed of light, and the circumference of the path becomes equal to c/f, where f is the frequency of oscillation. If now the frequency is held constant while B increases, the electron gains energy without leaving the circular path determined by f. Engineers and physicists of the General Electric Company have succeeded in building a synchrotron which produces 70-Mev electrons, and are now designing one for 500-Mev electrons.

239. Nuclear Reactions. The systematic study of *nuclear reactions*, that is, changes in nuclei which are brought about in one manner or another, is one of the most important of the many-fold purposes for which these special electrical machines have been developed. The chief advantages possessed by them, as compared with natural sources of high-energy particles, are (1) very much larger quantities of particles are available; (2) these may be focused into a narrow beam; (3) the energy in the beam may be controlled to any desired value, and this energy is accurately measurable; and (4) the larger machines provide energies greatly in excess of those provided radioactively.

Many examples of nuclear reactions have been considered already in Chapters XV and XVI, including radioactive transformations, the experiments of Rutherford, the production of neutrons, neutron-produced reactions, and the production of artificially radioactive elements. Some of these reactions, like radioactive transformations, occur spontaneously, whereas others result from the introduction into a nucleus of extra particles, such as protons, neutrons, deuterons, and alpha particles. Any reaction may be represented by an equation of the type already used in Chapters XV and XVI for the examples cited. For example

$$ {}_{7}\mathrm{N}^{14} + {}_{2}\alpha^{4} = {}_{8}\mathrm{O}^{17} + {}_{1}p^{1} \qquad (17 \cdot 13) $$

is used in Sec. 221 to represent Rutherford's historic experiment. This same reaction may also be represented by the following, more condensed, notation.

$$ \mathrm{N}^{14}(\alpha, p)\mathrm{O}^{17} \qquad (17 \cdot 14) $$

The original nucleus and the new nucleus are represented to the left and the right, respectively; only the mass-number is written, since the atomic number is indicated by the chemical symbol for the element. The symbols representing the two particles involved are written inside the parentheses, the emitted particle following the captured one. This part of the formula represents the *type* of the reaction. For example, this is an (α, p) reaction. The reacting particles may be represented by the following symbols: proton, p; neutron, n; alpha particle, α; photon, γ; electron, β^-; positron, β^+. A variety of other symbols will be found for electrons and positrons, but the others are quite generally accepted as standard notation.

The reaction products may be identified in various ways. First of all, their energies may be determined by measuring their ranges in air or other absorbing media, as has already been described. They may be detected by scintillations on a fluorescent screen (Sec. 221), the Geiger-Müller counter (Sec. 264), the proportional counter (Sec. 265), or the Wilson cloud chamber. The cloud chamber, when it may be used, yields the most detailed information; it supplies not only the range of the particles, but also the angles between the tracks of the several particles involved. Simultaneous photographs, taken from two different directions, provide stereoscopic views from which these angles may be determined even when the paths do not lie in the plane of either photograph. Figure 172 shows such cloud-chamber pictures. The cloud chamber is especially useful when placed in a strong magnetic field, with the lines of force of the field perpendicular to its face. The magnetic field then deflects the particles into paths which appear as curved tracks, as shown in Figs. 148 and 151, and the curvature of this track then gives a direct measure of the *momentum* of the particle, provided its charge is known. The equation which gives the momentum is the mass-spectrograph equation (2·5), which may be written

$$mv = \frac{BqR}{c} \qquad (17 \cdot 15)$$

For particles having the same charge q, the momenta are directly proportional to BR, the product of the flux density and the radius of curvature of the track. From such data the masses and energies of the reacting particles may be determined, often with high precision.

240. Nuclear Species. Among the results obtained from study of nuclear reactions are the discoveries of many more nuclear species.[4]

---

[4] There is serious need for a name to designate nuclei of any one nuclear species, that is, nuclei having the same atomic number Z and mass number A. There is a tendency to use "isotope" for this purpose, but "isotope" has quite a different meaning. In its proper sense "isotope" should refer only to one of two or more nuclear species of the *same* element. The name *nuclide* has been proposed by T. P. Kohman.

The number of species of stable nuclei exceeds 270, and the number of known species of unstable (radioactive) nuclei total around 500 when we include those discovered as fission products (see Sec. 250). The smallest number of isotopes for any one element is now known to be three, with the average between seven and eight. The manner in which the isotopes of the lighter elements are built is shown in Fig. 162, which charts the number of protons and neutrons in each isotope of the first eighteen elements. The dots and circles representing these nuclei form a diagonal band across the chart which shows that, for these light elements, the protons and neutrons are about equal in number. For all stable nuclei, after hydrogen, the number of neutrons is never less than the number of protons, although it may be greater. Indeed, beyond the first twenty elements the neutrons increase more rapidly than the protons, until there are over 50 per cent more neutrons than protons in the heaviest elements. This is shown in Fig. 163. The neutrons perform a very

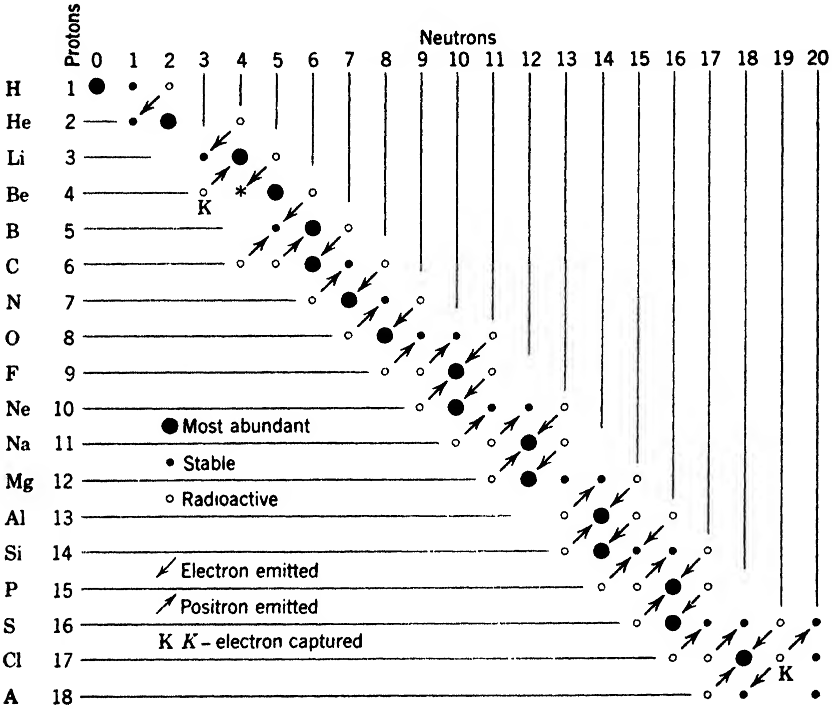


FIG. 162. PROTON-NEUTRON COMPOSITION OF LIGHT NUCLEI. NOTE: The Be^8 nucleus, indicated by (*), always breaks up into two He^4 nuclei. See footnote 2, p. 288.

important part in holding the nucleus together, and, as the size of the nucleus increases, a greater proportion of neutrons is required to make the nucleus stable. In larger nuclei the number of neutrons is not very critical, and this explains why they have so many isotopes. Tin, for example, has ten stable isotopes.

The unstable or radioactive isotopes form, for the most part, fringes along the upper and lower borders of the band of stable isotopes.

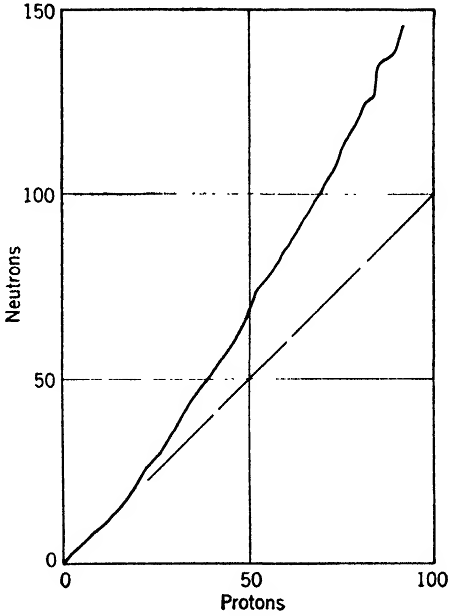


FIG. 163.

Those along the upper fringe break down by emitting an electron; the nature of the transition is indicated by arrows drawn in Fig. 162, pointing to the stable nucleus which is produced. The radioactive nuclei represented by the lower fringe transmute generally by emission of a positron, although a few accomplish the same final result by capture of a *K* electron, that is, by swallowing one of the innermost or *K* group of electrons outside the nucleus. When an unstable nucleus finds itself in the middle of the band of stable nuclei, it may not quite know which way to go. Chlorine 36, for example, sometimes becomes argon 36, by electron emission, and sometimes sulfur 36, by positron emission or by *K*-electron capture. Potassium 40 (not shown in the figure) finds itself in a similar predicament; this nucleus is further remarkable as being the only light radioactive element which occurs naturally. More and greater irregularities appear among the heavier elements.

241. "Practical" Values. Sooner or later the scientist who is engaged in what is often called "pure" scientific research is confronted with the question: "What is the practical value of all this?" Such questioners should not be turned aside, nor should the question be scorned as basely commercial. Far from meaning "What is the dollar-and-cents value of this work?" such questions may mean, "What is the value to human comfort and happiness?" or "How will these discoveries help to elevate man's state socially and spiritually?" The scientist can answer always that, although he may not know now the specific answers to such questions, he is confident that abundant value will follow from all true scientific discovery. Indeed, it is an axiom of scientific research that such work is most successful when carried on without any thought to its possible applications or to its ultimate broader usefulness. Scientists hold to the belief that research which has as its primary purpose the extension of our knowledge of the fundamental laws of nature will never fail to justify itself, in practical ways as well as in the reward of discovery itself, and this faith has never been betrayed.

The discovery of X-rays is an often-cited example of this. Not only was Roentgen *not* searching for an aid to surgery when he made this discovery, but it is also very doubtful whether anyone who was interested in surgery at that time would have given the slightest thought to the possibility of being able to see through flesh and bone, let alone have wasted any time in searching for means to do so. And yet X-rays are of the utmost importance to surgery. Many other such examples might be cited, among which not the least are the discoveries in nuclear physics.

The practical applications which have arisen from nuclear physics fall into roughly two groups. One group concerns the utilization of nuclear reactions as sources of power; these applications, which involve nuclear reactions on a large scale, will be considered in the next chapter. The other group comprises a variety of applications to other sciences, such as medicine, biology, and chemistry, and even to other branches of physics, applications in which the artificially produced radioactive elements play principal parts. The reactions involved in these applications are on a scale which is exceedingly small as compared with those in the first group; their great importance lies in the unique services they render. Not so long ago it was considered possible that practical values in nuclear science might be limited to this second group. Even so, thousands of dollars were spent for the construction of cyclotrons and for the support of large nuclear research laboratories, solely because of the importance of these applications. A large part of this money came from the various foundations for medical research.

242. Applications to Radiotherapy. The most obvious use for artificial radioactive elements is to replace natural radioactive elements in radiotherapy, or the irradiation of diseased body tissues, and for this application the new materials possess several distinct advantages over the older ones. First is their short lives. The half-lives vary from a small fraction of a second upwards, and there are a considerable number of useful elements with half-lives which are of convenient length. For example, iodine 128 is half gone in 25 minutes, sodium 24 in 15 hours, and phosphorus 32 in 14 days. All these lives are long enough to be useful, but so short that the radioactive material need not be recovered to prevent overdosage. Indeed, the dose is determined simply by the quantity of radioactive material administered.

The second advantage follows out of the first. It is possible to administer these radioactive materials in the most convenient manner, either externally or internally, and it is also possible in some cases to choose the radioactive material so that naturally selective chemical or physiological processes will deposit it in the diseased area to a greater extent than elsewhere. Furthermore, materials whose half-lives are such as to make them suitable for this use emit radiations of only moderate penetrating power, so that the treatment may be quite localized.

Neutrons also possess therapeutic value, since they are found to produce biological effects similar to those produced by X-rays and gamma rays, including a similar selective lethal effect upon diseased tissue. They may prove of even greater therapeutic value, since they are more strongly absorbed in fleshy than in bony tissue. This is just the opposite of the behavior of X-rays and gamma rays. The actual biological effects of all these radiations are of secondary origin, being caused by the charged particles which they energize by their passage. For neutrons these charged particles are principally protons, set into motion by neutron impacts in the hydrogen-rich fleshy tissue.

These new tools for radiation therapy will not come into general service until after much medical research has determined the best ways to utilize them, and particularly not until medical men generally have become proficient in their use. Progress in both these directions may be time consuming.

243. "Tracer" Applications. The application for artificially radioactive substances which so far has been most developed is less directly practical, although one which, for that very reason perhaps, is farther reaching in its ultimate importance. This is their utilization as "tagged" atoms or "tracers." The great ease with which the tiniest amount of radioactive material can be detected makes these substances ideally useful in many fields of scientific investigation, to "tag" or label certain groups of atoms so that they may be "traced" or followed through subsequent physical, chemical, or biological processes.

A simple example of this application is to be found in the study of diffusion in solids. Many years ago an experiment of this kind was performed by placing a lead plate in very close contact with a plate of gold. After these plates had been left in close contact for several years, traces of gold were found beneath the surface of the lead plate; some diffusion of the two metals had taken place across the surfaces. Now the experiment may be performed in a few days, with lead alone. If one lead plate contains some radioactive lead, any diffusion from one plate into the other will carry over a proportionate number of radioactive atoms. The amount of diffusion then may be measured by testing the second plate for radioactivity. Because of the great sensitivity of this test the time required for a successful experiment is comparatively short. Radioactive tracer atoms thus have made possible studies in this field of diffusion in solids which never could have been undertaken before.[5]

Another example is a study made of the utilization of calcium by rats, in which the tracer element was radioactive phosphorus. A very small portion of this tracer substance was added to the calcium phosphate which was fed the rats, so that progress of the calcium through the rat's system could be followed with a Geiger-Müller counter. Radioactive phosphorus has been employed also to study the metabolism of sugar. It had been suspected that molecules containing phosphorus occurred as intermediary steps in the burning of sugar in the body, and this was verified in these experiments. Similarly, important information concerning hemoglobin and the storage and metabolism of iron has been obtained with radioactive iron as the tracer element, and radioactive iodine has served in studies of thyroid physiology. In all these fields the work may still be considered as only just begun. Much more rapid progress may be expected now that much greater quantities, as well as greater varieties, of radioactive materials are available from the great nuclear piles. (See Sec. 254.)

Tracer atoms do not need to be radioactive; any atom which may be distinguished from others which are chemically the same (isotopes of the same element) will serve. For example, carbon 13, which forms about 1 per cent of natural carbon, may be distinguished from the commoner C^{12} by means of the mass spectrograph (Sec. 21). If now

---

[5] For a survey of this work, see Hevesy, *Transactions of the Faraday Society*, August, 1938, pp. 841–845.

all, or even a large fraction, of the C^{12} atoms in a carbon compound are replaced by C^{13} atoms, the changes which these compounds undergo may be followed by mass-spectrographic examination of the products. Ordinary carbon may be enriched with C^{13} by any of the methods described in Sec. 253 for the concentration of U^{235}. Carbon 13 in carbon dioxide has served in this manner for the study of the photosynthesis of starch in growing plants. It has served likewise, in sugar, for studies in animal metabolism.

244. Fundamental Research. In summing up activities in the fields of nuclear science we must not overlook the most important one, namely, fundamental research. Research is the frontier of science, and in nuclear physics the frontier is still all around us. There are countless nuclear reactions yet to be discovered, new elements to be found, and better knowledge to be obtained concerning the properties of all elements. Innumerable applications are yet to be discovered, as well as better methods still to be learned for utilizing those already known. Improvements in apparatus, and the invention of new apparatus, must accompany all this work. But the investigation of fundamental theory to explain these phenomena must be placed first in importance. For example, what is the nature of the forces exerted among protons and neutrons, which enable them to combine so as to form nuclei? And what are the properties of these particles which account for the existence of such forces? These and many other questions are now answered only in part, if at all. The quest for better answers involves research in many fields, both theoretical and experimental, and is fundamental to all other research. The results of some of the more recent investigations are given in the next two chapters.

PROBLEMS

1. Assume that the charging belt of the Carnegie Institution Van de Graaf machine carries a charge of 300 stat-coulombs per square centimeter of its surface. (*a*) What current in milliamperes does it carry? (*b*) What power is required to drive this belt (neglecting friction) when the potential of the cap is 3 million volts above ground potential?

2. Prove that, in order to operate a cyclotron, the frequency of the a-c. voltage applied to the dees must be

$$f = \frac{Bq}{2\pi mc}$$

where q and m are the charge and the mass respectively of the ions being accelerated.

3. In order to compensate for the increase in mass with speed, the magnetic field in a cyclotron must vary slightly from the center to the rim of the pole-piece faces. Where must it be strongest? Explain why.

4. A cyclotron has poles 60 inches in diameter, with a field strength between the poles of 12,000 oersteds. (*a*) Compute the *energy* given to (i) protons, (ii) deuterons, (iii) alpha particles. (*b*) Compute how many times each of these ions must circle around before escaping, if the voltage across the gap between the dees at the time of crossing is 30 kilovolts.

5. Compute the frequency of oscillations for the accelerating circuit of the cyclotron described in problem 4, for each type of ion there specified.

6. (*a*) Compute the electromotive-force acting around a betatron orbit of 160-cm diameter, when the flux-density *at the orbit* is increasing at the rate of 1,500 gauss per millisecond. (*b*) Compute the value for the electric field E along this orbit under these same conditions. (*c*) Compute the force exerted upon an electron by this field.

7. An electron is circling in a betatron orbit of 160-cm diameter when the flux-density *at the orbit* is 4000 gauss. (*a*) Compute the momentum of this electron. (*b*) Compute its energy. (*c*) Compute the frequency of its revolution about the orbit.

8. 70-Mev electrons are produced by a synchrotron. (*a*) Compute the maximum energy for X-rays produced when these electrons strike a target. (*b*) Compute the wavelength of these X-rays.

9. A cloud chamber is operated in a magnetic field of 1000 oersteds. A track is observed which has a radius of curvature of 4 meters. (*a*) Compute the *momentum* of the particle which made this track. (*b*) Compute the energy of this particle if it is a proton. (*c*) Compute its energy if it is an electron.

10. Carbon 12 is bombarded by deuterons, and a proton is emitted. Write the symbolic equation for this nuclear reaction and identify the new nucleus.

11. Potassium 39 is bombarded with protons and alpha rays are emitted. Write the symbolic equation for this nuclear reaction and identify the new nucleus.

12. Selenium 75 is a radioactive nucleus which converts to a stable nucleus by capturing one of its own K electrons. To what element does the stable nucleus belong? Explain.

13. Write the nuclear-reaction equations found in Secs. 216, 221, 225, 227, 229, and 234, in the condensed form of equation (17·14).

14. Carbon 14 is formed when a certain stable nucleus captures a neutron and ejects a proton. Identify the parent nucleus and write the nuclear-reaction equations in both forms.

15. Carbon 14 is radioactive. What particle will it emit, and what will be the resulting new nucleus?

REFERENCES

FOR ADDITIONAL READING

MAX BORN, *The Restless Universe*, Chapter V.
J. B. HOAG, *Electron and Nuclear Physics*, Chapters XIII and XIV.
G. F. HULL, *Modern Physics*, Chapter XII.
J. A. ELDRIDGE, *Physical Basis of Things*, Chapter XXV.
R. A. MILLIKAN, *Electrons, Protons, Photons, Neutrons, and Cosmic Rays*.
J. D. STRANATHAN, *The "Particles" of Modern Physics*, Chapters 9, 10, and 11.
A Symposium on the Use of Isotopes in Biology and Medicine (the University of Wisconsin Press).
D. W. KERST, "The Betatron," *American Journal of Physics*, Vol. 10, 1942, pp. 219–223.
D. W. KERST, "The Betatron: Development and Applications," *American Scientist*, Vol. 35, January, 1947, pp. 56–84.

W. F. Westendorp and E. E. Charlton, "A 100-Million-Volt Induction Electron Generator," *Journal of Applied Physics*, Vol. 16, 1945, pp. 581–593.

For Advanced Reading

E. Pollard and W. L. Davidson, Jr., *Applied Nuclear Physics.*

N. Feather, *Nuclear Physics.*

F. Rasetti, *Nuclear Physics.*

G. Gamow, *The Structure of Atomic Nuclei and Nuclear Transformations.*

L. B. Loeb, *Atomic Structure*, Chapters IV and V.

J. A. Wheeler, "Elementary Particle Physics," *American Scientist*, Vol. 35, April, 1947, pp. 177–193.

P. Morrison, "Introduction to the Theory of Nuclear Reactions," *American Journal of Physics*, Vol. 9, 1941, pp. 135–161.

"Physics in 1946," *Journal of Applied Physics*, Vol. 18, February, 1947.

H. A. Bethe, *Elementary Nuclear Theory.*

F. K. Richtmyer and E. H. Kennard, *Introduction to Modern Physics*, Chapter XI.

CHAPTER XVIII

NUCLEAR ENERGY

245. Nuclear Energy. We shall now consider more carefully the energy changes which accompany nuclear reactions. The basic principle underlying all such considerations is the *equivalence of mass and energy*, which was first established for photons (Sec. 107) and for the kinetic energy of electrons (Sec. 212). Later (Sec. 224) it was found to include the creation and the destruction of electrons as well. In each case the equivalence of mass and energy is expressed by the relation

$$\text{mass} \times c^2 = \text{energy} \tag{18·1}$$

The laws for conservation of mass and for conservation of energy both are incomplete unless we include the *mass of energy* in the one, and the *energy of mass* in the other.

246. Mass of Potential Energy. A further illustration of this law, one which extends it to include all forms of energy, is supplied by the experiment of Cockcroft and Walton which is described in Sec. 234. In this experiment a proton was captured by a lithium nucleus, and the combination immediately split up, with explosive violence, into two alpha particles. The masses of all these particles are known very precisely from mass-spectrograph measurements; expressed in atomic units they are 1.0076 for the proton, 4.0028 for each alpha particle, and 7.0165 for the lithium nucleus (the most abundant isotope, of mass number 7). The total mass before reaction was

$$1.0076 + 7.0165 = 8.0241$$

After reaction it was

$$2 \times 4.0028 = 8.0056$$

A mass of 0.0185 units is lost, and the range of the alpha particles, as observed in the cloud-chamber photographs, shows exactly what has become of this mass. From this range the kinetic energy of these alpha particles can be computed to be 8.5 Mev for each, or 17 Mev, equivalent to 0.0183 mass unit,[1] for both. When two particles, such as these alpha particles, fly apart because of mutual repulsion, they lose potential energy in amount equal to the kinetic energy gained, and this experiment proves that this loss of mass is just the *mass of the potential energy* which has been lost. (The difference, less than 2 per cent, is within the experimental error.)

The mass equivalence for potential energy has likewise been verified by all other nuclear reactions for which accurate data are available, and many have now been investigated. **The law of equivalence between mass and energy of all kinds may now be considered one of the fundamental laws of physics.**

247. Binding Energy. This law enables us now to explain away the only remaining objection to the proton-neutron hypothesis for nuclear structure, namely, the loss in mass which results when the component protons and neutrons are combined to form a nucleus. See Sec. 228. This loss now may be explained as being the mass of the potential energy lost by the nuclear particles when they combine and are held together by mutual attractions. Consider, for example, the helium nucleus, which is built up with two protons and two neutrons. The masses of these particles are 1.0076 for each proton, 1.0089 for each neutron, and 4.0028 for the helium nucleus. The total mass before combination is 4.0330, and only 4.0028 afterwards. The loss of mass is 0.0302, and this is the mass of the potential energy lost when the four particles are brought together. It is equivalent to 29 Mev or 45×10^{-6} erg. The magnitude of this lost energy is better appreciated if we compute the energy released in forming one gram atom, or 4 grams, of helium. The mass lost is 0.030 gram, large enough to measure on an ordinary balance, and the energy released is $0.030\ c^2$ ergs, or 27×10^{18} ergs. This is equivalent to 750,000 kilowatt-hours of energy per gram atom, or 188,000 kilowatt-hours per gram!

The energy lost when a nucleus is formed is called its **binding energy,** since it represents the work which must be done to break it apart again. In other words, the greater the loss of mass, the tighter the nuclear particles are bound together, and the tougher is the nucleus.

248. Binding Factor. If we divide the binding energy for a nucleus by its mass number, we obtain the *binding energy per nuclear particle,*

---

[1] One mass unit equals 1.66×10^{-24} gram, or the mass of an atom of unit atomic weight. According to the law of mass-energy equivalence, this mass is equivalent to $1.66 \times 10^{-24}\ c^2$ erg, or 1.49×10^{-3} erg. Since 1 Mev equals 1.60×10^{-6} erg, the energy equivalent to 1 mass-unit equals 935 Mev, or 1 Mev equals 0.00107 mass unit. Consequently the 17 Mev of lost energy in this nuclear reaction is equivalent to 0.0183 mass unit.

which we shall call its **binding factor.** It is usually expressed in mass units.[2] Figures 164 and 165 show how this binding factor varies with the number of nuclear particles, or *nucleons.* It reaches a maximum value when the nucleus contains 50 or 60 nucleons, then remains fairly constant, except for small irregularities not shown, until another 100 or more nucleons have been added. Beyond 180 nucleons it begins to decrease more noticeably. Since the binding factor is an index of nuclear stability, this plot shows that the stability of nuclei is highest in the middle group and falls off as heavy nuclei are approached.

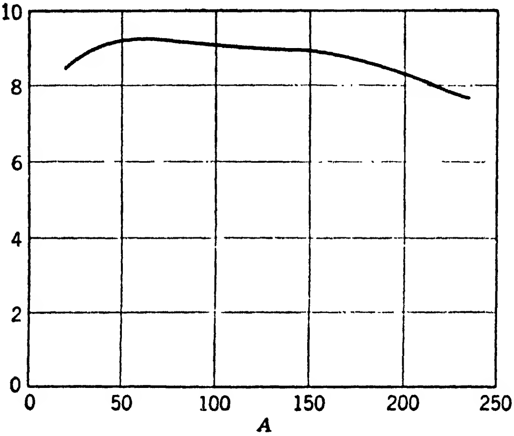


FIG. 164. BINDING FACTORS. Ordinates—Binding factor × 10^3 grams.

Binding factors for the lighter nuclei exhibit the periodic variations shown in Fig. 165. Since data for these nuclei are obtained with high precision, the peaks at He^4, Be^8, C^{12}, and Ne^{20} must be taken to indicate that these nuclei are considerably more tightly bound than their neighbors, and this is in general agreement with other experimental data. Thus the He^4 nucleus, better known as an alpha particle, is

---

[2] Commonly the binding factor is represented by the *packing fraction,* which is defined as the ratio of the *mass defect* to the mass number, *A*. Mass defect, in turn, is defined as the difference between the nuclear mass and the mass number. The binding factor, when expressed in mass units, is equal to the difference between the average mass defect of the protons and neutrons in the nucleus, and this packing fraction. For example, nitrogen 14 has a nuclear mass of 14.0075, and contains equal numbers of protons and neutrons. The mass defect of N^{14} is then 0.0075, and its packing fraction equals 0.0075 divided by 14, or 0.00054. The masses of the proton and the neutron are 1.00758 and 1.00893, respectively, giving an average mass defect of 0.00825. The binding factor for N^{14} is then $0.00825 - 0.00054 = 0.00771$ mass unit.

known to be an especially tough little particle. On the other hand, Be^8 does not exist naturally, and whenever it is produced artificially it immediately blows apart into two alpha particles. This anomaly may perhaps be explained by saying that the protons and neutrons in the Be^8 nucleus tend to group together as two alpha particles. Since the binding factors for He^4 and Be^8 are almost exactly the same, the formation of two alpha particles would use all the binding energy of the Be^8 nucleus, leaving none over to hold the two alpha particles together.

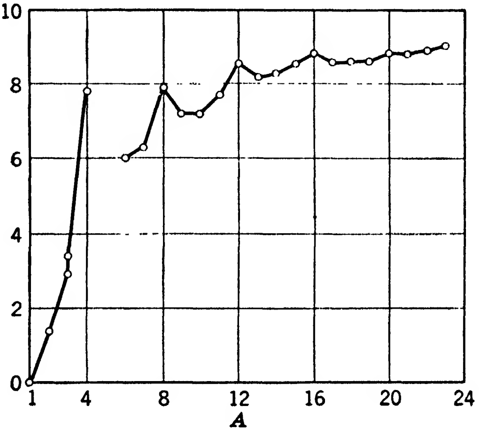


FIG. 165. BINDING FACTORS FOR LIGHT ELEMENTS. Ordinates—Binding factor × 10^3 grams.

249. Nuclear Power. It has been pointed out in Sec. 247 that the building of one gram of helium from protons and neutrons releases a quantity of energy which staggers the imagination. What are the possibilities for utilization of this or perhaps other nuclear reactions as sources of commercial power? Ever since the discovery of radioactivity the scientists have been besieged with questions of this kind, and up until 1940 the answer had to be that no nuclear reaction was known which would meet the requirements for such a power source.

In order to understand what these requirements are, let us consider the practical sources of power provided by chemical reactions, such as the combustion of wood, coal, or gasoline. Two requirements must be met to make such power sources practical. First, the reaction must be one in which the energy of the chemical products of the reaction is less than that of the substances reacting, such as the fuel and oxygen. The difference in energy is released as thermal energy to operate the furnaces, engines, or other industrial machines. Second, the reaction must be one which will maintain itself once it is started. Such reactions are called "chain reactions": the reaction with one molecule of the fuel sets off the reaction with another molecule, and that with still another until all the fuel is consumed. Furthermore, if the reaction is to be a practical one, it must also be capable of being controlled to proceed at a desirable rate; it must not die out, and in all but a very few applications it should not proceed with explosive violence.

The thermal energy given off by such a chemical reaction represents a loss of mass, just as for a nuclear reaction. The total mass of the reaction products must be less than that of the oxygen and fuel before combustion, *by the mass of the energy released.* For chemical reactions, however, this loss of mass is entirely too small ever to be detected. Indeed, the law of conservation of mass was based in part upon the failure of all experiments made to discover such a loss of mass for the most violent of chemical reactions. This again emphasizes the incomparably greater effectiveness of nuclear sources of energy.

Many nuclear reactions are known which meet the first requirement as stated above, but even now few are known which meet the second requirement, under conditions which may be realized on earth. All nuclear reactions which have been described so far are ones which are induced to go only by the expenditure of far more energy than may be recovered from the reaction. Consider, for example, the reaction produced in the famous experiment of Cockcroft and Walton. The amount of energy released by each nuclear explosion is relatively enormous, but the process will not go of itself. Each explosion must be touched off by a bombarding proton, and the number of protons which succeed in doing this is a vanishingly small percentage of the total proton beam. Neutron reactions offer greater possibilities for success, but all known neutron reactions, other than the ones about to be described, require neutrons supplied from an external source, with an expenditure of energy greatly exceeding that produced by the reactions.

250. "Fission" of Uranium. Early in 1939 O. Hahn and F. Strassman, in Germany, discovered traces of radioactive barium in uranium which had been bombarded by neutrons, and L. Meitner and O. R. Frisch hit upon the correct explanation: *this barium represented fragments of uranium nuclei which had split apart spontaneously upon capture of neutrons.* News of this discovery reached the nuclear physicists of this country while they were attending the annual Conference for Theoretical Physics at Washington, in January, 1939. (It arrived by way of a private letter to Professor Bohr.) Immediately workers in nuclear physics laboratories all over the country undertook to check this report, and before the conference was ended the discovery had been confirmed by direct observations of the fragments at the laboratories of the Carnegie Institution of Washington, Johns Hopkins University, Columbia University, and the University of California. Soon afterward it was shown that thorium and protoactinium, as well as both isotopes of uranium, may undergo this process of nuclear splitting, which is called *nuclear fission*.[3]

When such a nucleus, U^{235} for example, captures a neutron it becomes unstable and breaks apart into two fragments which are of roughly the same size, although the exact sizes are a matter of chance and may vary greatly from one fission to the next. A big nucleus rather closely resembles a drop of very dense liquid, and the energy released by the captured neutron may be considered to set this droplet nucleus into vibrations which are violent enough to break it up. Initially the two fragments are highly unstable, since together they may contain between six and ten neutrons more than they could hold as stable nuclei. Some of these neutrons, probably two for U^{235} fission, are released during the fission process, being ejected at high speeds from the cracking nucleus. Those which stick with the fragments generally convert themselves into protons, by a series of electron emissions, until a stable proportion of neutrons and protons is achieved. These radioactive emissions have enabled scientists to identify the chemical natures of the fragments; almost all elements having atomic numbers between 34 and 57 have been identified among these fragments and their subsequent transition forms.

The origin of the energy released by this fission process may be discovered by examining the binding-factor plot of Fig. 164. Here it is seen that the binding factor for uranium is considerably less than for the elements which constitute the fragments. This means that the sum of the binding energies of the two fragment nuclei is greater than the binding energy for the uranium nucleus. Remembering now that binding energy is actually energy *released* when the binding takes place, we see that this excess equals the energy which is released during the fission process (and during the subsequent radioactive processes). We may estimate the amount of this energy from the data given in Fig. 164. The binding factor is about 7.4 Mev per nucleon for uranium, and about 8.8 Mev per nucleon for elements of half this size. The difference is 1.4 Mev per nucleon. Since there are 235 nucleons in a U^{235} nucleus, the total energy release upon splitting a U^{235} nucleus in half would be approximately 235×1.4 Mev, or 330 Mev. The actual process is not as efficient as this (the immediate fragments are radioactive nuclei

---

[3] More recently Lawrence has shown that nuclei of thallium, lead, and bismuth also undergo fission when bombarded with neutrons from the 184-inch cyclotron.

whose binding factors are less than those of the stable ones formed ultimately), but even so the average energy release is found to be about 200 Mev for each uranium atom, or 22,000 kilowatt-hours per gram of uranium.

251. Nuclear Chain Reactions. The discovery of this process of nuclear fission again raised strong hopes for the development of practical sources of nuclear power ("atomic" power is the popular term). One factor is especially significant: *each fission releases several new neutrons.* If these new neutrons can in turn cause fission of other nuclei, then a nuclear chain reaction may be possible. To maintain this reaction, an average of at least one of the neutrons produced in each fission must be able to effect another fission. If the average is just one, the reaction will proceed at a steady rate, and a steady power source will result. If the average falls below one, the reaction will die out; if it is much greater than one, it may go with explosive violence.

The several fissionable nuclei differ greatly in respect to the conditions which produce fission. Uranium 238 nuclei, for example, will split up only when they capture fast neutrons, for which the probability of capture is quite low. This restricts the number of effective neutrons to so low a percentage of those produced that a chain reaction is difficult if not impossible. Similar conditions exist for thorium and protoactinium nuclei. On the contrary, fission of U^{235} nuclei is produced best by very slow neutrons,[4] and the process is then so effective that a chain reaction is possible, as everyone now knows. This does not mean, however, that once fission is started in a mass of U^{235} nuclei it will keep on until all have been split up. Before the fission-produced neutrons have been captured by other U^{235} nuclei so as to produce new fissions, they may escape from the mass entirely, or they may be captured by impurities in the mass. Unfortunately, uranium 238 is such an impurity. Indeed, U^{238} nuclei show a very great probability for capture without fission of neutrons having speeds intermediate between the speeds of emission and the very low thermal speeds which are most effective for fission of uranium 235.

252. The "Atom" Bomb. The state of world history at the time of the discovery of nuclear fission made it inevitable that nuclear energy should have its first practical utilization in a super-bomb. This might have occurred even without the pressure of wartime conditions; it is to be recalled that chemical energy also saw its first utilization in explo-

---

[4] As has been explained in Sec. 227, neutrons which have not been captured earlier will eventually slow down to speeds determined by the temperature of the material in which they find themselves—that is, so that their average kinetic energy equals $\frac{3}{2}kT$. See Sec. 161. Hence these slow neutrons are called *thermal* neutrons.

sives. Furthermore, a bomb is relatively easier to produce than a power plant. To produce an "atom" bomb it is necessary only to bring together a sufficiently large mass of fairly pure uranium 235; there are always present enough stray neutrons to touch it off. If the mass is large enough to insure an average of more than one useful neutron for each fission produced, the reaction will proceed with explosive violence. This is not a military secret, but an obvious fact. The actual means for accomplishing this is another matter. Further details concerning the atom bomb are to be found in references listed at the end of the chapter, including the Smyth Report.

FIG. 166. ATOMIC BOMB EXPLOSION AT BIKINI ATOLL. U. S. Signal Corps photograph.

This Report is a semipopular but official account of the work of a wartime agency, called the Manhattan Project, which was set up by the government to explore the possibilities of developing atomic power and atomic bombs. Hundreds of scientists, including many of the foremost physicists of the world, were engaged in this Project. The secrecy of the work, as well as its cooperative nature, make it difficult to assign credit to any individuals. In addition to such well-known nuclear physicists as Bohr, Einstein, and Fermi, there could be mentioned many more of outstanding accomplishment. The success of this work in producing an atomic bomb is a matter of history. But it is perhaps even more important to the future that, out of the very necessity of the immediate purpose of developing a bomb, enormous amounts of purely scientific data were taken and many fundamental problems were solved. Some of these results have been made public and may be mentioned in the following pages, but others may not be revealed until much later.

253. Purification of Uranium 235. The biggest problem to be solved was that of obtaining relatively large quantities of uranium 235 in sufficient purity. Uranium itself is a comparatively rare element, and the light 235 isotope constitutes only seven-tenths of one per cent of the whole. Since the chief impurity is then the 238 isotope, the separation is a physical and not a chemical problem. Several processes are possible, although none is very efficient and all are very costly. They will be mentioned only briefly here; more details are given in the Smyth Report and other references.

The most promising method employs what is essentially an enormous mass spectrograph (see Sec. 21) called a Calutron. A beam of charged uranium atoms is shot between the poles of a huge magnet,[5] where it is bent into a semicircular path and separated into two beams of slightly different radii, one for each isotope. Suitable collectors placed at the ends of these curved beams catch the separated isotopes. The separation is close to 100 per cent, but the mass which may be collected in a day is quite small.

Other methods depend upon differences in density of suitable gaseous compounds of the two uranium isotopes. They include centrifuging in ultra-high-speed centrifuges, diffusion through porous walls (see Sec. 162), and thermal diffusion. None of these methods will produce more than a slight separation, so that they must be repeated many many times to build up the concentration.

Even after the separation of uranium 235 from uranium 238 had been accomplished, there was much to be done before a bomb could be made. Foremost was the problem of *critical size,* that is, the minimum mass of uranium 235 needed to set off the chain reaction. There is no such thing as a little atomic bomb. Too small a mass will not explode. The best data available were not good enough to predict this critical size with any certainty, but they did indicate that it would be very large in terms of energy released and devastation wrought. Obviously this critical size could not be worked out by cut-and-try methods; better fundamental data were needed.

254. The Atomic Pile. One possible way to produce a nuclear chain reaction is with uranium 235 which has been purified by one of the

---

[5] The first magnet to be employed for this purpose is the one now serving in the 184-inch cyclotron at the University of California. The name, Calutron, is derived from this circumstance.

methods described in the previous section, and this method has proved very successful in the atomic bomb. A chain reaction might also be possible with the natural mixture of the two uranium isotopes, if a means could be found to prevent the capture of neutrons without fission by uranium 238, or at least to reduce it sufficiently. This method, although it appeared impractical for an atomic bomb, did offer many advantages for the production of an atomic power plant. For this reason, as well as for other even more significant reasons which will appear later, a large part of the activity of the Manhattan Project (an entire section of the work, under the direction of E. Fermi and L. Szilard) was concerned with this process.

If neutrons are not to be captured by U^{238} nuclei, they must be slowed down to thermal speeds by collisions with other nuclei *which will not capture them.* Very pure graphite, purified to a degree never before attained, was found suitable to this purpose. After the effectiveness of graphite for this purpose had been determined experimentally, an **atomic pile** was built at Columbia University to test out the practicability of this idea. This first atomic pile was a cube of graphite 8 feet on a side, containing 7 tons of uranium oxide in the form of lumps which were distributed throughout the cube like plums in a pudding. When a neutron source was placed near the bottom of the pile, it was found that fission was taking place in all the uranium lumps, and the amount of this fission showed that neutrons resulting from the fission processes accounted for a considerable part of it. This is what was anticipated in the design of the pile by Fermi; fast neutrons, escaping from one uranium lump, are slowed down in the graphite before reaching another lump, so that they are there able to produce more fissions. Only the U^{235} nuclei are split apart in this process. The fission-produced neutrons were not sufficient to maintain the process in this pile with the primary neutron source removed. The data indicated, however, that if neutrons could be prevented from escaping through the walls of the pile, the neutrons produced by fission would be more than sufficient to keep the process going. If the pile is made larger, the neutron losses will be proportionally smaller. Hence a nuclear chain reaction is possible in such a pile if it is made large enough.

From these data it was possible to compute the critical size for a pile which would support a chain reaction, and on December 2, 1942, such a pile was successfully operated at the University of Chicago. It started to operate of itself as soon as it reached critical size. *This was the first atomic-energy power plant.* Its power output was controlled and regulated by means of cadmium-steel strips, which absorb neutrons very strongly, pushed into holes running through the pile. The operation could be checked by pushing these control strips farther into the pile, or accelerated by pulling them farther out. The power of this first pile was kept down to a few hundred watts. Since then much more powerful piles have been built at the Oak Ridge plant, near Clinton, Tennessee, and at Hanford, in the state of Washington. The Hanford plant, which is the larger, develops thousands of kilowatts of power.

Strangely enough, none of these atomic piles has yet been run as a power plant. Instead, they are being operated for the production of certain by-products whose present value is far greater than the value of the power developed, and the power is thrown away. Thus the primary purpose of the Hanford plant is the production of *plutonium* (see Sec. 255), and the power developed is dissipated by flowing water through the piles. This power is sufficient to produce a sensible rise in the temperature of the Columbia River, which supplies this cooling water.

FIG. 167. ATOMIC ENERGY PLANT. This U. S. Signal Corps photograph shows a portion of the Hanford Engineering Works, in the state of Washington, where great atomic piles are operated and plutonium is produced.

255. Plutonium. Plutonium is element 94 in the periodic table. It does not occur naturally but must be produced from uranium 238 by the process to be described below. It is important because, like uranium 235, its nuclei will fission upon the capture of slow neutrons.

The story of the discovery of plutonium begins before the discovery of nuclear fission. Shortly after the discovery of the neutron, Fermi reported the discovery of radioactive elements of atomic numbers 93 to 97, which he believed had been created by neutron capture in uranium nuclei. The discovery of fission cast serious doubt upon the existence of these transuranic elements; it was considered possible that Fermi had observed instead some of the radioactive products of fission. The verification of the existence of elements 93 and 94, now known as neptunium (Np) and plutonium (Pu), and the development of a method for the large-scale production of plutonium, are among the prominent scientific accomplishments of the Manhattan Project. The nuclear reactions occurring in the process for the production of plutonium are represented by the following equations:

$$
\begin{aligned}
{}_{92}\mathrm{U}^{238} + n &= {}_{92}\mathrm{U}^{239} + \text{gamma radiation} \\
{}_{92}\mathrm{U}^{239} &= {}_{93}\mathrm{Np}^{239} + \text{electron} \\
{}_{93}\mathrm{Np}^{239} &= {}_{94}\mathrm{Pu}^{239} + \text{electron} \qquad (18\cdot 2)
\end{aligned}
$$

Translated, these equations state that first the U^{238} nucleus may capture a neutron, the binding energy being radiated away as gamma rays. This is the neutron capture without fission which has been discussed in Secs. 251 and 254, for which there is a large probability when neutrons have certain intermediate speeds (in the neighborhood of 25 ev of energy). The U^{239} nucleus which results is radioactive, emitting a beta-ray electron and becoming a nucleus of neptunium. The Np^{239} nucleus in turn emits a beta-ray electron and becomes plutonium. Both these radioactive processes have quite short half-lives, so that plutonium is produced with little delay.

From theoretical considerations it was predicted that plutonium, like uranium 235, should be fissionable by neutrons of all speeds, with slow neutrons most effective. Experiments verified this prediction, which placed plutonium in strong competition with uranium 235 as a primary bomb material, provided that a process could be developed which would produce it in sufficient quantity. The chief advantage over uranium 235 which is possessed by plutonium is that it can be separated from uranium 238 by comparatively easy *chemical* means, in place of the very difficult *physical* processes required for the separation of uranium 235.

256. Plutonium Production. The atomic pile provides a very satisfactory means for the large-scale production of plutonium; some of the neutrons released by fission of U^{235} nuclei are captured by U^{238} nuclei, which then convert into plutonium nuclei by the sequence of reactions shown in equations (18·2). In a pile designed for the production of plutonium the uranium metal is made into rods which are inserted into holes running through the graphite structure. Periodically these rods are withdrawn, the plutonium extracted chemically, and the rods then replaced. All operations, including the chemical processes, must be carried out by remote control, to protect the workers from the intense radiations.

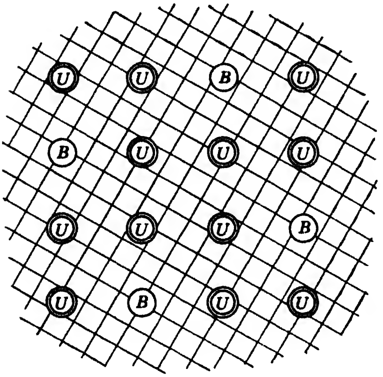


Fig. 168. Cross-Section (schematic) through an Atomic Pile. *U*—Uranium rods encased in tubular aluminum boxes. *B*—Cadmium steel rods for controlling the operation. The cross-hatching represents the graphite pile.

257. Atomic Power Plants. When atomic piles are operated for the production of plutonium, the power generated is a waste product, as has been stated in Sec. 254. By making certain modifications in an atomic pile it should be possible to employ this power for useful purposes, thus making it an atomic power plant. It is still a long way from being the popular conception of an atomic power plant, however. Since such enormous quantities of power may be derived from such very small amounts of nuclear "fuel," the public generally envisions smallness of size as one of the primary advantages of atomic power plants. On the contrary, atomic piles must be of a certain critical size before they will operate at all, and the critical size for raw atomic fuel (natural uranium) is very large, requiring tons of uranium metal as well as graphite. Furthermore, the operation of a pile generates much radioactive radiation, for which bulky and heavy protective shielding must be provided. The completed pile is far too large for anything smaller than a central power station. For such purposes it may eventually prove practical. Already engineers have computed the relative costs per kilowatt-hour for fuel and operation of such a plant. At the present time and for as long as fuels such as coal and oil are plentiful the atomic power plant is unable to compete economically with older sources. Since, however, we must look forward to a time when ordinary fuels are exhausted, it is very important and comforting to know that this other source of energy is available.

The critical size of a pile may be reduced by increasing the proportion of uranium 235 or of plutonium in the metal; such piles are called "enriched" piles. This will increase also the efficiency of operation, but the process of enriching the uranium fuel adds to its cost. Furthermore, even when the plant itself has been made small, it is still necessary to find some way to reduce the bulk of the protective shielding. All these problems are being given serious consideration by scientists and engineers, and the time may come when we shall have practical atomic power plants of small size. There is an immediate need for such an atomic power plant to drive experimental rockets; here the shielding problem is not serious, unless human passengers are to be carried, and a practical atomic-power motor for rockets may be developed fairly soon. There exists also the possibility that another nuclear reaction may be discovered which will serve even better as a power supply.

In closing this section it should be noted that, in the final analysis, all other sources of terrestrial energy, including water power, wind power, oil, and coal, are derived from solar radiation, and that the sun is able to supply that radiation to us year after year because its energy is supplied by another nuclear process, which will be discussed in Sec. 259.

258. The Sun. Speculation concerning the physical nature of the sun and of the other stars has always intrigued the human mind. Little by little, science has been turning speculation into knowledge. All that we know concerning celestial bodies must be brought to us by the telescope. When, however, this instrument is combined with the spectrograph, the radiometer, and the interferometer, this information becomes amazing in its extent. From such observations we have learned not only the distances, motions, diameters, and masses of a very large number of stars in addition to the sun, but also their surface temperatures and chemical compositions. In some ways we know more about the stars than we do about the earth itself. A few of the methods whereby this has been accomplished are discussed in earlier chapters of this book; the others will be found in the books on astronomy and astrophysics.

Observation must stop at the surface of the sun, but speculation, which meets with no such limitation, may attempt to interpret the nature of the interior of the sun from the observed nature of its surface. Much has been accomplished in this direction. For example, all theories agree that, although the temperature of even the surface of the sun (about 6000° C) is very high in comparison with temperatures attainable on earth, the temperature of its interior must run up to millions of degrees. All theories likewise agree that the sun is a mass of gas: a very dense gas, to be sure, but still a gas, since nothing can exist in the solid or liquid state at the temperatures to be found in the sun. Beyond these two conclusions, theories are not always in agreement.

One question with which such speculation has greatly concerned itself, ever since the law of conservation of energy was first recognized, is this: "What is the origin of the energy of stars—or more specifically, of our sun?" Various hypotheses have been considered in attempts to answer this question. The work done in compressing this gaseous sun, as the forces of solar gravity cause it slowly to contract, will furnish a continuous supply of power, and one which has in the past been considered seriously as capable of being the principal source of solar radiation. But it is now known that this source of power is entirely inadequate to account for the constancy of solar radiation over the whole period of time for which the natural sciences give us records. Next, the discovery of radioactivity, together with the large amount of energy which is released by radioactive transformations, provided an even more promising hypothesis; but even this hypothesis could not be adequate unless, for some unknown reason, the sun contains a much larger proportion of naturally radioactive materials than does the earth. Here theories to explain solar radiation rested until very recently.

Our present knowledge concerning the structure of matter makes plausible an even more detailed picture of the deep interior of the sun. At the temperatures and pressures which in all probability exist there, matter can no longer exist as molecules, or even as atoms in the ordinary sense. Collisions between normal atoms moving at speeds corresponding to such temperatures would strip off all electrons from almost every atom, leaving only bare nuclei. (Some of the heaviest atoms might retain their K electrons.) Thus we must consider the deep interior of the sun as being a *gas* in which the particles are not molecules or atoms, but these thousands-of-times-smaller nuclei and electrons. Nevertheless these particles, because of their smallness, are still free to move as particles of a gas, even though they are crowded so close together by the enormous pressures in the sun's interior as to make the density many times greater than that of the densest solid to be found on earth.

259. Solar Energy. We may now look to nuclear transformations induced by collisions among these solar-gas particles for the source of solar energy. Even at 20 million degrees centigrade, the estimated temperature at the interior of the sun, only a small fraction of these collisions are capable of inducing nuclear reactions. But the particles are crowded so closely together that the very frequency of their collisions may compensate for this relative inefficiency; nuclear reactions induced by collisions may be sufficient to account for all solar radiation.

It was shown by R. Atkinson and F. Houtermans in 1929 that the principal reaction in the sun could be the conversion of protons into alpha particles, or helium nuclei, through interactions of protons with other light nuclei. They estimated that, at a temperature between 18 and 20 million degrees centigrade, sufficient energy would be released in this manner to account for the entire solar radiation. As long as the hydrogen lasts, the rate of supply of this energy would be practically constant. According to this theory the *fuel* that keeps the sun going is nuclear hydrogen! The detailed analysis of particular types of nuclear reactions leading to the transformation of hydrogen into helium, and related problems concerning stellar stability and evolution, have been considered in recent times by G. Gamow, E. Teller, H. Bethe, and C. V. Weizsäcker.

The series of reactions considered at present to be most probable is:

$$
\begin{aligned}
C^{12} + H^1 &= N^{13} \\
N^{13} &= C^{13} + \text{Beta ray positron} \\
C^{13} + H^1 &= N^{14} \\
N^{14} + H^1 &= O^{15} \\
O^{15} &= N^{15} + \text{Beta ray positron} \\
N^{15} + H^1 &= C^{12} + He^4 \qquad (18{\cdot}3)
\end{aligned}
$$

This brings us back to the starting point, with 4 protons converted into 1 helium nucleus and 2 positrons. As usual, the 2 positrons disappear with 2 electrons in a suicide pact.

If this theory be correct, then eventually all the hydrogen in the sun will be used up, and a profound change in its brightness and size may be expected. There exist in the heavens various types of stars as regards size, total brightness, and temperature; this theory indicates that certain of these stars, called white dwarfs, represent what our sun will be like when it has exhausted its stock of hydrogen, some ten billion years or more from now.

260. By-Products. All this emphasis upon power production must not be permitted to conceal other important aspects of recent nuclear physics work. First of all comes the advancement of fundamental knowledge concerning the structure of matter, the properties of elementary particles, and the laws which govern the interactions of these particles. A large part of the work of the Manhattan Project was devoted to obtaining more information of this sort. In this case it was a matter of sheer necessity; cut-and-try methods cannot be used in developing an atomic bomb! It was realized more than ever before that practical applications must be based upon complete fundamental knowledge. Some of this knowledge was already at hand, the result of so-called "pure" research which had been carried out earlier without any practical end in view. Much was still unknown when the Manhattan Project started; this had to be discovered by an extensive program of fundamental research, both theoretical and experimental.

Second to this fundamental research stands the development of new and very powerful tools for experimental research. Chief of these is the atomic pile. In many ways it surpasses the cyclotron for producing nuclear reactions and nuclear-reaction products. The artificially radioactive materials which are produced as by-products of the fission process are produced in the atomic pile in such quantities as to make them available at relatively low cost for all kinds of physical, chemical, biological, and medical experiments. Still other radioactive materials may be produced by exposing the parent substance to the intense slow-neutron radiations in the pile. For example, the very useful radioactive carbon 14 may be produced very readily in this manner, although its production otherwise is quite difficult. The parent substance is nitrogen, introduced into the pile in the form of some nitrogen compound. The reaction is

$$N^{14} + n^1 = C^{14} + p^1 \tag{18·4}$$

261. Trans-Uranic Elements. One significant contribution of wartime research to fundamental knowledge was substantiation of the discovery of elements 93, 94, 95, and 96, now called neptunium, plutonium, americium, and curium (Np, Pu, Am, and Cm). It will be recalled that, before the fission of uranium was discovered, Fermi announced the discovery of a series of *trans-uranic* elements, that is, elements beyond uranium in the periodic table. (See Sec. 255.) He based this discovery upon analysis of certain radioactive radiations emitted from uranium which had been bombarded with neutrons. After the discovery of fission the existence of these elements was questioned, since the fission products seemed able to account for the radiations studied by Fermi. Secrecy then shrouded all nuclear research until after the war, when it was revealed that the existence of at least four of these trans-uranic elements had been confirmed, and that the first two had been manufactured in large quantity. This has been described in Secs. 255 and 256.

All these elements are radioactive, and none exists naturally. Various nuclear reactions serve to produce them. Chemically they form a series in the periodic table which begins with actinium and is analogous to the rare-earth series of elements. For these elements, as for the rare-earth elements, the electron configuration in the two outer shells (*P* and *Q* shells) is the same for all; as the atomic number increases, the additional electrons go into the next shell below, or the *O* shell. All these elements may prove sometime to be of practical value, as has plutonium.

Their discovery makes its most important contribution, however, in the further information which it supplies concerning the structure of matter. For many years scientists had wondered why the list of natural elements ended with element number 92. This question now has been answered quite fully. If these elements ever existed naturally, they disappeared long ago by radioactive disintegration.

PROBLEMS

1. Compute the density of an oxygen 16 nucleus, if its radius is 3×10^{-13} cm. (All nuclei have about the same density.)

2. Tin 120 has an atomic weight of 119.91. (*a*) How many protons and how many neutrons does its nucleus contain? (*b*) Compute the loss of mass which results when this tin nucleus is built up from these nucleons. (*c*) What binding energy, in Mev, does this represent?

3. Boron 11 has an atomic weight of 11.0129. (*a*) Compute the binding energy of a boron nucleus. (*b*) Compute its binding factor.

4. Assume the mean diameter of the Calutron beams to be 180 inches. Compute the separation in inches of the two beams at the collector.

5. Each alpha particle emitted by radium has 4.74 Mev of energy. Compute the energy released in this manner by the decomposition of 1 gram of radium. (The total energy released by 1 gram of radium and its subsequent radioactive product elements would be roughly 10 times this figure.) Express the result in kilowatt-hours.

6. Compute the energy released by fission of 1 gram of uranium. (See Sec. 250.)

7. Compute the energy released when 1 gram of hydrogen is converted into helium in the interior of the sun. Express the result in kilowatt-hours.

8. The earth receives about 10^{12} kilowatts of power from the sun. Compute the mass of hydrogen which must be converted into helium in the sun to supply this power to the earth for one day.

REFERENCES

H. DeW. Smyth, *Atomic Energy for Military Purposes.* (This is the Smyth Report.)
G. Gamow, *Atomic Energy.*
G. Gamow, *Birth and Death of the Sun.*
J. A. Wheeler, "Elementary Particle Physics," *American Scientist,* Vol. 35, April, 1947, pp. 177–193.
P. Morrison, "Physics in 1946," *Journal of Applied Physics,* Vol. 18, February, 1947, pp. 133–152.
M. A. Tuve, "The Forces Which Govern the Atomic Nucleus," *Scientific Monthly,* Vol. 47, pp. 344–363, October, 1938.
Physics Staff, University of Pittsburgh, *Atomic Physics,* Chapter XIV (Astrophysics).
See also the references at the end of Chapter XVII.

CHAPTER XIX

COSMIC RAYS

262. Cosmic Rays. For further knowledge concerning the elementary particles of physics we shall now turn to a consideration of cosmic rays. Indeed, we have already had occasion to do this, briefly, in connection with the discovery of the positron, as well as in discussing the mass of energy. Now we shall do so more systematically.

Although the study of cosmic rays now ranks as one of the major fields of scientific research, it had, as often happens in scientific discovery, a very humble beginning. In 1911, not long after the discovery of radioactivity, Hess became curious to know why a perfectly insulated electroscope would slowly but surely lose its charge. It generally was assumed that this loss was due to very weak radioactive radiations coming from the earth, and this was indeed the explanation in part. Hess put this explanation to the test by carrying an electroscope away from the earth in a balloon and thereby discovered that, instead of decreasing, the loss of charge increased! When care was taken to inclose the electroscope in an airtight chamber, so that the density of the air in this ionization chamber did not diminish as it was carried to higher altitudes, the ionization increased many times. Quite evidently the cause was some sort of radiations coming from *outside* the earth, not from within. Since further experiments indicated that these radiations were not coming from the direction of the sun, or indeed from any one part of the heavens more than from any other part, and that they were exceedingly constant in their intensity, they were called *cosmic rays.*

263. Electroscopes. The first cosmic-ray studies were made with ionization chambers and electroscopes, in essentially the same manner as described in Chapter XV for the study of radioactive radiations. Since cosmic rays are all very penetrating, the ionization chamber and the electroscope may be completely inclosed and made airtight where necessary. They are usually provided with a device, introduced by Neher, which automatically recharges them at regular intervals. The record is made photographically, and appears as a zigzag or sawtoothed line, as is indicated in Fig. 169. The vertical rises represent the periodic recharging, and the slanting portions, the discharge due to cosmic rays. The steepness of this slanting portion measures the intensity of the cosmic radiations.

Electroscopes still form an important part of the equipment for cosmic-ray studies, although they are now supplemented, and in many cases supplanted, by other devices, such as the Geiger-Müller counter and special Wilson cloud chambers.

FIG. 169.

264. Geiger-Müller Counter. The Geiger-Müller counter (or G-M counter, as it will hereafter be called) is shown in diagram in Fig. 170. The radiation-sensitive part is the *counter* tube, *A*, a metal tube closed at its ends by plugs of insulating material, with a fine wire, *W*, stretched along its axis. This closed tube contains air or some other gas, usually at a reduced pressure. The metal tube *A* is charged to a high potential by means of the battery *B*, while the wire *W*, which serves as the other electrode, is connected through the condenser *C* to the grid of the amplifier tube *D*. *W* is also connected to the ground through the very high resistance *R*, so that normally it is at zero potential.

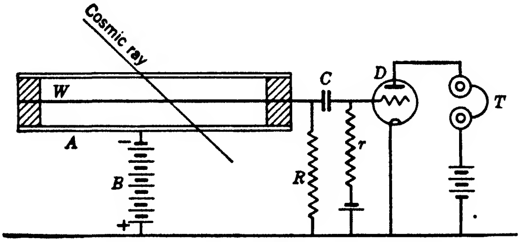


FIG. 170. GEIGER-MÜLLER COUNTER.

The battery potential is close to that at which a discharge will pass between *A* and *W*. The potential required varies from 500 to 3000 volts, depending upon the dimensions of the tube, the kind of gas, and its pressure. When a cosmic-ray particle passes through the tube, it produces ions in the inclosed gas, and a momentary discharge does pass. (Since the cosmic rays are so penetrating, they may pass through the counter tube in any direction, as for example the direction indicated in Fig. 170.) This discharge lowers the potential of *W*. Since *W* is connected to the grid of the amplifier tube *D*, by the condenser *C*, the potential of this grid likewise falls momentarily; and the resultant sudden decrease in the current flowing to the plate of D causes a click to be heard in the telephones T. The discharge is but momentary, since C soon charges up and the current through the very high resistance R is too small to support it. The charge on C then leaks off through R and r, and the apparatus is ready to record the arrival of another cosmic-ray particle. The circuit may be designed so that the whole process takes but a few thousandths of a second. Even though the average cosmic-ray intensity, as indicated by an electrometer, may be quite steady, the clicks in the telephone will be heard at very irregular or random intervals, the deviations from the average being in accord with the laws of probability or chance. The circuit here shown has been chosen as the easiest one to describe. In actual practice, the circuits are more sensitive but more complicated; a considerable number of electron tubes are employed, and the current through the final tube may be great enough to operate automatic recording apparatus.

Gamma rays and X-rays will also operate a G-M counter, as has been mentioned in Sec. 213; if the walls of the tube are made thin enough, beta rays may also be detected and counted. G-M counters may also be built which will detect neutrons, as explained in Sec. 226. Since neutrons do not themselves ionize the inclosed gas, neutron counters are built with a cylinder of silver, whose atoms are rendered radioactive by capture of neutrons. The electrons which are subsequently emitted by the radioactive atoms then actuate the counter.

265. Linear Amplifier. In the G-M counter, the ionization produced by the radiations passing through it acts as a *trigger* to set off a momentary discharge through the tube, and both strong and weak rays produce equal effects. The simple ionization chamber is far less sensitive, but the lower sensitivity may be compensated for by using with it a very sensitive electron-tube amplifier circuit. This combination of ionization chamber and electron-tube amplifier, which is known as a linear amplifier, has the advantage that it not only counts the particles passing through it, but measures their intensity as well. The record is usually made photographically, on a strip of photographic paper or film. Each ray passing through produces a sharp peak in the record, and the more intense the ionization produced, the higher the peak.

The *proportional counter* is a modification of this linear amplifier which in some way resembles the G-M counter. Like the G-M counter, it has a high potential across its ionization chamber, so that the current is amplified many times because of ions formed by collision. In the G-M counter, however, this potential is so high that every particle counted produces the same output voltage, whereas the potential across the ionization chamber of the proportional counter is lowered to a point where the output voltage pulses are approximately proportional to the ionizing power of the particles counted. These voltage pulses are further amplified by an electron-tube amplifier, so that they may be recorded photographically or reproduced on the screen of a cathode-ray oscilloscope.

266. Cloud Chambers for Cosmic Rays. In the first few years of cosmic-ray studies, when the only evidence concerning them was obtained by electroscopes and G-M counters, it was largely a matter of conjecture whether the cosmic-ray particles were photons, electrons, protons, or even some (at that time) unknown particles. The Wilson cloud chamber was turned to for further information. For cosmic-ray studies the cloud chamber, which is essentially the same as the one shown in Fig. 16 and described in Sec. 25, is made quite large, and is set up edgewise. At first, success in obtaining tracks for cosmic-ray particles was left to chance. Many successive expansions were made and photographed, and the photographs were then searched for evidence. Later, Blackett and Occhialini devised a scheme whereby the cosmic rays were induced to take their own photographs. Two G-M counter tubes were arranged, one above and one below the cloud chamber, and the circuits operated by them were so arranged that the chamber was expanded and the camera shutter tripped when, and only when, both tubes "fired" simultaneously, thus indicating that a cosmic ray had passed through the chamber and the two G-M tubes.

Anderson, in 1931, added two more features to the equipment of the cosmic-ray cloud chamber. First he placed his cloud chamber inside a very strong and uniform magnetic field, whose lines of force were perpendicular to the face of the chamber. Next he placed a thick lead plate edgewise across the middle of this chamber. This is clearly seen in Fig. 151, which is a photograph taken with this chamber. A cloud chamber thus equipped provides three different sorts of evidence concerning the particles which pass through it. First, there is the density of the ionization which appears along the path of the particle. For singly charged particles this density is approximately inversely proportional to the speed of the particle. Second, there is the curvature which the magnetic field produces in the path of every charged particle. If the particle is singly charged, as practically all of them are, its momentum is directly proportional to $B\rho$, the product of the magnetic flux-density, B, and ρ, the radius of the curve made by its path (as has been explained in Sec. 239). Finally, the loss of momentum which occurs when a particle passes through the lead plate may be used to differentiate between different types of particles, as will be discussed below.

An exceedingly important function also of the lead plate is to determine the *direction* of passage of the particles through the cloud chamber, since this makes it possible to distinguish between positively and negatively charged particles. It was this property of the plate which made possible, as has been explained in Sec. 223, the discovery of the positron by Anderson in the course of these cosmic-ray studies.

267. Cosmic-Ray Particles. With all this varied equipment the discovery of the exact nature of cosmic rays would seem to be a straightforward matter, involving merely the making of enough observations. But this has not proved to be the case. To be sure, cloud chambers demonstrate clearly that the cosmic-ray particles which affect electroscopes and counter tubes are electrically charged particles, and not the photons which, because of their great penetration, they were at first assumed to be. But, whenever the tracks which appear in cloud-chamber photographs may be attributed to cosmic-ray particles, they are almost always very similar in appearance. Almost always they are thinly ionized tracks, such as are produced by very high-speed electrons or beta-ray particles, and differ one from another only in the different curvatures which appear in a strong magnetic field.

From the curvatures of these tracks it may be seen that the particles have very high momenta indeed. In magnetic fields which are strong enough to curl the cathode-ray electrons in a powerful X-ray tube into circles a few millimeters in diameter, these cosmic-ray particles make tracks of comparatively small curvature. Some of them indeed have so much momentum that they show no observable curvature at all. On the assumption that these particles are the electrons which they appear to be, their energies range from several ten-millions of electron-volts upward; and some possessing energies as high as ten billion electron-volts have been observed in cloud chambers under conditions where their energy could be estimated with some accuracy.

However, it is not safe to assume that these particles are all electrons (or positrons).[1] When particles possess energies as great as do these particles, distinctions which depend upon rest-mass, that is, upon mass as measured at comparatively low speeds, become quite small. At ten million electron-volts (10 Mev), of energy, the effective mass of an electron is twenty times its rest-mass, the increase being due to the mass of its kinetic energy, as explained in Sec. 212. To repeat equation (15·3) given there,

$$m = m_0 + \frac{\text{Kinetic energy}}{c^2} \tag{19·1}$$

---

[1] Both positively and negatively charged particles appear with about the same frequency. To avoid awkwardness, "electron" will be used hereafter as a generic term for both electrons and positrons.

If V is the kinetic energy in electron-volts, this may be written

$$m = m_0 + \frac{Ve}{300c^2} \tag{19·1a}$$

The relative masses of high-speed electrons may quickly be figured if it is remembered that kinetic energy equivalent to about half a million electron-volts (0.51 Mev) has a mass equal to the rest-mass of the electron. If this law, which has been found to hold for beta rays, still applies when extrapolated to a billion electron-volts, a one-billion-volt electron will have a mass equal to that of a proton at rest!

Distinctions which depend upon differences in speed are even more completely wiped out at such high momenta and energies. As is explained in Secs. 211 and 212,[2] electrons having but a half million electron-volts of energy are moving with 87 per cent of the speed of light, while at 3.5 Mev (corresponding to very energetic beta rays) the percentage is 99, and at 10 Mev it is 99.9. Even though the rest-masses of cosmic-ray particles might be considerably greater than the rest-mass of the electron, their speeds and energies would be almost the same as those figured for electrons having the same momenta. To a first approximation, the speed of any cosmic-ray particle is the speed of light, c, and its energy is mc^2, or c times its momentum. Since the momentum of a particle observed in a cloud chamber placed in a magnetic field is proportional to $B\rho$, this first approximation to the energy is also proportional to $B\rho$. Indeed, it can easily be shown that the approximate energy in electron-volts equals 300 $B\rho$.

268. Showers. One of the most astonishing of the early discoveries was the paradoxical increase in ionization which might be produced by surrounding an ionization chamber with a sheath of a few millimeters of lead. G-M counters likewise showed increased counts under similar conditions, and the effect in both cases is now explained as due to *showers* of secondary particles being produced by the cosmic rays which impinge upon the lead sheath. Cloud-chamber photographs, such as Figs. 171, 172, and 173, indicate what is occurring. In Fig. 171, a group of such shower particles apparently has its origin in the upper right-hand part of the apparatus just outside the camera field. Some of the particles are stopped by the dense metal plate placed across the middle of the cloud chamber, but a portion of them penetrate it with an increased

---

[2] Since (Sec. 211)

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}$$

$$v = c\sqrt{1 - m_0^2/m^2} \tag{19·2}$$

Fig. 171. Cosmic-Ray Shower. Cloud-chamber photograph by Professor P. Auger of the University of Paris.

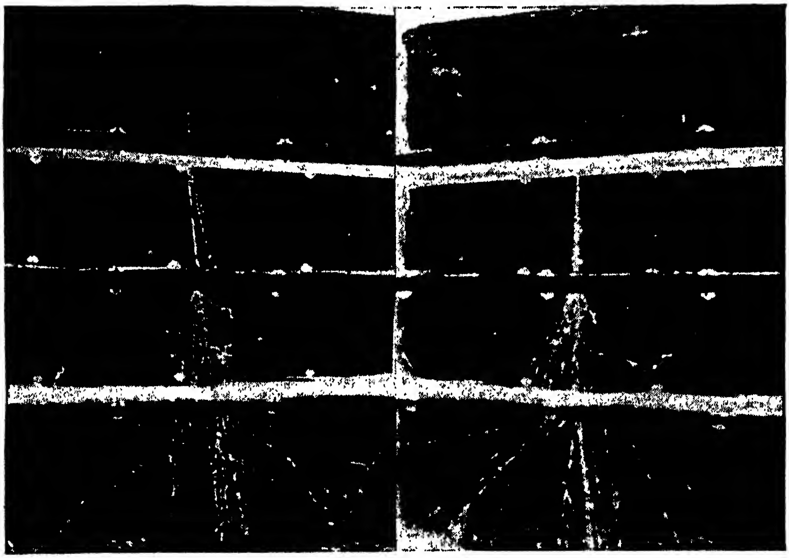


Fig. 172. Cosmic-Ray Shower. (See footnote 3.) The two photographs show the same shower from different angles.

multiplication of particles appearing below it. The first detailed studies of this phenomenon were made by Street and Fussell, with a cloud chamber which had three horizontal lead plates placed across it, one above the other like shelves in a cupboard; with this arrangement the development of showers is shown in a most beautiful manner. Figure 172, which is reproduced from one of their photographs,[3] shows a single particle incident upon the top plate, accounting for, by successive multiplication within the plates, the score or more of particles which emerge from the bottom plate.

Sometimes a shower fills the cloud chamber with rays, but the most informative showers are the small ones. Street and Fussell showed that, when the plate is quite thin, the effect which occurs most often is the production of a *pair* of particles. Figure 173 shows two such pairs. One pair appears to be produced by a charged particle which itself passes straight through the plate from the top, while the other pair has no apparent cause at all.

269. Shower Theory. Shower production, and especially pair production, are phenomena which are predicted by the quantum theory of the electron as developed by Dirac. This theory predicts that, when an electron passes through the strong electric field close to the nucleus of an atom, it should sometimes lose instantaneously part or all of its energy, the lost energy appearing as a photon of electromagnetic radiation. This is the explanation for the continuous spectrum of X-rays (Sec. 144), and the experimental data for X-rays are in good agreement with this theory. When the photons themselves are of very high energy (in excess of 1 Mev), they may in turn interact with the electric field near a nucleus to produce an electron-positron pair, as has been described in Sec. 223. Figure 173 shows both of these processes very beautifully. According to the theory, both of the pairs appearing below the plate are produced by photons. The initial particle, which seems to be the cause for one of the pairs, does so indirectly. Somewhere in the plate it has lost energy so as to produce a photon, and the photon in turn has reacted within the plate to produce the pair out of its own energy. The same primary particle evidently is responsible for the second pair of particles, but the photon in this case has been produced somewhere above the cloud chamber and has crossed the upper part of the chamber before "materializing" as a pair of particles. This figure shows also how, when photons are produced in this manner by a charged particle of high energy, they are driven forward in the direction of the

---

[3] "Cloud-Chamber Studies of Cosmic-Ray Showers and Penetrating Particles," J. C. Street, *Journal of the Franklin Institute*, Vol. 227, pp. 765–788, June, 1939. This article shows many other beautiful shower photographs.

motion of the initiating particle; this is also in accord with the theory, especially as it has been developed by Euler. The narrow cone formed by the X-ray beam from the betatron target (see Sec. 237) is a beautiful verification of this theory. Figure 172 shows especially well how the most energetic parts of the beam drive forward in the direction of the original particle, the particles which diverge very much from this direction being of low energies and easily stopped by the metal plates.

All these phenomena, and many others as well, indicate that these shower-producing particles are electrons, and that the quantum theory for the electron remains valid for such high-energy electrons.

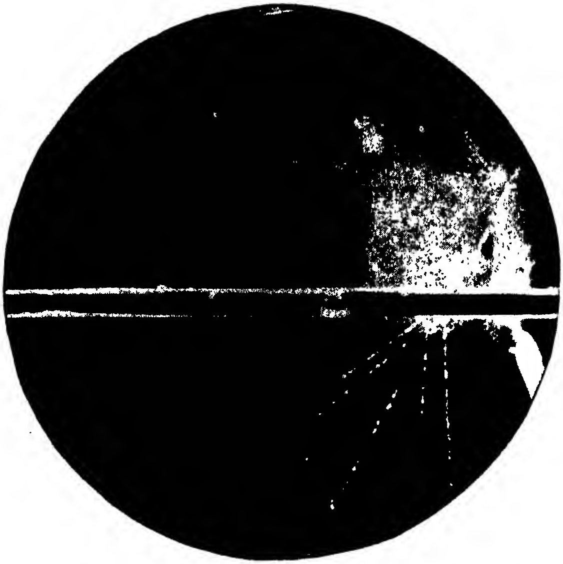


FIG. 173. PAIR PRODUCTION IN COSMIC RAYS. Cloud-chamber photograph by Professor P. Auger of the University of Paris.

270. Atmospheric Showers. This same phenomenon of showers may be produced by interactions between the shower-producing electrons and atomic nuclei in the atmosphere, as has been demonstrated by many experiments. Auger, in France, made an especially thorough study of such showers, using coincidence counters. Two G-M counters may be so connected to the recording apparatus that they will respond only when both of them are "fired" simultaneously (that is, as close together as the response time of the counter circuits, which was about a millionth of a second in Auger's apparatus). When the two counters are placed on the same level and some distance apart, a coincidence will indicate that they have been fired by two *separate* particles originating in the *same* shower. Auger found that the main part of the shower was concentrated in a narrow beam, just as was predicted theoretically by Euler and as is shown by the cloud-chamber pictures of showers produced in dense materials. See Figs. 171 and 172. For these atmospheric showers this beam was perhaps 20 feet in diameter at the surface of the earth. But his counters showed some coincidences at much greater distances than this. His best observations were made high on the side of Jungfrau mountain, in the Swiss Alps, where the showers were much more intense than at sea level; here he found coincidences due to shower particles when his counters were as much as 1000 feet apart. From these data he estimated that some atmospheric showers cover as much as a million square feet of area and may contain as many as a million particles. Thence he estimated that the primary particle which originates such a shower must have at least 10^{14} or 10^{15} electron-volts of energy.

271. Hard and Soft Components. Early in the study of cosmic rays, measurements of their absorption by dense substances revealed the existence of at least two components, the one called "soft," the other, "hard" or "penetrating" rays. Even the soft rays are penetrating enough that some can pass through as much as 5 cm of lead, but the hard rays are only partially absorbed by as much as a meter of lead. At first it was thought that both components might consist of electrons, the hard rays differing from the soft only by reason of possessing greater energies. Then it was shown that showers are produced almost entirely by the soft component, and this presented a serious difficulty to the first explanation. Since the shower theory (Sec. 269) predicts that the production of showers by electrons should *increase* with their energy, the hard component could be electrons only if the theory does not hold for such very high-energy electrons.

However, when magnetic analysis was applied to both kinds of particles, these explanations were shown to be impossible. Anderson and Neddermeyer, for example, using their big cloud chamber in a strong magnetic field, with a platinum plate one centimeter thick across the middle of it, showed that the *momenta* of both kinds of particles were comparable. Throughout the range of momenta studied (momenta measured by $B\rho$ values between 2×10^5 and 15×10^5 gauss-cm) they found both soft and hard cosmic rays: both shower particles, which were easily absorbed in the plate, and particles which did not originate in showers and did not produce showers, but which on the average lost comparatively small amounts of energy in the plate. Street and his co-workers, using quite different methods, substantiated these findings, as did various other workers.

272. The Meson, or Mesotron. Since the particles forming the soft component can be identified as electrons, by the good agreement between theory and experiment in the production of showers, the simplest way to account for the penetrating particles is to assume that they are more *massive* than the shower-producing electrons.[4] But no such a heavy particle was then known. Protons would be entirely too massive. Protons having even the highest of these momenta would have speeds less than half the speed of light, and would leave much more densely ionized tracks in the cloud chamber than those which are observed. Neither would their energies be high enough to enable them to penetrate the platinum plate.

Although no particle was then known which had a mass between the mass of the electron and that of the proton, these and various other experiments seemed to indicate the existence of such a particle, and "heavy electrons" began to be talked about. The "heavy-electron" tracks appeared very little different from those produced by electrons, but some of them, especially those which showed the most curvature, appeared also to be somewhat more densely ionized; from this increased density it was estimated that the mass of this new particle might be very roughly one-tenth the mass of the proton, or something like 200 times the electron mass. It has been given several names: *barytron* was the first, but this was quickly followed by *mesotron* and *meson*. Both *meson* and *mesotron* are in current use.

273. Meson Theory of Yukawa. It is highly probable that these early estimates for the mass of the meson were influenced by a theory which had been advanced by a Japanese physicist, Yukawa, to account for the forces among nuclear particles (protons and neutrons) which hold the nucleus together. In this theory, which is too abstruse to review here, Yukawa postulated the existence, by production within the nucleus, of particles having between 100 and 300 times the electron mass. According to this theory, these particles would have also a very brief existence, of the order of magnitude of 10^{-6} to 10^{-7} sec on the average. Neutral particles, as well as charged particles of both sign, were required. It seemed probable that mesons were these Yukawa particles, especially so as more and more of his predictions were found to be in agreement with experimental data.

274. Mass of the Meson. The apparent agreement with theory of the first rough values for the mass of the meson supplied experimenters

---

[4] According to the theory of shower production, ability to produce showers decreases very rapidly with increase in rest-mass of the particle (it is in inverse proportion to the square of the rest-mass). Since the absorption of electrons by a metal plate is mostly due to the shower-producing process, any particle which does not produce showers would be comparatively penetrating.

with an added inducement to devise special apparatus and methods for more precise measurements of this mass. In general these measurements involve determination of the *momentum* of a fairly slow-moving meson, from the curvature of its track in a magnetic field, plus determination of its *energy* in some manner, such as its range in the gas of a large cloud chamber. In a recent determination [5] the energy for fairly fast mesons was determined by observing, in an apparatus devised by Street, the number of lead plates they could penetrate after passing through the cloud chamber in which their momentum was measured. The lead plates were in a second cloud chamber, placed below the first one, and coincidence counters operated both cloud chambers simultaneously when a meson passed through both. The results of this experiment seemed to indicate that all electrically charged mesons have the same rest-mass, the value being close to 200 times the rest-mass of the electron. Other experimental results appeared for the most part to be consistent with this value, although a few data indicated the presence of some mesons of lesser mass, and some of greater mass. Other evidence for the existence of different kinds of mesons will be considered in Sec. 277.

275. Life of the Meson. A considerable number of experiments indicate that the absorption of mesons is greater in the atmosphere than in an equivalent thickness of water or earth or lead, and this may be explained by assuming that mesons have very short lives. According to this explanation, the loss in going through the atmosphere is partly true absorption loss, and partly loss due to *radioactive disintegration.* In passing through an equivalent thickness of denser material, the absorption loss is just as great, but the time of passage is smaller and hence the loss by disintegration is less. The slight decrease in cosmic-ray intensity in the summer, as compared to the winter, may be explained in this same way. Although the atmosphere has the same total absorption summer and winter, it is thicker through in summer, and more mesons would die on the way through it.

This explanation for these phenomena is pretty commonly accepted at present, partly because of its simplicity, and partly at least because it represents another point of agreement between these experimentally discovered particles and the heavy nuclear particles postulated by Yukawa. As mentioned above, Yukawa predicted that his particles would disintegrate in the same manner as radioactive nuclei (see Sec. 214) with a half-life of around one-millionth of a second, and experimental data indicate a half-life value of about two-millionths of a second. This is the half-life of a meson at rest, or moving very slowly. Relativ-

---

[5] W. B. Fretter (University of California), "Mass of Cosmic-Ray Mesotrons," *Physical Review*, Vol. 70, pp. 625–631, November, 1946.

ity theory (see Appendix V) predicts that the faster a meson moves the longer it will live on the average, in the ratio of 1 to $\sqrt{1 - v^2/c^2}$, and this very surprising prediction is verified by experiments! The fate of the meson upon disintegration appears generally to be the production of an electron and a neutrino; this is predicted by theory, and evidence in support of it is found in many cloud-chamber photographs. E. J. Williams[6] first obtained a cloud-chamber photograph showing such a disintegration. The meson track comes to an end within the cloud chamber, and from this point starts the track of a very energetic electron.

When slow mesons come to rest within a block of some solid substance the decay electrons may be detected by G-M counter tubes placed around the block. Since the appearance of each electron marks the death of a meson at rest, this apparatus provides another and much better means for measuring the half-lives of mesons.[7] Furthermore, this apparatus may be modified so as to observe positively and negatively charged mesons separately. One way to do this is to pass a meson beam through strongly magnetized blocks of iron, at right angles to the lines of flux, so that the magnetic deflections will separate mesons of one charge from those of the other. Then it is found that the positively charged meson is the one which has a half-life of about 2 microseconds.[8] Negatively charged mesons are found to behave differently. In air, and in solids of low atomic weight such as carbon, they generally disintegrate radioactively, as do positively charged mesons, but with a shorter half-life.[8] In solids having large atomic weight a negatively charged meson is far more likely to be captured by the nucleus of an atom before it disintegrates in this fashion. This is what might be expected as a result of the attraction between its negative charge and the positive charge of the nucleus; the strange thing is that this is not so with the lighter nuclei.

276. Stars. Sometimes a cloud-chamber photograph may record a really cataclysmic nuclear event as a set of several, in rare instances a dozen or more, densely ionized tracks radiating from a common center. The dense ionization indicates that the particles involved are protons or even heavier particles; the "star," as this record is called, shows the tracks of the charged fragments of a nucleus which has disintegrated with explosive violence. Few stars appear at sea level, but the number increases rapidly with altitude.

---

[6] *Nature*, January 20, 1940. The photograph is there reproduced.

[7] See F. K. Richtmyer and E. H. Kennard, *Introduction to Modern Physics*, p. 719, for a description of such experiments.

[8] G. E. Valley and B. Rossi give 2.19 ± 0.24 microsec for positive mesons and 0.74 ± 0.17 microsec for negative mesons. *Physical Review*, Vol. 73, p. 177, January 15, 1948.

Special photographic plates provide even simpler means for recording such nuclear explosions. If a specially prepared photographic plate, carefully wrapped to exclude all light, is left for several weeks on a high mountain top, after development it will show many stars, microscopic in size but otherwise similar to those seen in a cloud chamber. See Fig. 173*A*. For some time it has been known that very energetic charged particles will leave tracks in a photographic emulsion which appear as microscopic lines of silver grains after the plate has been developed, but at first this phenomenon was considered of little importance. Now, largely through its use by Professor Powell and the group of physicists working with him at the University of Bristol, in England, it has become of great value to the study of nuclear processes, such as these stars. The length of each line appearing in the emulsion provides a record of the energy of the particle which made it, while the density of the silver grains along the line supplies evidence concerning its speed; thus tracks may be identified as having been made by protons, mesons, or electrons.

These star records show that the nuclear disintegrations which produce them may have been touched off in a variety of ways. Often the cause appears to have been a cosmic ray proton, and at other times it appears to have been a neutron. In February of 1947 Occhialini and Powell reported six stars, found in plates exposed for six weeks on Pic du Midi in the Pyrenees Mountains, in which it seemed quite certain that the nuclear explosion was set off by a particle *lighter* than a proton. Grain density and other characteristics of its track indicated that it was probably a meson. If so, this was a very important discovery. With a staff increased to include C. M. G. Lattes and H. Muirhead, they undertook an intensive search of their plates for more evidence concerning mesons, and by May, 1947, they had found 65 meson tracks which ended in the emulsion. Most of these tracks just came to an end, but 15 produced stars, thus confirming the earlier observation. In 10 cases a single new particle started out from where the meson track ended. For some of these the new particle was a heavy one; these could be interpreted as representing explosions in which only one charged particle was emitted. But in 2 cases the new particle was quite definitely a meson!

277. Heavy Mesons. More plates were exposed, this time on a higher mountain in the Bolivian Andes. By October, 1947,[9] they had discovered over 600 meson tracks which ended in the emulsion. Forty of

[9] G. P. S. Occhialini and C. F. Powell, *Nature*, Vol. 159, pp. 93–94, January, 1947; pp. 186–190, February 8, 1947.

C. M. G. Lattes, H. Muirhead, G. P. S. Occhialini, and C. F. Powell, *Nature*, Vol. 159, pp. 694–697, May 24, 1947.

C. M. G. Lattes, G. P. S. Occhialini, and C. F. Powell, *Nature*, Vol. 160, pp. 453–456, October 4, 1947; pp. 486–492, October 11, 1947.

these showed secondary mesons produced at the point where the primary meson came to rest, and of these 40 secondary mesons, 11 came to rest without leaving the emulsion. All these 11 secondary mesons had essentially the same range, indicating that the process which produced them was a single "radioactive" process.

It is an unescapable conclusion that the primary meson is a different particle from the secondary meson, and a heavier particle, and the grain densities along the tracks of the two particles bear out this conclusion. Powell and his associates named the heavy primary mesons *π-mesons*, and the lighter secondary ones, *μ-mesons*. Their very thorough study of these records indicates that the μ-meson is the ordinary meson, with a rest-mass about 200 times that of the electron, and that the π-meson is a newly discovered particle which disintegrates radioactively into a μ-meson and one other particle. This other particle is uncharged and hence undetected and unidentified by these experiments.

If we knew the mass and energy of this neutral particle the rest-mass of the π-meson might be determined from these observations. Or it might be measured directly by taking the apparatus described in Sec. 274 to a high elevation where π-mesons are more abundant. Physicists at the University of California have built an apparatus of this type which is light enough to be carried in an airplane, and plan to carry out this experiment in the near future.

278. Creation of Mesons. By this time (October, 1947) Powell and his group had over 100 clear records of stars produced by mesons; mesons which can do this they call σ-mesons. They have observed also stars in which mesons are among the ejected particles, and in several instances they have observed that the newly created meson produces a new star. In one instance the newborn meson apparently disintegrates into a μ-meson. From these data they deduce tentatively that both π-mesons and σ-mesons are heavy mesons, the σ-mesons being negatively charged and the π-mesons, positively charged. In the emulsion the σ-mesons always are able to enter a nucleus before they disintegrate. In free space they might be expected to disintegrate into a negatively charged μ-meson and a neutral particle. If this hypothesis is correct, then all μ-mesons may have had their origin in the disintegration of heavy mesons.

No cosmic-ray evidence for the production of mesons is quite as satisfactory as would be their production by laboratory methods. Theory indicated that they might be created by sufficiently energetic gamma rays, and hope was high that they could be produced by the very hard X-rays from a betatron. So far this has proved to be a false hope. But in February, 1948,[10] two physicists at the University of California, E. Gardner and C. M. G. Lattes, did succeed in creating mesons by shooting high-energy alpha particles against a carbon target. The alpha particles were given 380-Mev energy by means of the great cyclotron, and the mesons were detected by tracks which they produced in a photographic plate placed near the target, the method with which Lattes had become so familiar in his work in Powell's group. The photomicrograph reproduced in Fig. 173*A* shows one of these meson tracks. In the case shown here the meson ends its track by capture in a nucleus, which then explodes and forms a star. Measurements at the University of California indicate a rest-mass for these mesons of about 300 times that of the electron, indicating that these are heavy mesons. Other target materials and other charged particles serve also to produce mesons in this manner. The production of mesons artificially has all the usual advantages over natural methods. Not only can they be produced under controlled conditions but they can be produced in far greater quantity than by natural methods. A whole new field of experimentation is thus opened up.

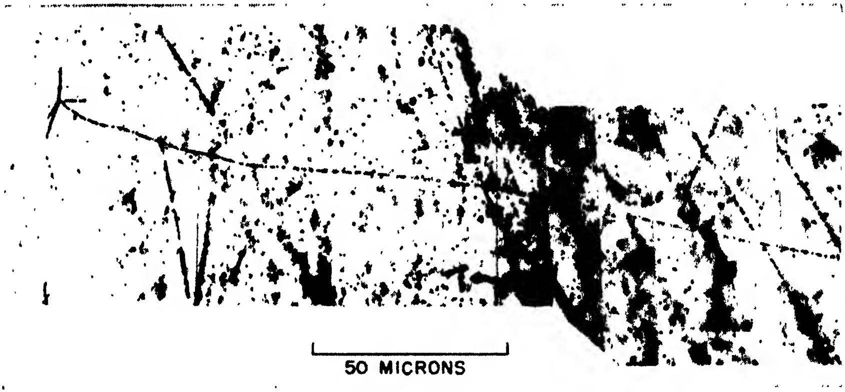


FIG. 173A. PHOTOMICROGRAPH OF THE TRACK OF AN ARTIFICIALLY CREATED MESON. See text, Sec. 278. This track starts at the right edge of the figure and ends in the star near the left edge. This picture first appeared in *Science News Letter* of March 20, 1948, and is reproduced here through the kindness of the publishers and of Dr. Gardner and Dr. Lattes, the discoverers.

The discovery of heavy mesons offers considerable aid to the theoretical physicists who are trying to bring about agreement between the experimental evidence concerning mesons and the meson theory for

---

[10] "How Mesons Were Made," *Science News Letter*, Vol. 53, pp. 181-182, March 20, 1948.

nuclear forces. The reluctance of ordinary mesons to be captured by light nuclei, which has been noted in Sec. 275, indicates that the interaction forces among nucleons and ordinary mesons are too weak to be the forces postulated by Yukawa to bind the parts of a nucleus together. Now it appears possible that heavy mesons may be the Yukawa particles. The properties of these heavy mesons must be known much better before this can be certain.

279. Cosmic Rays Are Secondary Particles. At sea level about 20 per cent of the observed cosmic rays are electrons, the remainder being mesons. Cloud-chamber photographs show, in addition, an occasional heavily ionized track representing a proton or an even heavier ion, as well as evidence for the presence of neutrons; presumably these are fragments resulting from nuclear explosions. None of these particles is a primary particle, however, in the sense that it has come from without the earth's atmosphere. Even the most penetrating mesons are short-lived particles which must have been created somewhere in the upper atmosphere. And the electrons found at sea level must be shower particles many times removed, in the genealogical sense, from the primary particles which initiate the atmospheric showers.

What, then, are the primary particles, and how great energies do they possess? If we could take our instruments out into interstellar space, we might be able to obtain a direct answer to these questions. Lacking means for making this direct experiment as yet, the cosmic-ray scientists have done what they could by exploring the atmosphere from sea level to as far above the earth as possible, at all accessible points over the earth, year in and year out. Let us now consider what is revealed by these explorations.

280. Cosmic Rays at High Altitudes. As was mentioned at the beginning of this chapter, the first cosmic-ray explorations were made by sending the measuring instruments as far as possible above the earth. The first observations were made from stations on high mountain peaks, or by occasional trips with manned balloons. These were followed by more frequent observations made with self-recording instruments, carried by strings or clusters of small balloons about 3 feet in diameter. At first these instruments with their records were dropped by parachute and recovered by offering a reward to the finder. Later on the record was transmitted directly to the ground by radio, in the manner now employed for radiosonde weather data (see Sec. 82). A number of ascensions with giant balloons have yielded valuable data up to a height of 12 miles, big airplanes have been employed for some special studies, and by 1947 a few explorations to far greater heights by means of rockets had been carried out. The first instruments to be sent aloft were simple ionization chambers and electroscopes, but, as time went on, more and more elaborate equipment was designed for this purpose. In recent work complete cloud-chamber equipment has been carried aloft by airplane.

These observations all show that the total intensity of cosmic rays increases rapidly with altitude, reaching a peak value at a height of 10 or 12 miles, then begins to fall off. The soft component increases with altitude more rapidly than the hard component, becoming a major part at high altitudes. Relative magnitudes vary with magnetic latitude in the manner to be described in the next section, but the general trend is the same everywhere. These data are consistent with the hypothesis, stated above, that all cosmic-ray particles at lower elevations are *secondary* particles, created within the atmosphere. The primary particles, whatever they may be, are few in number as compared with their progeny. As they enter the atmosphere, they produce secondary particles, which produce other secondaries in their turn; thus the number of particles multiplies rapidly at the top of the atmosphere. As the atmosphere becomes denser, however, the average energy per particle becomes less and less, and losses by absorption multiply to such an extent that, below the peak at about 11 miles up, the intensity begins to fall off again.

281. Latitude Effect. In 1927–28, Clay, of Holland, took an electroscope with him on a long sea voyage, and he discovered that the cosmic radiation was noticeably less in equatorial regions than elsewhere. Since then it has been shown by the measurements of Compton, Millikan, and numerous others that the cosmic-ray intensity is least at the *magnetic* equator of the earth, and increases towards the poles until about 50° magnetic latitude is reached. From 50° to the magnetic poles the sea-level value remains remarkably constant. The explanation for this latitude effect is not hard to find, and directly concerns the *primary* particles. The earth is a huge magnet, whose magnetic field extends to many times the earth's diameter before it becomes of negligible intensity. Electrically charged particles approaching the earth from outer space will travel in curved paths through this magnetic field; and if they do not possess sufficient energy, they may make a U turn and go away again without reaching the earth. See Fig. 175, ray *A*. This difficulty in reaching the earth will be greatest at the magnetic equator, where the particles must move at right angles to the lines of force of the earth's field, and hence must possess relatively high energies in order not to be turned away. It should vanish at the magnetic poles, where the weakest of rays might spiral around the lines of force and follow them down to the pole.

According to this explanation, the intensity should increase to a very large value at the poles, and the constancy of the sea-level intensity above latitude 50° must be explained in another manner, namely, by the absorption of the earth's atmosphere. From the magnetic equator to 50° magnetic latitude, the low energy limits are set by the earth's magnetic field; above 50°, they are set by the absorption of the atmosphere. Above 50°, softer rays may reach the top of the atmosphere, but they cannot get through it to the earth's surface. This latitude effect is most positive evidence that the primary rays are electrically charged particles.

The vertical component of the cosmic rays may be measured, separately from that coming to the earth from other directions than vertical, by using a vertical array of G-M coincidence counters as a cosmic-ray telescope. This "telescope" consists of two or more G-M counter tubes laid parallel to each other, as is indicated schematically in Fig. 174, and so connected to the recording circuit that the only cosmic rays which are counted are those which pass through all the counter tubes, so as to "fire" all of them simultaneously. Since the angles which are of importance are angles east and west of the vertical (as is explained in Sec. 280), this array of tubes is placed with the tube axes pointing north and south. When only the vertical component is measured in this manner, the intensity is found to vary in a manner similar to that found for the total radiation, but to be constant from about magnetic latitude 40° to the poles.

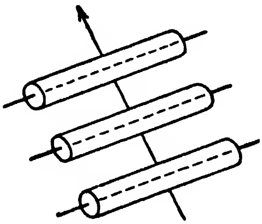


FIG. 174. COSMIC-RAY "TELESCOPE." The "telescope" points in the direction of the arrow.

282. Energies of Primary Cosmic Rays. It is possible to compute, by a method due to Störmer, the minimum energy which a charged particle must possess in order to penetrate the magnetic barrier at any latitude. The computations are very tedious, and are made practicable only by use of a marvelous computing machine invented by Bush, at Massachusetts Institute of Technology. The computations have been made by Lemaitre and Vallerta. The energy values are computed for electrons, but since all of them are in billions of electron-volts, they are practically the same for heavier particles. (See Sec. 267.)

These computations give 6 Bev (billion electron-volts) as the minimum energy for a charged particle which is able to enter the atmosphere at magnetic latitude 40° (vertical incidence), and this, as is explained in Sec. 281, determines the lower limit of energy for the primary rays which are able to penetrate the earth's atmosphere. The upper limit is not so easily determined. From absorption measurements on the penetrating component it is estimated that the corresponding primary rays might have energies as high as 100 Bev, while the evidence given by atmospheric showers (Sec. 270) indicates that the primary shower particles may possibly possess energies as great as 10^5 to 10^6 Bev. Some estimates go as high as 10^8 Bev!

283. East-West Asymmetry. The analyzing power of the earth's magnetic field may be utilized in still another way. A positively charged particle approaching the earth at the equator is deflected towards the east, as is shown in Fig. 175. The particle, *A*, has too little energy to reach the earth, and is turned away from it by the magnetic field. Particle *B* has just enough energy to reach the earth, and must enter the atmosphere from the *west*, as shown. The computations of Lemaitre and Vallerta show that its energy is about 10 Bev. Particle *C*, which comes vertically down to the earth, and particle *D*, which comes in from the east, must have still higher energies. Particle *C* must have at least 18 Bev, and *D*, at least 75 Bev of energy. The same minimum energy values will apply to a negatively charged particle, but the directions will be reversed. It will be deflected towards the west, and the particles having the least energy will come in from the east.

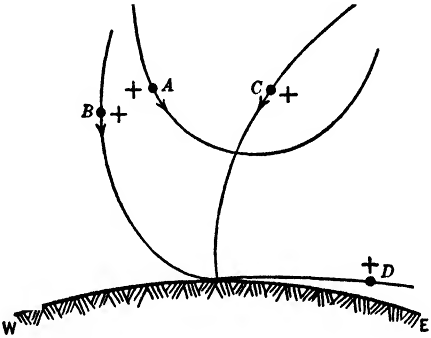


Fig. 175. The East-West Effect. At the equator, looking north.

This provides a possible means for determining whether the incoming particles are predominantly charged positively or negatively. If a cosmic-ray telescope, such as is diagrammed in Fig. 174, gives equal readings whether it is pointed to the east or to the west, there are equal numbers of primary particles of each sign. A greater intensity reading when pointed east indicates a greater abundance of negatively charged primary particles, whereas the reverse is true if the intensity is greater from the west. The particles entering the cosmic-ray telescope are, of course, not the primary particles but secondary particles of both signs. Nevertheless, the telescope indicates the direction of approach for the *primary* particles, since the secondaries, for the most part, travel in the direction taken by the primaries which produced them.

T. H. Johnson, of the Bartol Research Foundation, has in this manner made extensive studies of the east and west intensities at various locations near the equator, both at sea level and at high elevations, and he always finds a greater intensity from the *west* than from the east. In other words, the positively charged primary cosmic-ray particles which come to the earth through its magnetic field at the equator exceed those negatively charged. The magnitude of this excess is sufficient to account for the production of mesons by some kind of *positively* charged primary particle.

284. Origin of Mesons. It is now believed quite possible that the positively charged particles which are the progenitors of mesons are *protons*. This was suggested first by T. H. Johnson, who reasoned as follows: (1) It seems impossible to conceive of an interstellar space occupied by more positively charged than negatively charged particles; the electric fields which would result from such an excess would be appalling in their magnitudes! It seems more reasonable, therefore, to suppose that interstellar space contains equal numbers of both positively and negatively charged particles, but that for some reason only the positively charged particles are able to reach the earth through the barrier set up by its magnetic field. (2) The ability of positively charged particles to penetrate this barrier, whereas negatively charged ones are excluded, could be explained by assuming that particles of both signs of charge have the same average speed,[11] but that the positively charged particles are the more *massive*. More massive particles could possess enough momentum to penetrate the magnetic barrier, while lighter ones having the same speeds would be turned away. The proton is the lightest positively charged particle which could satisfy these requirements. Mesons might be heavy enough, but they could not exist in interstellar space, because of their very short lives.

More recently, W. F. G. Swann, director of the Bartol Research Foundation, has analyzed very carefully the energy and momentum requirements for the particle which could serve as a producer of mesons, and he arrives at the conclusion that this particle has just the mass of the proton. The current theory is that, in a nuclear reaction involving the nucleus of an atom of the gas forming the upper atmosphere, the primary proton creates primary or heavy mesons, which shoot downward, in the general direction taken by the incoming proton. Most of this process takes place near the top of the atmosphere; few if any of these cosmic-ray protons get far below the level at which the peak intensity of cosmic rays occurs. The primary mesons are then assumed to

---

[11] This follows as a necessary corollary of the requirement that there be equal numbers of both kinds of charged particles in interstellar space. If, for example, the particles having negative charges moved faster than those of opposite charge, the two kinds of particles would tend to become separated.

be so very short-lived that practically all of them disintegrate into ordinary mesons, or are captured by nuclei, long before they reach the surface of the earth.

285. Origin of Showers. It is more difficult to find a theory for the production of the shower electrons which form the soft component. At first it was assumed that the primary particles were very high-energy electrons. Later on, when it was realized that mesons can start showers, the possibility that *all* cosmic-ray particles arise from protons alone has been given serious consideration. Evidence that mesons can start showers is given by showers observed deep down under the earth, in mines, at depths to which electrons from the atmospheric showers cannot penetrate. There are several ways in which this might be brought about. Theory indicates that, very rarely, a meson can produce electrons by the regular shower process, in close approach to an atomic nucleus. It is more possible for it to give high energy to an electron in a direct collision. Finally, every meson, when it dies, is believed to convert into a high-speed electron and a neutrino. Any of these processes may start a shower.

The possibility that all shower electrons originate from mesons is suggested by the finding of mesons in considerable numbers in the great atmospheric showers. This was first discovered by Auger, and has since been confirmed by many observations. Some serious difficulties confront this theory. In particular, shower electrons are moving too much in random directions, especially at the top of the atmosphere, to be accounted for by the same mesons which form the hard component. It may be that several types of primary mesons are produced by the incoming protons, and that mesons of some type not yet discovered, possibly neutral mesons, are the progenitors of the shower particles. Various other suggestions have been made. This is still one of the important unsolved problems.

286. Source of Primary Cosmic Rays. From the very beginnings of the investigation of cosmic rays, speculations have been made concerning whence they come, but as yet no one has been able to do more than speculate in this regard. The name "cosmic" is itself a speculation of this kind. Extensive as does our own galactic system appear, comprising as it does all of the milky way in addition to all of the visible stars, it is a very small part of the vast reaches of space which may properly be called "cosmic," and but one of a host of similar "island universes." (Of these, all but our own are called nebulae, since, because of their almost inconceivable distances from one another, they appear in all but the most powerful telescopes as mere faint patches of light.) As the name "cosmic" indicates, speculation at first put the origin of cosmic rays in the cosmic void outside of these relatively insignificant nebulae. More recently, it has been considered possible that their origin might be inside our own galaxy. Still more recently, slight changes in intensity observed at times of strong sun-spot activity have suggested the possibility that some, at least, of the cosmic rays might have their source in the sun, as well as in other stars.

287. Cosmic Rays and Elementary Particles. These provoking questions concerning the origins of cosmic-ray particles, and the source of the primary cosmic rays, are themselves ample justification for further intensive study in this field. But the study of cosmic rays is of even greater importance to the furtherance of our knowledge concerning the *elementary particles* of physics, and the laws which determine how the rest of the physical world is built up from these particles. This aspect of cosmic-ray study has been apparent throughout this chapter; the discoveries concerning mesons are outstanding examples. The importance of cosmic rays to such studies derives chiefly from the enormous energy concentrations which they make available—up to 1000 Bev or more in a single particle. This provides experimental means for testing the theories of the theoretical physicists at energy levels far beyond the present limitations of all other experimental methods. For example, the *bremstrahlen* theory for the exchange of energy between charged particles and photons of electromagnetic radiation has been verified at these high-energy levels, where it accounts for showers (Sec. 269). On the other hand, the experimental physicist is provided with means for producing entirely new nuclear reactions, as well as old reactions under quite new conditions.

PROBLEMS

1. The average incidence of cosmic rays at sea level is approximately one charged particle per minute per square centimeter. How many counts per second will be registered by a G-M counter whose cylinder is 30 cm long by 5 cm in diameter?

2. Assume that a charged cosmic-ray particle will produce an average of 30 ion-pairs per centimeter. Compute the current flowing between two metal plates, as a result of cosmic radiation at sea level (see problem 1), if the plates are 40 cm in diameter and 6 cm apart. The plates are connected to a potential source of sufficient voltage to collect all ions formed.

3. Review problem 7, Chapter XV.

4. Review problem 9, Chapter XVII. In addition, compute the energy of this particle if it is a meson.

5. Compute the "life" of a meson which is moving with energy equal to 100 Bev.

6. Assume that a meson at rest explodes into an electron and a neutrino, of equal energies. Compute the energy of the electron.

REFERENCES

For Additional Reading

J. D. **Stranathan**, *The "Particles" of Modern Physics*, Chapters 12 and 13.

R. A. **Millikan**, *Electrons, Protons, Photons, Neutrons, and Cosmic Rays*, Chapters XIII, XIV, XV, and XVI.

K. K. **Darrow**, "Particles in Cosmic Rays," *Bell Technical Journal*, Vol. 18, pp. 190–217, January, 1939.

W. F. G. **Swann**, "What Are Cosmic Rays?" *Journal of the Franklin Institute*, Vol. 226, pp. 601–622, November, 1938.

W. F. G. **Swann**, "Cosmic Rays," *Journal of the Franklin Institute*, Vol. 236, pp. 521–540, December, 1943.

S. H. **Neddermeyer** and C. D. **Anderson**, "Cosmic-Ray Particles of Intermediate Mass," *Physical Review*, Vol. 54, pp. 88–89, July 1, 1938. (Contains photograph showing slow meson track.)

J. C. **Stearns** and D. K. **Froman**, "Cosmic Rays," *American Physics Teacher*, Vol. 7, pp. 79–99, April, 1939.

J. A. **Wheeler**, "Elementary Particle Physics," *American Scientist*, Vol. 35, pp. 177–193, April, 1947.

For Advanced Reading

J. C. **Street**, "Cosmic Ray Showers and Penetrating Particles," *Journal of the Franklin Institute*, Vol. 227, p. 765, June, 1939.

T. H. **Johnson** and J. G. **Barry**, "The East-West Symmetry of Cosmic Rays," *Physical Review*, Vol. 56, pp. 219–225, August 1, 1939.

W. B. **Fretter**, "Mass of Cosmic-Ray Mesotrons," *Physical Review*, Vol. 70, pp. 625–631, Nov., 1946.

J. B. **Hoag**, *Electron and Nuclear Physics*, Chapters XVII and XVIII.

H. J. J. **Braddick**, *Cosmic Rays and Mesotrons*. Cambridge University Press.

W. **Heisenberg**, *Cosmic Radiation*. (Translated by T. H. Johnson.) Dover Press.

F. K. **Richtmyer** and E. H. **Kennard**, *Introduction to Modern Physics*, Chapter XII.

EPILOGUE

In closing let us survey briefly some of the salient points in our present knowledge of fundamental physics. Foremost among the recent discoveries we have a number of new particles—positron, neutron, meson, and neutrino—which appear to complete the list begun by the electron and the proton. The proton itself has acquired a new identity, as well as the name, proton, as a part of these discoveries. And we have a very good start towards a consistent plan for the structure of the physical universe from these particles. There is, of course, very much more to be done.

In fundamental physical theory, first place is still occupied by the conservation laws for electric charge and for momentum. These laws have remained unchanged from their earliest conceptions. At one time it became necessary to invent the neutrino to save the momentum law, but this invention now appears quite justified by experiment. The conservation laws for mass and for energy had equally good foundations in all the phenomena of the older physics, but it has been found necessary to merge these laws in their applications to modern physics. The key to this merger is Einstein's law for mass-energy equivalence. The new conservation law for mass plus energy has stood all tests, including even the high-energy range encountered in cosmic rays; at the same time, of course, the earlier separate laws remain equally good for the range of phenomena in which the mass of energy is negligible.

Finally, to these conservation laws must be added an entirely new law. This law, which may be called rather inadequately the *quantum* law, is best stated in terms of a dual existence of all physical entities as both particles and waves, including on the one hand "material" particles such as electrons and neutrons, and on the other hand particles of pure energy such as photons of light. Neutrinos may fall into this latter classification; we do not yet know. For the most part the full statements of this quantum law are quite mathematical and even now subject to changes in detail. For most practical purposes, however, they may be reduced to the *quantum rules* which have been stated in earlier chapters. One aspect of this theory which is becoming of increasing importance is that characteristic of particles which is called *spin* (see Sec. 150).

The general problem of physical forces is one of the oldest and yet one concerning which our knowledge is most incomplete. We have exact laws to describe the forces of gravitation, electrostatics, magnetism, and electromagnetism, but we are almost as far as ever from explaining these forces, that is, accounting for them in a more fundamental manner. The forces of interaction among electrons and nuclei in atoms and molecules, among molecules in a liquid, and among nucleons in a nucleus cannot be described nearly as well, yet the clue for explanation of all physical forces may come first from study of these forces. Here is a field of research which holds the greatest interest.

Research laboratories are hard at work on all these problems, as well as many others not mentioned here. In addition to the laboratories of many universities there are the great government-sponsored laboratories, such as the Clinton Laboratory in Tennessee, the Argonne Laboratory in Chicago, and the Brookhaven Laboratory on Long Island. The commercial laboratories, such as the General Electric, Westinghouse, and Bell Telephone Laboratories, also are doing their share of fundamental research. Some unusually interesting analyses of the current status of physical research have been written by Professor Wheeler of Princeton University, Professor Morrison of Cornell University, and Professor Hull of Dartmouth College.[1]

A celebrated scientist once received a letter from his publisher which expressed the plaintive wish that science would not progress so fast; too many changes had to be made while his book was in proof. This complaint emphasizes the *incompleteness* of all physical science, which has been pointed out above and elsewhere throughout this book. It is not to be deplored that this is so, nor is it to be considered a weakness of science that, no matter how far it may be developed in any branch, it will *always* be incomplete. In science there are always new frontiers; when the fields of research which we can see before us now have been fully explored we may confidently expect that other fields will have made their appearance.

So far these comments have related to the fundamental aspects of physics. The practical or applied aspects are even more diverse and difficult to catalog. We shall not attempt it here. There is hardly a field of the newer physics which has not already found useful applica-

---

[1] J. A. Wheeler, "Elementary Particle Physics," *American Scientist*, Vol. 35, 1947, pp. 177–193.

P. Morrison, "Physics in 1946," *Journal of Applied Physics*, Vol. 18, 1947, pp. 133–152.

G. F. Hull, "Fifty Years of Physics," *Science*, Vol. 104, 1946, pp. 238–244.

tion, and new applications continue to be found for many of the older fields.

These aspects of physics concern also the relations of physics to human activities generally, and this topic has been the subject of much comment and dispute among scientists, economists, politicians, and "humanists" of all kinds. Some of these comments, such as the naive suggestions that all scientific research be halted until other human activities have caught up with science, and pious exhortations that scientists limit their discoveries to those capable of beneficent applications only, appear totally ridiculous to scientists. Comments such as these result from an ignorance of the basic principles of science that is not only deplorable but actually dangerous; they constitute a very strong argument for better education of people generally in the fundamentals of science.

Another point of view which appears often in such discussions, in many diverse forms, may be expressed by the following question: "Why does not the physicist, who can accomplish so much in the incomprehensible realms of electrons, light quanta, and nuclei, turn this ability to the more important and *simpler* problems of human behavior?" The answer to such mistaken ideas has been emphasized throughout this book; the great successes in physical research have resulted from the relative *simplicity* of the phenomena involved. Physical experiments may be repeated time and time again, and depend upon a relatively small number of *controllable* factors. The results may be expressed in terms of physical laws which are amazingly simple also, when the extent of their applicability is considered. The problems of human behavior are incomparably more complicated. Far simpler are the problems of biological science, and even these are far more difficult and uncertain than physical problems, involving many more variables, some of which may be uncontrollable or even unknown.[2] The physicist has attained a relatively greater measure of success because his science is limited to phenomena which are simple enough to be fully controlled.

This does not make the physicist inhumanly unconcerned with the services which may be derived by mankind from the results of his researches. He is always encouraged in his work by the firm belief that beneficial results will develop from it, not only in the form of practical applications but also in the benefit to be gained by all science from a better understanding of the laws of the physical world. As has been emphasized earlier, he has never yet been disappointed. At the same time there is no possible safeguard against the destructive use of sci-

---

[2] In *The Scientists Speak*, the editor, Dr. Warren Weaver of the Rockefeller Foundation, has made a beautiful analysis of these differences among the sciences in an essay entitled "Science and Complexity."

ence—or any other agency, for that matter. And such misuse is not to be blamed upon the scientist. It is not difficult to prove that the benefits have always in the long run far exceeded the damage done by any scientific discovery. Salvation of the world, however, is not to be won or lost by science, but by those human activities which concern the souls of men. Towards these questions the scientist makes a very humble approach because he appreciates better than anyone else the complex and difficult nature of the problems there encountered; he knows how hard the relatively simple problems of physics can be.

APPENDIX I

UNITS

In dealing with atomic and molecular phenomena it is most convenient to use electrostatic units (e.s.u.) for all electrical quantities, and electromagnetic units (e.m.u.) for magnetic quantities only. This system of units is called the *Gaussian* system. The conversion factors for the most important quantities are tabulated below:

QUANTITY	PRACTICAL		GAUSSIAN
Charge	10 coulombs	=	3×10^{10} stat-coulombs
Current	10 amperes	=	3×10^{10} stat-amperes
Potential	300 volts	=	1 stat-volt
Energy	1 joule	=	10^7 ergs
Magnetic field, H	1 oersted	=	1 oersted
Magnetic flux-density, B	1 gauss	=	1 gauss

Observe that electrostatic units are designated by the prefix "stat-." For the most precise work (better than $1/10$ per cent) the factor 3×10^{10} above must be replaced by the value of c given in Appendix II.

For atomic dimensions, wavelengths of light, etc., the *angstrom unit A* is frequently used.

$$1\text{A} = 10^{-8}\text{ cm}$$

The electron-volt (ev) often serves as a convenient unit of energy. One electron-volt equals the energy which may be given by a potential difference of 1 volt to a particle carrying the elementary charge, e.

$$1\text{ ev} = 4.80 \times 10^{-10} \times \frac{1}{300} = 1.60 \times 10^{-12}\text{ erg}$$

The electron-volt may of course be used to measure the energy of any particle, regardless of how its energy is produced. For example, it may be used to measure the energy of a neutron, which does not obtain its energy from the direct action of electrical forces. For larger amounts of energy one million electron-volts (1 Mev) and one billion electron-volts (1 Bev) are often employed as units.

$$1\text{ erg} = 625\text{ Bev}$$

Mass-energy equivalence (see Secs. 212, 222, and 246) may be expressed by any of the following ratios:

$$1 \text{ gram} = 9 \times 10^{20} \text{ ergs} = 5.62 \times 10^{23} \text{ Bev}$$

$$1 \text{ mass unit} = 1.66 \times 10^{-24} \text{ gram} = 1.49 \times 10^{-3} \text{ erg} = 935 \text{ Mev}$$

$$1 \text{ Bev} = 1.07 \text{ mass unit}$$

APPENDIX II

UNIVERSAL PHYSICAL CONSTANTS

The values of universal physical constants such as the elementary electric charge, e, cannot be determined with certainty by a single experiment, but only by agreement between the results of experiments representing several quite different methods. Furthermore, the determination of many of these physical quantities is dependent upon the values of others. If all physical experiments were free from systematic error, then all values of each physical constant as obtained by different methods would agree with one another, at least within the limits of precision of the measurements. In other words, all experiments for the determination of these physical constants would be consistent with one another.

Now it happens that, as has been explained in the body of this book, these experiments are consistent to a degree which leaves no doubt as to the correctness of the basic principles underlying them. Unfortunately, however, small inconsistencies still exist which are greater than may be ascribed to the chance errors of the experiments involved, although they are now far smaller than they were a few years ago. Physicists will not be satisfied until the causes of these discrepancies have been apprehended and eliminated. Professor R. T. Birge, of the University of California, has studied this problem for many years, and the values given below are his latest estimates for the best values at present obtainable. They were published by him in *Reports on Progress in Physics*, Vol. 8, 1942, pp. 90–134. He reviewed the values for e, N_0, and F in a later article in the *American Journal of Physics*, Vol. 13, 1945, pp. 63–72.

As a matter of convenience, these constants have in general been carried to but three significant figures in the body of the book.

$c = 2.99776 \times 10^{10}$ cm/sec = ratio of e.m.u. to e.s.u. of electric charge;
± 4 also the speed of light in a vacuum.
$e = 4.8021 \times 10^{-10}$ stat-coulomb = elementary electric charge.
± 6

$F = 2.89247 \times 10^{14}$ stat-coulombs = Faraday's constant (electric charge transported by a gram atom of a univalent element *).
± 30

$N_0 = 6.02338 \times 10^{23}$ = Avogadro's number (number of atoms in a gram atom, or molecules in a mole *).
± 43

$P_0 = 1.013246 \times 10^{6}$ dynes/cm^2 = normal atmospheric pressure.
± 4

$T_0 = 273.16°$ K = freezing point of water on the absolute or Kelvin temperature scale.
± 1

$V_0 = 22.4146 \times 10^{3}$ cm^3 = volume of a mole of a perfect gas at 0° C and normal atmospheric pressure.*
± 6

$R_0 = 8.31436 \times 10^{7}$ ergs/degree mole = gas constant for one mole.*
± 38

$k = 1.38047 \times 10^{-16}$ erg/degree = Boltzmann's constant (gas constant per molecule).
± 26

$h = 6.624 \times 10^{-27}$ erg sec = Planck's "quantum" constant.
± 2

$M_0 = 1.66035 \times 10^{-24}$ gram = mass of atom of unit atomic weight.*
± 31

$m = 9.1066 \times 10^{-28}$ gram = mass of the electron.
± 32

$M_p = 1.67248 \times 10^{-24}$ gram = mass of the proton.
± 31

The "probable error" (an *index* of the range of uncertainty) is written (with the ± sign) under each value. It applies to the figure or figures under which it is written.

* Two scales of atomic weights are now in use: the "chemical" scale and the "physical" scale. The chemical scale is based upon the value of 16 for the atomic weight of oxygen as determined by chemical analysis, whereas the physical scale assigns the value of 16 to the most abundant isotope of oxygen. Atomic weights in the physical scale are larger than those on the chemical scale in the ratio of 1.00027 to 1. The values of F, N_0, V_0, R_0, and M_0 are given above in terms of the chemical scale. In terms of the physical scale, F, N_0, V_0, and R_0 should be increased in this ratio, and M_0 should be correspondingly decreased. On the physical scale

$$M_0 = 1.65990 \times 10^{-24} \text{ gram}$$
$$\pm 31$$

APPENDIX III

PERIODIC TABLES OF THE ELEMENTS

The first periodic table or chart of the elements was made by Mendelyeev, and was based upon chemical similarities between adjacent elements in his chart. The same relationships are found if the tabular arrangement is made in accord with the *electronic structures* of the atoms, since the chemical properties of atoms are determined by the configurations of their electron shells, principally by the outer shells. Tables I to III are arranged according to this electronic plan. In these tables the integer under each symbol is the value of Z, the atomic number of the element represented, which is also the number of elementary charges upon its nucleus and the number of electrons in its normal atom. Symbols inclosed in brackets (e.g., [Pu]) represent radioactive elements which do not exist naturally but must be produced artificially by some form of nuclear reaction. Table IV lists the elements in alphabetical order, together with their atomic numbers and atomic weights, and the symbols which represent them.

Table I represents those elements in which electrons are added to the *outside* shell as Z increases from one element to the next. This table corresponds to the "short periods" of the usual periodic table chart, and these elements exhibit best the chemical relationships among elements lying in the same column which formed the basis for Mendelyeev's table. For the elements here represented the number of electrons increases from 1 to 8, as indicated by the numbers at the top of the table, while the numbers in the underlying shells remain constant for each row, as indicated at the right of the table. In the first two columns there is an interesting alternation of chemical properties, which may be accounted for by differences in the next-to-outer shells. For example, Na, K, Rb, and Cs, which have 8 electrons in this next-to-outer shell, resemble one another much more closely than they resemble the alternate elements in this column, namely Cu, Ag, and Au, which have 18 electrons in this next-to-outer shell. This difference is better represented in Table II, in which these two sets of elements appear at opposite ends of a row.

TABLE I

Electrons in Outer Shell								Electrons in Inner Shells						
1	2	3	4	5	6	7	8	*K*	*L*	*M*	*N*	*O*	*P*	SHELLS
H 1	He 2							—						
Li 3	Be 4	B 5	C 6	N 7	O 8	F 9	Ne 10	2						
Na 11	Mg 12	Al 13	Si 14	P 15	S 16	Cl 17	A 18	2	8					
K 19	Ca 20			(*a*)				2	8	8				
Cu 29	Zn 30	Ga 31	Ge 32	As 33	Se 34	Br 35	Kr 36	2	8	18				
Rb 37	Sr 38			(*b*)				2	8	18	8			
Ag 47	Cd 48	In 49	Sn 50	Sb 51	Te 52	I 53	Xe 54	2	8	18	18			
Cs 55	Ba 56			(*c*)				2	8	18	18	8		
Au 79	Hg 80	Tl 81	Pb 82	Bi 83	Po 84	[At] 85	Rn 86	2	8	18	32	18		
[Fa] 87	Ra 88			(*d*)				2	8	18	32	18	9	

Table II represents those elements in which electrons are being added to the *next-to-outer* shell, while the number in the outer shell remains between 0 and 2, with 2 most frequent. The letters (*a*), (*b*), and (*c*) in Table I indicate where the corresponding parts of Table II have been omitted from Table I. In Table II elements having the same number of outer electrons are placed in the same row, while those having the same number in the next-to-outer shell are arranged in the same column. The elements from the first and second columns of Table I are added for completeness.

TABLE II

	Electrons in Next-to-Outer Shell											Electrons in Outer Shell	
	8	9	10	11	12	13	14	15	16	17	18		
(*a*)	K 19					Cr 24					Cu 29	1	*N* Shell
	Ca 20	Sc 21	Ti 22	V 23		Mn 25	Fe 26	Co 27	Ni 28		Zn 30	2	
(*b*)											Pd 46	0	
	Rb 37				Cb 41	Mo 42	[Tc] 43	Ru 44	Rh 45		Ag 47	1	*O* Shell
	Sr 38	Y 39	Zr 40								Cd 48	2	
(*c*)	Cs 55									Pt 78	Au 79	1	
	Ba 56	(*d*)	Hf 72	Ta 73	W 74	Re 75	Os 76	Ir 77			Hg 80	2	*P* Shell

TABLE III

	Electrons in Third from Outside Shell															Electrons in Next-to-Outer Shell	
	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32		
(*d*)					Nd 60	— 61	Sm 62	Eu 63			Dy 66	Ho 67	Er 68	Tm 69	Yb 70	8	*O* Shell
	La 57	Ce 58	Pr 59					Gd 64	Tb 65						Lu 71	9	
(*e*)								Am 95								8	*P* Shell
	Ac 89	Th 90	Pa 91	U 92	Np 93	Pu 94		Cm 96								9	

TABLE IV

Element	Symbol	Z	Atomic Weight	Element	Symbol	Z	Atomic Weight
Actinium	Ac	89	227.05	Lead	Pb	82	207.21
Aluminum	Al	13	26.97	Lithium	Li	3	6.940
Americium	[Am]	95	241	Lutecium	Lu	71	174.99
Antimony	Sb	51	121.76	Magnesium	Mg	12	24.32
Argon	A	18	39.94	Manganese	Mn	25	54.93
Arsenic	As	33	74.91	Mercury	Hg	80	200.61
Astatine	[At]	85	211	Molybdenum	Mo	42	95.95
Barium	Ba	56	137.36	Neodymium	Nd	60	144.27
Beryllium	Be	4	9.02	Neon	Ne	10	20.18
Bismuth	Bi	83	209.00	Neptunium	[Np]	93	237
Boron	B	5	10.82	Nickel	Ni	28	58.69
Bromine	Br	35	79.92	Nitrogen	N	7	14.008
Cadmium	Cd	48	112.41	Osmium	Os	76	190.2
Calcium	Ca	20	40.08	Oxygen	O	8	16.00
Carbon	C	6	12.01	Palladium	Pd	46	106.7
Cerium	Ce	58	140.13	Phosphorus	P	15	30.98
Cesium	Cs	55	132.91	Platinum	Pt	78	195.23
Chlorine	Cl	17	35.46	Plutonium	[Pu]	94	239
Chromium	Cr	24	52.01	Polonium	Po	84	210
Cobalt	Co	27	58.94	Potassium	K	19	39.10
Columbium	Cb	41	92.91	Praseodymium	Pr	59	140.92
Copper	Cu	29	63.57				
Curium	[Cm]	96	242	Protoactinium	Pa	91	231
Dysprosium	Dy	66	162.46				
Erbium	Er	68	167.2	Radium	Ra	88	226.05
Europium	Eu	63	152.0	Radon	Rn	86	222
Fluorine	F	9	19.00	Rhenium	Re	75	186.31
Francium	[Fa]	87	223	Rhodium	Rh	45	102.91
Gadolinium	Gd	64	156.9	Rubidium	Rb	37	85.48
Gallium	Ga	31	69.72	Ruthenium	Ru	44	101.7
Germanium	Ge	32	72.60	Samarium	Sm	62	150.43
Gold	Au	79	197.2	Scandium	Sc	21	45.10
Hafnium	Hf	72	178.6	Selenium	Se	34	78.96
Helium	He	2	4.003	Silicon	Si	14	28.06
Holmium	Ho	67	164.94	Silver	Ag	47	107.88
Hydrogen	H	1	1.008	Sodium	Na	11	22.997
Indium	In	49	114.76	Strontium	Sr	38	87.63
Iodine	I	53	126.92	Sulfur	S	16	32.06
Iridium	Ir	77	193.1	Tantalum	Ta	73	180.88
Iron	Fe	26	55.85	Technetium	[Tc]	43	99
Krypton	Kr	36	83.7	Tellurium	Te	52	127.61
Lanthanum	La	57	138.92	Terbium	Tb	65	159.2

TABLE IV (*Continued*)

Element	Symbol	Z	Atomic Weight	Element	Symbol	Z	Atomic Weight
Thallium	Tl	81	204.39	Vanadium	V	23	50.95
Thorium	Th	90	232.12	Xenon	Xe	54	131.3
Thulium	Tm	69	169.4	Ytterbium	Yb	70	173.04
Tin	Sn	50	118.70	Yttrium	Y	39	88.92
Titanium	Ti	22	47.90	Zinc	Zn	30	65.38
Tungsten	W	74	183.92	Zirconium	Zr	40	91.22
Uranium	U	92	238.07	( )	[—]	61	147

Table III represents elements in which there are always 2 electrons in the outer shell, and always either 8 or 9 in the next-to-outer shell. Except when the population of the next-to-outer shell jumps from 8 to 9 electrons, the additional electrons go to build up the third shell from the outside. The series of elements which is labelled (*d*) in this table is the *rare-earth* series, which, in numerical sequence, belongs at (*d*) in Table I. These elements are also called the *lanthanides*, from lanthanum, the first of the series. All these rare-earth elements have two *P* electrons and either eight or nine *O* electrons; this explains their great chemical similarity. The building of their electronic structure in the third-from-outer or *N* shell is revealed by their X-ray spectra. (See Sec. 146.) Below the *N* shell all these elements have the same electronic structure: 2, 8, and 18 electrons in the *K*, *L*, and *M* shells respectively. The elements designated by (*e*) in Table III form an analogous series at the very end of the periodic table, as indicated by (*e*) in Table I. For all but one of these elements the *P* and *Q* shells contain 9 and 2 electrons respectively, the differences being in the *O* shell only. The exception is americium (Am), which has only eight *P* electrons. Below the *O* shell all contain 2, 8, 18, and 32 electrons in the *K*, *L*, *M*, and *N* shells respectively. This series may also be called the *actinides*, from actinium, which heads it.

The most recent data for the electronic arrangements have been taken from *Periodic Chart of the Atoms*, 1947 Edition, by H. D. Hubbard and W. F. Meggers, with the permission of W. M. Welch and Co., the publishers, and Dr. Meggers. This chart and the booklet "key" which accompanies it contain many more useful and interesting data concerning properties of atoms.

APPENDIX IV

PERIOD OF OSCILLATION FOR AN OSCILLATING ELECTRIC CIRCUIT

The oscillating electric circuit has been described in Sec. 62, and its period of oscillation has there been given as

$$P \text{ (seconds)} = 2\pi \sqrt{L \text{ (henries)} \times C \text{ (farads)}} \tag{1}$$

The proof for this equation may be made as follows:

When the current flowing through a coil changes, there is induced in that coil an e.m.f. which is proportional to the rate of change of the current, that is, to the change of current per unit time. The proportionality factor is called the coefficient of self-induction, L. If we let $\dot{I}$ (I with a *dot* over it) stand for the *change of current per unit time*, this law may be represented by the equation

$$\text{E.m.f.} = -L\dot{I} \tag{2}$$

In this equation the negative sign, of course, represents the fact, stated by Lenz's law, that the induced e.m.f. is always in such a direction as to oppose the *change* of current.

When this e.m.f. is used to charge a condenser whose capacitance is C, as is done in the oscillating electric circuit shown in Fig. 45, then (if the resistance of the circuit is negligibly small, as it usually is) the potential difference across the condenser at any time is equal to this e.m.f., or

$$\frac{Q}{C} = -L\dot{I} \tag{3}$$

whence

$$\dot{I} = -\frac{1}{LC}Q \tag{4}$$

If the electrical oscillations which take place in this circuit are now compared to the mechanical oscillations of a bob oscillating up and down at the end of a spring, it is seen that the charge, Q, is analogous to the displacement of the bob; the current flowing in the circuit is analogous to the velocity of the bob; and the change of current per unit time, $\dot{I}$, is analogous to its acceleration. Since, for *all* simple harmonic motions (such as that of the oscillating bob),

$$\text{Acceleration} = -\frac{4\pi^2}{P^2} \times \text{Displacement} \tag{5}$$

it is evident that (4) is the equation of a simple harmonic motion: it follows that

$$\frac{4\pi^2}{P^2} = \frac{1}{LC} \tag{6}$$

or

$$P = 2\pi\sqrt{LC} \tag{7}$$

APPENDIX V

RELATIVITY

The theory of relativity has not been considered previously in this book, since modern physics has here been considered principally from an experimental point of view, whereas relativity theory is not related specifically to any one group of experimental phenomena, but is rather a way of looking upon all of them. It is in a sense a mathematical or philosophical point of view in contrast to the experimental one. A very brief outline of its physical basis, together with some of its interpretations, will be given below.

Relativity in an ordinary sense is a very familiar concept. Every motion that we describe is a relative motion in the strictest sense of the word. The surface of the earth is not at rest, although we usually refer velocities and accelerations to it as if it were. The sun is not at rest, although we may quite satisfactorily describe the motions of all parts of the solar system as if it were. In general, each and every motion in one system of bodies may be described with reference to some other body or system of bodies which may itself be assumed at rest. Furthermore, we have no experimental means for ascertaining the motion of a closed system *from within the system.* In a closed elevator we have no means whatsoever of telling whether the car is moving or not, unless we can see out, or watch the controller lever (which represents a connection with the outside), or remember how we were moving when we last looked out. Our sensations, as well as the results of any mechanical experiments which we might perform within the car, would be the same whether the car were at rest or moving in either direction at constant speed. The momentary increases or decreases in weight which we occasionally experience might be due to either a starting or a stopping of the car (depending upon the direction of its motion, which we do not know); or, if our outside experience did not tell us it was impossible, they might be due to alterations in the force of gravity.

This type of relativity is sometimes called *Newtonian relativity.* The discovery of the electromagnetic nature of light waves seemed to suggest a real possibility for the detection of absolute motion, since this theory seemed to require the existence of a medium, called the *electromagnetic ether,* through which these waves could travel. Although not a material medium, in the sense of one possessing mass or inertia, this electromagnetic ether must possess certain electrical properties, and it seemed almost essential to the theory that it be fixed rigidly in position throughout space, with all bodies moving freely through it. This electromagnetic ether should then serve as the absolutely stationary system to which all motions might be referred. The comparisons could be made by means of light (or radio) beams.

According to this early idea, light or radio waves should always travel (in a vacuum) with the same speed, *relative to the medium* (ether) in which they are traveling, whether the source be moving or not. This is our experience with respect to sound waves traveling in air, and it is the deduction arrived at by any straightforward wave theory. In a very considerable sense this is also true for light, as is evidenced very clearly by the Doppler effect: if the source of light is a star which is moving towards us, the successive waves are piled together (see Sec. 120), and all of the frequencies of the spectrum lines in its light are raised. A star receding from us exhibits the reverse phenomenon—the frequencies are lowered. Here again, however, it is only the motion of the star *relative to us* that we perceive. By observing similarly the light from many stars, we may of course infer how much of such motions belong to our system, and how much to the motions of the other stars, assuming the average motion of all of them to be zero; but that is an assumption which we may not verify in any way by such observations.

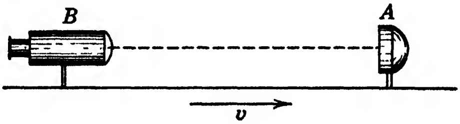


FIG. 176.

Another type of experiment, however, seems to promise successful results. If we are moving with respect to the ether, say in the direction indicated by v, Fig. 176, then light, traveling from A to B, would travel through the ether with the speed c, but would approach B from A with the *relative* speed $(c + v)$. If the distance between A and B is D, the time required for this trip would then be

$$t' = \frac{D}{c + v} \tag{1}$$

rather than

$$t_0 = \frac{D}{c} \tag{2}$$

the time required had the system been at *rest* in the ether. On the other hand, if the light were traveling in the opposite direction, the relative speed would be $(c - v)$, and the time for the trip

$$t'' = \frac{D}{c - v} \tag{3}$$

Offhand, it would seem a simple matter to check this theory by timing the travel of light between two points, first in one direction and then in the other. Unless we were so unfortunate as to chose our two points along a line at right angles to the direction of our motion relative to the ether, and this may easily be guarded against by trying several different directions, the times should come out different. Practically, grave difficulties are involved in such an experiment. Some of them are experimental difficulties, resulting from the very short times involved; these presumably may be overcome. But a very fundamental difficulty still remains: in order to time the flight of light, we must know at B the instant of time at which the light signal starts out from A. And our only means for communicating this information from A to B is by means of the very thing which we are studying, namely, a light or radio signal. One way out of this difficulty might be found. If we could have two clocks keeping perfect time, we might set them to read the same time at A and then, leaving one of them at A, carry the other one to B.[1] But even this scheme is not perfect. Our only way of rating these clocks as keeping perfect time is to compare them while they are running side by side at A. And it may be that, for some *fundamental* reason, the very act of moving one of them may cause its rate to change.

This latter scheme has not been tried, because of the experimental difficulties involved, and because there is already a simple and more sensitive method possible. All measurements of the speed of light are made by sending the light on a round trip—across and back again. If this occurs in the direction in which the earth is moving through the ether, the time for the complete trip is

$$\begin{aligned} T_1 = t' + t'' &= \frac{D}{c + v} + \frac{D}{c - v} = \frac{2Dc}{c^2 - v^2} \\ &= \frac{2D}{c} \frac{1}{1 - v^2/c^2} \end{aligned} \tag{4}$$

---

[1] This would of course require a clock whose rate was independent of gravity force or of any other local conditions, as well as one capable of far higher precision than has yet been attained in clocks. It is conceivable, however, that such a clock could sometime be made. There already exists a most remarkable quartz-crystal clock, whose rate is determined by the mechanical vibrations of a specially shaped quartz-crystal ring; the vibrations are maintained and counted electrically, and as long as the temperature is kept sufficiently constant, these clocks will keep time with a precision as high as that of the best pendulum clocks which have ever been made.

or greater than the time $2D/c$, which would be required had the earth been at rest relative to the ether.

Had the velocity of light been measured along a line perpendicular to the motion of the earth, the time would also have been increased by the motion, but not so much. This time, T_2, is obtained as follows:

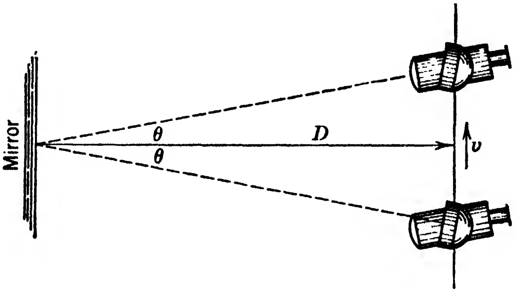


FIG. 177.

In this case the light travels along the sides of an isosceles triangle as shown in Fig. 177. From the geometry of the figure, the distance traveled by the light is

$$2 \times \sqrt{D^2 + \left(\frac{vT_2}{2}\right)^2} = cT_2 \tag{5}$$

whence

$$T_2 = \frac{2D}{c\sqrt{1 - v^2/c^2}} \tag{6}$$

Since, to judge from all known motions of the earth, v^2/c^2 is probably very small, approximate values for T_1 and T_2 may be obtained by expansions of $1/(1 - v^2/c^2)$ and $1/\sqrt{1 - v^2/c^2}$ into series. This gives

$$T_1 = \frac{2D}{c}\left(1 + \frac{v^2}{c^2} + \frac{v^4}{c^4} + \frac{v^6}{c^6} \cdots\right) \tag{7}$$

and

$$T_2 = \frac{2D}{c}\left(1 + \frac{v^2}{2c^2} + \frac{3v^4}{8c^4} + \frac{5v^6}{16c^6} \cdots\right) \tag{8}$$

When v^2/c^2 is very small, only the first two terms of the expansion are of measurable value, and the difference in these two times is then

$$T_1 - T_2 = \frac{Dv^2}{c^3} \tag{9}$$

Michelson's interferometer provides an ideal means for simultaneously timing two such light journeys at right angles to each other. In this apparatus, whose essential parts are diagrammed in Fig. 178, the light beam is divided into two by the half-silvered mirror, M_1. The reflected beam travels to M_2 and back again to the telescope T, while the transmitted beam, moving in a direction perpendicular to the first one, goes to M_3 and back, by reflection at M_1, to T. If the two distances, from M_1 to M_2, and M_1 to M_3, are made equal, and if the apparatus is at rest in the ether, the two beams coming together in the telescope will be in step, a condition which will be apparent as a set of interference fringes[2] in the middle of the telescope field. If the apparatus is moving with respect to the ether, say along the line $M_1 - M_3$, this motion will slow up one beam more than the other one, and the fringes will be shifted to the one side or the other of the field. We cannot of course start or stop the motion of our apparatus through the ether to demonstrate this shift, but we can rotate the whole apparatus so that at one time the line $M_1 - M_3$ is parallel to the motion through the ether, and later perpendicular to it. This was the procedure in the famous experiment first performed by Michelson and Morley in 1887,

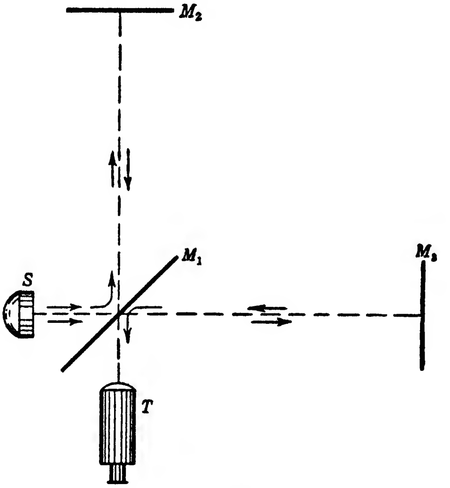


FIG. 178.

---

[2] The interferometer is adjusted so that the wave fronts of the two waves enter the telescope at a very slight angle to each other. At the middle of the field, where these wave fronts cross each other, the paths are exactly equal, and a bright fringe appears. To either side of this bright central fringe, at the places where the wave fronts are separated by distances of λ, 2λ, 3λ, etc., other bright fringes appear, as explained in Sec. 30 and illustrated more fully by Fig. 18.

and since repeated by many others. None of these experiments shows a shift of the fringes that is anywhere near as big as the theory predicts should occur as a result of the known motions of the earth in space. Most of the experimenters report that the shift is less than the experimental errors in their apparatus. D. C. Miller, who has made more observations than anyone else, finds a small shift which is far too small to support the theory given above, and which has not yet been interpreted. There exist also a considerable number of phenomena of other types in which an effect due to motion of the apparatus through the ether might be expected, but in none of them has any such effect been observed.

It is not entirely impossible to account for these negative results by quite a number of theories other than Einstein's relativity theory. None of them, however, is so simple in its fundamental concepts or so broad in its scope as the relativity theory.

Einstein's theory is based upon two fundamental postulates: (1) All motion is relative. Not only is it impossible to *measure* anything but relative motion—*absolute motion does not exist.* (2) The velocity of light, measured by any system of apparatus whose parts are at rest with respect to each other, will always be the same. This means that, if two sets of apparatus are calibrated to read the same when they are at rest with respect to each other, they will give the same value for the velocity of light, even when they are in motion with respect to each other!

If these postulates are to be reconciled with the observations that may be made upon bodies which are moving relative to the observers and the observing apparatus, the following paradoxes must exist (direct mathematical proof for them may be found in special books on this subject):

(*a*) All bodies moving relative to the observer will *appear* to be shortened in the direction of motion in the ratio of

$$\sqrt{1-\frac{v^2}{c^2}} \tag{10}$$

To be specific, let us consider two "systems," A and B, (that is two independent groups of observers, measuring apparatus, etc.) which are moving with respect to each other. Further, we shall assume that each set of observers has standards of length, time, mass, etc., which agree exactly with each other when B is at *rest* relative to A. Let the two systems be in motion with respect to each other. If we are located in system A, then we may consider that we are at rest, and that B moves with respect to us with a velocity v. This *point of view* is shown by Fig. 179(*a*). Looking at B as it goes by, we observe that all lengths in

B are *shortened* in the direction of motion in the ratio $\sqrt{1 - v^2/c^2}$. This is not apparent to observers in B. On the contrary, *they* observe that *our* scales are all shorter than theirs by the same ratio! From their point of view [shown by Fig. 179(b)] they are at rest, and our system is moving, relative to theirs, with the same relative speed but in the opposite direction.

(b) All events, such as the vibrations of pendulums, etc., when they occur upon bodies which are moving relative to the observer, appear to require more time than they do when their position is stationary with respect to him. Thus, in the special case represented by Fig. 179(a), we, in A, observe that all clocks in B are running slow (observers in B again reciprocate, and make the same observation concerning our clocks), in spite of the exact agreement of the clocks in both systems when A and B are at rest relative to each other. The ratio of the time interval as we measure it with our clocks, to that which it appears to be (to us) according to the clocks in B, is

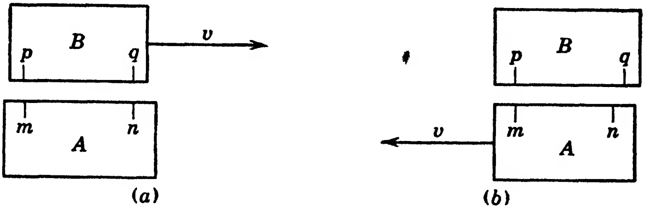


FIG. 179.

$$\frac{1}{\sqrt{1 - v^2/c^2}} \tag{11}$$

(c) An even more fundamental and intuitive concept which is destroyed by this theory is that of *simultaneity*. Two events which happen at the same spot may coincide, and thereby be simultaneous. But what about two events occurring at different places? To make this more specific, consider four points—m, n, in A; and p, q, in B—so chosen that when to us, in A, m coincides with p, n *simultaneously* coincides with q. This *point of view* is shown in Fig. 179(a). To observers in B, the situation is shown in Fig. 179(b). When p coincides with m, q does not simultaneously coincide with n. Indeed, the coincidence of q with n occurred earlier as they reckoned time!

(d) The masses of moving bodies appear to be increased in the ratio of

$$\frac{1}{\sqrt{1 - v^2/c^2}} \tag{12}$$

This is deduced mathematically as a direct consequence of (*a*) and (*b*). From this deduction there follows directly the concept of mass of kinetic energy which has been introduced in Sec. 212 as an *experimental* concept. This variation of mass with speed is the most important connection between the relativity theory and the experimental aspects of modern physics which have been considered in this book.

Whenever the relativity theory has been taken into account, it has been found to be consistent with the experimental results. In theoretical physics it plays a most important part. For example, it underlies Dirac's theory for the spinning electron and for the existence of the positron.

The theory here considered is the so-called *special* or restricted relativity, to distinguish it from the more complete theory called *general* relativity. In special relativity only steady relative velocities are considered, whereas in general relativity all types of relative motion are considered. Although general relativity is important theoretically, it has no connections with the topics considered in this book, and will not be mentioned further here.

REFERENCES

FOR ADDITIONAL READING

J. A. ELDRIDGE, *The Physical Basis of Things*, Chapters I and II.
C. P. STEINMETZ, *Relativity and Space.*
G. E. M. JAUNCEY, *Modern Physics*, Chapter XIX.
PHYSICS STAFF, UNIVERSITY OF PITTSBURGH, *Atomic Physics*, Chapter XIII.
G. GAMOW, *Mr. Tompkins in Wonderland.*
H. E. IVES and G. R. STILLWELL, "Experimental Study of the Rate of a Moving Atomic Clock," *Journal of the Optical Society of America*, Vol. 28, pp. 215–228, July, 1938. (A popular account of this work also appears in *Electronics* for June, 1938.)

FOR ADVANCED READING

R. D. CARMICHAEL, *Theory of Relativity.*
J. RICE, *Relativity.*

APPENDIX VI.

THEORY FOR RADIOACTIVE DECAY

The decay of radioactivity for any radioactive element follows a *probability* law, or law of chance, which has been stated in Sec. 214 and is illustrated by Fig. 149. The significance of this law for a single nucleus may be stated as follows: the life of any single nucleus is quite indefinite; it may decay immediately, it may exist without change for many years, or it may break up at any time in between. All we can say is that there is a certain *probability* that it may break up in any specified interval of time. This probability is proportional to the *length* of the time *interval*, and does not depend at all upon how long the nucleus has existed already. If we take the time interval to be one "half-life," as defined in Sec. 214, then there is a fifty-fifty chance for any single nucleus to survive for this interval of one half-life. In other words, it is just as likely to survive for one half-life as to disintegrate within that interval of time. If it survives for one half-life, it still has a fifty-fifty chance to survive another half-life. When an enormous number of identical nuclei are present together, the law of averages applies; approximately half of them survive and half break up in the time of one half-life, as shown by Fig. 149.

Why should this be so? Consider a nuclear transformation in which an alpha particle is ejected radioactively. If the alpha particle has enough energy to escape now, why does it wait sometimes for many years before escaping? And if it does not have enough energy to do so now, where within the nucleus can it obtain the energy with which to escape later? Gamow, in 1928, pointed out how beautifully the wave-mechanical theory answers these questions. The same explanation was proposed simultaneously and independently by Gurney and Condon. Inside the nucleus the alpha particle may have several million electron-volts of energy more than it needs to have, once it gets outside the nucleus. Indeed, once it gets far enough outside the nucleus to be beyond the range of the forces which hold nuclear particles together, the repulsion between its own electric charge and that of the nucleus will carry it away with dispatch. On the other hand, as has been suggested in Sec. 228, the forces within the nucleus are apparent only near to the surface of the droplet-like nucleus, so that the energy required for escape need be manifest for only a very short distance near the outside boundary of the nucleus.

This situation is represented graphically by Fig. 180, which shows the potential energy, V, of an alpha particle at the distances out from the center of the nucleus which are represented by r. The radius of the nucleus is indicated by r_0. The diagram resembles a high, thin wall about the nucleus, and indeed the analog is not a bad one, and is often made. The outside, curved portion of this "wall" is the potential due to the electrical forces of repulsion, while the almost straight inside portion represents the potential due to the nuclear forces. According to the old mechanical theory, in which particles are particles only, an alpha particle cannot in any manner escape unless its total energy is at least as great as the value of V represented by the top of the "wall." *If only a particle*, an alpha particle having no more energy than that represented by the horizontal line drawn at A could never escape, even though the "wall" is very thin at that level.

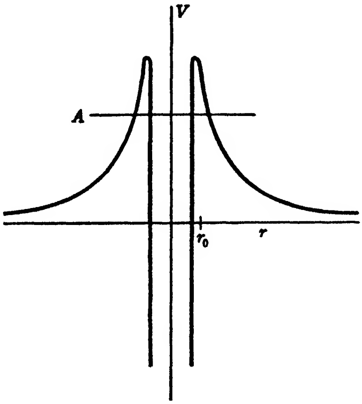


FIG. 180.

We now know, however, that an alpha particle is also a wave, and it may be that these waves are able to penetrate this potential "wall." Thus it is well known that light waves may penetrate to some degree into regions where the simple ray theory of light would not account for their going. For example, if light falls upon a polished glass surface from inside the glass, at an angle greater than the critical angle, as shown in Fig. 181(a), it will be totally reflected. If, however, a second polished glass surface is brought close to the first one, but without actually touching it, as is shown in Fig. 181(b), a small but observable amount of light will escape through the thin air film and into the second piece of glass. Figure 181 (*c*) is a photograph of this phenomenon. (The glass surfaces are slightly convex, so that the transmission of light through the optical "barrier" is limited to a small circular patch, as is seen in the photograph.)

In like manner, said Gamow, the waves representing the alpha particle whose energy is represented by the level at *A* in Fig. 180 may be able to penetrate through the potential "wall," and this small portion of the alpha-particle wave which penetrates the "wall" represents a small *probability* that the alpha particle itself may escape from the nucleus, through this thin region of high potential which encompasses it. Thus this theory is able to account exactly for the probability law which has been found experimentally. Furthermore, since this theory assumes the existence of wave-characteristics for the alpha particle, its beautiful agreement with experiment establishes the validity of the wave-particle concepts for this heavy particle.

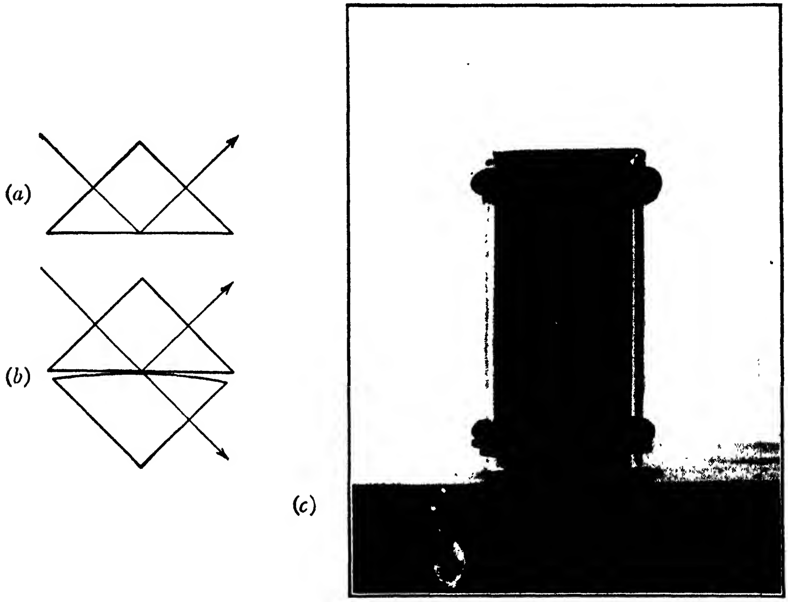


FIG. 181.

The radioactive emission of beta rays also follows a probability law which likewise may be accounted for as a consequence of the wave characteristics of the particles involved, although, as explained in Sec. 230, the fundamental processes of emission are quite different.

Index