Cambridge
Physical Series

MODERN ELECTRICAL THEORY

CAMBRIDGE PHYSICAL SERIES

MODERN ELECTRICAL THEORY

CAMBRIDGE UNIVERSITY PRESS
London: FETTER LANE, E.C.
C. F. CLAY, MANAGER

Edinburgh: 100, PRINCES STREET
London: H. K. LEWIS, 136, GOWER STREET, W.C.
London: WILLIAM WESLEY AND SON, 28, ESSEX STREET, STRAND
Berlin: A. ASHER AND CO.
Leipzig: F. A. BROCKHAUS
New York: G. P. PUTNAM'S SONS
Bombay and Calcutta: MACMILLAN AND CO., LTD.

MODERN ELECTRICAL THEORY

by

NORMAN ROBERT CAMPBELL, Sc.D.
Honorary Fellow for Research in Physics in the University of Leeds

SECOND EDITION

Cambridge:
at the University Press
1913

Cambridge:
PRINTED BY JOHN CLAY, M.A.
AT THE UNIVERSITY PRESS

PREFACE TO THE SECOND EDITION

THIS volume, nominally a second edition, is really a new book. The great changes in our knowledge of Radiation which have been effected in the last six years by the experiments of Barkla and Bragg and by the theories of Planck and Einstein have made it necessary to treat entirely afresh those portions of the first edition which now form Part II. The Principle of Relativity and Stark's work on atomic structure have altered Part III completely. Part I has been rewritten, not in order that the title of the book might be justified, but in order that the more obvious faults of the first edition might be removed.

The new book, unlike the old, has been written to suit the needs of one definite class of students. All mistaken attempts at being 'popular' have been abandoned, and it has been assumed that the reader has an adequate acquaintance with the older physics. The book is addressed to those who, having such an acquaintance, desire to study the more modern developements of the science. Otherwise the purpose of the work is unchanged. Its primary concern is with theories and not with facts; only those facts are described which are essential to the theories discussed and are not described adequately in other text-books with which the reader may be supposed to be familiar. As before 'I have attempted to expound the subject in its logical order, to analyse the arguments by which the various phenomena are correlated, to draw special attention to the assumptions that are made.' And while the foolish restriction of analysis to so-called 'elementary methods' is no longer maintained (for how can anyone understand modern electrical theory who is not acquainted with the calculus?), my principle has still been 'to resort to calculation only when, and so far as, it is necessary to show the nature of the argument.' None of the proofs given pretend to mathematical rigidity, but I hope that none of them are physically misleading.

It is perhaps interesting to inquire how far the electrical theory of today differs from that of six years ago. There has been little change or progress in the ideas of that time. The electronic theory is as firmly established as ever today, but, while we have more knowledge of where to seek a solution of the difficulties which attend it, few of those which were discussed in the earlier edition have been definitely resolved. These difficulties concern chiefly the theory of metallic conduction and to these a new and very grave difficulty concerning specific heats has been added. Our view that positive electrification is inseparably connected with atoms is confirmed, but we seem further from rather than nearer to a complete knowledge of the internal structure of those atoms. On the other hand the centre of interest has been changed by a pure addition to our ideas; it is the problem of radiation that is the chief concern of modern physics. No one can be more conscious than myself of the inadequacy of the treatment of that fascinating problem which is given here; but I think I may claim that it is the best attempt in the English language to deal generally with the matter, for, so far as I know, there is no other.

To make room for the newer theories, several large omissions from the first edition have been made; these omissions have been rendered possible by the introduction of modern work on refraction and conduction in gases into text-books of a more ordinary type. A complete excision of all matters which are treated adequately elsewhere appeared undesirable, for some of them are intimately connected with the theories which it is the special purpose of the volume to discuss.

A word should be said as to the object of the References at the end of each chapter. It is hardly necessary to explain that the list of memoirs given is intended not as a complete bibliography, still less as an indication of those which have been most important in the developement of the subject. They are intended to inform the reader where he may find discussions, fuller than, but on the same lines as, that given in the chapter concerned, on those points which have been treated with undue brevity.

The work claims to be nothing more original than a compilation, but nevertheless it is my duty to make acknowledgements. I have received the greatest possible help from Dr H. Dember and Dr U. Meyer, who have translated the book into German. Not only have they pointed out innumerable mistakes, both large and small, but have rendered invaluable service by drawing my attention to very recent work with which I was unacquainted. It will be obvious to all that Part II owes much to the inspiration of Prof. Bragg, who, with Mr W. L. Bragg, has read through all the manuscript and offered most useful suggestions for improvement. Appendix II is the result of a discussion with Prof. Turner, who was so good as to enter on a long correspondence with me on the subjects discussed in it. Once more 'if any reader shall learn from studying the book half as much as its author has learnt from writing it, no further justification of its production will be needed.'

N. R. C.

LEEDS,
March 1913.

CONTENTS

PART I.

THE ELECTRONIC THEORY.

PART II.

RADIATION.

PART III.

ELECTRICITY AND MATTER.

ERRATA.

P. 53, last line, *for* 'Kilching' *read* 'Kilchling.'

P. 139, eqn. (163), *for* 'm' *read* 'm^2.'

P. 172, line 20, *for* 'ρ' *read* 'if ρ is.'

" " 22, *before* '$\frac{H \sin\theta}{c^2 v} . e/m$' *insert* '$1/\rho =$.'

" " 25, eqn. (176), *for* 'ρ' *read* '$1/\rho$.'

" " 26, *for* '$\rho/\sin^2\theta$' *read* '$\rho \sin^2\theta$.'

P. 177, last line but one, to p. 178, line 3, *omit from* 'but it may be noted' *to* 'of the same energy.'

P. 311, eqn. (248), *omit* '$d\nu$' on right hand side.

NOTE ON THE DIRECTION OF VECTORS

A statement is necessary as to the conventions adopted in this volume for the directions in which vectors are counted positive and negative.

'Positive' and 'negative' rotations about an axis are related to the positive direction of that axis as rotation is related to translation in a right-handed or left-handed screw respectively. A quantity which is of the nature of a rotation is counted positive or negative accordingly as the rotation is positive or negative.

When rectangular coordinates are employed they are placed so that positive rotation about the positive direction of x turns the positive direction of y towards the positive direction of z; similar statements are true if the letters x, y, z are commuted cyclically. Thus if x is positive towards the north, y towards the east, z is positive downwards. This convention is universally adopted in continental treatises; unfortunately the opposite convention is adopted in many well-known English text-books.

The direction of the magnetic intensity defined by (5) turns out to be that of the acceleration caused by it in a 'north' (i.e. north-seeking) pole. The relation of the direction of the magnetic intensity due to a straight current circuit to the direction of the (positive) current is that of positive rotation and so is that of the direction of a circular current to that of the magnetic intensity at its centre. The direction of the electric intensity is, of course, that of the acceleration of a positive charge.

PART I

CHAPTER I

THE PROPERTIES OF ELECTRICITY

The old and the new science.

1. THE electrical science of the last century was concerned with the properties of electricity; modern electrical science is concerned with the relations between electricity and matter. The older workers investigated primarily the phenomena which were exhibited by all charged bodies, whatever their nature in other respects; to-day we are more interested in the phenomena which are peculiar to certain bodies; by means of a knowledge of the electrical properties of substances we attempt to establish a connection between their other properties.

Of course the distinction cannot be maintained as applicable to every detail. Measurements of some of the more important electrical properties in which bodies differ were made more than a century ago, and, on the other hand, our views of the properties of electricity have been considerably modified by experiments and theories essentially modern. But the distinction can be clearly traced and it is of logical as well as of historical importance; for, until we know the properties of electricity, we cannot enter on the task of explaining in terms of those properties the laws which relate the qualities of material bodies. The final achievement of the older science is the foundation on which its successor builds.

We must start then with an account of the properties of electricity, the properties, that is to say, which bodies possess merely in virtue of the fact that they are charged. Here only a very summary account need be given, for the subject is discussed in innumerable treatises with some of which the reader is supposed to be acquainted.

It must be clearly understood that no attempt is made here to prove the various propositions concerned, which are treated as assumptions on which the subsequent discussions are to be based. In so far as the discussions represent the experimental facts they establish the truth of the assumptions.

Electrostatics.

2. The properties of electricity are conveniently divided into two classes: those of electricity at rest, which form the subject matter of electrostatics, and those of electricity in motion, which are studied by electromagnetism.

A set of charged bodies, all at rest relative to each other, exhibit certain attractions and repulsions; for the purpose of studying these forces it is convenient to imagine the bodies divided into particles of infinitesimal dimensions, which may be numbered 1, 2, 3,.... The following laws express the results of the observations which have been made on such systems:

The attraction exerted on the particle 1 can be expressed as the product of two factors, one of which depends only on the nature of that particle and the other only on the nature of the others and their relation to the first. The first factor, e_1, is termed the charge on the particle 1, the second, D, is called the electric displacement due to the remainder of the system. e_1 is a scalar, D a vector quantity.

Since the electrostatic attractions are forces in the Newtonian sense, we may suppose that the force on 1 due to the whole system is compounded of forces due to each of the particles composing it, and that the forces between any two particles are mutual. Hence the contribution of the particle 2 to the value of D must again be the product of two factors, one the charge on 2, e_2, and the other, f_{12}, determined by the relation of 2 to 1. It is found that it is consistent with experiment to suppose that f_{12} is proportional to $\frac{1}{r_{12}^2}$, where r_{12} is the distance of 1 from 2; f_{12} is a vector and its direction is in the line joining 1 and 2.

These are the laws of electrostatics in the form in which they are employed by modern electrical theory. If F_1 is the force on

1 and ρ the volume density of the charges at any point, they may be expressed in the analytical form

$$F_1 = ke_1 D \ldots\ldots(1), \qquad \operatorname{div} D = \rho \ldots\ldots(2).$$

k is a quantity determined by the choice of the unit of charge; the choice will be made so that $k = 4\pi$[1]. It must be noted that nothing has been said about the nature of the medium between the charged bodies, for all material bodies in the field are supposed to be included in the system of charged bodies which is investigated. There is no distinction made between 'true' charges and 'induced charges'; if a piece of material, uncharged when it was without the field, is introduced into the field, the system of charged bodies is supposed to be changed by the addition to it of the charges which are sometimes said to be 'induced' in the material.

3. Suitable experimental conditions for the study of the properties of electricity in motion are most easily realised by producing the motion by joining two or more charged bodies by means of conductors. It appears at first that the phenomena in such cases are dependent on the nature of the conductor and that, since they vary with the properties of material bodies, they are not determined wholly by the properties of electricity which alone are being described at present. But investigations, most of which were undertaken with the ostensible object of 'establishing the identity of frictional and voltaic electricity,' have shown that the phenomena are wholly determined by the rate at which a charge of electricity is conveyed, by the direction in which it is conveyed and by the values of D along its path. The quantities can be measured by methods which are independent of the particular properties of any material substance.

Electro-magnetism.

Again, in treating the theory of electromagnetism, it is convenient to introduce a quantity termed the 'magnetic intensity,' H, which is measured experimentally by the couple exerted on a magnet. But, if we limit ourselves to the properties of electricity only, we can know nothing of magnets, and H has to be

---

[1] Since we shall often be concerned with numerical values of electrical and magnetic quantities, it is better to adopt ordinary electrostatic units than the less familiar units of Heaviside which abolish the constantly recurring factor '4π.'

regarded as a quantity introduced for mathematical convenience only and defined in terms of electrical quantities. Accordingly we start by defining H as a quantity determined by the equation

$$\int H\,ds = 4\pi \int C\,dS \qquad \text{(3)},$$

where the integral on the left side of the equation is taken round the boundary of the surface over which the integral on the right side is taken, and C is the 'current' which, for the present, will be defined as the charge conveyed per unit of time through unit area of that surface and along the normal to it. H is a vector parallel at each point of the boundary to the boundary at that point. The equation is often written in the equivalent form

$$4\pi C = \text{curl } H \qquad \text{(4)}.$$

By means of H we can now express simply two experimental laws. The first concerns the force exerted between two charged bodies which are moving relatively to each other. If v is the velocity of a body carrying a charge e relatively to the observer who is measuring the value of H at points along its path, then, in virtue of its motion, the body is subjected to a force eHv/c^2 in a direction at right angles both to H and v. Hence we may express in analytic form the total force to which any charged body is subjected by electrostatic and electromagnetic forces as

$$F = e\left(4\pi D + \frac{1}{c^2}[v \,.\, H]\right) \qquad \text{(5)}.$$

The second law is that of 'induced currents,' discovered by Faraday. If dS is an element of an area in a region where the value of H is varying, ds an element of the boundary of that area, then

$$\int D\,ds = -\frac{1}{4\pi c^2}\frac{d}{dt}\int H\,dS \qquad \text{(6)},$$

where D is the component of the electric displacement parallel to ds, and H the component of the magnetic intensity normal to dS. By a consideration of the dimensions of the quantities in (1), (2), (3) it will be seen that c has the dimensions of a velocity. The value of c has been determined experimentally to be 3×10^{10} cm./sec. The equation is often written in the equivalent form

$$-\frac{1}{4\pi c^2}\frac{dH}{dt} = \text{curl}\, D \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots (7).$$

Once more it will be noted that no quantity has been introduced representing the properties of the medium between the moving charges. When we come to consider magnetic substances the treatment which has been adopted will force us to adopt the view that they are the seat of currents determining values of H by means of (3).

4. We have now stated the fundamental properties of electricity as they are known to modern electrical theory. The form in which these laws has been expressed has been chosen for its suitability for mathematical developements. To most of the older students of physics the science which they studied was interesting chiefly because it gave rise to such developements, and when they had enunciated the propositions which have been given, the goal of their endeavours had been attained. But there are other students, much more numerous among the younger than among the older generations, who are by no means content to regard such propositions as final. They could not derive from analytical exercises alone the intellectual satisfaction which is the object of all scientific research; they desired theories which would 'explain' the mathematical laws.

Theories of electricity.

When a theory is said to explain certain laws two things are usually meant. In the first place it is meant that the theory establishes a connection between the laws; it replaces a set of apparently independent propositions by a single proposition from which they may all be deduced. In the second place it is usually meant that the theory enables the laws which it explains to be 'visualised'; it traces an analogy, more or less close, between the phenomena expressed by the laws and some other phenomena, usually of a mechanical nature, with which we are familiar in everyday experience, so that from our knowledge of what would happen in any given circumstances of the familiar phenomena we can make, by means of the analogy, some estimate of what will happen in the less familiar phenomena to which they are analogous.

A well-known example of such a theory is the dynamical theory of gases. It consists in the first place of a set of propositions from which such laws as those of Boyle and Gay-Lussac can be deduced; apart from the theory there would be no reason whatever for believing that a gas which obeys one of those laws must also obey the other. In the second place it traces an analogy between the expansion of the walls of a vessel under the pressure of the gas which it contains and the completely familiar movement of a body under the impact of moving masses.

Faraday's Theory.

5. The only important theory which has ever been proposed to explain the properties of electricity is that of Faraday, and it is as the framer of this theory that Faraday must be proclaimed the founder of modern electrical science. Not only was the discovery which led to the developement of that science a direct consequence of the consideration of his theory, but the spirit which led him to frame it is typical of the modern physicist rather than of his predecessors. As is well known, the theory was suggested to Faraday by his discovery that all bodies are not alike in respect of electrostatic action, that if charged bodies are separated by a solid dielectric they attract each other less than if they are separated by air. Faraday was the first to study the relations between electricity and matter rather than the properties of electricity itself.

The main features of Faraday's theory have been described so often that the most summary account of them will suffice in this place. Faraday supposed that the region surrounding charged bodies was traversed by 'lines of force,' the ends of which were rigidly attached to bodies charged with electricity of opposite signs. The lines were analogous to elastic strings in that they constantly tended to contract and so caused the opposite charges at their ends to approach. They differed from elastic strings in that their tension was independent of their length and that neighbouring lines tended to repel each other. These properties of the lines are sufficient to explain, at least qualitatively, all the laws of electrostatics.

Faraday was less successful in applying his theory of electromagnetism. He showed that, when a current flows through a wire, lines of force from the region surrounding the wire must be travelling into the wire, where they disappear, while fresh lines of force are sent out by the source maintaining the current. To this motion of the lines of force through the region surrounding the wire the magnetic phenomena displayed in that region were to be attributed, but no analogy was suggested whereby the magnetic action of the moving lines might be visualised as their electrostatic action could be by means of the analogy with elastic strings.

Developement of Faraday's theory.

6. Faraday's theory underwent little alteration or developement at the hands of subsequent workers. The all-important contributions of Maxwell to electrical science were, as we shall see, inspired by that theory, but no attempts were made for a long time to show that it could be adapted to the quantitative treatment of electrical phenomena. This extension was ultimately made by J. J. Thomson. Even in the form which he gave to the theory it is not so well adapted as the simple laws in the form which have been given for the mathematical deduction of their consequences, but some of his results will be useful later and will be quoted without proof.

In order to explain quantitatively the electrostatic and electromagnetic laws it is sufficient to add to Faraday's theory the following hypotheses:

(1) The number of tubes of force[1] having their ends on a body is equal to the charge on that body.

(2) The tension along the length of a single tube of force is $2\pi N$, where N is the number of tubes passing through unit area drawn perpendicular to the tube at the point where the tension is considered. The lateral pressure which a tube exerts on its neighbours at right angles to its length is $\frac{2\pi N}{\omega}$ per unit area of the tube, where ω is the area of the cross-section of the tube.

---

[1] Thomson supposed that the 'lines of force' had a finite volume and accordingly substituted the term 'tubes' for 'lines.'

(3) A tube of force travelling with a velocity v relative to the observer produces a magnetic intensity $4\pi v \sin\theta$, where θ is the angle between the direction of the tube and the direction of its motion. The magnetic intensity is perpendicular to both these directions.

A difficulty.

7. Before proceeding to consider the consequences of Faraday's theory it will be well to notice briefly a difficulty which may have occurred already to the reader.

We have attributed both electrostatic and electromagnetic action to the properties of the same tubes; the former has been attributed to the presence of tubes at rest, the latter to the presence of tubes in motion. It would seem, then, that if any region is by any means cut off from the electrostatic influence of a system, it must also be cut off from its electromagnetic influence. But it is well known that any region surrounded by a closed conductor is isolated completely from the electrostatic action of any system outside the conductor, while the presence of the closed conductor has no influence on the magnetic action of steady currents outside it.

To overcome this contradiction we must alter slightly our notion of the relation of a tube to its charge. Hitherto we have regarded the tube as joining two opposite charges to both of which it belongs equally. The difficulty which has been pointed out forces us to the conclusion that each charge carries its own tubes with it independently of the presence of other charges and that tubes attached to charges of opposite sign have oppositely directed properties.

The use of an analogy, to which no special importance is attached, will make this matter clearer. Let us suppose that the electrostatic properties of a tube are determined by a rotation of the contents of the tube round its axis, and that the direction of rotation is related differently to that of the tube according as the tube is attached to a positive or negative charge (see Fig. 1). In Fig. 2 the thick lines represent infinitely large condenser plates bearing opposite charges; it will be seen that between the plates the rotation of the positive tubes is in the same direction as that of the negative tubes, while outside the plates the rotations are in opposite directions. Since we have attributed the electrostatic action to the rotation of the tubes, the two sets of tubes will act in the same direction between the plates and in opposite directions outside them; the existence of an electric field between the plates and of no field outside them is explained. On the other hand, magnetic action must be supposed to be determined not by the rotation of the tubes but by the translatory motion of the two sets of tubes relatively to each other. If one of the condenser plates moves relatively to the other carrying its tubes with it, any action determined by this translatory motion will take place outside as well as between the plates.

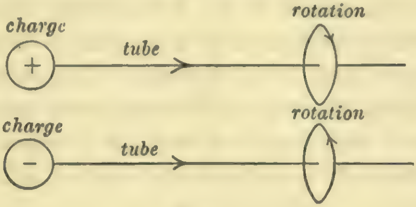


Fig. 1.

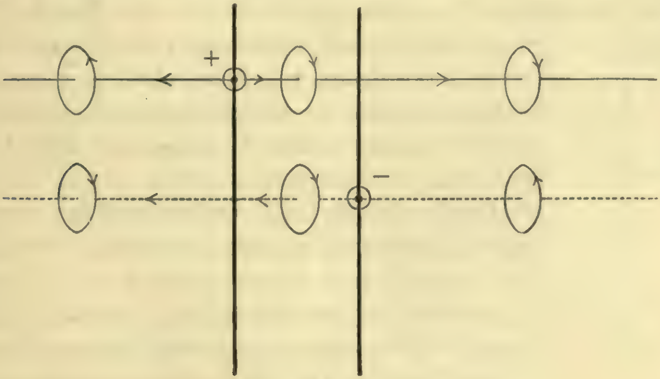


Fig. 2.

This modified view necessitates a slight revision of the quantitative relations just given, but since the alterations are unimportant for any of the calculations given here, it is undesirable to introduce further complications by making them. The only difference which is made by the modification of the theory which has just been introduced is that the connection between a positive and a negative charge no longer appears so fundamental. It is now quite possible to conceive of the production of a positive charge without a negative charge; the fact that such a production does not actually occur requires further explanation.

The mass of the tubes of force.

8. It is a consequence of the fundamental equations that the energy contained in unit volume of a uniform magnetic field H is $H^2/8\pi c^2$. According to Faraday's theory the value H is connected with the presence in that volume of unit length of each of N tubes travelling with a velocity v, where $H = 4\pi N v \sin\theta$. Hence the energy due to each of these tubes is $\frac{2\pi N}{c^2} v^2 \sin^2\theta$. The occurrence of the factor v^2 immediately suggests that the energy may be expressed as kinetic energy of the tubes; the desired result is attained if to each unit length of the tubes is attributed a mass

$$\frac{4\pi N}{c^2}\sin^2\theta.$$

The conception of a mass connected with the Faraday tubes is one of the most important consequences of the theory. We have not been able to obtain any satisfactory analogy which makes it possible to visualise the production of a magnetic field by the tubes and the associated action of the tubes on each other due to their motion, but the new conception enables us at once to visualise the process of the induction of currents, that is, the production of a current in one circuit by a variation in the intensity of the current in a neighbouring circuit.

Fig. 3.

Let $PQRS$ and $P'Q'R'S'$ (Fig. 3) be two circuits lying one inside the other in the same plane. $PQRS$ contains a battery or some other source of the production of Faraday tubes so that a current flows in it accompanied by the motion of tubes. These tubes may be imagined to stretch across from PQ (+) to SR (−) and to move down the circuit from top to bottom with their axes parallel to the portion QR

into which they collapse at the end of their passage. Under the influence of these tubes there will be 'induced electrification' in $P'Q'R'S'$, or, in the terms of Faraday's theory, the tubes of the metal of the circuit (see Chap. III) will be orientated in such a way that there is no electric intensity in the substance of the conductor. Now suppose that the conductor QR is removed and the circuit broken. The tube which was just about to collapse into QR will now be unable to do so: it will remain stretching across the gap and exerting a lateral pressure on the tubes above it tending to prevent them from moving down the circuit. But it must be remembered that the tubes possess mass and that a finite force will require a finite time in order to reduce them to rest. For a short interval the tubes will continue to move down the wires PQ and RS, but, being unable to escape at the ends, they will remain piled up about the region QR. The distribution of the tubes over the area of the circuit is changed: there are, relatively to the former condition, more tubes in the lower part of the circuit than in the upper.

This change in the distribution of the tubes will cause a readjustment of the 'induced electrification' in the circuit $P'Q'R'S'$. It was in equilibrium previously, but now the number of tubes in the region $Q'R'$ is greater relatively to that in the region $P'S'$. Accordingly there must be a movement of electricity in the sense determined by the action of the tubes in the lower part of the circuit: that is to say, positive electricity will move in the direction $Q'R'S'P'$ and tubes having their positive ends on the part $P'Q'$ must move down the circuit. The motion of these tubes will give rise to a magnetic intensity, and since the motion of the tubes is in the same direction as that of the tubes in $PQRS$ before the circuit was broken the magnetic intensity will be in the same direction as that due to the current in the inducing circuit. But this latter intensity has disappeared by the breaking of the circuit: hence the intensity due to the induced current will be such as to counteract the change in the field due to the stoppage of the inducing current.

The reader may work out for himself the processes which occur when the current is started in $PQRS$, and he will find that they are such as to give rise to the effects which are observed experimentally. It will be noted that no attempt has been made to prove that the quantitative relations obtained are the same as those given by the usual theories: the calculations would be exceedingly complicated, and since they are not required for future reference and would not throw further light on Faraday's conceptions of the electromagnetic field, they may be omitted without detriment.

Consequences of the theory.

9. So far we have only regarded Faraday's theory as an alternative means of expressing the laws which were adequately expressed by the mathematical theory of §§ 1–3. But it is the most valuable and the most remarkable feature of all true theories, such as that of Faraday, that they not only express the laws to explain which they were designed, but also suggest new laws which were not suspected at the time they were framed. We must now turn our attention to an instance of this kind, the feature of Faraday's theory which renders it of such supreme importance for modern electrical theory.

According to the older mathematical theory of electromagnetism (mainly due to Ampère) the existence of a magnetic field is inseparably connected with the motion of charged bodies or at least of charges; there is nothing in the theory to suggest that there could be a magnetic field apart from the motion of charges. But according to Faraday the magnetic field is associated, not with the motion of charges, but with the motion of the tubes attached to them, and, since the tubes are flexible, there is every reason to suppose that the tubes might move without the charges. It is true that if we started with a system of tubes and charges at rest it would be impossible to produce a motion of the tubes without producing a motion of the charges, but since the theory leads us also to suppose that the tubes possess inertia, it is possible that the tubes, set in motion by means of the charges at their ends, might persist in motion after the motion of the charges had ceased.

No action of this kind has ever been detected by direct experiment, and we know now that the experimental difficulties of direct detection are quite insuperable, but we can obtain indirect evidence as to the correctness of the suggestion that magnetic fields may exist without the motion of charges or charged bodies[1].

Velocity of disturbances.

10. If a uniform elastic string under a finite tension and possessing a finite mass is slightly distorted at any part and then released, a disturbance reproducing the form of the distortion travels along the string in each direction with a uniform finite velocity. It is proved in any treatise on vibrations that, if the small disturbance is 'transverse,' any point being displaced only in a direction perpendicular to the length of the string, the velocity with which the disturbance is propagated is $(T/m)^{\frac{1}{2}}$, where T is the tension of the string and m its mass per unit of length. A longitudinal disturbance will be propagated with a different velocity determined by the variation of the tension of the string with its length.

The Faraday tubes, to the action of which we have attributed the electrostatic and electromagnetic action, resemble such an elastic string in being subject to a finite tension and in possessing a finite mass. If our analogy is trustworthy, a transverse disturbance impressed on a tube at any point should be propagated along that tube unchanged in form with a finite velocity; but a longitudinal disturbance will not be propagated at all, because the tubes are not elastic in the sense of resisting alterations in length. It is easy to calculate the magnitude of the velocity of propagation

---

[1] Some writers appear to think that the experiments of Rowland and Roentgen on the magnetic effect of electric convection afford direct evidence of the correctness of the suggestion of Faraday's theory (as interpreted by Maxwell). What Rowland's experiments prove is that a moving charge is equivalent in its magnetic action to a current flowing in a conductor along the path of the charge and conveying electricity at the same rate; Roentgen's experiments add to this proof the discovery that in magnetic action an 'induced charge' is precisely equivalent to a 'true charge.' These propositions had to be proved before the old question of the 'identity of frictional and voltaic electricity' could be definitely answered in the affirmative, for that question asked (among other things) whether electricity conveyed from one body to another through a conductor had the same properties as electricity conveyed through a dielectric. But they have nothing to do with the question whether the chief feature of Maxwell's view of the nature of the current which determines the magnetic intensity is correct; the feature which differentiated Maxwell's from all previous theories and led to his theory of light is the hypothesis that any change in 'electric displacement,' even if it is unaccompanied by the motion of charges, is to be regarded as a current determining a magnetic field.

if the distortion is so small that it does not disturb the uniformity of the field. We must remember that the tube is subject not only to a tension in the direction of its length, but also to a uniform pressure on its sides. The effect of this pressure can readily be shown to be equivalent to an increase in the tension of the tube.

For, if the pressure were the same over the ends of the tube as over its sides, the tube would be subject to an uniform hydrostatic pressure and would be in equilibrium; accordingly, the pressure on the ends must be in equilibrium with the pressure on the sides, so that the resultant of the latter pressure must be equal and opposite to that of the former; that is to say, the resultant of the pressure on the sides must be a pull on each of the ends of the tube, equal to the product of the area of the ends and the pressure per unit area on the sides. Since this pressure is $\frac{2\pi N}{\omega}$, the tension on the ends of the tubes due to the pressure on its sides must be $2\pi N$ and equal to the tension along the length of the tube. Hence the total tension of the tube is $4\pi N$.

The mass of the tube per unit of length is $\frac{4\pi N}{c^2}$, since the tube moves at right angles to its length; hence the velocity with which the disturbance is propagated along the tube is

$$\left(4\pi N \middle/ \frac{4\pi N}{c^2}\right)^{\frac{1}{2}} = c.$$

We must now inquire what will be the physical effect of the propagation of such a disturbance along a set of adjacent parallel tubes. The disturbance at any point may be regarded as being produced by the superposition on the undisturbed tubes (in number N) of a set of tubes (in number N') at right angles to the direction of the undisturbed tubes. The propagation of the disturbance consists in an alteration of the undisturbed field by the passage through it of the N' tubes with velocity c. As they pass over any point they will produce an electric displacement $4\pi N'$ in the direction of their length and, since they are moving at right angles to their length with velocity c, they will produce in a direction at right angles to the electric displacement and to the direction of propagation a magnetic intensity $4\pi N'c$; thus the electric displacement and magnetic intensity due to the disturbance will always be at right angles and proportional to each other.

Hence we see that Faraday's view of the nature of electrostatic and electromagnetic fields leads directly to the suggestion of the possibility of processes which could not have been anticipated by the consideration, unenlightened by that theory, of the quantitative laws which have been enunciated. It is suggested that if a charged body is moved in any way the portions of the field at some distance from it should not be disturbed until a time has elapsed sufficient to allow a disturbance travelling with a velocity c from the charged body to reach them, and that the disturbance when it arrives should consist of electrostatic and magnetic fields proportional to each other, the former being in the direction in which the charged body was moved, and the latter at right angles to this direction and that of the line joining the disturbed point to the original position of the charged body. Further since N, the number of tubes attached to any charged body passing through unit area at right angles to their length, varies inversely as the square of the distance from the charged body (at points outside the electrostatic field due to the body), and since the fields of the disturbance are proportional to N, the intensity of those fields should vary also inversely as the square of the distance.

The reader is, of course, aware that the view is now universally accepted that it is such disturbances which appear to our senses as light.

Maxwell's work.

11. These consequences, though they follow naturally from Faraday's theory, were not perceived by Faraday himself. It is curious to observe, when reading his works, how near at times he came to the modern view and yet how far he was from ever grasping it. He perceived that an 'electric current' possessed some property entirely analogous to inertia, and yet, with curious inconsistency, in considering that inertia he abandoned his theory and attributed the inertia to the electricity flowing in the wire rather than to the lines of force outside it; so he failed to realise fully the analogy between those lines and elastic strings along which disturbances could be propagated with a finite velocity. Though he was continually searching for some connection between electric and magnetic phenomena and light, he deliberately contrasts 'the uses of the aether in conveying magnetic force and in transmitting radiations,' suggesting that 'the aether' may have a double use rather than that the two uses may be the same. Probably he was hampered in his speculations by a difficulty in breaking through the traditional division of the branches of physics into the separate studies of different agencies affecting different senses; he found it as hard to conceive of any relation between electricity and light as his contemporaries found it to imagine how 'heat,' affecting one sense, could be 'a mode of motion,' usually affecting another.

The suggestiveness of Faraday's theory remained undiscovered until it was brought to light by Maxwell. But though Maxwell's work was professedly inspired by Faraday's ideas, the terms in which he expressed them were entirely different. Maxwell had the analytical training which Faraday lacked and he preferred to modify the work of the great mathematicians, Ampère, Gauss, Neumann and the rest, rather than to abandon it completely. In that work the idea of a 'current circuit' is fundamental, while in Faraday's theory it appears only as a special instance of the motion of lines of force; instead of expressing magnitudes connected with circuits in terms of magnitudes connected with lines of force, Maxwell preferred to adopt the converse process. It was the fact that he was thinking in one language and writing in another that made his work so difficult to understand and delayed its general acceptance until it had been remoulded by Hertz.

Maxwell modified the Amperean theory of electromagnetism by introducing on the right hand side of equation (3) a term representing a current which determines a magnetic intensity in the absence of moving charges. The quantity C, according to Faraday's theory, is determined by the number of tubes entering the area considered; if these tubes are attached to charges which cross the area, the appropriate term can easily be shown to be ρv, where ρ is the density and v the velocity of the charges; and if the tubes are attached to charges which do not cross the area the appropriate term is $\frac{dD}{dt}$. It is the introduction of this second term which is characteristic of Maxwell's theory. We now write

$$\int H ds = 4\pi \int \left(\rho v + \frac{dD}{dt}\right) dS \text{ ....(8),}$$

or

$$4\pi \left(\rho v + \frac{dD}{dt}\right) = \text{curl } H \text{.....................(9).}$$

If Maxwell's theory is equivalent to Faraday's it must, of course, lead to the same consequences; and indeed any treatise on the theory of electromagnetism will show how from equations (2), (7) and (9) the same conclusions may be drawn as to the finite velocity of propagation of the disturbances set up by the motion of charges, and as to the relation between the electrostatic and electromagnetic fields in those disturbances. As was said before, the analytical theory is much better suited than Faraday's for the mathematical developement of the consequences of the various laws, just as Faraday's theory is much the more valuable to those who wish to 'visualise' the processes taking place.

Maxwell's Theory of Light.

12. When Maxwell suggested that light might be considered as a special periodic form of the electromagnetic disturbances which we have just considered, he could bring to the support of his hypothesis evidence which was very striking but not entirely complete. The disturbances, like light, were transverse and therefore capable of being polarised; they were also, like light, propagated through vacua and media optically similar to vacua with a velocity independent of the frequency; and the results of the measurements of this velocity did not differ from the results of the measurements of c more than the results of either set differed amongst themselves. If account had had to be taken of none but such media, it would have been hardly possible to conceive of any more conclusive evidence; the only properties attributed to light other than a definite velocity and polarity are intensity and frequency, and the electromagnetic disturbances could be imagined to have any such properties as might be determined by optical experiments. It would have hardly been necessary to wait until the brilliant experiments of Hertz had produced by electrical means disturbances differing from light only in frequency, in order to accept the hypothesis with complete confidence.

But when the propagation of light through denser media was taken into consideration, the agreement between the calculated properties of the electromagnetic waves and the observed properties of light was somewhat less satisfactory. The statement of electromagnetic theory which has been given in this chapter omits all reference to the properties of special material bodies and, of course, could not predict anything about the propagation of disturbances through them. But it will be convenient here to make a brief excursion from the proper limits of the chapter and to sketch briefly (although it is probably well known to the reader) the manner in which the properties of material bodies entered into the Maxwellian theory.

Material Media.

13. On this theory the electromagnetic properties of a substance are supposed to be completely determined by three magnitudes, the dielectric constant K, the magnetic permeability μ, and the conductivity σ. By means of the first two quantities two new vectors are defined, the electric intensity, $E=\frac{4\pi D}{K}$ and the magnetic induction, $B=\mu H$. The third quantity is used to define, by means of Ohm's Law, a part of the current C which is peculiar to conductors and equal to σE. The fundamental equations are then stated to be

$$F_1=e_1\left(E+\frac{1}{c^2}[v\,.\,B]\right)\text{.........}(10);\qquad \operatorname{div} D=\rho\text{............}(11),$$

$$4\pi\left(\sigma E+\rho v+\frac{dD}{dt}\right)=\operatorname{curl} H\text{...}(12);\qquad -\frac{1}{c^2}\frac{dB}{dt}=\operatorname{curl} E\text{...}(13).$$

In a region where no material bodies are present, K and μ are given the value 1, σ the value 0, and the equations reduce to the form given in (2), (5), (7), (9).

From these equations the following conclusions may be deduced, for a proof of which the reader is referred to any modern treatise on light:

(1) An electromagnetic disturbance of any form travels in

a medium for which σ is zero with a velocity $c/(\mu K)^{\frac{1}{2}}$. The frequency of a periodic disturbance is unchanged by transition from one medium to another, but at the surface separating media with different values of K and μ part of the disturbance is reflected, while another part is refracted. The argument of Huyghens (which is quite independent of the nature of the wave considered) shows that, when a plane wave is incident on such a surface, the angle of reflection (r) is equal to the angle of incidence (i), while the angle of refraction (θ) is related to i by the relation $\frac{\sin\theta}{\sin i}=\frac{v_1}{v_2}$, where v_1 and v_2 are the velocities of the disturbance in the first and second medium respectively. If the incident plane wave is plane polarised with the electric displacement in the plane of incidence, the relation between the intensities of the incident, reflected and refracted waves is that found experimentally between the intensities of these waves when light, polarised with its plane of polarisation perpendicular to the plane of incidence, is incident upon the surface of separation of two media; if the magnetic intensity is in the plane of incidence the relation is that found for light with its plane of polarisation coincident with the plane of incidence.

(2) When an electromagnetic disturbance is incident on a medium for which σ is finite, again part of it is reflected and part refracted. The ratio of the reflected to the incident wave is very nearly $1-\frac{2}{\sqrt{\sigma\tau}}$, where τ is the period of the wave; it becomes unity if σ or τ is infinite. The refracted wave does not travel through the medium, but is absorbed in a distance determined chiefly by the value of σ. Such a medium behaves towards an electromagnetic disturbance in much the same way as a 'metallic reflector' behaves towards light.

(3) Through whatever medium the disturbance travels its behaviour is unaltered by placing that medium in an electrostatic or magnetic field.

Comparison with experiment.

14. Let us now compare these predicted properties of the electromagnetic disturbances with the observed properties of light.

In some respects prediction and observation agree well. Every material medium has a value of μK which is greater than 1; we should expect, as we find, the velocity of light to be less in all material media than in a vacuum. Again the reflection and refraction of polarised disturbances are the same as those of polarised light, if we suppose that the magnetic intensity of the disturbance lies in the plane of polarisation and the electric displacement perpendicular to it. And we find that usually it is those media which have a large value of σ which behave towards light as metallic reflectors, and that the reflecting power of the better conductors is on the whole greater than that of the worse.

But there are many serious discrepancies. The theory predicts that in any medium the velocity of all disturbances should be the same, where it is found that in all material media the velocity of light varies with the frequency. Again there is a large class of bodies (electrolytic solutions) which are generally classed as conductors and yet are as transparent as media for which the value of σ is a million times smaller; and there are many highly insulating bodies which are almost as opaque as metals.

The theory predicts that for bodies for which the value of μ is 1 (i.e. all except magnetic bodies) the refractive index, n, should be related to K, by the equation $n^2 = K$. This relation is approximately valid for a large number of substances, especially hydrocarbons, if the value of n chosen is that appropriate to waves of low frequency; but for other substances, such as glass and water, it is not approximately fulfilled, nor can any clear distinction be made on the ground of their other properties between the substances for which the relation holds and those for which it does not. The relation between the conductivity and the reflecting power is approximately fulfilled in the case of such metals as silver and mercury, but not in the case of all metals; moreover the theory offers no explanation of the fact that some metals, such as copper and gold, reflect waves of some frequencies so markedly more than waves of others. But here again the agreement between the theory and experiment is better the lower the frequency of the light with which the disturbances are compared. And lastly there is no such difference in the optical behaviour of magnetic and non-magnetic metals as the theory which regards only the value of μ would lead us to anticipate. Such differences as there are (the occurrence of the Faraday and Kerr effects) are not of the kind which is expected.

The source of the discrepancy is obvious and was suggested by Maxwell himself. The constants which are introduced into the electromagnetic equations to represent the properties of material bodies are all determined by observations on the behaviour of those bodies in steady electrostatic or magnetic fields, and they are not applicable to fields changing so rapidly as those in the electromagnetic disturbances which constitute light. The failure of Maxwell's theory to deal adequately with the propagation of light through material media drew attention to the inadequacy of the prevalent view of the electromagnetic properties of those media; it directed attention from the properties of electricity, independent of particular materials, to those relations between electricity and matter which are different for different materials. Once more a theory has been almost more valuable on account of the phenomena which it failed to explain, than on account of those which it explained successfully; it was the deficiencies of Maxwell's theory which led to the foundation of modern electrical science, which in a single generation has remodelled almost every branch of chemistry and physics.

The electron theory.

15. In an earlier paragraph it was mentioned that this new science has reacted upon that which forms its foundation and has changed and extended our views of the properties of electricity. Since the nature of this change is now familiar even to elementary students of physics, it will be convenient to give a brief account of it here, leaving to later chapters any account of the considerations which led to the various hypotheses.

The fundamental hypothesis of the newer theory is that electric charges are not distributed continuously through the bodies which contain them, but are concentrated on particles the total volume of which is small compared with that of the charged body[1]. The

---

[1] It is not certain that a definite meaning can be attached to the expression 'the volume of one of these particles.' It would be more accurate to say that the greater part of the electrical energy is concentrated within volumes the sum of which is small compared to that of the whole charged body.

negative charges are concentrated on particles which have the same properties from whatever body they are derived. It is only necessary to attribute to them two properties, a charge and a mass; the charge on each of these particles (ϵ) is about $4{\cdot}7 \times 10^{-10}$, the mass of each of them (m_0) is about $8{\cdot}8 \times 10^{-28}$ grm., the ratio of the charge to the mass being about $5{\cdot}3 \times 10^{17}$.* The mass of a hydrogen atom is about $1{\cdot}6 \times 10^{-24}$ grms. These particles are termed 'electrons,' a name originally invented by Johnstone Stoney to denote the indivisible unit of electric charge, the existence of which appeared to be indicated by the phenomena of electrolysis.

On the other hand the particles on which positive charges are concentrated are, as far as we know, never smaller than atoms, so that the discontinuity in the distribution of positive electricity is of the same nature as that of the distribution of matter. The positive charge on an atom is always a multiple (probably always a small multiple) of the negative charge on an electron. An electrically neutral atom is regarded as a complex structure positively charged to which are attached a number of electrons sufficient to neutralise that charge; a positively charged atom is an 'ordinary' atom deprived of one or more electrons.

The motion of a charged particle.

16. The introduction of these assumptions does not make it necessary to change the fundamental equations of electromagnetic theory (2), (5), (6), (9), except by the introduction of Σe in place of ρ, where Σe is the sum of all the charges on the particles in unit volume. But some special cases of those equations become of such great importance that it is desirable to state them explicitly.

The first case is that of the magnetic intensity H at a point P distant r from a particle carrying a charge e and travelling with velocity v in a line making an angle θ with r. Let $l = r \sin\theta$ be the radius of a circle perpendicular to v, passing through P and having its centre on the line of v. Then, by symmetry, the value

---

* By an unfortunate convention, which can only be justified as a means of setting traps in examinations, the electronic charge is usually expressed in electrostatic units, but the ratio of the charge to the mass in electromagnetic units/grammes. Here, as everywhere in this book when the contrary is not explicitly stated, electrostatic units on the C.G.S. system are employed.

of H in the plane of the circle is the same everywhere along the circumference, so that the left-hand side of (8) is $2\pi l H$. In the right-hand side, $\rho=0$, since there are no charges passing through the plane of the circle at the moment under consideration; but D, the surface integral of the normal electric displacement over the circle, is

$$\int_0^\theta 2\pi r^2 \sin\theta\, d\theta \,.\, \frac{e}{4\pi r^2} = \tfrac{1}{2}e\,(1-\cos\theta).$$

Hence we have

$$2\pi H l = 2\pi e \frac{d}{dt}(1-\cos\theta) = 2\pi e \sin\theta \frac{d\theta}{dt}.$$

But
$$v = -\frac{d}{dt}(l\cot\theta) = \frac{l}{\sin^2\theta}\frac{d\theta}{dt}.$$

Hence
$$H = \frac{ev\sin\theta}{r^2} \qquad\text{..................(14).}$$

The component of H in any plane containing v is zero, since both ρv and $\frac{dD}{dt}$ are zero along the normal to any surface drawn in that plane. Consequently we conclude that the magnetic intensity produced at P by the moving charge is $\frac{ev\sin\theta}{r^2}$, and is perpendicular to the plane containing r and v.

The second case is that of the motion of a charged particle in a region where the electric intensity and the magnetic intensity are both finite. Without loss of generality we may take the magnetic intensity H along the axis of Z, while the electric intensity has components X, Z in the plane xz. Then, since by the force on the particle we mean the product of its mass, m, and its acceleration, (5) becomes

$$m\ddot{x} - \frac{eH}{c^2}\dot{y} - eX = 0 \qquad\text{........................(15),}$$

$$m\ddot{y} + \frac{eH}{c^2}\dot{x} \qquad = 0 \qquad\text{........................(16),}$$

$$m\ddot{z} - eZ \qquad = 0 \qquad\text{........................(17).}$$

The solution of these equations is

$$x = \left(\frac{c^2X}{H} + v_0\right)\frac{1}{\omega}(1-\cos\omega t) + \frac{u_0}{\omega}\sin\omega t \qquad\text{..............(18),}$$

$$y = -\frac{u_0}{\omega}(1 - \cos \omega t) - \frac{c^2 X}{H} t + \left(v_0 + \frac{c^2 X}{H}\right)\frac{1}{\omega} \sin \omega t \ldots(19),$$

$$z = \tfrac{1}{2} z \frac{e}{m} t^2 + w_0 t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(20),$$

where u_0, v_0, w_0 is the initial velocity of the particle starting from the origin and $\omega = H/c^2 . e/m$.

The following cases are of special importance in experimental investigation:

(1) If the electron starts from rest in a purely electrostatic field Z, it moves, of course, in a straight line along the direction of Z and its velocity at any point (z) is $\left(\frac{2\epsilon V}{m_0}\right)^{\frac{1}{2}}$ and its energy ϵV, where $V = Zz$ is the difference of potential through which it has fallen in moving from 0 to z.

(2) If the electron starts with a velocity u_0 at right angles to the magnetic intensity and there is no electrostatic field, we have

$$x = \frac{u_0}{\omega} \sin \omega t, \quad y = -\frac{u_0}{\omega}(1 - \cos \omega t), \quad z = 0 \ldots\ldots(21).$$

That is to say, the electron travels in a circle of which the radius is $\rho = \frac{u_0}{\omega}$ in a plane perpendicular to the magnetic intensity; the speed of the electron remains unaltered since its acceleration is always perpendicular to its velocity. If the initial velocity has a component w_0 in the direction of the magnetic intensity, this component remains unchanged and the electron describes a spiral of constant radius and pitch about the axis of z. If Z is finite, so that there is an electric field in the same direction as the magnetic field, the radius of the spiral is constant, but the pitch continually increases. Thus in the case when there is no electric field perpendicular to the magnetic field the electron tends on the average to travel along the lines of magnetic intensity; if this intensity is great the electrons travel in almost straight lines along the direction of the magnetic field, as they travelled along the lines of the purely electrostatic field.

(3) If $Z = 0$ while X is finite, so that the electric and magnetic fields are at right angles and u_0 and w_0 are zero, the motion of the electron is given by

$$\left.\begin{aligned} x &= \left(\frac{c^2X}{H} + v_0\right)\frac{1}{\omega}(1-\cos\omega t);\\ y &= -\frac{c^2X}{H}t + \left(v_0 + \frac{c^2X}{H}\right)\frac{1}{\omega}\sin\omega t; \qquad z = 0\end{aligned}\right\} \quad \ldots(22).$$

If $c^2X/H = -v_0$, this becomes $x = 0$, $y = v_0t$, $z = 0$. Thus if the fields are at right angles and the electron is projected perpendicular to both of them, it will travel in a straight line unaffected by the fields, if their strengths are so chosen that $c^2X/H = -v_0$.

The charge on an electron.

17. It will also be convenient to indicate here the principles of the methods by which the charge and the mass of electrons are determined.

To determine the charge on an electron the only method available is to measure directly the total charge, $N\epsilon$, on a collection of electrons, of which N, the number, is known. The problem thus reduces itself to that of determining N. In one of the best methods the conditions of experiment are so arranged that N is known to be 1, in others the value of N is calculated by means of some theory. The first method and most forms of the second depend upon phenomena which will be considered at length later, but there are some forms of the second method which depend only on the results of the dynamical theory of gases, which does not require special treatment in this volume. These methods will be considered briefly here.

There is every reason to believe that the charge carried by a univalent electrolytic ion, such as the hydrogen ion, is ϵ. The hydrogen ion when it is discharged almost certainly becomes a hydrogen atom. In 1 gramme-molecule of hydrogen there are 2ν atoms, where ν is the number of molecules in a gramme-molecule of any gas. Hence in order to liberate 1 gramme-molecule ($= 2$ gms.) of hydrogen a quantity of electricity equal to $2\nu\epsilon$ must pass through an electrolyte; this quantity has been determined with considerable accuracy to be $2 \times 2{\cdot}894 \times 10^{14}$, and in order to know ϵ we have only to find the value of ν.

Several lines of argument, based partly on the relation between the viscosity of a gas and the properties of its molecules and partly on estimations of the diameter of a molecule derived from the study of very thin films, have long been known which show that ν is not very different from 10^{24}, but the uncertainties attaching to the various hypotheses made in arriving at this value are so serious that such arguments could never lead to an estimate which was known to be correct even within 100 °/₀. But the ingenuity of Perrin, who investigated the properties of a 'gas,' the molecules of which were individually visible, has given us much more reliable information concerning the fundamental quantities of molecular theory.

The 'gas' which he used consisted of a suspension of very fine particles of mastic in water. It has long been known that such fine particles in suspension, having a diameter of the order of 10^{-4} cm., when viewed under the microscope, are seen to be affected by a continual irregular motion, known from the discoverer as 'Brownian motion.' Recent investigations have confirmed the view which has been held for some time that this motion is of the same nature as the thermal agitation of the molecules of a gas and is caused by the impact on the visible particles of the invisible molecules, the momentum of which is yet sufficient to set the larger bodies into a motion which can be observed. On this view the mean kinetic energy of one of the suspended particles is equal to the mean kinetic energy (E) of the molecule of a gas at the same temperature and is simply related to ν by the equation $E = \frac{3}{2}\frac{RT}{\nu}$, where R is the gas constant of which the value is known. Accordingly if the mean kinetic energy of one of the suspended particles can be determined, the value of ν and therefore of ϵ can be deduced.

This value of the mean kinetic energy was measured by Perrin by observations on the distribution of the particles in a vertical plane. If the particles had no kinetic energy they would all sink under gravity to the bottom of the vessel; but since they possess such energy, while some sink others are carried upwards; an equilibrium is attained in which the number of the particles in any horizontal layer decreases with increase of the distance of that layer from the bottom. The greater the mean kinetic energy, the greater is the proportion of the particles found in the higher layers; calculation by the methods of the dynamical theory of gases leads to the conclusion that

$$\log \frac{n_2}{n_1} = \frac{2}{E} \pi a^2 . d . h,$$

where d is the difference in the densities of the particles and of the water, n_1 and n_2 are the numbers of the particles in two planes of which the first is a distance h above the second and a is the radius of the particles. These three quantities can be determined by direct observation and lead at once to values for E, ν and ϵ. By this method ϵ was found to have the value $4{\cdot}2 \times 10^{-10}$; the other methods described later give values 10 °/₀ to 20 °/₀ higher, but it is unnecessary in this volume to enter on a criticism of the relative weights to be given to the different determinations.

The mass of an electron.

18. The mass of any body is measured by observing its acceleration when it is subjected to known forces. It is a fundamental assumption of electronic theory that the only forces which can act upon an electron are the electrostatic and magnetic forces related to the electric displacement and the magnetic intensity by (5); these latter quantities can be measured by means of (2) and (9), but since the force determined by them always contains the factor ϵ, it is clear that observations of the acceleration of an electron can never give directly a value for m_0, but only a value for ϵ/m_0; in order to deduce m_0, the value of ϵ must be ascertained by one of the methods just mentioned.

Two principal methods, based on the results which have just been given, have been used for measuring the quantity ϵ/m_0. In the first method the electrons starting from rest are caused to acquire a known velocity by falling through a known potential difference (V) in the absence of a magnetic field, as in case (1). The electrons are subsequently made to travel, as in case (2), perpendicular to a magnetic field of known intensity in the absence of an electrostatic field; ρ, the radius of curvature of their path, is measured, and this quantity together with the known values of H and V are sufficient to determine ϵ/m_0. For we have

$$\epsilon/m_0 = vc^2/\rho H \quad\text{.........................}(23),$$

where $$v = \sqrt{\frac{2\epsilon V}{m_0}}.$$

The second method differs from the first only in the means adopted for obtaining electrons travelling with a known velocity. A bundle of electrons travelling with unknown speeds are made to travel, as in case (3), between two parallel plates a very small distance apart; an electric field X is established perpendicular to the plates and a magnetic field H parallel to them. Only those electrons for which $v_0 = -c^2X/H$ travel in straight lines and, consequently, only these electrons emerge from between the plates; electrons with higher or lower velocities strike one or other of the plates. The emergent electrons of known uniform velocity are then made to travel through a pure magnetic field and the value of ϵ/m_0 found from (23). This second method is that which yields the most accurate results; recent determinations show that the value of ϵ/m_0 for electrons moving with velocities which are a small fraction of the velocity of light is $5{\cdot}31 \times 10^{17}$.

When we have determined the mass of an electron and the charge carried by it, we know all those properties of it which are independent of the nature of any material body. We have completed our survey of the properties of electricity and must turn our attention to the relations between electricity and matter.

References.

An admirable account of general electromagnetic theory and of the relation of the fundamental equations to the experimental evidence is given by Lorentz in his *Theory of Electrons.* I know of no other English treatise which can be recommended with confidence. Drude's *Physik des Aethers* is excellent.

Faraday's theory is described in his *Experimental Researches in Electricity and Magnetism*, of which the following sections are the most important:—1231, 1613–1616, 3070–3299.

J. J. Thomson's developement of the theory is given in his *Recent Researches in Electricity and Magnetism*, chap. I and in his *Electricity and Matter*, chap. I.

The complete theory of the motion of an electron in electric and magnetic fields (§ 16) is given in J. J. Thomson's *Conduction of Electricity through Gases*, chap. IV.

CHAPTER II

DIELECTRICS

Classification of materials.

19. In the earliest period of electrical science material bodies were classified in respect of their electrical properties as dielectrics and conductors, in respect of their magnetic properties as magnetic or non-magnetic. At the time when the older electrical theory culminated in the work of Maxwell it had been discovered that the classification was not of universal application, but it was thought that all the electrical and magnetic properties of any body could be sufficiently represented by three constants, the dielectric constant, the conductivity and the magnetic permeability. We have seen that this view is also inadequate and that a knowledge of these three quantities is not sufficient to determine completely the electrical behaviour of a body, but we shall find it convenient to adopt the classification and to make use of the constants. We shall suppose that there are dielectrics which have no conductivity and bodies of which the permeability is unity, and a consideration of the nature of the three constants will occupy a large part of our attention.

The problem which will concern us in the next three chapters is that of forming a new and more perfect theory as to how dielectrics differ from conductors and magnetic from non-magnetic bodies, and how the various members of each of these four classes differ amongst themselves. In the last chapter were enunciated the fundamental assumptions on which any such theory is to be built; all electric and magnetic phenomena are to be regarded as the effects of changes in the distribution or the motion of the electrons, the properties of which have been stated. Accordingly we must seek for the difference between a conductor and a dielectric or between a magnetic and a non-magnetic body in the circumstances of the electrons contained in that body.

The second difference will not concern us at present, for we shall suppose that all the bodies considered are non-magnetic. The nature of the first difference is suggested to us by a very simple observation, the oldest interpretation of which has only to be translated into the terms of our theory in order to be satisfactory for our purpose.

Dielectrics and conductors.

20. If a charged body is brought near to one end of a bar composed either of dielectric or of conducting material, it is observed that there is 'induced' on that end a charge opposite in sign to that borne by the charged body and on the other end a charge similar to that borne by the charged body. If the bar is cut in half while under the influence of the external charge, then, if it is composed of conducting material, it is observed that the two halves are permanently charged; but, if it is composed of dielectric material, the two halves when removed from the neighbourhood of the charged body are found to be uncharged. The 'fluid' theory of electricity offered an easy explanation of these observations. The neutral bar was supposed to contain equal quantities of electricity of opposite signs which normally counterbalance each other exactly. But when the bar is brought into the neighbourhood of a charged body, the electricity on that body attracts the opposite fluid in the neutral bar and repels the similar fluid; these two fluids accordingly tend to move to opposite ends of the bar. In a conductor the fluids are supposed to be able to move freely, so that a complete separation of them is effected, but in a dielectric they are not so free to move and the separation is not complete. The experiment of cutting the bar gives us information as to how far the electricities can be separated in a dielectric. It might seem that, if they were separated at all, it ought to be possible to find some place where a cut across the bar would leave an excess of opposite electricities on the separated portions. But, according to the molecular theory of matter, there are limits to our powers of mechanical division; we cannot cut across a molecule. The fact that no section of the dielectric bar separates the opposite electricities, while any section of a conducting bar separates them, shows that in a dielectric the opposite electricities, though separated, are always contained in the same molecule, while in a conductor they can be separated into different molecules; the electricities in a dielectric are only free to move within the molecule, those in a conductor are free to move from one molecule to another. Mosotti expressed this conclusion when he pointed out that a dielectric behaves like a collection of conducting molecules separated by a vacuum.

Lorentz's theory.

21. We have only to replace the 'fluid' electricity by a collection of electrons to obtain the modern view of the difference between dielectrics and conductors. Dielectrics are bodies in which the electrons cannot move outside the molecules in which they are contained; conductors are bodies in which the electrons can move freely throughout the body and are not confined within molecules. Dielectrics differ mainly in the forces which act upon the electrons within the molecules, preventing them from leaving the molecules; conductors differ in the forces which act upon the electrons when they have escaped from the molecules. This is the simple explanation which the electron theory has to offer of the distinction between dielectrics and conductors.

The explanation of the properties of dielectrics by means of discrete charged particles confined within the molecules was first offered by H. A. Lorentz in 1878. In his earliest memoirs, Lorentz made no assumption as to the magnitude or the sign of the charge on these particles, for no assumption was necessary for his purpose; but he assumed that the charge was always of the same magnitude. The idea of an atom of electricity common to all substances was not new; it had been suggested by several earlier writers as a natural consequence of the atomic theory of matter applied to the facts of electrolytic conduction; and as soon as any question was raised as to the magnitude of the charge on the particles imagined by Lorentz the answer was immediately given that it was the charge carried by the hydrogen ion. Nevertheless the application of the conception to other electrical phenomena was an immense advance in electrical theory and made possible all the greater advances that followed as a natural consequence.

22. We may now proceed to calculate a relation between the dielectric constant of a substance and the circumstances of the electrons contained in its molecules. Let us imagine a slab of the substance placed between two parallel plates charged with opposite signs. Before the slab was placed between the plates the electric intensity in the region between them was determined by the surface density of the charge on the plates. If this electric intensity is E_0, the force acting on a charge e between the plates is E_0e. When the slab is not acted on by an external charge the electrons in its molecules take up such a position that the molecule exerts little or no external electrostatic force; but when it is brought between the plates, each electron is displaced a distance x towards the positive plate, leaving the rest of the molecule positively charged. Accordingly on the face of the slab nearest the positive plate there will be a layer of electrons and on the face nearest the negative plate a layer of positively charged molecules; in the interior of the slab there will still be no charge on the average. There is now added to the electric intensity due to the charges on the plates that due to the surface charges on the faces of the dielectric, and the latter quantity will be in the opposite direction to the former. Hence, if the electric intensity in the region between the plates is now E_1, E_1 will be less than E_0. The dielectric constant of a medium is defined as the ratio of the electric intensity due to a given charged system in a region where there is no material substance to the electric intensity due to the same charge in the same region if occupied by that medium. Hence $K = \frac{E_0}{E_1}$.

The dielectric constant.

The distance x through which the electrons will be displaced is determined by the equilibrium of the force $E_1\epsilon$ acting on them with the force, due to the positive charge on the molecule, which restrains them from leaving it. We will suppose that the restraining force is proportional to x, say fx. Then $E_1\epsilon = fx$. If N is the number of electrons in unit volume of the substance, the number to be found in unit area of the charged layer on the face of the slab is Nx; the surface density of the charge on either of the faces is $Nx\epsilon$, and, by a well-known theorem, the electric intensity due to these surface charges is $4\pi Nx\epsilon$. Accordingly

$$E_1 = E_0 - 4\pi Nx\epsilon = E_0 - 4\pi N\epsilon^2 \cdot \frac{E_1}{f} \quad \text{............(24)}$$

and

$$\frac{E_0}{E_1} = K = 1 + \frac{4\pi N\epsilon^2}{f} \quad \text{....................(25).}$$

The expression deduced for K contains two unknown quantities, N and f; we know that N cannot be less than the number of molecules in unit volume, for each molecule must, on our theory, contain at least one electron, but we know nothing whatsoever about f. A direct comparison of the observed and calculated values is impossible, but it is possible to obtain some test of our theory by means of observations on the variation of the dielectric constant of the same substance with a change in its density due to a change of pressure or temperature. Since in such cases the density d is simply proportional to the number of molecules in unit volume, N is proportional to d; f is independent of d, if the restraining force on an electron is determined wholly by the nature of the molecule in which it is contained and not by its relation to other molecules. In this case

$$K - 1 \propto d \quad \text{..........................(26).}$$

This relation has been found to hold very accurately between the dielectric constant and the density of a gas.

When solids or liquids are investigated the relation is found no longer to be accurate; in these denser bodies the molecules are so near together that f is determined partly by the forces exerted on an electron by the fields of the neighbouring molecules. The field inside the dielectric, though uniform on a large scale, is not uniform on a molecular scale; each electron is displaced from the position in which it would neutralise the positive charge on the molecule, and the system has a finite electrostatic field which is appreciable at distances comparable with the displacement of the electron. If we assume that the other molecules are distributed on spherical shells round the molecule under consideration, the electric intensity acting on this molecule due to the others is the same as that inside two equal spheres, uniformly charged with equal and opposite densities $\pm N\epsilon$, the centres of which are a distance x apart. This electric intensity is $\frac{4}{3}\pi N\epsilon x$ and tends to drag the electron out of the molecule. Accordingly, if the force exerted on the electron by the molecule to which it belongs is $f'x$, the whole force on it will be $(f' - \frac{4}{3}\pi N\epsilon^2)\,x$. f' is, of course, independent of the density, so long as the nature of the molecule is not affected; hence

$$K = 1 + \frac{4\pi N\epsilon^2}{f' - \frac{4}{3}\pi N\epsilon^2} \quad\text{.....................}(27),$$

whence

$$\frac{K-1}{K+2} = \frac{4\pi N\epsilon^2}{3f'} \propto d \quad\text{.....................}(28).$$

The following table shows that when the temperature is changed the relation between the dielectric constant and the density predicted by (28) agrees satisfactorily with that determined experimentally.

TABLE I.

	d	K (observed)	K (calculated)
Water			
	0·9971	75·7	
	0·9954	73·7	72·6
	0·9988	78·7	79·2
	0·9997	80·2	81·2
	1·0000	82·3	81·8
Benzol			
	0·883	2·336	
	0·841	2·243	2·246
Alcohol			
	0·8009	25·7	
	0·8129	27·5	29·6

It should be remarked that (28) can be deduced also from Mosotti's theory of the nature of the dielectric constant. So far the only advantage which Lorentz's theory has over its predecessors (it is certainly a great advantage) is that the latter makes the distinction between dielectrics and conductors absolutely fundamental and incapable of explanation, while the former attribute the properties of substances of both kinds to agencies essentially similar.

The refraction of light.

23. It is when we turn our attention to the rapidly varying electric fields which constitute light that the immense advance made by Lorentz becomes apparent. We have assumed so far that the electrons take up such positions that they are in equilibrium under the forces acting on them, and, of course, so long as those forces are constant, this assumption must be true. But if the fields change, it is by no means certain that the electron will occupy at every instant the position in which it would be in equilibrium with the forces acting on it at that instant. Bodies which possess a finite mass require a finite time to reach a position of equilibrium, and if the force is reversed before that position is attained, equilibrium will never be attained at all. In general, the displacements of a system possessing mass, acted upon by rapidly alternating forces, decrease as the rapidity of the alternations is increased. Now it has been pointed out that the value of the dielectric constant of a body appears to be less for alternating than for steady forces. On our theory the magnitude of that constant is determined by the distance through which the electrons are displaced by a given force; if the electrons possess mass, this distance will be less when the force alternates than when it is steady, and we should expect just such a variation of the dielectric constant as has been found. In dealing with optical phenomena the most important feature of the electron theory is not the assumption that charges are distributed discontinuously, but the assumption that the particles on which they are distributed possess a finite mass.

It is possible to predict from general principles and without any detailed calculation the nature of the influence which the electrons contained in a dielectric will exert upon a wave of light passing through it. The alternating electric field in the light will set the electrons in the dielectric into 'forced vibrations' of its own period, and consequently the lines of force attached to these electrons will be made to vibrate and along them will travel disturbances such as we considered in § 10. These secondary vibrations will be emitted equally in all directions, but they will 'interfere' with each other and with the primary vibrations of the lines of force which form the incident light. The application of the well-known Huyghens principle shows that outside the medium the result of this interference is the propagation of a reflected wave making the same angle with the normal to the surface of the dielectric as the incident wave. Inside the medium there will be a refracted wave, the direction of which is determined by the speed with which light travels in the medium. This speed is determined by the tension and the mass of the lines of force, and both these quantities depend on the number of lines of force around the line along which the disturbance is travelling. The displacements of the electrons, as we have just seen, will cause the developement of secondary lines of force connecting the displaced electrons with the positive molecules; accordingly the tension and the mass of the lines of force along which the primary disturbance is travelling will be changed and the speed of propagation of the disturbance will change also.

The refractive index.

24. To calculate the new speed of propagation the fundamental equations of the last chapter may be employed. We will consider the simple case of a plane light wave travelling in the medium along the axis of z; the electric displacement in the wave-front is taken parallel to x, and consequently the magnetic intensity is parallel to y. The disturbance is supposed to be simply periodic in time and space; consequently D, the electric displacement, may be assumed to have the form

$$D = D_0 e^{\iota(pt - qz)} \quad \text{.......................(29)}.$$

If ν is the frequency of the vibrations[1] and λ their wave-length, v their velocity,

$$\nu = p/2\pi \text{......(30)}; \quad \lambda = \frac{2\pi}{q} \text{......(31)}; \quad v = \lambda\nu = p/q \text{......(32)}.$$

Inside the medium the electric displacement in the wave-front, D_1, is given, as in (24), by the equation

$$D_1 = D - N\epsilon x \quad \text{.......................(33)},$$

where x is the displacement of the electron.

The equation of motion of the electron is

$$m_0 \ddot{x} + fx - 4\pi\epsilon D_1 = 0 \quad \text{.................(34)}.$$

---

[1] I.e. the number of complete vibrations in 1 second; the word is used by some writers to denote the number of vibrations in 2π seconds.

(33) and (34) are satisfied by

$$x = \frac{4\pi\epsilon}{f - p^2 m_0 + 4\pi N\epsilon^2} D \quad \text{..................(35)},$$

$$D_1 = \frac{f - p^2 m_0}{f - p^2 m_0 + 4\pi N\epsilon^2} D \quad \text{..................(36)}.$$

If in (34) we put $D_1 = 0$, we obtain the equation of the 'free vibration' of the electron; its 'natural frequency,' ν_0, is given by

$$\nu_0^2 = \frac{f}{4\pi^2 m_0} \text{..........................(37)}.$$

The magnetic intensity in the wave-front, β, is given by (9), which in this case reduces to

$$-\frac{d\beta}{dz} = 4\pi\left(\frac{dD_1}{dt} + N\epsilon\frac{dx}{dt}\right) \text{..................(38)}$$

since $\rho = N\epsilon$.

On the other hand (7) becomes

$$\frac{d\beta}{dt} = -4\pi c^2 \frac{dD_1}{dz} \quad \text{.....................(39)}.$$

Hence $$\frac{d^2\beta}{dt\,dz} = -4\pi\frac{d^2}{dt^2}(D_1 + N\epsilon x) = -4\pi c^2 \frac{d^2 D_1}{dz^2} \quad \text{......(40)},$$

or $$\frac{d^2 D}{dt^2} = \frac{\pi m_0(\nu_0^2 - \nu^2)}{\pi m_0(\nu_0^2 - \nu^2) + N\epsilon^2} c^2 \frac{d^2 D}{dz^2} \quad \text{............(41)}.$$

But from (29)

$$\frac{d^2 D}{dt^2} = -p^2 D \text{...........(42)}; \quad \frac{d^2 D}{dz^2} = -q^2 D \quad \text{.........(43)},$$

so that $$\frac{p^2}{q^2} = v^2 = c^2 \frac{1}{1 + \dfrac{N\epsilon^2}{\pi m_0(\nu_0^2 - \nu^2)}} \quad \text{..................(44)}.$$

Now n, the refractive index of the dielectric, is $\frac{c}{v}$; hence, finally,

$$n^2 = 1 + \frac{N\epsilon^2}{\pi m_0(\nu_0^2 - \nu^2)} \text{.....................(45)}.$$

According to the older theories of light the optical properties of a transparent medium are completely determined by its refractive index; this quantity determines not only the direction of the refracted ray, but also the intensity of the reflected and refracted rays. A very brief examination would show that our theory agrees in this respect with others. The relative intensities of the incident, reflected and refracted rays can be calculated from a consideration of the relations between the various electric and magnetic quantities on either side of the boundary of the medium which must hold according to general electromagnetic theory, but the calculation need not be detailed here. In this matter the theory which has just been developed agrees completely with Maxwell's theory, the conclusions of which were enumerated in § 13.

25. Let us now compare our theory with the experimental facts.

Dispersion.

In the visible spectrum most transparent bodies show 'normal dispersion,' that is to say the refractive index increases with the frequency of the incident light. Many formulae have been proposed at various times to represent this dispersion. Of these two are of especial importance for our purpose. The first is purely empirical and unfounded on any theoretical basis :—

$$n^2 = -A'T^2 + A + BT^{-2} + cT^{-4} \quad \ldots\ldots\ldots\ldots\ldots (46),$$

where $T = 1/\nu$, the period of the light.

The other formula was originally propounded by Sellmeier and extended by Helmholtz and Ketteler, by whose names it is generally known. It is founded upon a theory similar in some respects to that which has just been propounded. It is assumed that the dispersion is determined by the presence in the refracting medium of systems having definite 'free periods' of vibration which are set into 'forced vibrations' by the incident light. It can then be shown on general dynamical reasoning that the relation between n and λ, the wave-length in vacuo of the incident light, should be of the form

$$n^2 = 1 + \frac{M\lambda^2}{\lambda^2 - \lambda_0^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (47),$$

where λ_0 is the wave-length in vacuo of light having the same frequency as the free vibrations of the vibrating systems in the medium. There are only a few substances the dispersion of which is represented with sufficient accuracy by this formula throughout the visible spectrum; but if an additional constant is introduced and we write

$$n^2 = M_0 + \frac{M\lambda^2}{\lambda^2 - \lambda_0^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(48),$$

then the dispersion of almost all transparent substances is represented with sufficient accuracy throughout that region. M_0 and M are positive quantities.

The necessity for the third constant was explained by Helmholtz on the assumption that the medium contains more than one kind of vibrating system. If there are several such systems with different periods of free vibration, then (47) becomes

$$n^2 = 1 + \frac{M_1\lambda^2}{\lambda^2 - \lambda_1^2} + \frac{M_2\lambda^2}{\lambda^2 - \lambda_2^2} + \ldots + \frac{M_n\lambda^2}{\lambda^2 - \lambda_n^2} = 1 + \Sigma\frac{M_n\lambda^2}{\lambda^2 - \lambda_n^2} \ldots(49),$$

where the suffixes 1, 2, ... n denote quantities referring to the different systems. This relation may also be written

$$n^2 = B_0 + \Sigma\frac{B_n}{\lambda^2 - \lambda_n^2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(50),$$

where $B_0 = 1 + \Sigma M_n \ldots\ldots\ldots(51)$ and $B_n = M_n\lambda_n^2 \ldots\ldots\ldots(52)$.

Both (46) and (48) may be regarded as special cases of (49). Thus, if the frequency of one of the vibrating systems is very much greater than that of the incident light, so that λ_n is very much smaller than λ, the term representing the effect of this system becomes M_n. If the medium contains one other system executing free vibrations with a frequency comparable with that of the incident light, but still greater than that frequency, (49) reduces to (48), if $M_0 = 1 + M_n$.

Again, suppose that the medium contains systems of two kinds, one executing vibrations of a frequency considerably greater and the other executing vibrations of a frequency considerably less than that of the incident light. Then since for the former system (suffix v) λ_v is less than λ, and for the latter (suffix r) λ_r is greater than λ, we may write (49) in the form

$$n^2 = 1 + M_v\left(1 - \frac{\lambda_v^2}{\lambda^2}\right)^{-1} - M_r\lambda^2 \cdot \frac{1}{\lambda_r^2}\left(1 - \frac{\lambda^2}{\lambda_r^2}\right)^{-1} \ldots(53).$$

If we neglect the higher powers of $\frac{\lambda_v}{\lambda}$ and $\frac{\lambda}{\lambda_r}$, (53) becomes equivalent to (46), if we write

$$A' = \frac{c^2 M_r}{\lambda_r^2};\ A = 1 + M_v;\ B = M_v \lambda_v^2/c^2;\ C = M_v \lambda_v^4/c^4 \ \ldots(54).$$

It appears then that, on the Helmholtz-Ketteler theory, a substance which shows 'normal dispersion' for visible light is one containing systems the free vibrations of which differ very greatly in frequency from visible light; the frequencies of these free vibrations are those of light far in the ultra-violet or far in the infra-red. The best test of the theory will be obtained by examination of the refractive index of a medium for light the frequency of which is near that of the vibrating systems contained in it. For it will be seen that the refractive index should undergo remarkable changes in such regions of the spectrum; when frequency of the incident light is just less than that of the vibrating system the refractive index should be very large, when it is just greater, the refractive index should be very small[1]; in such a region the substance should refract light of less frequency more than light of greater frequency. Such a phenomenon, when it occurs in the visible spectrum, has been termed 'anomalous dispersion'; we arrive at the conclusion that a substance should exhibit normal dispersion when the frequency of the incident light is very different from that of the free vibration of any system contained in the substance, and should exhibit anomalous dispersion when the frequency of the incident light is near that of one of those vibrations. Normal and anomalous dispersions are not separate phenomena characteristic of different substances, but inseparable phenomena which will always be exhibited by the same substance in different parts of the spectrum.

To test this prediction of the theory by direct determination of the refractive index of some substance throughout the whole spectrum is impossible on account of the difficulties which attend the examination of ultra-violet and infra-red radiations. But the test may be made by other methods, to understand which it will

---

[1] Between these values there is a range of frequencies for which n is imaginary; the meaning of this result is discussed in § 29.

be well to compare this theory with the electronic theory expounded in § 24.

Comparison with the theory.

26. The electronic theory differs from the Helmholtz-Ketteler theory only in making definite assumptions as to the constitution of the vibrating systems; the formulae of the former must be special cases of those of the latter. If we had supposed the medium to contain several kinds, and not only one kind, of vibrating electrons, we should have obtained in place of (45),

$$n^2 = 1 + \Sigma \frac{N_n \epsilon^2}{\pi m_0 (\nu_n^2 - \nu^2)} \quad \text{.................(55).}$$

(The electrons of different kinds are supposed to have the same charge and mass, but to be attached differently to the molecules, so that the quantities f_n, determining the restraining forces on them, are different.)

Since $\lambda = c/\nu$, (55) reduces immediately to (49) if we put

$$M_n = \frac{N_n \epsilon^2 \lambda_n^2}{\pi m_0 c^2} \quad \text{..........................(56)}$$

and reduces to (50) if we put

$$B_0 = 1 + \Sigma \frac{N_n \epsilon^2 \lambda_n^2}{\pi m_0 c^2} \text{.........(57)}; \qquad B_n = \frac{N_n \epsilon^2 \lambda_n^4}{\pi m_0 c^2} \text{.........(58).}$$

Accordingly all the conclusions which we have drawn as to the nature of normal and anomalous dispersions follow directly from the electronic theory of dielectrics.

Before we proceed to examine more nearly these conclusions we may revert for a moment to the relation between the dielectric constant and the refractive index, the failure to explain which was the main deficiency of the Maxwellian theory.

If we had supposed that the medium contained more than one set of electrons, the electrons of different sets being attached to their molecules by different forces, we should have obtained in place of (25)

$$K = 1 + \Sigma \frac{4\pi N_n \epsilon^2}{f_n} = 1 + \Sigma \frac{N_n \epsilon^2}{\pi m_0 \nu_n^2} \quad \text{...........(59).}$$

By comparison of (59) with (57), we see that $K = B_0$. Now if λ is very large compared with any of the λ_n's, (50) reduces to $n^2 = B_0$.

Hence we see that $K = n_\infty^2$, where n_∞ is the refractive index of the medium for very long waves. Since long waves correspond to small frequencies, we arrive at the conclusion which we anticipated before, that the Maxwellian theory, according to which $K = n^2$, is quite adequate so long as the electric fields considered by it change slowly, and that it fails only when the changes are very rapid.

When the values of the constants in (50) are determined from observations of the refractive index for visible light it is still found that the relation $K = B_0$ is not accurately true in all cases. And this result is to be expected. For if a substance contains a vibrating system the frequency of which corresponds to light far in the infra-red, this system will exert hardly any influence on the refractive index for visible light and the observations of visible light will not lead us to suspect the presence of such systems. Since these systems will affect appreciably the value of the dielectric constant, the value of K calculated on the assumption that the systems are not there will be incorrect.

The approximate formula (46) is better adapted than (50) for testing our theory by comparisons of the refractive index and the dielectric constant. We have from (59), (56), (54)

$$K = 1 + M_v + M_r; \quad K - A = M_r = A'\lambda_r^2/c^2 \quad \text{......(60)},$$

or

$$\frac{K - A}{A'} = \frac{\lambda_r^2}{c^2} \quad \text{........................(61)}.$$

Accordingly from the difference between the dielectric constant and the constant term A in the formula for the refractive index, we can calculate the frequency of the more slowly vibrating systems in the medium. It is, of course, these systems which make the Maxwellian theory inapplicable, for their vibrations are slow compared with those of visible light, but fast compared with those of the fields employed for measuring K, and they will influence the values of n^2 and K in opposite directions. One numerical example may be given. For water, a substance showing a very large discrepancy between K and the value of n^2 for visible light,

$$K = 76, \quad A = 1{\cdot}77, \quad A' = 1{\cdot}15 \times 10^{27},$$

whence

$$\lambda_r^2 = 9 \times 10^{20} \times \frac{74{\cdot}2}{1{\cdot}15} \times 10^{-27},$$

or

$$\lambda_r = 7{\cdot}6 \times 10^{-3} \text{ cm}.$$

(The wave-length of yellow light is about 5×10^{-5} cm.) Experiment shows that water possesses several systems the frequencies of which lie in the region of the infra-red indicated. The most marked of them has a frequency corresponding to a wave-length 4×10^{-3} cm. The approximate agreement of the calculated and observed values is very strong evidence for the electronic theory of dielectrics.

27. Now let us revert to the consideration of what happens when a medium is acted upon by light of which the frequency is very near that of one of the vibrating systems in the medium. If $\nu = \nu_0$, so that $p^2 = f/m$, we see from (36) and (37) that $D_1 = 0$, and $x = \frac{D}{N\epsilon}$. The electron appears no longer to be acted upon by a restraining force, but occupies at each instant such a position that it neutralises the electrostatic field completely. But according to our theory a conductor is a substance in which the electrons are not subject to a restraining force and in which they can move so as to make the electric field in the substance zero. A substance acted upon by light of which the frequency coincides with that of one of its own free vibrations should behave towards that light as a conductor.

Selective reflection.

In the next chapter we shall examine more closely the optical behaviour of conductors; but it is well known that the most striking feature peculiar to conductors is the intense reflection which is shown by metals. We should expect then that a substance would reflect metallically light which has the same frequency as one of its free vibrations. And this expectation is confirmed by experiment. In a region of the spectrum in which a body shows anomalous dispersion it also possesses very great reflecting power.

The power of a dielectric to reflect selectively light of a frequency equal to that of the electrons in its molecules has been used by Rubens and his collaborators to isolate such light. If a beam of light of all frequencies is reflected from the surface of a dielectric, the reflected light contains a much larger proportion of those frequencies which are reflected metallically than of any others, and if the reflection is repeated several times the beam ultimately consists only of these frequencies. The rays obtained by this method of repeated reflections from surfaces of the same dielectric were termed by Rubens "reststrahlen" (or "remaining rays"); their wave-lengths could be measured in the ordinary way by the use of a diffraction grating. These wave-lengths are to be identified with the λ_n's of formula (50); if all the λ_n's could be determined in this way the validity of (50) could be tested with great accuracy by discovering whether values of B_n could be assigned which would make the calculated and observed values of the refractive index agree for all possible wave-lengths of the incident light. Unfortunately, however, it is possible by this method to ascertain only those values of λ_n which lie in the infra-red; the isolation of the ultra-violet reststrahlen is more difficult, partly on experimental grounds and partly because the reflecting power of a substance in the neighbourhood of a free period, like that of a metal, decreases as the refractive index increases (see p. 63) and is less if the free period is in the ultra-violet than if it lies in the infra-red. The values of λ_n in the ultra-violet have to be ascertained by discovering what value suits best the experimental results; but it is found that a complete agreement can be attained by assuming that there is only one λ_n in the ultra-violet and that it lies in a part which is inaccessible to direct observation. Among the substances which have been investigated in this way by Rubens is quartz (the 'ordinary ray'). By the reststrahlen method it was found that the values of λ_n in the infra-red should be

$$\lambda_1 = 8{\cdot}85 \times 10^{-4} \text{ cm.}; \quad \lambda_2 = 20{\cdot}75 \times 10^{-4} \text{ cm.};$$

the value of λ_n in the ultra-violet and the values of B_n corresponding to the three values of λ_n which accorded best with observation were found to be

$$\begin{aligned} \lambda_3 &= 0{\cdot}10495 \times 10^{-4}, \\ B_0 &= 4{\cdot}34048, \\ B_3 &= 1{\cdot}05354 \times 10^{-10}, \\ B_1 &= 4{\cdot}7453 \times 10^{-7}, \\ B_2 &= 5{\cdot}9298 \times 10^{-6}. \end{aligned}$$

In the following table the calculated and observed values for n are compared.

TABLE II.

	$\lambda \times 10^4$	n obs.	n calc.
	0·185	1·67571	1·67570
visible spectrum	0·394	1·55846	1·55839
	0·656	1·54189	1·54200
	1·160	1·5329	1·5331
	4·20	1·4569	1·4571
	6·45	1·274	1·234
	51·2	2·46	2·15
	61·1	2·12	2·13
	∞	$2·08 = \sqrt{K}$	2·08

The agreement between the observed and calculated values leaves room for no doubt whatever that the theory of dispersion which is based on the electronic theory is substantially accurate in its application to dielectrics.

The theory has also been tested by careful examination of the variation of the refractive index with the wave-length in the neighbourhood of one of the free periods. It is convenient to choose for this purpose a substance which has a free period in the visible spectrum. The most extensive measurements of this nature are those of R. W. Wood, and their results accord well with the predictions of the theory. However, in general the propagation through a medium of light the frequency of which coincides nearly with that of one of its free vibrations is complicated by phenomena which we shall consider shortly, so that the formulae which have been developed so far are not applicable.

Consequences of the electronic theory.

28. We must now notice briefly the most important consequences of the electronic theory of dispersion. In (57) and (58) relations are stated between the quantities B_n, λ_n which may be measured experimentally and the quantities N_n, ϵ, m_0 which determine the properties of the electrons. It is easy to see that a knowledge of the former quantities may be used to give valuable information about the relations of the vibrating systems to the molecules in which they are contained.

We have

$$\frac{B_n}{\lambda_n^4} = \frac{N_n \epsilon^2}{\pi m_0 c^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots(62),$$

where N_n is the number of electrons in unit volume which have a free period corresponding to λ_n. N_n must be proportional to N', the number of molecules in unit volume, since all molecules are supposed to be the same. Let $N_n = p_n N'$, where p_n is now the number of such vibrating systems in each molecule. Let W be the molecular weight of the substance referred to the atom of hydrogen as unity, and let the mass of this atom be m_h. Then Wm_h is the mass of each molecule and, if d is the density of the substance,

$$N' = d/Wm_h \quad \text{......................} (63).$$

Now (see p. 25) we know that the total charge on all the atoms of hydrogen in one gramme of the substance is $2{\cdot}894 \times 10^{14}$. Hence $m_h = (2{\cdot}894 \times 10^{14})^{-1}\,\epsilon$, since each atom bears the charge ϵ. So that

$$\frac{B_n}{\lambda_n^4} = \frac{p_n d}{W\epsilon} \times 2{\cdot}894 \times 10^{14} \times \frac{\epsilon^2}{\pi m_0 c^2} \quad \text{............} (64),$$

whence

$$p_n \,.\, \epsilon/m_0 = 9{\cdot}8 \times 10^6 \,.\, \frac{B_n}{\lambda_n^4} \,.\, \frac{W}{d} \quad \text{..............} (65).$$

We will apply (65) to the figures given for Quartz (SiO_2) on p. 45. We have $W = 60{\cdot}4$, $d = 2{\cdot}65$. Hence, using the suffixes 1, 2, 3 to denote respectively the two systems having free periods in the infra-red and the one system having a free period in the ultra-violet, we have

$$p_3 \,.\, \epsilon/m_0 = 2{\cdot}04 \times 10^{18},$$
$$p_2 \,.\, \epsilon/m_0 = 7{\cdot}19 \times 10^{13},$$
$$p_1 \,.\, \epsilon/m_0 = 1{\cdot}73 \times 10^{14}.$$

The value of ϵ/m_0 deduced by other methods is $5{\cdot}31 \times 10^{17}$ (see p. 28).

We readily conclude that p_3 is 3 or 4[1]; that is to say, that in each molecule there are 3 or 4 electrons which vibrate with the same period, corresponding to a wave-length, $0{\cdot}105 \times 10^{-4}$ cm., far in the ultra-violet. But the values corresponding to the free vibrations in the infra-red do not accord at all with our theory;

---

[1] The value 4 is more probable than 3, for it is found that the value of ϵ/m_0 which gives the best agreement in these calculations is about $4{\cdot}5 \times 10^{17}$, considerably smaller than that given by other methods. The reason for the discrepancy is not accurately known; it is probably due less to the deficiencies of the theory which are noted below than to our ignorance concerning the structure of the atom. A plausible hypothesis has been suggested by Koenisberger (*Ann. d. Phys.* 32. 845, 1910).

since all the molecules are the same, p cannot be less than 1[1], and yet our calculations make it about 10^{-4}.

A satisfactory explanation is easily found. The electrons are not the only charged systems in the medium which are subject to elastic forces and so have free periods which may affect the refractive index; there are also the molecules or atoms in which the electrons are contained. These systems are left positively charged when the electrons are displaced and they are subject, as we saw in § 22, to forces arising from the charges on the neighbouring molecules; under the action of these forces they may possibly vibrate in periods not very different from those of light.

Let us suppose that part of the molecule, having a molecular weight W' and bearing a charge $v\epsilon$, where v is some integer, is capable of executing such vibrations with a frequency corresponding to a wave-length λ_n. It will add to the value of n^2 a term $\frac{B_n}{\lambda^2 - \lambda_n^2}$; and the value of B_n may be obtained from (59) by replacing N_n by N', ϵ by $v\epsilon$, and m_0 by $W'm_h$, the mass of the part of the molecule.

Hence
$$\frac{B_n}{\lambda_n^4} = \frac{N'v^2\epsilon^2}{\pi W' m_h c^2} = v^2\left(\frac{\epsilon}{m_h}\right)^2 \frac{d}{W}\cdot\frac{1}{\pi c^2}\cdot\frac{1}{W'} \quad \text{.........(66),}$$

or
$$W' = v^2 \cdot \frac{\lambda_n^4}{B_n}\cdot\frac{d}{W} \times 2{\cdot}95 \times 10^7 \text{...............(67).}$$

Putting in the numerical values of λ_1, B_1 and λ_2, B_2 for Quartz, we obtain from the former

$$W' = v^2 \times 1{\cdot}68 = 26{\cdot}9, \text{ if } v = 4,$$

from the latter

$$W' = v^2 \times 4{\cdot}05 = 16{\cdot}2, \text{ if } v = 2.$$

Since the atomic weights of silicon and oxygen are 28·4 and 16·0, we can account for the free vibrations of Quartz in the infra-red on the assumption that one of them represents an atom of silicon

---

[1] There are certain substances, apparently chemically pure, in which the value of p is very much less than 1 for vibrating systems which are undoubtedly electronic, e.g. salts of neodidymium. No adequate explanation has been offered of these cases, which are entirely exceptional.

bearing a charge equal to that on 4 electrons, and the other an atom of oxygen with a charge equal to that on 2 electrons.

In all other cases for which measurements are available similar results are obtained. It is always found that the vibrations in the ultra-violet are due to electrons and those in the infra-red to portions of the molecules. Such calculations throw valuable light on the constitution of molecules and the fundamental problems of chemistry; we shall consider some of the results obtained in a later chapter. It appears that the number of electronic free vibrations is determined by the nature of the atoms concerned rather than by that of the molecule as a whole; the electrons must be considered to belong to the atoms and to be an essential part of them, a conclusion which will be confirmed by evidence from all sources.

Unfortunately the evidence as to the number of electrons belonging to each atom which can be derived from such considerations is not as definite as could be desired. It has been noted that the value of λ_n for the ultra-violet free periods cannot be determined by direct experiment; it can only be deduced from the form of the dispersion curve; if it should happen that there is more than one ultra-violet free period, the values of B_n and λ_n deduced from the dispersion on the assumption that there is only one will not be correct.

Now there is certainly more than one free period in the ultra-violet. From (57), (58) we see that

$$B_0 - 1 = \Sigma \frac{B_n}{\lambda_n^2} \qquad \text{.....................(68).}$$

But when the values of B_0 and B_n determined experimentally are inserted in this equation, it is always found that the left-hand side is greater than the right-hand side, and this discrepancy can be explained readily if we suppose that there are in the atom electrons which have no influence on the dispersion because their free periods are very short compared with the frequency of any light with which investigations can be made. For a set of electrons (s) contributes to (50) two terms. Firstly there is the term $\frac{N_s \epsilon^2 \lambda_s^2}{\pi m_0 c^2}$, which is independent of λ and is included in the constant term B_0; secondly there is the term $\frac{B_s}{\lambda^2 - \lambda_s^2}$, where $B_s = \frac{N_s \epsilon^2 \lambda_s^4}{\pi m_0 c^2}$, which varies with λ and influences the dispersion. If all our experiments are made with light for which λ is very much greater than λ_s, the second term will be very small compared with the first; B_s will be put equal to 0 and no term representing this group of electrons will appear on the right-hand side of (68). If we deduce the values of our B_n's from experiment we must take account of the group (s) with such very short periods by writing

$$B_0 - 1 - \Sigma \frac{N_n \epsilon^2 \lambda_n^2}{\pi m_0 c^2} = \Sigma \frac{N_s \epsilon^2 \lambda_s^2}{\pi m_0 c^2} \quad \text{............(69).}$$

It is found that the term on the right hand of (69) is usually not less than a tenth of the third term on the left. But since λ_s is very much smaller than any λ_n, ΣN_s must be very much greater than ΣN_n. We arrive at the very interesting conclusion that the number of electrons which have free periods very much less than that of the furthest ultra-violet which we can investigate is considerably greater than the number of electrons which have free periods long enough to influence the dispersion in any possible experiment. The importance of this conclusion will appear later.

29. **Absorption.** Hitherto we have supposed that all the light which falls on the medium is either reflected or refracted. There is a third alternative which we have not considered; the light may be absorbed. In general, homogeneous dielectrics do not absorb light in most regions of the spectrum, but every medium absorbs light in a region in which it shows anomalous dispersion. The refusal of a substance to transmit light in such a region is due not only to reflection, but also to absorption; when the frequency of the incident light is near to that of one of the free vibrations of the medium, part of its energy is converted into some other form and appears in neither the reflected nor the refracted beam.

We can easily account for this absorption by modifying slightly our assumptions concerning the motion of the electrons within the molecules. Hitherto we have supposed that all the forces acting on such electrons are 'conservative'; we will now suppose that there is also a force of the nature of friction, always acting in the direction opposite to that of the motion of the electron. For simplicity we will suppose that this frictional force is proportional to the velocity of the electron and will write for the equation of its motion, instead of (34),

$$m_0\ddot{x} + k\dot{x} + fx - 4\pi\epsilon D_1 = 0 \quad \dots\dots\dots\dots\dots(70),$$

where k may be regarded as a coefficient of friction.

Proceeding in our calculations precisely as in § 24 we find that if the electric displacement in the incident wave is again given by (29)

$$D_1 = \frac{\pi m_0(\nu_0^2 - \nu^2) + \frac{1}{2}\iota k\nu}{\pi m_0(\nu_0^2 - \nu^2) + \frac{1}{2}\iota k\nu + N\epsilon^2} D \dots\dots\dots\dots(71),$$

$$\frac{p^2}{q^2} = c^2 \frac{1}{1 + \dfrac{N\epsilon^2}{\pi m_0(\nu_0^2 - \nu^2) + \frac{1}{2}\iota k\nu}} \quad \dots\dots\dots\dots\dots(72).$$

The occurrence of complex quantities in the expressions for the amplitude and the velocity of the disturbance may seem at first sight to be difficult to interpret physically, but if we follow out our calculations consistently this difficulty will disappear. Let us write

$$A + \iota B = \frac{D_1}{D}; \quad P - \iota Q = \frac{q}{p} \quad \dots\dots\dots\dots\dots(73).$$

Then
$$D_1 = D_0(A + \iota B)\, e^{\iota p(t - \overline{P - \iota Q}\,.\,z)} \dots\dots\dots\dots\dots(74)$$

$$= D_0(A^2 + B^2)^{\frac{1}{2}}\,.\,e^{-pQz}\,.\,e^{\iota p(t - Pz + \delta)} \dots\dots(75),$$

where
$$\tan p\delta = B/A.$$

The presence of the term δ shows that there is a difference in phase between D_1 and D, which is unimportant for our present purpose. For the rest, we see that the light in the medium has the same frequency, $p/2\pi$, as the incident light and that it travels with a velocity $1/P$; its amplitude decreases as the distance from the surface increases, so that when the light has travelled a distance z in the medium the amplitude is reduced in the ratio of e^{-pQz} to 1. If we define the absorption coefficient of the medium, κ, by the relation that the amplitude of the light is reduced in the ratio of $e^{-2\pi\kappa}$ to 1 in travelling through one wave-length, we have

$$2\pi\kappa = pQ\lambda, \text{ or } \kappa = Q/P \quad \text{..................(76)},$$

since
$$\lambda = \frac{1}{\nu P} = \frac{2\pi}{pP}.$$

Since n, the refractive index of the medium, is cP, we have from (73)

$$\frac{q^2}{p^2} = \frac{1}{c^2} n^2 (1 - \iota\kappa)^2 \quad \text{....................(77)},$$

or, separating the real and imaginary portions of (72),

$$n^2(1-\kappa^2) = 1 + \frac{4\pi m_0 (\nu_0^2 - \nu^2) N\epsilon^2}{4\pi^2 m_0^2 (\nu_0^2 - \nu^2)^2 + k^2\nu^2} \quad \text{.........(78)},$$

$$n^2\kappa = \frac{N\epsilon^2 k\nu}{4\pi^2 m_0^2 (\nu_0^2 - \nu^2)^2 + k^2\nu^2} \quad \text{...........(79)}.$$

If k is small compared to $m_0\nu_0$, we see that the absorption coefficient is inappreciable except when ν_0 and ν are nearly equal; there will be appreciable absorption only in the region where there is anomalous dispersion and in that region we shall have a well-marked 'absorption band,' such as is always found experimentally.

Several theories have been suggested as to the nature of the frictional forces which damp the vibrations of the electrons and so produce absorption of the light. The most plausible of them is due to Lorentz, who suggests that when the molecules in which the vibrating electrons are contained collide with other molecules, the energy of the vibrating electron is transformed into translatory motion of the molecules, producing a rise in the temperature of the substance. On this view (78), (79) must be modified slightly, and it appears that the frequency of the incident light which is most strongly absorbed is near to, but not identical with, that which is most strongly reflected or refracted. Such a difference between the wave-lengths of maximum absorption and maximum reflection is found in practice; indeed the relations between n, κ and ν predicted by Lorentz's theory are completely confirmed by experimental measurements. Considerable interest attaches to such measurements made within the region of anomalous dispersion, for it has been shown recently that the relation between κ and ν can be used to deduce values for the number of electrons in each atom which have free periods in that region in much the same way as the relation between n and ν was used for the same purpose in § 28. The values deduced by both methods agree well and perhaps the former are the more reliable, for all the optical phenomena within an absorption band are determined almost entirely by the electrons having free periods in that region and there are no doubts arising from the possibility of the existence of unsuspected free periods in other regions of the spectrum.

Crystalline media.

30. We have supposed throughout that all the dielectrics considered are isotropic and have taken no account of the phenomena of double refraction and rotation of the plane of polarisation which are characteristic of some crystals. In a medium which refracts doubly we must suppose that the free periods of the electrons in the molecules are different according to the relation between the direction of the vibration and the optic axis; in a medium which rotates the plane of polarisation we must suppose that the free period is different when the electron performs circular vibrations in opposite senses. It is, of course, easy to imagine systems of forces acting on the electrons which would produce such differences in the free periods, but at present we can proceed no further. The problem of interest in connection with crystals is not to suggest a system of forces under which the electrons may be supposed to vibrate, but to suggest a connection between such a system and the chemical and other properties of the molecule which determine its crystalline form. Such speculations on these and other matters which concern the 'structure of the atom' will be found in a later chapter.

REFERENCES.

The original paper of Lorentz which may be regarded as the foundation of the electronic theory of dielectrics is reprinted from the *Proceedings of the Amsterdam Academy* for 1878 in Abraham and Langevin, 'Ions, Électrons, Corpuscles,' I. 430–435.

The treatment of refraction and dispersion given follows mainly that of Drude which is contained in his *Theory of Optics*. For a full discussion of the facts of dispersion and the theories concerning it the reader should refer to Kayser's *Handbuch der Spektroscopie*, Vol. 4. There is also an admirable account of the main experimental data in Wood's *Physical Optics*, which, like all modern text books on Light, gives the substance of the preceding chapter.

Important contributions to the interesting conclusions of § 28, which have appeared too recently for inclusion in any of these works, are to be found in papers by Erfle (*Ann. d. Phys.* 24. 672, 1907) and by Koenisberger and Kilching (*Ann. d. Phys.* 28. 889, 1909 and 32. 843, 1910).

CHAPTER III

CONDUCTORS: ELECTROLYTIC AND METALLIC

31. A CONDUCTOR is a body containing electrons which are not bound to the corresponding positive charges by elastic forces. If the conductor is surrounded by dielectrics and placed in an electric field, the opposite charges will separate so as to neutralise completely the field in the interior of the conductor and then all change will cease; but if by any means we produce a permanent electric field in a conducting circuit the electrons will flow round that circuit giving rise to a continuous current. The current density, s, across any section of the circuit is equal to the total charge conveyed across unit area in unit time. Hence, if N is the number of electrons in unit volume and v their velocity at the point considered,

The forms of conduction.

$$s = \epsilon v N \qquad \text{(80)}.$$

In some cases we shall find it necessary to suppose that the positively charged atoms can move as well as the electrons; we must then add to (80) a second term $e_1 v_1 N_1$, where e_1, v_1, N_1 are the charge, velocity and concentration of the positive particles. The direction of the current is taken as that in which the positive particles move, and v_1 will always be measured in the opposite direction to v.

There are three forms of conduction, clearly differentiated from each other:—metallic, electrolytic and gaseous. The last form is distinguished from the others by the fact that it does not follow Ohm's Law; the second is distinguished from the first by the fact that the passage of the electric current is accompanied by chemical changes in the conductor. In order to apply the electronic theory to conduction we have to show how the differences between the three classes of conductors may be attributed to differences in the number or the motion of the electrons, or in the part played by the positive particles.

Electrolytic conduction.

32. An adequate theory of electrolytic conduction is now so familiar that a few words of reference to it will suffice here. It was discovered by Faraday that the mass of any element liberated at one of the electrodes when a current passed through an electrolyte is proportional to the quantity of electricity which has passed, to the atomic weight of the element and inversely proportional to its valency. These laws are simply explained if the current flowing to the electrode where the element is liberated is carried by atoms of the element bearing a charge proportional to the valency. In other cases when the chemical changes consequent on the passage of the current are not quite so simple it is necessary to suppose that the carriers or 'ions' are combinations of atoms carrying charges which are still small multiples of that carried by a univalent atom. The ions in their passage through the electrolyte are subject to a viscous resistance, so that their velocity is uniform in a uniform field and proportional to that field. Since the concentration of the ions is independent of the field, (80) shows that Ohm's Law will be obeyed.

In the past this theory has been all-important for the developement of the electronic theory. Since there can be no portion of an element smaller than an atom, there can take part in electrolysis no charge of electricity less than the charge on a univalent atom, and this minimum charge, or a small multiple of it, is characteristic of all substances alike. In the recognition of this universal atom of electricity appears the first sign of the essential feature of the modern electronic theory. But otherwise that theory has little bearing upon electrolytic conduction; as yet it has afforded no clue to the outstanding problem as to why some and not all solutions are electrolytes. We may pass immediately to consider the other forms of conduction.

33. Many attempts have been made to apply an ionic theory to metallic conduction, but until recently none of them were successful. The difficulty arose in suggesting a suitable nature for the ions. Most electric circuits are composed of at least two kinds of material and we must suppose either that the ions can cross from one material to the other or that they cannot. If they cannot, then the ions must remain piled up at the boundary, and we should expect to find a change in the distribution of the mass of the conductor caused by the current; but no such change can be detected. If the ions can cross from one material to the other, then, since no mixing of the materials of the circuit can be detected, the ions must be of the same nature in all materials.

Metallic conduction.

Until the developement of the electronic hypothesis this conclusion would have invalidated necessarily any theory which led to it; not because it was thought impossible that there should be any common constituent of all elements, but because nothing was known about such a constituent. Ever since the conception of an element was introduced, physicists have probably entertained privately the belief that these elements would be found to be but many forms of one primordial substance, and the belief derived some support from such facts as those which gave rise to Prout's hypothesis. But whatever arguments there might be for the existence of such a primordial substance, nothing was known about its properties. All research in chemistry and physics had tended to emphasize the difference between different elements: there was no trace of a common constituent. Any theory of conduction based upon the properties of such a substance would be useless, for it would not connect the conduction of electricity with any other known phenomenon or enable relations to be established between the conductivity and any other characteristics of a body.

But now this objection is removed. We have already noticed in § 28 some evidence for the existence of electrons common to all substances and in later chapters we shall find much more conclusive evidence. We have only to imagine that the current in a metallic conductor is carried by the universal electrons to overcome the difficulties which have previously attended ionic theories;

no part of the current can be carried by the positively charged atoms which differ in nature in different conductors. It will be well to remark here that, since the electrons are negatively charged, the direction of the current will always be opposite to that of the carriers.

Motion of the free electrons.

34. In order to arrive at an estimate of the relation between the conductivity of a metal and the properties of the electrons in it, some assumption must be made as to the nature of the motion of the electrons. The assumption which is generally adopted is derived from the dynamical theory of gases and was introduced into electronic theory by Drude and J. J. Thomson. According to the theory of gases the molecules of any body are in a state of continual motion, colliding with each other, and the mean kinetic energy of a molecule is proportional to the absolute temperature (T) of the body; if a body contains molecules of two kinds the mean kinetic energy of each kind is the same. This doctrine is an instance of the principle of the equipartition of energy. Drude assumed that the electrons in the metal may be regarded as a kind of molecule, that an electron possesses the same mean kinetic energy as a molecule of a gas at the same temperature, and that electrons collide with each other and with the molecules of the metal in much the same way as the molecules of a gas collide with each other and with the walls of the vessel in which they are contained. This assumption forms the basis of all modern theories of metallic conduction.

Accordingly if u is the mean velocity of the electrons in a metal at a temperature T, we may write

$$\tfrac{1}{2} m_0 u^2 = \alpha T \quad \text{.......................(81)},$$

where α is the ratio of the energy of a molecule to the absolute temperature of the gas of which it forms part, a quantity which is a universal constant independent of the nature of the gas.

α can easily be calculated. The volume of one gramme-molecule of a gas at a pressure of one atmosphere ($1{\cdot}0132 \times 10^6$ dynes per cm.^{2}) and a temperature of 0° C. (273·1 abs.) is $2{\cdot}2412 \times 10^4$ cm.^{3} Now, according to the dynamical theory of gases,

$$pv = \tfrac{2}{3} \alpha v_0 T \quad \text{.......................(82)},$$

where ν_0 is the number of molecules in one gramme-molecule. Hence

$$\left(\frac{\alpha}{\epsilon}\right) = \frac{3pv}{2T} \cdot \frac{1}{\nu_0 \epsilon} \quad \text{.....................(83).}$$

But $\nu_0 \epsilon = 2{\cdot}894 \times 10^{14}$ (see p. 25).

Hence $$\underline{\frac{\alpha}{\epsilon} = 4{\cdot}3 \times 10^{-7}.}$$

We may note that

$$\begin{aligned} u^2 = \frac{2\alpha T}{m} &= 2T \cdot \frac{\alpha}{\epsilon} \cdot \frac{\epsilon}{m_0} \\ &= 582 \times (4{\cdot}3 \times 10^{-7}) \times (5{\cdot}3 \times 10^{17}) \text{ at } 18^\circ \text{ C.} \\ &= 1{\cdot}34 \times 10^{14} \end{aligned}$$

or $$u = 1{\cdot}16 \times 10^7 \text{ cm./sec.}$$

For experimental purposes it is sometimes convenient to express the velocity of an electron as the potential difference through which it must fall from rest in order to acquire that velocity. If V is this potential difference expressed in volts (1 volt $= \frac{1}{300}$ el. st. unit), we have from p. 23.

$$\begin{aligned} V = \frac{1}{2}\left(\frac{m_0}{\epsilon}\right) u^2 \times 300 &= 300 \times \frac{\alpha}{\epsilon} T = 1{\cdot}29 \times 10^{-4}\, T \\ &= {\cdot}038 \text{ volt at } 18^\circ \text{ C.} \end{aligned}$$

The accurate calculation of any of the properties of a gas from the fundamental hypotheses of the dynamical theory involves analysis of great complexity and difficulty; and the precise results attained are affected in most cases by the assumptions which are made as to the nature of a collision between two molecules, a matter concerning which we are still profoundly ignorant. The same difficulties, of course, attend any calculations concerning the motion of the electrons in a metal; we know even less of the nature of a collision between an electron and a molecule than we do of a collision between two molecules. And while we have no knowledge on this point, it is not worth while to enter upon any elaborate mathematical treatment. Accordingly in what follows, the calculations will be simplified greatly by assumptions which are known not to be accurately true, but which are shown by a more precise inquiry not to invalidate the general conclusions which we shall attain. For a more complete exposition the reader is referred to the sources mentioned at the end of the chapter.

The three chief assumptions of this nature which we shall make are:

(1) That the electrons, instead of possessing various velocities distributed about the mean u, all possess the velocity u.

(2) That the interval between two collisions by the same electron is the same for all electrons and for all collisions, so long as the temperature is the same.

(3) That the previous history of an electron is completely wiped out when it collides, so that, whatever happens before a collision, the velocity after a collision is always u.

We may now proceed to calculate the conductivity of a metal according to our theory.

The electrical conductivity.

35. Let X be the electric intensity in the metal and t the time (supposed constant) which elapses between two collisions of an electron. During the time t the electron is subject to an acceleration $\frac{X\epsilon}{m_0}$ in the direction of the electric intensity and moves a distance $\frac{1}{2}\frac{X\epsilon}{m_0}t^2$ in that direction. If this distance is small compared with the free path of the electron, t will be almost the same as if the field were not acting; the only effect of the field will be to superimpose on the irregular motion of the electrons a continual drift along the direction of the electric intensity, and the velocity of this drift will be

$$v = \frac{1}{t}\cdot\frac{1}{2}\frac{X\epsilon}{m_0}t^2 = \frac{X\epsilon t}{2m_0} \qquad \text{……………….(84).}$$

By (80) the current density in the metal is

$$s = \epsilon v N = \frac{X\epsilon^2 t N}{2m_0} \qquad \text{………………….(85).}$$

But σ, the electrical conductivity of a metal which obeys Ohm's Law, is given by

$$s = \sigma X \qquad \text{……………………….(86).}$$

Hence

$$\sigma = \frac{N\epsilon^2 t}{2m_0} \qquad \text{…………………….(87).}$$

If l is the free path which an electron traverses between two collisions, we have

$$t = l/u \quad \text{.........................(88)},$$

and, remembering (81),

$$\sigma = \frac{\epsilon^2 N l u}{4\alpha T} \quad \text{.........................(89)}$$

$$= \frac{\epsilon^2 N l}{2\sqrt{2\alpha T m_0}} \quad \text{....................(90)}.$$

As yet we are completely ignorant of the values of N and l and can apply no direct test of the validity of our theory; but an indirect test can be applied by considering the variation of the conductivity with the temperature. Experiment shows that the conductivity of a pure metal is nearly proportional to $1/T$; hence according to our theory we must have $Nlu =$ constant. This conclusion is not what we should expect; there is no reason to believe that N varies with the temperature and we should expect that l also was nearly constant and that, therefore, according to (90), σ should be proportional to $1/T^{\frac{1}{2}}$. For alloys, again, there is no simple relation between the conductivity and the temperature. If there was no other evidence of the truth of our theory, these discrepancies would probably lead to its rejection, but we shall find later such convincing proof of its essential correctness that we must attribute them to our ignorance of the nature of the collisions of an electron in a metal. It appears that in a pure metal $l \propto T^{-\frac{1}{2}}$, but in the present state of our knowledge there is no use in attempting to devise speculative explanations of this relation.

Optical properties of metals.

36. Good metallic conductors are all good reflectors of light. Of this relation any electromagnetic theory offers a sufficient explanation. When a conductor is placed in an electric field the induced electrification at its surface is such as to neutralise the electric intensity in the substance of the conductor; accordingly when such a body is placed in the alternating electric field which constitutes light, the disturbance cannot penetrate into the body, but must be either reflected or absorbed. Absorption will take place only if the developement of the induced electrification requires the expenditure of energy which is not subsequently restored to the beam of light, and, as we have seen in § 29, there will be such a waste of energy only if the motion of the electrons is hindered by frictional forces. Now the collisions of the electrons in a metal are equivalent to a frictional force; we have seen in (84) that an electron in a metal subject to a constant force X attains a steady velocity $v = \frac{X\epsilon t}{2m_0} = \epsilon X \cdot \frac{\sigma}{N\epsilon^2}$; hence by putting $\ddot{x} = f = 0$ in (70) we see that there must be a force on the electrons $k\dot{x}$, opposing their motion, where $k = \frac{N\epsilon^2}{\sigma}$. The greater the conductivity, the less the frictional force and, consequently, the less the energy absorbed.

Accordingly we may apply the formulae which we developed in § 29 to the calculation of the optical properties of a metal. Since the electrons are 'free' and subject to no elastic forces, $f = 0$ and consequently $\nu_0 = 0$; of course we must also remove the restriction that k is small compared to $m_0\nu_0$, which was applicable when we were considering the motion of 'bound' electrons under forces of a frequency near to that of their own free vibrations. Hence we find from (78), (79)

$$n^2(1-\kappa^2) = 1 - \frac{4\pi m_0 N\epsilon^2}{4\pi^2 m_0^2 \nu^2 + k^2} = 1 - \frac{4\pi m_0 N\epsilon^2}{4\pi^2 m_0^2 \nu^2 + \frac{N^2\epsilon^4}{\sigma^2}} \quad \text{...(91)},$$

$$n^2\kappa = \frac{1}{\nu}\,\frac{N\epsilon^2 k}{4\pi^2 m_0^2 \nu^2 + k^2} = \frac{1}{\nu}\,\frac{\frac{N^2\epsilon^4}{\sigma}}{4\pi^2 m_0^2 \nu^2 + \frac{N^2\epsilon^4}{\sigma^2}} \quad \text{......(92).}$$

There is nothing in our theory to make it necessary to suppose that all the electrons in a metal are free; if a sufficient number of them are free the phenomena of conduction in a steady field will appear, whether or no there are also present 'bound' electrons, attached to the molecules by elastic forces. The optical properties of metals make it necessary to suppose that there are, in fact, such bound electrons as well as free electrons. For it is well known that some metals, namely those which, like copper or gold, possess a distinct colour, show selective reflection superimposed on the general reflection which is characteristic of metals, and these selective effects may be attributed, as in the case of dielectrics, to the presence of electrons with definite free periods. Accordingly we shall suppose that there are present in the metal bound electrons (characterised by the suffix 1) as well as free electrons (characterised by the suffix 2); if we investigate only light which is not near the free periods of the former, we may write from (55), (91), (92)

$$n^2(1-\kappa^2) = 1 + \Sigma \frac{N_1\epsilon^2}{\pi m_0(\nu_1^2-\nu^2)} - \frac{4\pi m_0 N_2\epsilon^2}{4\pi^2 m_0^2\nu^2 + \frac{N_2^2\epsilon^4}{\sigma^2}} \quad \text{...(93)},$$

and

$$n^2\kappa = \frac{1}{\nu}\frac{\frac{N_2^2\epsilon^4}{\sigma}}{4\pi^2 m_0^2\nu^2 + \frac{N_2^2\epsilon^4}{\sigma^2}} \quad \text{.................(94).}$$

If we restrict ourselves further to light for which ν is very much less than ν_1, the term in ν_1 becomes independent of ν and we may write

$$q = 1 + \Sigma \frac{N_1\epsilon^2}{\pi m_0(\nu_1^2-\nu^2)} \quad \text{.................(95),}$$

where q is a constant independent of the wave-length of the incident light.

The quantities n and κ for metals have been determined by two methods, one, that of Drude, depending on the elliptic polarisation of light reflected from the surface, and the other, that of Hagen and Rubens, consisting in the determination of the reflecting power of the metal. The principles of these methods are independent of any special theory of the nature of light, and therefore require no special treatment here. It is sufficient for our purpose to know that the quantities have been determined with considerable accuracy for light of all wave-lengths from the extreme violet to the infra-red. We will compare the results of our theory with those of experiment.

If ν is very small, that is, if the wave-length of the incident light is very great, (94) reduces to

$$n^2\kappa = \sigma/\nu \quad \text{.........................(96).}$$

If R is the proportion of the incident light reflected at the surface of the metal, it can be shown that in this case

$$1 - R = 2\left(\frac{\nu}{\sigma}\right)^{\frac{1}{2}} \text{ ....(97).}$$

(96) and (97) can be deduced from the Maxwellian theory noticed in § 13. It will be seen that $R = 1$ and the metal is a perfect reflector, if the conductivity or the wave-length of the incident light is very great. This relation is found to be generally valid; on the whole the better conductors are the better reflectors, metals reflect most light of great wave-length and all conductors become nearly perfect reflectors when the wave-length is great enough. It is found, however, that (97) never holds accurately in the region of the visible spectrum, but that it holds for light the wave-length of which is greater than 25μ, i.e. 25×10^{-4} cm.

Again, from (94) we may conclude that if two substances have the same conductivity and the same number N_2 of charged carriers, while the mass of the carriers in one of them is very much greater than that of an electron, then this substance will reflect and absorb light very much less than the other. Such a conclusion is obvious, for if the mass of the carriers of the current is very great, they will fail to respond to the rapidly alternating fields constituting light and, though they may behave as conductors in steady fields, in such changing fields they will behave as dielectrics. This conclusion enables us to explain why electrolytes, in which the carriers are charged atoms, do not behave in optical phenomena like metals of equivalent conductivity.

Number of free electrons.

37. We can apply a more stringent numerical test to our theory in the following way. From (93) and (94) it is a mere matter of algebra to deduce the relation

$$\frac{N_2\epsilon^2}{\pi m_0 \nu^2} = n^2(\kappa^2 - 1) + q + \frac{4n^4\kappa^2}{n^2(\kappa^2-1)+q} = F(\text{say}) \text{ ......(98).}$$

Hence $F\nu^2$ should be the same for all values of ν, and we should have, since $N_2\epsilon = p_2 \frac{d}{w} \times 2{\cdot}89 \times 10^{14}$ (see (63)),

$$p_2\epsilon/m_0 = F\nu^2 . \frac{w}{d} \times 1{\cdot}08 \times 10^{-14} \text{ ....(99),}$$

where w and d are the atomic weight and density of the metal and p_2 is the number of free electrons liberated from each atom. It must be remembered that in drawing this conclusion it has been supposed that q is constant and that ν is small compared with ν_1. Consequently (98) should hold only for frequencies less than those of any of the bound electrons; we should not expect it to hold for a metal which showed selective absorption in the visible spectrum if the incident light were in that region. The following tables show that in this matter the predictions of our theory are in accordance with experiment; the relation holds throughout the spectrum for steel and silver but not for copper; it fails for all metals in the ultra-violet.

TABLE III.

Steel		Silver		Copper	
ν	$F\nu^2$	ν	$F\nu^2$	ν	$F\nu$
$1{\cdot}32 \times 10^{15}$	167	$1{\cdot}20 \times 10^{15}$	138	$1{\cdot}00 \times 10^{15}$	62
1·00 ,,	128	·93 ,,	26	·67 ,,	50
·54 ,,	118	·76 ,,	43	·55 ,,	31
·48 ,,	126	·60 ,,	47	·48 ,,	32
		·51 ,,	47		

The frequency of yellow light is about $\cdot 5 \times 10^{-15}$.

In these tables the values of F have been estimated on the assumption that $q = 3$. It is impossible to find directly the value of q, for it depends on the properties of the bound electrons, the optical effects of which are masked by those due to the free electrons. But fortunately the term in q is probably small; the value adopted for it is that which is found for most dielectrics, since there is no reason to suppose that the properties of the bound electrons are very different in conductors and dielectrics.

We can now calculate p_2 by means of (99). Assuming that ϵ/m_0 is $4{\cdot}5 \times 10^{17}$ (see p. 46) we find the following values for p_2 and N_2.

TABLE IV.

Metal ...	Cu	Ag	Ni	Fe	Pt	Al	Pb	Hg	Bi	Sb
p_2	0·47	1·06	1·00	1·70	2·00	2·29	3·27	3·39	3·66	7·54
$N_2 \times 10^{-23}$	0·42	0·60	0·90	1·44	1·32	1·38	1·08	1·74	1·02	2·52

It will be observed that p_2 is a small number (it is not always an integer because of the uncertainties in our calculations), so that the number of free electrons in a metal is of the same order as the number of atoms. It is found that p_2 is apparently independent of the temperature of the metal.

Having reached these conclusions let us return to (87) and (88) and calculate the mean free path of the electrons and the number of collisions they make in one second. The lowest value of this last number is that for silver; for this metal $\sigma = 5{\cdot}4 \times 10^{17}$ (in the electrostatic units which we use), the atomic weight is 108 and the density 10, so that, calculating $N_2\epsilon/m_0$ as in § 28, we find

$$t = 7 \times 10^{-14} \text{ secs.}; \qquad l = ut = 8{\cdot}4 \times 10^{-7} \text{ cm.};$$

t is the period of light of wave-length 21μ. If the wave-length is less than 21μ, the electron will not on the average make even one collision before the direction of the force acting on it is reversed, and it might appear that the conductivity calculated for steady fields could have nothing to do with the conductivity for such rapidly alternating fields; yet it is found that for wave-lengths considerably less than 21μ the optical properties of silver can be calculated approximately from the conductivity for steady fields.

But this difficulty is only apparent. We have assumed that t is not influenced materially by the action of the electric field, so that the electron makes collisions in much the same way whether the field is changing or not. The absorption of light in the substance is determined, like the resistance to a steady current, by the rate at which the energy of the electrons acquired under the action of the field is dissipated by collisions, and if the rate at which collisions take place is independent of the periodicity of the field, the rate at which energy is dissipated will also be nearly independent. The absorption varies with the frequency, not so much because the rate of dissipation of energy varies, but because the phase difference between the incident and refracted light varies; so far as the absorption is determined by the dissipation it will be found to be the same for all periodicities, so long as the time between two collisions is short compared with the *time of observation.*

But we have yet to see that our assumption that t is independent of the field strength is justified. We have

$$\frac{v}{u} = \frac{X\epsilon l}{2m_0 u^2} = X \cdot \frac{1}{4}\left(\frac{\epsilon}{\alpha}\right)\frac{l}{T}$$

$$= X/580 \text{ about at } 18^\circ \text{ C.}$$

Hence unless $X > 6$ the velocity of the electrons due to the field will be only 1% of that due to their heat motion and t will be practically unaffected by the field. Now $N = 6$ means a field of 1800 volts per centimetre, far greater than can be applied to a metallic conductor in measuring its resistance. It is possible however that at very low temperatures, v might become comparable with u for fields that are attainable experimentally.

It should be noted in passing that the calculation of the number of electrons does not involve the quantity σ or depend in any way upon the accuracy of (89); the conclusions at which we arrived in this matter do not depend upon the assumptions which we have made as to the nature of the frictional forces acting on the electrons, but only on the assumption that there are electrons subject to frictional forces.

Thermal conductivity.

38. It is a well-known generalisation that good conductor of electricity are also good conductors of heat. Till the advent of the electronic theory no explanation of the relation could be suggested, but it is an immediate consequence of the views which are expounded here. We suppose that the conduction of heat, like the conduction of electricity, is effected by the agency of the free electrons in the same way that the conduction of heat in a gas is effected by the agency of the molecules. The free electrons at the hotter end of the bar have a greater kinetic energy than those at the colder end; when collisions[1] take place the former electrons lose energy and the latter gain energy, so that heat is transferred from the hotter to the colder parts.

---

[1] It is to be observed that throughout our investigations it is unnecessary to decide whether the electrons collide with each other or only with the molecules; if they collide only with molecules a transfer of energy from one electron to the other might be effected by consecutive collisions with the same molecule, which received energy from one and gave it to the other.

The following theorem derived from the dynamical theory of gases will enable us to calculate the thermal conductivity. If the properties of the molecules vary with their distance along the axis of z, so that the molecules at a section through the point z on that axis each possess an amount g of some quantity, then, G, the amount of this quantity which is transferred by means of the collisions between the molecules in unit time across any point on the axis, is given by

$$G = \frac{1}{3} ul \frac{d(Ng)}{dz} \qquad \text{..................(100)},$$

where N is the number of molecules in unit volume. We are concerned with the transfer of energy, and we must put

$$g = \frac{1}{2} m_0 u^2 = \alpha T;$$

then the rate at which energy is transferred along a bar is

$$\frac{1}{3} ul\alpha N_2 \frac{dT}{dz},$$

if we suppose that N_2 is constant and does not vary with the temperature. Now k_e, the thermal conductivity of a substance measured in ergs/cm. sec. degree, is defined by the relation that the rate at which energy is conducted along the bar is $k_e \frac{dT}{dz}$; hence we have

$$k_e = \frac{1}{3} ul\alpha N_2 \qquad \text{..................(101)}.$$

(If the heat energy is measured in calories, k_c, the appropriate value of the thermal conductivity, is $k_e/4{\cdot}2 \times 10^7$.) Comparing (89) and (101), we see that

$$\frac{k_e}{\sigma} = \frac{4}{3}\left(\frac{\alpha}{\epsilon}\right)^2 T \qquad \text{..............................(102)}$$

$$= 7{\cdot}15 \times 10^{-11} \text{ at } 18^\circ \text{ C.};$$

that is to say, the ratio of the thermal to the electrical conductivity of a metal is a universal constant, independent of the nature of the metal and proportional to its absolute temperature.

This law has long been known as the result of observations and is generally called the "Wiedemann-Franz Law." Table V shows how nearly it is true for a large number of metals. The number in the last column should be equal to the ratio of the absolute temperatures corresponding to 100° C. and 18° C., that is 1·29. It will be observed that k/σ is not quite constant, but increases as the conductivity diminishes. However, since the conductivity of silver is 200 times that of bismuth, the close agreement between the values of k/σ is remarkable.

TABLE V.

Metal	k/σ at 18°C.	$\dfrac{k/\sigma \text{ at } 100°\text{C.}}{k/\sigma \text{ at } 18°\text{C.}}$
Aluminium	$\cdot 706 \times 10^{-10}$	1·32
Copper	·738 „	1·30
Zinc	·745 „	1·29
Silver	·760 „	1·28
Nickel	·776 „	1·30
Cadmium	·784 „	1·28
Lead	·794 „	1·31
Gold	·807 „	1·27
Mercury	·815 „	1·30
Tin	·816 „	1·26
Platinum	·836 „	1·35
Palladium	·837 „	1·35
Iron	·890 „	1·33
Bismuth	1·068 „	1·12

Still more remarkable is the agreement of the actual value of k/σ with that predicted by (102); the close agreement is doubtless partly accidental, for if we had adopted slightly different, and probably more accurate, assumptions concerning the nature of collisions we should have found a number rather less than 1 in place of the 4/3. But nevertheless the results seem to leave no doubt as to the correctness of the assumptions of our theory which are involved in these conclusions, namely that conduction of heat and conduction of electricity are both effected by the agencies of carriers bearing the charge ϵ and that the mean kinetic energy of these carriers is the same as that of a molecule of the gas at the same temperature[1]. No more striking evidence of the truth of our theory or of its importance could be offered; one of the greatest

---

[1] It should be noted that the conclusion (102) is independent of any assumption concerning the mass of the electrons.

merits a scientific theory can have is that it correlates very diverse phenomena, and we have been able to calculate a quantity connected with the conductivity of copper from data derived from the electrolysis of water and the compressibility of hydrogen.

The fact that k/σ does not appear to be quite the same for all metals requires some slight modification of our theory, but no completely satisfactory suggestion has been made as yet. It is known that bodies of which the electrical conductivity is practically nil possess a finite thermal conductivity; since such bodies possess no free electrons, we must suppose that in such cases the conduction of heat is effected by the molecules as in a gas. Now if the molecules of a metal take some part in the thermal conduction the value which we have calculated from (101) will be too low, and the discrepancy will probably be greater the less the electrical conductivity of the substance. Hence some part of the discrepancy may be accounted for in this way, but unless the molecules of metals are more efficient in conducting heat than those of dielectrics this part must be small. Perhaps the explanation is to be found in the fact that the value of the mean free path is not so independent of the action of the field as we supposed in § 34, but in the present state of our knowledge there is little value in such speculations.

Contact of two metals.

39. In § 37 we saw that the concentration of the free electrons is different in different metals. If two such metals are placed in contact, the electrons will tend to pass from the metal in which the concentration is greater to that in which it is less. We can apply (100) to find the rate of this diffusion. Since we are now concerned only with the number of electrons and not with any quantity characteristic of them, we must put $g=1$; then G, the rate at which electrons flow across a section of the metal at z, is given by

$$G=\frac{ul}{3}\frac{dN}{dz} \qquad \text{(103)}.$$

But this flow of electrons cannot continue indefinitely; the metal which is losing electrons acquires a positive charge and that which is gaining electrons a negative charge; an electric field will be set up across the junction of the metals tending to stop the flow of electrons and equilibrium will be attained when the flow of electrons in this field is equal and opposite to the flow due to diffusion. Let X be the electric intensity of the field when equilibrium is attained; then by (84) the rate at which electrons cross the boundary under this force is Nv or $\frac{N\epsilon luX}{4\alpha T}$, and the electrons must arrange themselves finally so that

$$\frac{N\epsilon lu}{4\alpha T} X = \frac{ul}{3} \frac{dN}{dz} \quad\text{.....................(104)},$$

or

$$X = \frac{4}{3} \frac{\alpha}{\epsilon} T \frac{d \log N}{dz} \quad\text{..................(105)}.$$

These processes will be confined to a very thin layer in the junction of the metals; in the body of the metals the number of electrons will be practically the same as if each was isolated, and the value of $\int X dz$ from a point in one metal (a) to a point in the other metal (b) where the number of electrons is unchanged by the processes at the junction will be the difference of potential between the metals. Hence

$$V_{ab} = \frac{4}{3} \frac{\alpha}{\epsilon} T \log \frac{N_a}{N_b} \quad\text{..................(106)},$$

where V_{ab} is the potential difference of the metals and N_a, N_b the number of electrons in unit volume of them when each is isolated. It is to be noted that the metal at the higher potential (i.e. charged positively) is that for which the value of N is the greater, for it is this metal which will lose electrons.

If several different metals, a, b, c, d, are joined in series, we see from (106) that the potential between the free ends depends only on the materials of those ends and not on the intervening metals, for

$$V_{ab} + V_{bc} + V_{cd} = \frac{4}{3} \frac{\alpha}{\epsilon} T \left(\log \frac{N_a}{N_b} + \log \frac{N_b}{N_c} + \log \frac{N_c}{N_d} \right)$$
$$= \frac{4}{3} \frac{\alpha}{\epsilon} T \log \frac{N_a}{N_d} = V_{ad} \quad\text{.........(107)}.$$

Consequently no current will flow in a closed circuit composed of different metals if all the junctions are at the same temperature for if the circuit is cut at any point where there is not a junction the ends thus separated will be at the same potential.

At first sight we might be inclined to identify this difference of potential due to different concentrations of the electrons with the Volta difference of potential which is shown in ordinary circumstances between the free ends of two dissimilar metals the other ends of which are in contact. But we shall see presently that, if the values of N for the different metals given in Table IV are correct, the potential differences deduced from (106) are very much smaller than the Volta potential differences. The electronic theory thus leads to the view which has long been held on other grounds that the Volta effect is not due to the mere contact of the dissimilar metals, but to some processes, probably of a chemical nature, taking place at their free ends. The differences of potential considered here may, however, be identified with those which give rise to the phenomena of thermoelectricity.

The Peltier effect.

40. Let us consider the heat which will be developed when a current flows across the junction of two metals containing different numbers of electrons. Let the electric intensity due to the external source which drives the current be X. Then, if there is no internal electric intensity, each electron during its free path drifts a distance d along the direction of X and acquires an amount of energy $X\epsilon d$; when it collides, this energy is communicated to the molecules of the metal and appears as heat energy. Since the electrons make $1/t$ collisions per second, the amount of heat energy Q developed in the metal in one second by the N electrons in unit volume is given by

$$Q = NX\epsilon\, d/t = NX\epsilon v = X^2\frac{\epsilon^2 Nlu}{4\alpha T} = \sigma X^2 \quad \ldots(108).$$

If l is the length of the metal, q the area of its cross-section, i the current flowing through it and V the potential difference between its ends, we have

$$i = sq = \sigma Xq;\ \ V = Xl;\ \ R = \frac{l}{\sigma q} \quad \ldots\ldots\ldots(109),$$

so that for the whole bar

$$Q = \sigma X^2 ql = iV = Ri^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots(110),$$

the ordinary expressions for the Joule effect.

Suppose now that there is in the metal an internal difference of potential V' arising from the diffusion of the electrons from a place where N is greater to a place where N is smaller. This potential difference will not affect the flow of electrons through the metal under an external electric field, for it is only just sufficient to counterbalance the tendency of the electrons to move in the opposite direction. Hence, even if V' is in the same direction as the external field, it will still require an external difference of potential V to send the current i through the metal. But now the difference of potential between the ends of the metal will be $V + V'$ and the loss of electrical energy will be $i(V + V')$; consequently the gain of heat energy must be $i(V + V')$ and, if Q_1 is the heat now developed in the metal, we shall have

$$Q_1 = i(V + V') = Ri^2 + iV' \quad \text{...............(111).}$$

Similarly, if V' is in the opposite direction to V, we shall have

$$Q_1 = Ri^2 - iV' \quad \text{.....................(112).}$$

By comparing (111) and (112) with (110) we see that if there is an internal difference of potential in the metal, such as exists where two dissimilar metals are in contact, there will be developed an extra amount of heat beyond that due to the Joule effect; this extra heat is to be added to that due to the Joule effect if the internal and the external potential difference are in the same direction (so that the current flows from the positively charged to the negatively charged metal) and is to be subtracted from that due to the Joule effect if the current flows in the opposite direction. The amount of this heat developed in unit time by unit of current flowing across the junction is V_{ab}.

The developement or absorption of heat at the junction of two metals when a current flows across the junction is known as the Peltier effect. The Peltier coefficient Π_{ab} of two metals, a and b, is defined as the quantity of heat developed in unit time by the passage of unit current across the junction. Consequently

$$\Pi_{ab} = V_{ab} \quad \text{.......................(113).}$$

We can now calculate the Peltier coefficient from the values of N_2 given in Table IV. The results do not accord completely with experiment either qualitatively or quantitatively. It is found experimentally that heat is developed when the current passes from antimony to any of the other metals in the list and absorbed when it passes from bismuth to any of those metals. Consequently if our theory is correct, the value of N_2 for antimony should be higher and that for bismuth lower than for any other metal; we see that the former but not the latter condition is fulfilled. Nevertheless the order in which the metals are arranged according to their thermoelectric power agrees generally with that of the numbers in Table IV, and the absolute magnitude of the observed Peltier coefficients is not very different from that given by (106). Thus for a junction between bismuth and antimony we find

$$V_{ab} = \frac{4}{3} \times 4{\cdot}3 \times 10^{-7} \log_e \left(\frac{2{\cdot}52}{1{\cdot}02}\right) T = 5{\cdot}2 \times 10^{-7}\, T$$
$$= 1{\cdot}6 \times 10^{-4}\, T \text{ volts};$$

the observed value of V_{ab} is about $\cdot 8 \times 10^{-4}\, T$ volts.

On the other hand, according to (106) the Peltier effect ought to be proportional to the absolute temperature of the junction, a relation which is not observed to hold universally. Indeed for some junctions the Peltier coefficient first reaches a maximum, subsequently decreases and then changes sign as the temperature is increased. These divergencies between the predictions of our theory and the results of experiment are doubtless due to the neglect of two factors. In the first place we have assumed on the basis of evidence provided by the optical measurements that N is independent of the temperature. But evidence of this nature is not very reliable owing to the difficulty of making accurate observations and the uncertainty, which has been noted already, of the theory according to which they are interpreted. It is possible that N changes with the temperature at a rate which is different for different materials (see p. 84). In the second place, it has been assumed that there are no forces opposing the passage of electrons from one metal to another. It is quite possible that such forces exist. In the interior of a homogeneous metal the positive charges on the molecules which have lost electrons are distributed uniformly, so that there is no resultant field due to them in the interior of the metal; but at a junction between two dissimilar metals it is not certain that the field due to the positive charges in one metal will be equal and opposite to that of the positive charges in the other; there may be a resultant field tending to drive the electrons out of one metal into the other.

Taking these sources of uncertainty into account, we may conclude that though our theory of metallic conduction does not lead at once to a complete and satisfactory explanation of the Peltier effect, yet there is nothing in the observed phenomena which is inconsistent with that theory.

The Thomson effect.

41. Closely associated with the Peltier effect is the Thomson effect. It is found that a current flowing through a homogeneous metallic conductor, the ends of which are at different temperatures, developes in the conductor different amounts of heat according to the direction of the current. To describe this phenomenon Kelvin introduced the conception of the 'specific heat of electricity.' If the current of electricity were replaced by a current of some fluid, the metal would gain heat if the fluid entered at a high temperature and left it at a low temperature, and would lose heat if the fluid flowed in the opposite direction. If T and T' are the temperatures of the ends of the metal, w the heat capacity of the quantity of fluid which enters and leaves the metal in one second, the amount of heat gained or lost by the bar in that time is $w(T-T')$. Now since the amount of heat gained or lost by the bar (above that due to the Joule effect) when a quantity of electricity E passes through it from T to T' is found to be $\mu E(T-T')$, we may, by analogy, call μ the 'specific heat of unit quantity of electricity.'

It might seem that this effect was easily explained by the electronic theory of metallic conduction. Since the direction of the current is opposite to that of the motion of the electron when the quantity of electricity E flows into the metal at a temperature T, E/ϵ electrons leave it, each of which has a kinetic energy αT; when the same quantity leaves it at a temperature T', the same number of electrons enter it, each of which has a kinetic energy $\alpha T'$. The amount of energy given to the metal must be

$$-E/\epsilon \,.\, \alpha(T-T'),$$

so that the 'specific heat of electricity' is given by

$$\mu = -\frac{\alpha}{\epsilon} = -4{\cdot}3 \times 10^{-7} \text{ ergs per el. st. unit}$$

$$= -3 \times 10^{-5} \text{ gm. cals. per coulomb.}$$

However, it is found that μ is not the same for all substances; in many substances it is positive and not negative as we have calculated. It appears that the Thomson effect is not determined by this action only.

The Thomson effect which is actually observed is probably largely due to the presence of internal electric fields in a metal the ends of which are at different temperatures, similar to those which give rise to the Peltier effect. If one end of a metal is hotter than the other, the electrons at different ends possess different kinetic energies and, possibly, are also present in different numbers; on both of these accounts there will be a diffusion of the electrons from one end to the other which will be counterbalanced, when equilibrium is attained, by an electric field opposing the diffusion. A current flowing through the metal will develope different amounts of heat according as it flows with or against this internal field. The magnitude of this field and, consequently, the specific heat of electricity in the metal may be calculated in the same way as the quantities determining the Peltier effect; but since the calculations are somewhat complicated and the results of them affected by the same sources of uncertainty as those concerning the Peltier effect, they will not be given here. It is sufficient to say that the agreement between theory and experiment is, again, by no means perfect, but the discrepancies are not so serious or so inexplicable as to throw serious doubt upon the main hypotheses of the theory of metallic conduction.

Thermodynamic reasoning is often employed to show that there must be a certain connection between the Peltier and Thomson effects, and such reasoning assumes necessarily that the phenomena concerned are reversible. It does not seem certain how far the results of the electronic theory are in accordance with such reasoning and, indeed, that theory must throw doubt upon the reversibility of the phenomena. For in order to apply thermodynamics it is necessary to assume that the Peltier and Thomson effects are wholly independent of the Joule effect, so that the first two may be finite while the last is infinitesimal. Our theory however regards all three effects as inseparable manifestations of the same mechanism, the collisions of the free electrons as they pass through the positive molecules.

The Hall effect.

42. An electron moving in a magnetic field is subject to a force at right angles to the magnetic intensity and to the direction of its motion. Consequently if a bar carrying a current is placed in a magnetic field at right angles to the current, the electrons, which have on the whole a motion in the opposite direction to the current, will tend to move towards one of the sides of the bar in the plane perpendicular to the magnetic field. Fig. 4 shows the relations which will obtain between the directions of the current, the magnetic field and the deflection of the electrons[1]. The electrons will accumulate on the upper side of the bar until the electric field due to the negative charge on the upper side and the positive charge on the lower is sufficient to prevent any further accumulation. The existence of this field might be detected by joining opposite quadrants of an electrometer to the upper and lower sides of the bar, or by joining these sides by a wire, through which a current should flow from the lower to the upper side.

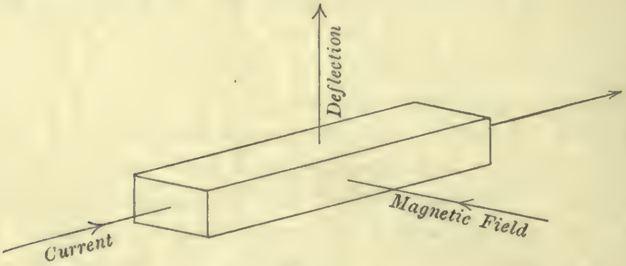


Fig. 4.

---

[1] The following rule may help the reader to remember the relation between the various vectors :—If an Electron moves towards the East in the earth's magnetic field, the Force on the electron will be towards the Feet. The relation may also be derived easily by remembering Ampère's rule for the relation between a current and the magnetic field excited by it together with the fact that parallel currents flowing in the same direction attract each other.

Such a transverse electric field has been detected in metals carrying currents in a magnetic field at right angles to the current and is known by the name of its discoverer as the Hall effect. Unfortunately however the direction of the transverse electric field is not always that predicted by our theory. The effect is said to be positive when the direction of the transverse field can be obtained by rotating the direction of the current through a right angle in the direction of a circular current which would give rise to the magnetic field acting; our theory, according to which the upper side of the bar in Fig. 4 should be negative, predicts a negative sign for the Hall effect. In bismuth the effect is negative, in antimony it is smaller and positive, in tellurium it is very much greater and positive, in other metals it is much smaller and of both signs. We shall discuss the cause of this discrepancy later when we have considered some other phenomena closely connected with the Hall effect; but while the discrepancy is unexplained no useful purpose would be served by attempting to deduce a quantitative relation between the magnitude of the effect and the circumstances of the electrons.

Since the electrons during their free flights tend to move towards the upper part of the bar, they will give to that part of the bar most of the energy acquired by moving under the electric field driving the current. Accordingly the upper part of the bar should be warmer than the lower. This effect is extremely difficult to observe, but it has been certainly detected in a few metals and in all of them is in the direction predicted by our theory.

Similar transverse electric fields and temperature gradients should be noted if the current of electricity is replaced by a current of heat due to a difference of temperature between the ends of the bar. When heat is being conducted through the bar there is no flow of electrons from one end to the other, but the electrons moving from the hot end to the cold are moving faster than those from the cold end to the hot, and are subject accordingly to a greater deflecting force. A little consideration will show that, if the current of heat is in the same direction as the current of electricity in Fig. 4, the electrons moving from the hot end to the cold are deflected towards the lower side of the bar, so that this side of the bar should be warmer than the other and should be at a lower potential. The transverse temperature gradient due to conduction of heat in a magnetic field has been detected in a few metals and shows the same discrepancies as the Hall effect; it is not, as our theory would predict, in the same direction in all metals; but the transverse electric field in the same circumstances has been detected in many more metals and always coincides in direction with the predictions of our theory.

In addition to these transversal effects there will be longitudinal changes in the electric field and the temperature gradient. Electrons moving in the vertical plane at right angles to the magnetic field will be deflected by that field along the length of the bar; those moving upwards will be deflected to the right, those moving downwards to the left. If a current is flowing as in Fig. 4, those moving upwards have a higher velocity; they will be more deflected and cause a general drift of the electrons to the left. This drift of the electrons is in the direction opposite to that representing the main current, so that the effect of the magnetic field will be to cause an apparent increase in the electrical resistance of the bar. Such a change has been detected in the case of bismuth; in some of the magnetic metals the magnetic field is found to cause a decrease in the resistance.

There will be similar longitudinal effects due to a current of heat through the bar, but the consideration of these and other complications connected with the Hall effect would serve no useful purpose. The chief interest of these phenomena for electronic theory lies in the fact that the direction of the various effects is not always in accordance with prediction; this discrepancy has been used as a support for an hypothesis which, though it is now generally abandoned, requires a brief notice.

Are there 'positive electrons'?

43. It has been argued that the difficulties connected with the sign of the Hall effect would vanish if it were assumed that positive particles, as well as the negative electrons, were instrumental in carrying the current of electricity or of heat through the metal. For those effects the direction of which for all metals is found to be in accordance with our theory are also those the direction of which is independent of the sign attributed to the carriers; while those effects the direction of which is different in different metals are also those the direction of which is dependent on the sign attributed to the carriers. If the sign of the carriers is reversed, the direction of their motion in a given current of electricity is reversed and, consequently, the direction of their deflection by a given field is unaltered; whatever the sign of the carriers they will accumulate at the top of the bar in Fig. 4; but the direction of the motion of the carriers in a given current of heat is the same whatever their sign; consequently if the carriers were positively charged they would accumulate at the top of the bar. Thus when a current of electricity flows, the direction of the electric field, but not the direction of the temperature gradient, depends on the sign of the carriers; when a current of heat flows, the direction of the transverse temperature gradient, but not the direction of the electric field, depends on the charge on the carriers.

One very serious objection to the hypothesis that positive as well as negative carriers are concerned in the current has been pointed out by Lorentz. If the carriers are all of one sign, there can be no accumulation of carriers unless there is also an accumulation of charge; but if carriers of both signs are present, it appears that there will generally be an accumulation of carriers of both signs at the boundary between two metals, even when they form part of a circuit in which the current is the same at every point. This accumulation of carriers in one part of the circuit should certainly produce some change in the properties of the metal which should be noticeable after the current has been flowing for a sufficient time.

But there is a much more fundamental objection to the hypothesis of positive carriers. If there are such carriers they must be the same in all metals, as we saw in § 33. Now there is some evidence from other sources of the existence of positive particles which are common constituents of all atoms, but there is no evidence whatsoever of the existence of positive particles which are at all comparable in size to the electrons; in all experiments which throw any light upon the nature of positively charged particles we always find that those particles are at least as large and as massive as an atom of hydrogen; carriers of this size could not move freely between the atoms of a metal; the resistance opposing their motion would be at least as great as that opposing the motion of an electrolytic ion in a solution, and the amount of current conveyed by them would be inappreciable compared with that carried by the electrons.

If any part of metallic conduction is to be attributed to the agency of positive carriers, these carriers must be utterly unlike any particles of which we can obtain evidence from any other source; it must be supposed that the mechanism of metallic conduction differs fundamentally from that of any other electrical phenomenon. The whole aim and object of the electron theory is to attribute all electrical phenomena to the same mechanism; to introduce the conception of positive electrons in order to explain a feature of metallic conduction only is not to expand the scope of that theory but to deny the fundamental assumption on which it rests. We must either deny the whole electronic theory or maintain our view that all particles at all similar to the electrons which we have imagined bear a negative charge.

The disagreement between theory and experiment in dealing with the Hall effect is not greater or more important than that which was noted in dealing with thermoelectricity; the discrepancies in both cases are most probably due to our ignorance of the nature of the collisions which an electron makes in a metal and of the molecular forces under which it moves. In the simple theory which has been outlined the only velocity attributed to the electrons is that of their drift through the metal; we have neglected altogether the much greater velocity of thermal agitation and the deflections of the electrons in moving through the magnetic field with this velocity. And again, in a later chapter we shall see reason for believing that the magnetic intensity in the region through which an electron moves between two atoms is not always at all the same as the magnetic intensity of the imposed field outside the metal. While these sources of uncertainty exist we may conclude that the phenomena of the Hall effect, though not at present completely explicable on our theory, are by no means necessarily inconsistent with it.

44. It will be convenient to consider here some phenomena which, though they are not concerned directly with the conductivity of metals, are closely connected with it.

Thermionics.

We have attempted to explain metallic conduction on the view that metals contain free electrons moving about among the atoms of the metal as the molecules of a gas move in the pores of a substance in which the gas is contained. At ordinary temperatures the electrons do not escape from the metal, and we must suppose that they are restrained within it by forces in the surface opposing their exit; such forces doubtless arise from the attraction of the positively charged atoms, which is equal in all directions in the interior of the metal, but is directed towards the interior at the surface. But since the kinetic energy of the electrons increases with the temperature, if the temperature of the metal were raised sufficiently, the energy of the electrons would be greater than the work required to carry them through the surface against the opposing forces, so that the electrons would emerge through the surface into the space beyond. The presence of such a stream of electrons would show itself as a current of negative electricity flowing from the metal to any surrounding conductors maintained at an equal or a higher potential.

Hitherto for the sake of simplicity we have regarded all the free electrons as possessing the same velocity; but if the electrons really behave like the molecules of a gas their velocities must be distributed according to the well-known Maxwell-Boltzmann law. Even at low temperatures we should expect a few electrons to possess sufficient kinetic energy to escape from the metal; the effect of an increase of temperature should be to increase continuously the number thus escaping. By the ordinary methods of the dynamical theory of gases, it can be shown that, if w is the work required to carry an electron through the surface of the metal and N the number of free electrons in unit volume of the metal, n, the number escaping from unit surface in unit time, should be given by

$$n = N\left(\frac{\alpha T}{2\pi m}\right)^{\frac{1}{2}} e^{-w/\alpha T} \dots\dots\dots\dots\dots\dots (114);$$

i, the current of negative electricity flowing from the metal, should be given by $i = n\epsilon$. In carrying out this calculation no assumptions are necessary as to the nature of the motion of the electrons within the metal; indeed the result can be obtained by purely thermodynamical reasoning of a general nature. It is assumed, however, that N is independent of the temperature.

We have no grounds for estimating *a priori* the magnitude of w; it might be so great that the current could not be detected at any temperatures attainable experimentally. As a matter of fact it is found that there is a current of negative electricity, increasing rapidly with the temperature, flowing from all metallic conductors at a temperature of more than 1000° C. to surrounding conductors, even if the intervening space is completely vacuous. The existence of this current appears to have been detected first by Edison, but our present knowledge of it is chiefly due to Richardson, who has given the name 'thermionics' to the study of the subject.

The investigations confirm very completely our theory of the phenomenon. By the application of the principles of § 18 to the study of the magnitude and direction of the current in a magnetic field it has been shown that the carriers of the current have a value for e/m coinciding closely with that found for electrons by the most accurate methods, and that the velocities with which these electrons leave the metal are distributed according to the Maxwell-Boltzmann law, so that the proportion of the electrons having a speed u is proportional to e^{-ku^2}, where $k = \frac{3m}{4\alpha T}$[1]. It is found, moreover, that the relation between the current leaving unit area of the surface and the temperature of that surface is completely represented by the formula $i = aT^{\frac{1}{2}} \cdot e^{-b/T}$, where a and b are constants characteristic of the metal. This formula is identical with (114) if

$$a = N\epsilon \left(\frac{\alpha}{2\pi m}\right)^{\frac{1}{2}} \quad\text{....................(115)};$$

$$b = \frac{w}{\alpha} \quad\text{..........................(116)}.$$

---

[1] According to the theory of gases the Maxwell-Boltzmann law should hold for the velocities of molecules after they have performed a finite amount of work in escaping through the surface as well as for the molecules in the interior of the metal.

By means of (115) and (116) we may deduce values of N and w if we assume that ϵ has the value $4{\cdot}7 \times 10^{-10}$. Thus for platinum it is found that $N = 5 \times 10^{21}$, $w = 8 \times 10^{-12}$ ergs. According to Table IV, N for platinum should be $1{\cdot}3 \times 10^{23}$; the agreement is not very good. Moreover the value of w is considerably greater than would have been anticipated. 8×10^{-12} ergs is the work necessary to carry an electron through a potential difference of about 5 volts, so that a potential difference of this magnitude must exist between the interior and the exterior of the metal. Similar measurements on tantalum indicate that the potential difference at the surface of that metal is under 4 volts; accordingly there should be a potential difference of at least 1 volt between the interiors of two pieces of platinum and tantalum when the exteriors are in contact. A potential difference of this magnitude at the surface of separation of the metals is quite irreconcilable with the theory of the Peltier effect which we considered in § 40; if there is such a potential difference at the surface we should expect it to be of the same order of magnitude as that due to the difference in the values of N, i.e. about 10^{-4} volt.

However all conclusions as to the magnitude of N and w from the study of thermionics are rendered uncertain by the fact that the thermionic current is influenced very markedly by the presence of very small amounts of impurity in the metals investigated. If the surface of platinum is not freed from the small quantity of hydrogen which usually adheres to it, the thermionic current at very high temperatures may be increased in the ratio of 10^5 to 1; it appears that this increase is due to a decrease in w and not to an increase in N. It has been suggested that this change is due to the presence of positively charged atoms of hydrogen on the surface of the metal, tending to drag out the electrons, but the matter is probably not quite so simple. For experiments with carbon also show that the very large thermionic current obtained with the ordinary material is reduced until it becomes almost inappreciable when the material is specially purified. It has accordingly been suggested that the whole theory on which our arguments are founded is erroneous and that the emission of electrons is not due merely to an increase in their thermal energy with the temperature, but to some chemical reaction involving the impurities. Perhaps this view is rather too extreme in face of the agreement between the measured velocities of the electrons and those calculated from Richardson's theory; still the difficulties mentioned are sufficiently serious to diminish the reliance which can be placed on any reasoning based on thermionic phenomena. It may be observed in passing that, if experiments on thermionics may be taken as giving any information concerning the number of free electrons, the agreement of (114) with experiment would confirm the view that N is almost if not quite independent of the temperature.

Thermionic currents are observed not only from metals but also from other substances, such as metallic oxides, which are conductors at high temperatures; and at temperatures too low for the negative thermionic current to be appreciable there is a positive thermionic current from the metal, which is carried by charged atoms usually derived from impurities present in the substance. But these matters are too complex for our theory in the present state of our knowledge.

45. In conclusion we must notice a somewhat serious difficulty which affects the most fundamental assumptions of the electronic theory of conduction.

The specific heat of metals.

The kinetic energy of the n electrons in any piece of metal has been assumed to be $n\alpha T$; hence the amount of energy which must be given to the electrons when the temperature of the metal is raised 1° is $n\alpha$. If the mass of the piece of metal in grammes is equal to the atomic weight of the metal, the number of atoms contained in it is $1/m_h$; hence, if there are p free electrons for each atom, the amount of energy which must be communicated to the electrons when the temperature of the metal is raised 1° is

$$p\frac{\alpha}{m_h} = p\left(\frac{\alpha}{\epsilon}\cdot\frac{\epsilon}{m_h}\right) = p \times 4\cdot 3 \times 10^{-7} \times 2\cdot 9 \times 10^{14}$$
$$= 1\cdot 24 \times 10^{8} p \text{ ergs or about } 3p \text{ calories.}$$

Experiment shows that the amount of heat required to raise the temperature of this mass of a metal through 1° (usually called the 'atomic heat' of the metal) lies between 5 and 7 calories at ordinary temperatures. Since the atomic heat of a metal can clearly not be less than the energy required to give the requisite energy to the electrons in it, p cannot be greater than 2·3 for any metal. The values for p calculated in Table IV are sometimes less and sometimes greater than 2·3; we must certainly conclude that the arguments which lead to the greater values are incorrect.

This conclusion in itself need not shake our faith in the theory, for the calculations of the precise values of p are affected by many sources of uncertainty; the arguments given could not be relied upon to show more than that the number of free electrons was of the same order of magnitude as the number of atoms. However even this general conclusion raises difficulties.

On our view a dielectric differs from a conductor in the fact that the latter does and the former does not contain free electrons. Accordingly the heat which is necessary to raise the temperature of a dielectric cannot be expended in increasing the energy of the free electrons in it: it must be expended in raising the energy of the atoms. Now the atomic heat of dielectrics is, in general, the same as that of metals[1], so we must conclude either that the free electrons add nothing material to the heat capacity of the substance or that the atoms in solid conductors have a much smaller energy than those in solid dielectrics. There appears no reason for adopting the latter conclusion, but the former would lead necessarily to the view that the number of free electrons is very much smaller than that of the atoms, for if p were as great as 1 the atomic heat of a conductor should be 1·5 times that of a dielectric.

The difficulty is undoubtedly very serious. The evidence derived from the agreement of the predicted and observed values of the Wiedemann-Franz ratio and the direct determination by Richardson of the velocities of the electrons in the thermionic

---

[1] It is true that the elements of which the atomic heat at ordinary temperatures is less than that given by Dulong and Petit's Law (C, B, Si) are all dielectrics, but recent investigations, which will be considered later, show that the difference between dielectrics and conductors in this respect is only apparent at ordinary temperatures. Other elements which are dielectrics have normal atomic heats; for compounds, which are almost all dielectrics at ordinary temperatures, the molecular heat is usually about five times the number of atoms in the molecule, that is to say, the atoms in the compound molecules possess nearly the same energy as the atoms in a substance obeying Dulong and Petit's Law.

current seem to provide a convincing proof that the kinetic energy of a free electron is indeed that which we have attributed to it; and yet it is quite impossible to explain the observed values of the conductivity of metals if the value of p is as low as the consideration of specific heats would suggest. We shall return to the question of specific heats in a later chapter and shall see that perhaps the discrepancy is not quite insuperable.

REFERENCES.

The first notable attempt at an electronic theory of metallic conduction is that of Riecke (*Wied. Ann.* 66. pp. 353 and 545, 1898). The fundamental assumptions of the modern theory were introduced simultaneously by Drude (*Ann. Phys.* 1. 566, 1900 and 3. 369, 1900) and J. J. Thomson (*Rapports du Congrès Internat. à Paris*, 1900, Vol. 3. 138); Drude's work is the most complete attempt to cover the whole ground. A much more elaborate mathematical treatment of the subject was given by Lorentz (*Proc. of the Amsterdam Academy*, 1905). Subsequent research has not really added much that is important to the results attained by these writers. A paper by Koenisberger and Weiss (*Ann. d. Phys.* 35. 1, 1911) gives an interesting comparison of the results of theory and experiment.

The treatment of the optical properties of conductors given here is based on the paper of Drude (*Ann. d. Phys.* 14. 936, 1904). Later writers on the subject have usually been more concerned with matters which will be considered in chap. X.

The treatment of thermoelectric phenomena by Baedecker (*Phys. Zeit.* 11. 809, 1910) and Krüger (*Phys. Zeit.* 11. 800, 1910) is interesting, the facts on which the thermodynamical arguments are based being derived from the study of thermionics. The former author gives a general summary of the state of the problem in his *Die elektrischen Erscheinungen in metallischen Leitern*, Braunschweig, 1911.

The older work on thermionics is summarised completely in J. J. Thomson's *Conduction of Electricity through Gases*, 2nd Edition, chap. VIII. The most important of the later papers is that of Richardson and Brown (*Phil. Mag.* 16. 890, 1908).

CHAPTER IV

GASEOUS CONDUCTION

Gases as conductors.

46. A GAS in its normal condition in an electric field of small intensity is an almost perfect insulator. If, however, the electric intensity is increased to a certain value, which depends on the properties and pressure of the gas and the nature of the electrodes, a current passes through the gas accompanied by the luminous phenomena of the 'spark.' In the first part of this chapter only the behaviour of the gas in electric fields much less intense than those required to produce the luminous discharge will be considered.

Processes have long been known by which the conductivity of gases in small electric fields might be increased enormously; it has long been known, for instance, that the gases of a flame and the gases liberated by the action of acids on metals possess a considerable conductivity for a short time after they are removed from the scene of action. But the conductivity conferred by these processes is not suitable for experimental study and nothing was known of the nature of the conductivity of gases until other and quite different means of producing it were discovered. The gaseous conductivity which has been, and still is, of fundamental importance for modern electrical theory is that which is produced by the passage of certain rays through the gas. The nature of these rays and the actions peculiar to each class of them will be treated in the next part of this volume; for the present we shall confine our attention to phenomena which are produced by all classes.

Theory of gaseous conduction.

47. The conductivity of gases differs from that of metals or electrolytes in two striking features:

(1) The increase in current produced by a given increase in the electric field is less the greater the field already acting; after a certain electric intensity (which is different in different circumstances) has been reached a further increase produces no change in the current. When the electric field is so great that the current is independent of the electric intensity, the current is said to be 'saturated.'

(2) In a given field the current is constant only while the intensity of the rays acting on the gas is constant. If the gas is removed from the action of the rays its conductivity gradually falls to zero. The rate at which the conductivity decays is greatly increased if an electric field acts after the action of the rays has ceased; if the field is great enough to produce a saturation current when the rays act, the conductivity vanishes almost instantaneously when the action of the rays ceases. The conductivity may also be destroyed completely by filtering the gas through cotton-wool.

A theory adequate to explain these features was proposed by J. J. Thomson and Rutherford as soon as the facts were known. It is supposed that the rays produce in the gas carriers or ions, charged some positively and some negatively, at a rate which is independent of the electric intensity and determined only by the properties of the rays and of the gas. The total charge on all ions of one sign is equal and opposite to that on all ions of the other sign. The current through the gas is carried by the ions which move in the electric field and give up their charges to the oppositely charged electrodes.

When no external field is acting the oppositely charged ions attract one another and coalesce (or 'recombine'), forming neutral particles which are no longer capable of carrying the current. The number of ions recombining in a given time increases with the number present in the gas; consequently, while the rays are acting, a state of equilibrium will be reached in which the number of ions recombining in a given time is equal to the number produced by the rays. If the rays cease to act, recombination proceeds until all the ions have disappeared and the gas ceases to conduct.

Some of the ions will also disappear by diffusing to the sides of the vessel containing the gas and giving up their charges there. If the volume of the gas is large the number of ions disappearing by diffusion will be small compared to the number disappearing by recombination; but if the volume of the gas is small the effect of diffusion will be great. The disappearance of the ions when the gas is filtered is to be attributed to the rapid diffusion of the ions to the sides of the pores of the filter.

If there is an external field, the ions will disappear not only by recombination and diffusion but also by attraction to the electrodes, and the rate at which the ions disappear will be increased. The greater the field, the greater will be the proportion of the ions disappearing by attraction to the electrodes rather than by recombination or diffusion and, since it is only ions disappearing in the first manner which carry the current, the greater will be the current. But a limit to the current will be reached when each ion, immediately it is produced, is attracted to the oppositely charged electrode and has no chance of recombining with other ions, except in the unlikely event of its colliding directly with an oppositely charged ion moving in the same straight line between the electrodes. In this state the 'saturation current' will be obtained; we see that if q is the number of ions of one sign produced by the rays in unit time and e is the charge on each ion, the saturation current C is given by $C = qe$.

According to this theory a steady current through a given gas is completely determined by the following quantities: X, the electric intensity at various points, q the rate at which ions are produced, e_+, e_- the charges on the ions, u_+, u_- their velocities (which will be functions of X), D_+, D_- their diffusion coefficients, and a a quantity which measures the rate of recombination. D is defined by the relation that, if N is the number of ions in unit volume of the gas, then n, the number diffusing in unit time across unit area perpendicular to x, is given by

$$n = D\frac{dN}{dx} \quad \dots\dots\dots\dots\dots\dots\dots(117);$$

a is defined by the relation that the number of ions recombining in unit time is aN_+N_-[1]. We shall not here be concerned with the quantitative developement of the theory or with the experiments which show that values may be assigned consistently to the various quantities concerned, so that the current under all circumstances is in agreement with that predicted by the theory. We shall only consider these values as far as they throw light on the important questions which concern the nature of the ions and the mechanism by which they are produced.

The charge on an ion.

48. The measurement of the charge on an ion was first made possible by a discovery due to C. T. R. Wilson, which enabled the individual ions present in a given volume of gas to be observed. It had long been known that a gas containing a condensable vapour might be cooled to a temperature at which the vapour was greatly 'supersaturated' without condensation taking place, so long as the gas was free from dust; while, if dust was present to serve as nuclei for the formation of drops, condensation occurred as soon as a temperature was reached at which the tension of the vapour in contact with the liquid was less than its pressure in the region occupied by the gas. Wilson showed that in this respect gaseous ions behave like dust; in the presence of ions it is possible to produce some supersaturation, but not so much as in the absence of ions. If there are positive ions present, condensation round those ions as nuclei begins when the pressure of the vapour in the gas is about seven times its equilibrium tension at the same temperature; if negative ions are present, condensation on those ions begins when the supersaturation is about fourfold. Now the ions in their normal state are quite invisible, just as are the molecules of a gas, but when the drops of liquid are condensed around them they are visible in a suitable illumination. Accordingly the ions may be made visible by saturating the gas in which they are contained with a suitable vapour and then cooling it; if the cooling is sufficient to produce fourfold supersaturation, only the negative ions will become visible; if it is sufficient to produce sevenfold super-

---

[1] It is not certain that a thus defined is independent of N_+ and N_-; according to Sutherland, if $N_+ = N_- = N$ the number of ions recombining in unit time is proportional to $N^{\frac{5}{3}}$ and not to N^2.

saturation, the positive as well as the negative ions will be rendered visible.

When the ions have thus been rendered visible, it is conceivable that their number might be determined by direct counting; the total charge on the ions may be found easily by an examination of the conductivity of the gas, and so a value for the charge on each ion might be deduced. The methods which have actually been used for the determination are rather more complicated in principle, but much better suited for experiment. It can be shown that if a sphere of radius a moves through a medium of a viscosity η under a constant force P, it will ultimately attain a steady velocity v, where[1]

$$v = P \cdot \frac{1}{6\pi a\eta} \quad\text{.......................(118).}$$

It is found that the drops after their formation fall with a velocity which is uniform after a very small fraction of a second, and this velocity may be determined experimentally. Since the force acting on the drops as they fall is $\frac{4}{3}\pi a^3\rho g$, where ρ is the density of the liquid of which the drops are composed, we have

$$v = \frac{2}{9} \cdot \frac{a^2\rho g}{\eta} \quad\text{.......................(119).}$$

Since η can also be measured, a knowledge of v leads to a value for the radius and, consequently, the volume of the spherical drops. Now if the quantity of liquid condensing when the vapour is cooled is sufficient to form drops on all the ions, the volume of it must be $V = N \cdot \frac{4\pi}{3} a^3$, where N is the number of ions in the gas. V may be determined from a knowledge of the properties of the vapour and of the temperature through which the gas is cooled and hence a value of N may be deduced. Ne is known from measurements of the conductivity, so that e is also known.

This method of determining e was employed by J. J. Thomson in 1898 and led to the value $e = 6{\cdot}5 \times 10^{-10}$. It was found that e

---

[1] This formula is only strictly accurate if the sphere is very large compared with the molecules of the gas; if it is comparable in size with the molecules the velocity is given by a more complicated formula, but is still a function of P, a, η.

was independent of the gas in which and the method by which the ions were produced. The method is interesting because it was the first that was ever employed for the measurement of this supremely important quantity, but it is not very accurate. The assumptions are involved that each drop contains only one ion and that the volume and charge of each drop are the same; neither of these assumptions has subsequently been proved to be true. A more accurate method and one which enables the charge on a single ion, instead of the average charge on a large number of ions, to be determined was devised by H. A. Wilson. If the drops after their formation are made to fall in a vertical electrical field of intensity X, the force acting on an ion is $mg + Xe$, and its velocity, v', is given by

$$v' = \frac{1}{6\pi a\eta}\left(Xe + \frac{4}{3}\pi\rho a^3 g\right) \quad\text{..............(120).}$$

Hence if we determine the velocity of fall of the same drop, first when there is no field acting and second when a field of intensity X is acting, we shall have from (119) and (120)

$$e = \frac{9\sqrt{2}\pi}{X} \cdot \left(\frac{\eta^3 v^3}{g\rho}\right)^{\frac{1}{2}} \cdot \frac{v' - v}{v} \quad\text{..............(121).}$$

By this method H. A. Wilson found for the smallest charge on a drop $e = 3{\cdot}1 \times 10^{-10}$. He could show also that the smallest charge on a negatively charged drop was the same as the smallest charge on a positively charged drop. But all drops had not the same charge; some of them had charges $2e$ and $3e$; but of course it was impossible for him to determine whether these multiple charges were carried by single ions or whether the drops carrying them were condensed round several ions.

Both Wilson's and Thomson's methods are liable to the objection that the drop, while it is falling, is always evaporating, so that its volume and its velocity are not strictly constant. In the last few years great improvements in the application of Wilson's principle to experiment have been made, chiefly by Millikan. In his latest work the condensation method of making the ions visible is abandoned; a few non-volatile particles (such as very small drops of oil or mercury) are introduced into the ionised gas and the rates of their fall with and without an electric field are observed. The particles when they are introduced usually bear a charge which is a small multiple of that carried by an ion and this charge is changed from time to time by the collision of an ion with a particle and the transference of the charge from the former to the latter. From such observations Millikan concludes that the charge carried by a particle is always a multiple of $e = 4{\cdot}8 \times 10^{-10}$, and that when it is changed (whether by the addition of positive or of negative electricity) the magnitude of the charge is always e and not a multiple of e. It appears then that positive and negative ions carry charges of the same magnitude and that ions carrying charges which are multiples of e are either extremely rare or non-existent. This conclusion is contradicted by evidence adduced by Townsend from experiments on the diffusion of ions similar to those mentioned below, which would prove that, at least in some circumstances, positive ions (but not negative ions) can carry a charge $2e$. This discrepancy is as yet not satisfactorily explained.

Gaseous ions and electrons.

49. The value for e is, of course, near to that which has been assigned to ϵ, the charge on an electron and, indeed, the methods which have been described are among the best for determining ϵ. We will consider briefly the evidence which led to the identification of e with ϵ.

Before the charge on an ion was measured evidence, arising from the experiments described in § 53, had been obtained of the existence of negatively charged particles of the same nature from whatever substance they were derived. The charge on these particles could not be measured, but, since the ratio of the charge to the mass was always the same, it was suspected that both the charge and the mass were the same. It had long been known that there was concerned in electrolytic conduction a charge which was the same for all substances, for the charge carried by a univalent atom is the same for all such atoms and that carried by any other atom is a small multiple of that charge. Since it is intrinsically improbable that there are two different charges characteristic alike of all substances, the charge upon the universal negatively charged particles (now called electrons) was immediately identified with that on a univalent electrolytic ion. When the charges on gaseous ions were studied and it was again found that the charge was independent of the material from which the ions were produced, the same argument led to the identification of the charge on a gaseous ion with the charge on an electron and the charge on a univalent atom. Moreover the absolute magnitude of the charge confirmed the hypothesis. Considerations derived from the theory of gases placed the value of ν_0, the number of molecules in a gramme-molecule of a gas, between $4{\cdot}7 \times 10^{23}$ and $2{\cdot}2 \times 10^{24}$; consequently if we adopt the earliest value for e, the charge on an ion, we find $\nu_0 e$ lies between $3{\cdot}1 \times 10^{14}$ and $1{\cdot}4 \times 10^{15}$. The value of $\nu_0 \epsilon$, where ϵ is the charge on a univalent atom, is $2{\cdot}9 \times 10^{14}$; though it lies outside these limits the uncertainties of $\nu_0 e$ are sufficient to account for the discrepancy.

Shortly afterwards Townsend produced a direct proof of the identity of the charge on an ion with that on a univalent atom based on measurements of the velocities of the ions in an electric field and their diffusion coefficients. It is found that the velocity of an ion is proportional to the electric intensity acting on it; let the velocity in a field of intensity X be $u_0 X$; then the force acting on the ion is eX and the ratio of the velocity to the force is $\frac{u_0}{e}$. On the other hand according to (117) the velocity of each of the Ndx ions contained in a layer dx of a gas in which the ions are unequally distributed is $\frac{D}{N}\frac{dN}{dx}$. If we regard the ions as forming a gas mixed with the unionised gas, this velocity may be considered to be due to a force $\frac{dP}{dx}dx$ acting on them, where P is the partial pressure of the gas formed by the ions. Now $P = \lambda N$, where λ is a universal constant involving the temperature only and equal to P_0/ν_0, where P_0 is the pressure of any gas which contains ν_0 molecules (or a mass of one gramme-molecule) in unit volume. Hence the force acting on each of the Ndx ions and giving them a velocity $\frac{D}{N}\frac{dN}{dx}$ is $\frac{1}{N}\frac{P_0}{\nu_0}\frac{dN}{dx}$, and the ratio of the velocity to the force is $\frac{\nu_0 D}{P_0}$. Comparing this result with that just obtained we see that we must have

$$\frac{u_0}{e} = \frac{\nu_0 D}{P_0} \text{........}(124), \qquad \text{or} \qquad \nu_0 e = P_0 \frac{u_0}{D} \text{........}(125).$$

But P_0, u_0, D can all be measured, so that $\nu_0 e$ can be determined. Townsend found that the value of $\nu_0 e$ was between $2{\cdot}25 \times 10^{14}$ and $3{\cdot}04 \times 10^{14}$, agreeing well with the value of $\nu_0 \epsilon$. Hence he confirmed the conclusion that the charge on a gaseous ion is nearly (or quite) the same in different gases and equal to the charge carried by a univalent electrolytic ion.

The charge on a gaseous ion, then, is that on an electron (or possibly sometimes a small multiple of it), and since ions of both signs are produced in equal numbers by the processes we are considering (or at least the total charge on ions of opposite sign is equal and opposite) it is naturally supposed that the process of ionising a gas consists in splitting off from the atoms of the gas of one (possibly two or three) electrons, leaving the remainder of the atoms positively charged. If no further action took place we should expect to find an immense difference in the properties of the negative and positive ions; the former would be free electrons, the latter charged atoms or molecules more than 1000 times more massive. We have already noticed some difference in the properties of the ions of different signs shown in their powers of acting as condensation nuclei, but the theory of this action is very obscure[1] and throws no light on the constitution of the ions. Evidence as to this constitution is obtained most readily by an examination of the velocity with which the ions move through the gas in an electric field, and we shall now consider this evidence, leaving until we deal with the nature of the ionising rays all hypotheses as to the action by which the initial separation of the electrons from the atoms is effected.

The velocity of the ions.

50. Methods dependent on two principles have been employed for the determination of the velocity of the ions in an electric field. The principle of the first method is to compare the velocity so produced with the velocity of the

---

[1] It can be shown that vapour should condense more easily on a charged than on an uncharged sphere of considerable size, but this action is connected with the decrease of energy consequent on the distribution of the charge on the sphere over a larger area. But of course the charge on an ion cannot be subdivided or distributed over a larger area by the condensation of vapour upon it.

gas as a whole; if a blast of air is directed so as to carry the ions away from the electrode, while the electric intensity drives them to it, the current will be zero, unless the velocity due to the field is greater than that of the blast. In methods based on the second principle the ionised gas is placed in an alternating electric field; if the alternations are so rapid that the electrons have not time to reach the electrode before the field is reversed, the current through the gas will be less than if a steady field of the same magnitude were applied. The results obtained by the alternating field method are much more accurate than those given by the other method, but within the limits of experimental error different methods give the same result. It is often forgotten that the agreement of measurements of some quantity introduced by a theory by very different methods provides very strong evidence in favour of the substantial truth of the theory.

It is found that at all pressures of a gas greater than 2 or 3 mm. of mercury the velocities of the positive and negative ions are proportional to the electric intensity acting on them. In the following table a few characteristic numbers are given; the velocities are those attained in a field of 1 volt per cm., and the gases are under a pressure of one atmosphere and at about 15° C.

TABLE VI.

	u_+	u_-
Hydrogen	6·7 cm./sec.	7·9 cm./sec.
Air	1·40	1·78
CO_2	0·79	0·87
SO_2	0·44	0·41
CCl_4	0·30	0·31

It will be observed that for the lighter gases the velocity of the negative ion is greater, but not very much greater, than that of the positive. The difference between the velocities is not nearly so great as would be expected if the former were a free electron and the latter a molecule. When we inquire into the absolute magnitude of the velocities we again obtain unexpected results. On the assumption that the positive ion is a molecule bearing the charge ϵ its velocity can be calculated exactly as in § 35; all the assumptions made there are as applicable to the molecules of a gas as to the free electrons in a metal, and we may write from (84)

$$u = \frac{X\epsilon t}{2M} \quad \text{.......................}(126),$$

where M is the mass of a molecule, and t the time between two collisions by the same molecule. t is known from observations on the viscosity of gases and $\frac{\epsilon}{M} = \frac{2{\cdot}89 \times 10^{14}}{W}$, where W is the molecular weight of the gas referred to the atom of hydrogen as unity; for a field of 1 volt/cm. $X = \frac{1}{300}$. Hence we have for

Hydrogen $t = 1{\cdot}06 \times 10^{-10}$, $\epsilon/M = 1{\cdot}44 \times 10^{14}$, $u = 25{\cdot}4$,
Oxygen $t = 2{\cdot}45 \times 10^{-10}$, $\epsilon/M = 10^{13}$, $u = 4{\cdot}1$,
Chlorine $t = 1{\cdot}67 \times 10^{-10}$, $\epsilon/M = 4{\cdot}5 \times 10^{12}$, $u = 1{\cdot}3$.

For the lighter gases it is always found that the velocities of ions of both signs are two or three times less than those calculated on the assumption that the ions are molecules of the gas.

Accordingly it was formerly believed that the ions must consist of aggregations of several molecules collected round the charged electron or molecule under the influences of the electric forces due to the charge; the charged particles were supposed to attract neutral molecules, as a charged rod attracts small pieces of paper. But there are difficulties in this view. For the same gas t varies inversely as the density, i.e. the number of molecules in unit volume; hence, if the nature of the ion remains the same at all densities, we should have from (126) $u\rho = \text{const.}$, where ρ is the density. For the positive ion this relation is found to hold when the density of the gas is changed by varying the pressure, and hence it would seem that the nature of the ion is constant. But if the ion were an aggregation of molecules, we should expect the number of molecules in the aggregation to be less at the lower densities, for the charged ion would have less chance of picking up neutral molecules and, consequently, we should expect the product $u\rho$ to increase as ρ decreased. Recent experimental evidence is still more strongly against the view and theory shows that it is unnecessary.

Sutherland has shown that the measurements are really quite consistent with the view that a positive ion is a single molecule bearing the charge ϵ. The presence of the charge on the molecule and the resulting attractions of the ion for neutral molecules and for oppositely charged ions decreases very markedly the rate at which it will move through the gas under a given electric field. The calculation of the effect of these attractions is very complicated and only the result will be quoted here. It is found that the velocity of an ion in an electric field X should be given by

$$\epsilon X = u \,.\, BN\,(a_1 + a_2)^2 \left(\frac{m_1 m_2}{m_1 + m_2}\right)^{\frac{1}{2}} (\alpha T)^{\frac{1}{2}} \left(1 + \frac{C}{T}\right) \ldots(127),$$

where N is the number of molecules (not ions) in 1 c.c. of the gas, a_1, m_1 the radius and mass of a molecule, a_2, m_2 the corresponding quantities for the ion, B and C quantities determined by the forces between an ion and a molecule, and αT the mean energy of a molecule at a temperature T. In the present state of our knowledge of the internal structure of the molecule B and C cannot be calculated directly, but we may obtain some test of the theory in an indirect way.

Since N varies as $1/T$, (127) may be written in the form

$$u_+ = A \,.\, \frac{T^{\frac{1}{2}}}{1 + C/T} \ldots\ldots\ldots\ldots\ldots\ldots\ldots(128),$$

where

$$1/A \propto (a_1 + a_2)^2 \left(\alpha \frac{m_1 m_2}{m_1 + m_2}\right)^{\frac{1}{2}} \ldots\ldots\ldots\ldots(129);$$

accordingly C may be determined from observations of the variation of u with the temperature. Suitable measurements for this purpose are available only for air; if C is put equal to 509·6, we obtain the following comparison of the observed and calculated values of u_+ for different values of T, the value for $T = 399$ being put equal to 1;

T	411	399	383	373	333	285	209	94
u_+ (obs.)	1·03	1·00	·95	·93	·86	·71	·49	·13
u_+ (calc.)	1·04	1·00	·95	·93	·86	·69	·49	·18

The agreement is excellent, except at the lowest temperatures, to which the theory is not strictly applicable.

This test in itself provides no evidence as to the nature of the ion, for the variation of u_+ with the temperature should be the same whatever the relative magnitudes of m_1 and m_2. But it is useful as showing that C/T is not very large compared with 1 and that, therefore, since C is not likely to vary very greatly in different gases, the term $(1 + C/T)$ will not be very different for different gases. Assuming, then, that $(1 + C/T)$ is the same for all gases, and that the positive ion is a single molecule bearing the charge ϵ, so that $m_1 = m_2$, $a_1 = a_2$, we see from (129) that the quantity $m_1^{\frac{1}{2}} a_1^2 u$ should be constant for all gases. Table VII compares this prediction with the facts, for all gases of which the values of both u_+ and a are known. The figures in the last column are certainly not constant, but the differences are not greater than can be easily attributed to the uncertainties in the quantities involved. It should be noted that nearly the same result would be obtained if it were assumed that the mass of an ion were a *constant* multiple of that of a molecule, but such an assumption, viewed in the light of our knowledge of gases, is intrinsically improbable. The matter could only be settled definitely by determining the constant B and comparing the observed and calculated absolute values of u_+. There is no reason to believe that B might not have a value equal to that required by experiment; the work of Sutherland, though it has not established the proposition that the positive ion is a simple charged molecule, has removed all the objections which were urged formerly against that very probable hypothesis.

TABLE VII.

Gas	m_1/m_h	$a_1 \times 10^8$	u	$m_1^{\frac{1}{2}} a_1^2 u / m_h^{\frac{1}{2}} \times 10^{16}$
H_2	2	2·17	6·7	44·5
He	4	1·92	5·1	37·5
O_2	32	2·71	1·33	55·0
CO	28	2·74	1·1	44·0
CO_2	44	2·90	0·79	44·3
N_2O	44	3·33	0·82	59·5
Cl_2	71	3·76	1·0 ?	119 ?

These considerations apply strictly only to the positive ion; on our theory the negative ion must be an electron immediately after its formation. Since later it resembles so greatly the positive ion, we must suppose that, soon after it is formed, it attaches itself to a molecule and thus becomes identical, except in the sign of its charge, with the positive ion. However, before it so attaches itself it will have had a short 'free life' during which it will acquire a much greater speed than it would have done if it had been attached to a molecule; this short period of freedom is probably the cause of the slightly greater average velocity of the negative ion.

After the electron has attached itself to a molecule the velocity of the ion thus formed will be, like that of the positive ion, inversely proportional to the density of the gas. But if the density of the gas is decreased, the distance through which the electron travels before it meets a molecule will be increased; it will spend a greater proportion of the time taken in travelling between the electrodes in a free state, and its average velocity will be higher. Accordingly we should expect the product $u_-\rho$ to increase as ρ is decreased and we find that in air it does increase very rapidly at pressures below 10 cm. And there is another way in which the velocity of the negative ion may be increased. The figures given in Table VI for air and hydrogen refer to those gases when they are 'dried' by sulphuric acid; if a little water vapour is added the velocity of the negative ion may become actually smaller than that of the positive ion; while if the greatest care is taken to remove all water u_- may be increased 50 times. Again, in argon and in nitrogen, u_- and u_+ are nearly equal unless the greatest care is taken to remove all traces of oxygen; when the gases are perfectly pure the velocity of the negative ion is 100 times greater than when they are slightly contaminated with oxygen or chlorine. These observations doubtless show that the electrons liberated when the gas is ionised attach themselves much more readily to some molecules than to others, and it is to be remarked that those atoms to which the electrons attach themselves most readily are also those to which we should attribute on electrochemical evidence the greatest affinity for negative electricity. The velocity of the positively charged ion is almost unchanged by the presence of these impurities, and hence it becomes somewhat improbable that it can also attract neutral molecules so as to form an aggregation. Our knowledge of the properties of the negative ion confirm our conclusion that the positive ion is a single molecule.

Attempts have been made from time to time to obtain information about the nature of the ions from the study of the conductivity of a mixture of gases. If there are two gases present, we should expect at first sight to find two sets of ions moving with different velocities, and from the magnitude of these velocities in different mixtures we might discover whether the ions collect round them neutral molecules. Extensive experiments have been carried out in this direction by Wellisch, who was unable to detect any sign of the presence of more than one kind of ion in any mixture; even when he introduced into hydrogen the ions formed in a very heavy gas (such as methyl iodide vapour), the velocity of the ions in the hydrogen was practically the same as that of hydrogen ions.

These results appear very remarkable, but a reference to (127) will show that the difference between the velocity in hydrogen of hydrogen ions (u_1) and that of the ions of some very heavy gas (u_2) should not be very great. Leaving the term $(1 + C/T)$ out of account, we have

$$\frac{u_2}{u_1} = \frac{(a_1 + a_2)^2}{4a_1^2} \cdot \left(\frac{2m_2}{m_1 + m_2}\right)^{\frac{1}{2}} \quad \text{..............(130).}$$

Even if m_2 is infinitely large compared with m_1, the value of the second factor will be only $\sqrt{2}$; and since the radii of different molecules do not differ very greatly, the first factor will not be very great. It is not to be expected that the ratio $\frac{u_2}{u_1}$ should be more than 3 or 4, and since u for methyl iodide ions in methyl iodide is only 1/30 of that of hydrogen ions in hydrogen, u for methyl iodide ions in hydrogen should be much nearer the latter than the former value. It does not appear necessary to alter our views of the nature of the ions on account of such experiments.

Information as to the nature of the ions might also be derived from a study of the quantities D and a (p. 89), but the experimental difficulties would be greater. Moreover, these quantities only measure the velocities of the ions under forces other than those of an external electric field and are necessarily connected with u by relations the form of which does not depend on assumptions concerning the mass of the ion. The conclusions derived from such study as to the nature of the ion must be the same as those derived from the investigation of u.

Ionisation by collision.

51. It was stated in § 47 that after the electric intensity acting on an ionised gas had attained a certain value, a further increase produced no change in the current through the gas. This statement is not accurately true. If the gas is air at atmospheric pressure, the electric intensity may perhaps be increased to 100 times that required to produce saturation without any consequent change in the current, but if it is increased indefinitely a stage is reached at last at which the current increases very rapidly with the electric intensity; ultimately a spark passes through the gas. The range between the attainment of saturation and the rapid increase of the current decreases with the pressure of the gas; at low pressures, indeed, saturation cannot be attained at all, because the increase leading to the spark discharge begins before the field is strong enough to make the effects of recombination negligible.

A theory of this phenomenon was put forward almost simultaneously by Townsend and J. J. Thomson. The ionisation of the molecule, involving the separation of opposite charges, requires the expenditure of energy. In the cases we have considered hitherto this energy is derived from the rays which ionise the gas; but since the electric field necessary to produce the spark discharge is practically independent of the action of any external ionising agent, it is clear that, if we are to attribute the passage of the current to the production of ions, we must seek the source of energy required for the production of those ions in the electric field and not in the rays. An ion moving between two collisions freely in an electric field of intensity X acquires an amount of energy $X\epsilon\lambda$ where λ is the free path of the ion[1]. If X or λ is increased sufficiently the energy of an ion when it makes a collision may be made as great as we please. According to the theory of Townsend and Thomson, if an ion when it makes a collision has

---

[1] In the experiments with which we are now concerned $X\epsilon\lambda$ is always *large* compared with the energy of thermal agitation. On p. 59 it was assumed that $X\epsilon\lambda$ was *small* compared with this energy; consequently the results obtained are different.

acquired by moving through the electric field a certain amount of energy (which must be not less than that required to ionise a molecule), it will ionise the molecule with which it collides. When the stage is reached at which $X\epsilon\lambda$ is equal to this energy, each ion as it is produced will produce in its turn more ions, and the current through the gas will increase very rapidly.

Of course the free paths of all the ions are not the same. Let λ be now the mean free path and $q = 1/\lambda$ the average number of collisions made by an ion in travelling 1 cm. Consider a large number of ions, n_0, starting from the origin and travelling under the action of the field along the axis of x. Let n be the number of these ions which arrive at the point x without having made a collision; then since the number of ions colliding in a layer dx is qdx, we have

$$-\frac{dn}{dx} = q \quad \text{........................(131)},$$

or

$$n = n_0 e^{-qx} \quad \text{........................(132)}.$$

Of the n ions which arrive at the point x without having previously made a collision the number which collide in that layer is $nqdx$. Hence, if n_0 ions start from any layer, the number which will travel freely a distance x and then collide in a layer dx is $n_0 q e^{-qx} dx$. The chance that an ion will travel freely a certain distance and then collide is independent of the distance which it has travelled previously, so that the previous sentence will be true if we substitute 'pass through' for 'start from.' Now the collision of an ion results in the production of a pair of ions if, and only if, x the distance which it has travelled freely previously to the collision is equal to or greater than b, where $X\epsilon b = \epsilon V_0$, the energy necessary to ionise by collision. Hence the number of pairs of ions produced by collision in any layer is $n_0 q e^{-qb} dx$, where n_0 is the number of ions passing through a layer distant b from the first layer in the direction from which the ions are coming. We shall write

$$a = qe^{-qb} \quad \text{........................(133)}.$$

Now consider a mass of gas contained between two electrodes a distance l apart and perpendicular to the axis of x (the positive electrode being at the origin) and ionised by rays which produce N pairs of ions in unit time in each layer dx. Let $f_1(x)$ be the number of positive ions crossing the layer at x in unit time from the positive to the negative side and let $f_2(x)$ be the number of negative ions crossing the same layer in the opposite direction in the same time; let the suffixes 1 and 2 refer in all cases to positive and negative ions respectively. Then from the result of the previous paragraph, the number of pairs of ions produced in unit time in the layer dx is

$$\{N + a_1 f_1(x - b_1) + a_2 f_2(x + b_2)\}\, dx.$$

But since the difference between the number of ions entering a layer and that leaving it must be the number produced in that layer, and since ions of opposite sign are always produced in equal numbers, we must have

$$\{N + a_1 f_1(x - b_1) + a_2 f_2(x + b_2)\}\, dx = \frac{df_1(x)}{dx}\, dx = -\frac{df_2(x)}{dx}\, dx \quad \text{......(134).}$$

Since no ions enter the gas from behind the electrodes we must have

$$f_1(0) = f_2(l) = 0 \quad \text{......................(135),}$$

and since the current i leaving an electrode must be equal to the number of ions arriving at it in unit time multiplied by the charge on each ion, we must have

$$f_1(l) = -f_2(0) = \frac{i}{\epsilon} \quad \text{......................(136).}$$

A solution of these equations is

$$\frac{i}{\epsilon} = N \cdot \frac{e^{cl} - 1}{a_1 - a_2 e^{cl}} \quad \text{......................(137),}$$

where

$$c = a_1 e^{-cb_1} - a_2 e^{cb_2} \quad \text{..................(138).}$$

(137) states a relation between i, l, N (which can be measured) and the quantities a_1, a_2, b_1, b_2 which are at present unknown. By studying this relation, the values of the unknowns can be ascertained. Elaborate experiments of this kind have been carried out by Townsend and his pupils and confirm the theory completely. The values of a and b depend upon those of X, q and V_0; q is proportional to the pressure of the gas and hence by finding the value

of a and b for different values of the field and the pressure the magnitude of V_0 can be ascertained. V_0 is the potential difference through which an ion must fall to acquire the energy necessary to ionise by collision.

It is found that for all gases the value of V_0 for the negative ions is between 8 and 20 volts, but V_0 for the positive ions is very much greater; in the matter of ionisation by collision there appears to be a very great difference between ions of opposite sign even in those gases in which the velocities of the two kinds of ions are practically equal. This fact is easily explained. It is proved in any treatise on dynamics that if two particles are allowed to approach each other under the influence of their mutual attraction, they will ultimately describe closed orbits round each other if their initial relative velocity is below a certain limit, but will fly apart again if that velocity is above this limit. When the velocity of the ions is being measured the electric field and, therefore, the velocity of the electron shortly after it has been ejected from the ionised molecule are very much less than when ionisation by collision is taking place; an attraction between an electron and a molecule which results in the first case in a permanent attachment between them may produce nothing but a deflection of the motion of the electron in the second. There is nothing contrary to our theory in the hypothesis that the electron attaches itself to a molecule when the field is insufficient to produce a saturation current, but does not attach itself when the field is great enough to produce ionisation by collision. We may suppose, therefore, that in the experiments which we are considering here the negative ion is always a free electron, while the positive ion is a charged molecule; the ions will be very different in their nature and it is not surprising that they have different powers of ionising by collision. But it must be remembered that the positive ion actually requires more energy than the negative ion before it can ionise by collision, and we must conclude that the energy which a system must possess in order to produce ionisation is not necessarily merely that required to separate the electron from the molecule but depends also upon the nature of the system producing the ionisation.

The spark discharge.

52. It will be seen from (137) that when $a_1 = a_2 e^{cl}$ the ratio i/N becomes infinite. In a given gas e^{cl} increases from zero with X/p; if we increase the field or decrease the pressure sufficiently this stage will be attained. Of course the current will not really become infinite; in our calculations we have neglected the effects of recombination, which, though they are very small under the conditions just considered, will set a finite limit to the current attainable[1]. But the conclusion is justified that when this stage is reached the current will have a large finite value however small is N, the ionisation produced by the external action. It is known that no gas is wholly free from all ionising agents, so that we should expect that it would be possible to maintain a large current through any gas by making the value of X/p sufficiently great. Of course this expectation is realised; the current which flows in the spark discharge can be maintained in a gas by an electric field which is practically independent of the action of an external ionising agent. Accordingly we may identify the field necessary to produce a spark with that necessary to make $e^{cl} = a_1/a_2$; the measured values of this field agree well with those predicted from the values of a_1, a_2 and c deduced from Townsend's experiments. Stark has introduced a useful nomenclature, according to which the spark discharge in any of its forms is called the 'unsupported' (selbstständig) current, and distinguished from the 'supported' (unselbstständig) current which is determined by the presence of an ionising agent and alone can exist when the electric field is small.

The unsupported discharge has been the subject of an immense amount of investigation both experimental and theoretical. When the pressure of the gas through which the discharge passes is not greater than a few millimetres of mercury it is seen that the luminosity is not uniform, but distributed in layers which are arranged differently round the two electrodes. But with the description and the explanation of these phenomena we are not immediately concerned; the study of them adds hardly anything to our knowledge of the nature of the ions carrying

---

[1] A finite limit to the current is also set by the fact that when a current flows through the gas the electric intensity in the gas produced by a given P.D. between the electrodes changes.

the current and the method in which they are produced. They are merely more complicated forms than we have considered hitherto of currents carried by ions produced by ionisation by collision, and so far as the various observations have been explained the explanations are based upon principles already enunciated. The only new point which we have to note is that it appears that the ions already present in the gas can produce fresh ions by colliding not only with the molecules of the gas but also with those of the solid or liquid electrodes. At the lower pressures most of the negative ions are produced by the impact of the positive ions on the kathode and not by their collisions with the molecules of the gas.

Kathode Rays.

53. It is only when the mean free path of the ions becomes comparable with the dimensions of the vessel in which they are contained that phenomena appear which require special consideration here.

If the pressure in a glass tube of the nature usually employed in investigations of the luminous discharge is reduced below ·001 mm. a green phosphorescence of the glass in certain regions may be noted. It was shown by Hittorf in 1869 that this phosphorescence is produced by the impact on the glass of 'rays' proceeding from the kathode. The term 'ray' denoted an influence travelling in straight lines, so that any region between which and the origin of the rays lies a suitable obstacle is shielded from the influence; accordingly Hittorf's proof of the existence of these kathode rays consisted in a demonstration that if a solid body were placed near the kathode a shadow of the obstacle was thrown on the phosphorescence in that part of the glass which lay on straight lines passing through points on the kathode and on the obstacle. Since it had been shown previously that the position of the phosphorescence was changed if the tube were placed in a magnetic field, Hittorf concluded that the rays were deflected from their straight course in such a field.

A long controversy raged for many years round the nature of these rays. Goldstein maintained that they were aethereal waves essentially of the same nature as light: Crookes held that they were streams of particles electrically charged. The former relied largely on an observation of Hertz, who showed that solid bodies were not absolutely opaque to the rays, which could pass through thin layers of gold or aluminium. The passage of particles even as small as atoms—and nobody had the temerity to suggest that the kathode particles might be smaller than atoms—through a solid body presented great difficulties, while the known properties of light showed that aethereal waves could traverse easily certain solid bodies. However there appeared to be no relation between the transparency of a body for light and that for kathode rays: gold leaf was more transparent to the latter than mica.

On the other hand the deflection of the rays in a magnetic field is an immediate consequence of the view that they consisted of charged particles, while the upholders of the opposite contention could only suggest vaguely that the deflection was in some way analogous to the unexplained rotation of the plane of polarisation of light in a magnetic field. Perhaps the chief reason for the support which the wave theory received is to be found in an unwillingness to accept an alternative resembling so closely the corpuscular theory of light which had proved fallacious.

Crookes sought to maintain his view by direct experiment. He placed in the path of the rays (which can be seen at low pressures as a faint streak of light stretching from kathode to anode) a little mill-wheel with mica vanes. When the current passed the wheel was rotated in the direction in which it would be rotated by the impact of massive particles travelling from kathode to anode. He also attempted to show that if two parallel beams of the rays were sent through the tube each was deflected by the presence of the other, as they might be if the rays consisted of particles carrying electrostatic charges. However, it is known now that neither of Crookes' experiments can be interpreted in the sense given by their author.

As regards the nature of the particles, it appears to have been held generally that they were charged atoms, but Crookes, by one of those flights of intuition which are only remembered when they prove successful, maintained that they consisted of matter in a 'fourth state'—neither solid, liquid, nor gaseous.

Many scientific controversies have ended in the comforting conclusion that both sides were right and both wrong. But in this case one of the alternatives has been definitely accepted and the other rejected. An experiment by Perrin was deemed conclusive, in which he demonstrated that when the kathode rays fell upon a body it received a charge[1]. This charge was found to be negative.

According to the theory that the kathode rays consist of charged particles, it is necessary to suppose that the particles possess a finite mass in order to explain the nature of their deflection in a magnetic field. It appeared almost impossible at the time to measure the mass of one of the particles directly, but attempts to measure the ratio of the charge to the mass in accordance with the principles of § 17 were made between 1890 and 1897 by Schuster and Wiechert. But these attempts were not very successful until J. J. Thomson devised in 1897 the method of the 'crossed' electric and magnetic fields (the 'second method' of p. 23). In his experiments a small bundle of the rays issuing from the kathode in a gas at a very low pressure was isolated by means of suitable screens and caused to excite a spot of phosphorescence on a screen at the other end of the vessel; during part of its path it passed through the crossed fields the strengths of which were adjusted so that the spot was undeflected. The velocity of the rays was thus ascertained; the value of e/m could then be found easily by removing the electric field and observing the deflection of the spot when the magnetic field alone acted.

By this method Thomson deduced values for e/m between $2{\cdot}0$ and $2{\cdot}7 \times 10^{17}$ and for the velocity values between $0{\cdot}3$ and $1{\cdot}2 \times 10^{10}$ cm./sec. It is now known that the former values were too low and the values of the velocity (which depends on the P.D. between the electrodes) probably too high, but the very important conclusion was reached that whatever the nature of the electrodes from which the rays proceeded and whatever the gas through which they passed, the value of e/m was always nearly the same; it certainly never differed in the ratio of 2 to 1 and there was no evidence that it differed at all. Moreover the value was from 700

---

[1] If all our present information as to 'secondary radiation' (see Chap. XI) had been available, Perrin's experiment might have seemed less convincing. But I fear that few of the 'crucial experiments' in the history of physics would stand a searching examination in the light of subsequent knowledge.

to 900 times greater than that for a hydrogen atom bearing the charge on a univalent ion. Unless the whole theory of kathode rays was wrong, one of two alternatives had to be accepted; either the carriers were atoms or aggregations of atoms carrying charges which were proportional to their mass and vastly greater than those carried by electrolytic ions, or the carriers were particles very much smaller than atoms, common constituents of all atoms and carrying the charge which was already known to be a universal constant. Thomson, of course, adopted the second alternative; he asserted that he had proved the existence of bodies which were common constituents of all atoms; and to these bodies he gave the name 'corpuscles.' He thus produced the first piece of experimental evidence in favour of the existence of the 'electrons' which Lorentz had introduced to explain the optical properties of dielectrics[1].

Thomson's measurements have been repeated by many investigators under much better experimental conditions and some of the most accurate measurements of ϵ/m_0 have been made on the electrons constituting the kathode rays. Of course these rays are now regarded as electrons, set free when the molecules of the kathode are ionised by the impact upon them of the positive ions; at such very low pressure these electrons have a free path so great that they can travel through the whole vessel without colliding with a gas molecule and so being deflected from a straight path. Their kinetic energy is derived from the field causing the discharge and is equal to the potential difference through which they have fallen[2]. Any further account of the properties of the kathode rays is postponed until Chap. IX.

---

[1] The name 'corpuscles' has been generally abandoned in favour of 'electrons,' but Thomson himself naturally retains it in his writings. It should be noted that it was not proved that these corpuscles were identical with Lorentz's electrons until some time after their existence had been demonstrated (see p. 148).

[2] Perhaps the reader may think it strange that, if the rays acquire their speed by falling through the electric field, they do not all strike the anode. The explanation of this apparent anomaly is to be found in a study of the electric field between the electrodes. When the discharge is passing, this field is by no means uniform; it is very intense quite close to the kathode in what is known as the 'Crookes dark space.' The field in the region between the boundary of this space and the anode is comparatively small. The kathode rays acquire their energy by falling through the dark space and are almost unaffected by the field in the other

Positive Rays.

54. The mean free path of the positive ions is not very much less than that of the electrons[1], and it is therefore to be expected that at low pressures the positive ions also should travel from one end of the vessel to the other and form positively charged rays moving in the opposite direction to the kathode rays. And such rays have been detected. Goldstein first observed that if the kathode were perforated there appeared in the region behind it rays which, like the kathode rays, produced luminescent and chemical effects on substances which they struck. He was unable to deflect the rays in a magnetic field, but Wien produced such a deflection and showed that it was in the direction opposite to that in which kathode rays, travelling in the same path, would be deflected; he concluded that the rays were positively charged, as was to be expected from their motion towards the kathode. Goldstein named the rays, from the method of their production, 'canal rays,' but subsequent investigation, in which Wien and J. J. Thomson have been especially active, has shown that the rays are very complex and the more general term 'positive rays' is now generally applied to them[2].

The method which has been employed to discover the nature of the positive rays is similar in principle to that employed with kathode rays, but since the phenomena are much more complicated than in the first case some modification of it has been necessary. The particles in a bundle of kathode rays all have the same value of e/m and all nearly the same speed; consequently they are all deflected to the same extent and the application of the fields merely moves the spot on the screen from one point to another. But the positive rays are not homogeneous and when the fields are applied they are differently deflected; the spot on the screen or the photographic plate which is used to detect them is spread out into a band. It is convenient to consider what will be the form of the band in two cases; firstly when all the particles have the same velocity but different values of e/m,

---

parts of the tube. The positive rays, however (see below), which are travelling in the other direction pass through the region occupied by the kathode, because they also acquire most of their energy close to the kathode.

[1] According to the elementary theory of gases the free path of an infinitesimally small particle should be four times that of a molecule of the gas.

[2] It will be seen however that all the 'positive rays' are not positively charged.

and secondly when they all have the same value of e/m but different velocities.

Let the particles travel with a velocity v along the axis of z, in an electric field of intensity X and a magnetic field of intensity H parallel to x. If f and g are the accelerations of the particle parallel to x and y, $f = X\frac{e}{m}$, $g = \frac{1}{c^2}H\frac{e}{m}v$. Let the deflection of the particle be so small that v may be considered to be always parallel to z, and let the distance it travels in the fields be l. Then since the time it is in the field is l/v, its coordinates when it emerges will be

$$x = \frac{1}{2}X\frac{e}{m}\frac{l^2}{v^2}, \quad y = \frac{1}{2}\frac{H}{c^2}\frac{e}{m}v\frac{l^2}{v^2}.$$

Hence

$$y/x = Av \quad \text{.........................(139)},$$

$$y^2/x = B\,e/m \quad \text{.....................(140)},$$

where A and B are constants independent of the properties of the rays. Consequently a bundle of rays which have the same value of v but different values of e/m will appear on the screen as a straight line; while one for which e/m is constant and v variable will appear as a parabola.

The appearance of the screen when struck by a bundle of the positive rays issuing through a fine hole in the kathode and passing through the parallel fields is very complicated and varies greatly with the pressure of the gas; but the figure traced by the rays has been analysed by Thomson into three simple types characteristic of three kinds of rays, all of which are sometimes present.

(1) There is always a spot in the position $x = y = 0$. This spot corresponds to undeflected rays which have no charge; no direct deduction as to their nature can be made.

(2) There are straight lines passing through the point $x = y = 0$ and extending on the side corresponding to a negative charge on the rays as well as on that corresponding to a positive charge. These lines denote the presence of rays which have a constant v but a variable e/m. Some of the particles must be negatively charged. It is remarkable that for these rays the

value of v, though it differs for different lines, appears to be independent of the potential difference between the anode and kathode.

(3) There are parabolic arcs corresponding to rays with constant e/m but variable v. The maximum value of v for these rays is determined by the fall of potential in the discharge tube; the range of values of v is not large.

The nature of the third group of rays is obvious; they consist of positive ions produced in the gas somewhere between the anode and kathode (probably quite near the kathode) which have acquired a velocity determined by the potential difference between the place where they were produced and the kathode. The rays are the positive analogues of the kathode rays; they appear to differ from the latter in having variable velocities only because they are produced between the electrodes and not at one of the electrodes, and so do not always fall through the full potential difference of the discharge.

The nature of the first and second groups, which are undoubtedly similar, is more obscure. The apparent variability of e/m is certainly to be attributed to a variation in the charge of the particles rather than in their mass. Some of the positive rays of the third class are travelling through the space behind the kathode with a velocity sufficient to enable them to ionise by collision, so that in this space there will be a certain number of free electrons. The more slowly moving positive rays may be able to combine with these electrons and so form neutral systems such as appear in the rays of the first class; and it is possible that the system so neutralised may be once more ionised by collision with an electron and so become positive, or it may pick up yet another electron and become negatively charged. If any of these changes in the sign of the particle take place when it has traversed part but not all of the fields, its deflection will be that corresponding to a particle which has a value for e/m intermediate between zero and the original value.

So far all is clear, but it has yet to be explained why the velocity of the rays of class 2 appears to be independent of the potential difference in the discharge; at first sight it would appear that these rays, being merely rays of class 3 which have undergone changes of sign, should have the same speed as rays of class 3. In his earlier experiments Thomson was only able to obtain rays of class 2 for which the maximum value of e/m was 3×10^{14}, corresponding to an atom of hydrogen bearing the charge ϵ. He suggested accordingly that such a particle was a 'positive electron,' a common constituent of all atoms, and that the process of ionisation consisted in splitting off from a molecule a 'neutral doublet,' formed of a positive electron with a negative electron; the doublet subsequently dissociated emitting the positively charged part with the velocity characteristic of rays of class 2. There were many difficulties in such a view, which was definitely disproved when rays of class 2 were detected which had different velocities and maximum values of e/m characteristic of atoms other than hydrogen which were present in the gas[1].

Thomson has therefore suggested another and very ingenious explanation of the constancy of v. In order that a particle should appear as a positive ray of class 2, it must have undergone two changes. Firstly it must have been neutralised by combination with an electron and secondly it must have been ionised again by collision with an electron[2]. If it is moving too fast, it will not be able to combine with an electron (see p. 105); while if it is moving too slow, it will not be split up again by collision with an electron, because the kinetic energy of the electron relative to the neutral system will be less than that required to produce ionisation by collision. The velocity of rays of class 2 must lie between the limits thus determined, and if the limits are close the velocity of the rays in the class would be almost constant; the limits would vary with the nature of the original positively charged atom, but not with the potential in the discharge tube. An explanation of the fact that rays of class 2 are formed more easily in hydrogen than in any other gas is found in the small mass of the hydrogen atom, which makes it acquire more readily a velocity relative to an electron great enough for ionisation by collision to occur.

---

[1] It is quite impossible to avoid the presence of hydrogen as an impurity in a discharge tube such as Thomson used.

[2] It must be assumed that *neutralisation* never occurs in the region affected by the fields; otherwise there would be no necessity for the second change. There is some evidence for this assumption.

The explanation is attractive, but it does not seem capable of explaining the existence of a lower limit to the velocity of the *negatively* charged rays; and it must be noted also that with some gases negatively charged rays of class 3 appear which seem to have velocities at least as great as those of class 2. The problems connected with rays of class 2 (and with some other less important phenomena of which no mention has been made) have not been fully solved.

Some of the most interesting results of Thomson's work concern the values of e/m which he finds for various rays. It appears that the particles are by no means always single atoms or molecules of the gases in the tube bearing a single charge ϵ, such as seem to form the positive and negative ions at high pressures. But since only the ratio e/m can be determined and not e or m separately, there is some latitude possible in the interpretation of his measurements, for the results of which the reader should refer to his original papers. These results, so far as they throw any light upon the problems with which we deal, will be discussed in a later chapter.

In addition to the positive rays just considered there are two other classes which deserve a brief mention.

Firstly, there are the 'retrograde rays,' which are positively charged rays of the same nature as the positive rays travelling in the same direction as the kathode rays. They probably consist of particles which acquired their velocity when they were negatively charged in the neighbourhood of the kathode and have subsequently lost electrons after they emerged from the intense field in that region.

Secondly, there are the 'anode rays,' discovered by Gehrcke and Reichenheim, which consist of positively charged atoms derived from the anode when the kathode rays are incident upon it. They are probably similar to Richardson's thermions (see p. 84) and are liberated by the intense local heating produced by the impact of the kathode rays.

These classes of rays have no special importance for the problems which are our chief concern.

REFERENCES.

The standard treatise on all the matters considered in this chapter is J. J. Thomson's *Conduction of Electricity through Gases*, in which an account of all important experiments before 1906 will be found. There is also a full account of the theory of the spark discharge which here has been dismissed so briefly.

An excellent summary of recent work on the velocity of the ions, both experimental and theoretical, is given by Franck in the *Jahrbuch der Radioaktivität und Elektronik*, 9. 235, 1912. The paper of Sutherland to which reference has been made will be found in *Phil. Mag.* September 1909, p. 341 ; in that paper the necessary references to the author's previous work will be found. Sutherland's papers are very difficult and I do not pretend to understand them sufficiently to offer any criticism of his results, but apparently those results have not been disputed. In most respects his conclusions agree with those of other writers.

Townsend has published his work in a convenient form in a small book entitled *The Ionisation of Gases by Collision* (Constable, London, 1910). A slight correction to his mathematical theory is made in *Phil. Mag.* March 1912, p. 385.

Summaries of the determinations of ϵ/m_0 and of ϵ are given by Seeliger and Pohl respectively in the *Jahrbuch der Radioaktivität*, 9. 28, 1912, and 8. 406, 1911.

The most important of Thomson's papers on positive rays are *Phil. Mag.* December 1909, p. 821, October 1910, p. 752 and February 1911, p. 225. Wien's work is to be found in *Ann. d. Phys.* 27. 1025, 1908; 30. 349, 1909; 33. 871, 1910; 39. 519, 1912.

CHAPTER V

MAGNETIC SUSCEPTIBILITY

Induced magnetisation.

55. WE have now considered in the light of the electronic theory the properties of material bodies which are connected with the constants K and σ; it remains to consider those connected with μ.

It will be well to recapitulate briefly a few of the chief results of the theory of magnetism. If any body is brought into a vacuous space between the poles of a magnet, the distribution of the magnetic intensity outside that body will be changed by the developement on the body of an induced magnetisation. The magnetic intensity inside the body cannot be measured without making a hole. But if a hole is made in the body induced magnetisation appears on the walls of the hole, and accordingly the magnetic intensity in the hole will depend upon its form. Hence the magnetic intensity in the body cannot be defined uniquely unless the form of the hole is specified. Two specifications are adopted. If the hole is a narrow cylinder such that the magnetic intensity in it is everywhere parallel to its axis, that magnetic intensity is termed the 'magnetic intensity in the body' and is denoted by H; if it is an infinitely narrow slit the sides of which are perpendicular to the magnetic intensity within it, that magnetic intensity is termed the 'magnetic induction in the body' and denoted by B. The ratio B/H is termed the permeability of the body and is denoted by μ. If I is the magnetic moment per unit volume of the magnet into which the body is converted by induction, then $k = I/H$ is called the susceptibility of the body; I is called the intensity of magnetisation. k is counted positive if the induced magnetisation is such that just outside the body the magnetic intensity due to it is in the same direction as that of the inducing field, that is, if the north pole of the induced magnet faces the south pole of the inducing magnet. From the definitions of μ and k it is easy to show that $\mu = 1 + 4\pi kc^2$.[1]

Para- and dia-magnetism.

56. In respect of the property of induced magnetisation all bodies may be divided into three classes, diamagnetic, paramagnetic and ferromagnetic. In the first class the value of k is negative, in the second it is positive, in the third k is positive, but is determined by the past history of the substance as well as by its actual condition. The vast majority of substances are diamagnetic; of the rest paramagnetic substances are almost all gases or liquids, while ferromagnetic substances are always solid. In diamagnetic substances k appears to be independent of H, while in most, if not all, paramagnetic substances k decreases as H increases so that a maximum magnetic moment is reached, when the body is said to be saturated. k for diamagnetic substances (except bismuth) is independent of the temperature; for paramagnetic substances k (for small values of H) is inversely proportional to the absolute temperature over a considerable range.

In addition to these qualitative differences between the classes there are quantitative differences. For any given substance k appears to be proportional to the density. If substances of equal density are compared, it is found that k for diamagnetic substances is generally very much less than that for paramagnetic, and that the range of variation of k is much less in the former than in the latter class.

These differences between the diamagnetic and paramagnetic susceptibilities naturally suggest that they represent the properties of entirely different mechanisms. It is generally believed that the mechanism which gives rise to diamagnetism is common to all substances, so that all substances would appear diamagnetic, if it were not that some of them contain an additional mechanism giving rise to paramagnetic effects which mask the much less intense diamagnetic effects.

---

[1] Unit magnetic pole is here defined by means of the equation $c^2 \frac{mm'}{r^2} =$ force between poles.

57. The possibility of developing an electronic theory of induced magnetism arises, of course, from the fact that in a region occupied only by current circuits and non-magnetic bodies[1] the magnetic intensity, as defined from the theory of magnetism by means of the force on a magnetic pole, is everywhere proportional to the quantity H, defined by (3); the factor of proportionality is 1 if unit pole is defined as in § 55. Consequently the magnetic field due to any distribution of magnetism, either permanent or induced, might be reproduced by the substitution for that magnetism of suitable current circuits; it can be shown that the current circuits necessary for this purpose would lie wholly within the region occupied by the magnetic bodies. The magnetic properties of substances can be explained on the view that they contain certain current circuits the magnitude or position of which may be changed by an external magnetic field. The problem before us is to frame hypotheses about the conditions of the electrons in substances subjected to a magnetic field from which suitable values for currents, produced by the motion of those electrons and giving rise to the phenomena of induced magnetism, may be deduced.

The older theories.

The hypotheses that are adopted are merely modifications of those by which Weber, Ampère, and Maxwell explained induced magnetisation. It was imagined that the molecules of all substances contained perfectly conducting current circuits. In paramagnetic bodies currents were flowing in these circuits, so that the molecule behaved as a magnet and under the influence of the external field tended to turn so as to place its axis along the direction of the field; this tendency was resisted by forces of an elastic nature, so that the magnetic moment of the substance increased with an increase in the field till all the magnetic axes coincided completely with the direction of the field. In other bodies the molecules had no finite magnetic moment, either because no current was flowing in the circuit, or because the molecule contained several circuits the effects of which at external points neutralised each other. If the magnetic field acting on such

---

[1] No bodies are wholly non-magnetic, but such bodies as air have so small a susceptibility that it may be neglected for all but the most delicate experiments.

a body were changed, currents would be induced in the conducting circuits, and, since the currents had no resistance, the currents would continue to flow until they were neutralised by an equal and opposite change in the magnetic field. In accordance with Weber's Law these currents would be in such a direction that the magnetic intensity due to them would be in the opposite direction to that of the inducing field, so that the body would appear diamagnetic.

Before proceeding to translate this hypothesis into the terms of the electronic theory one point connected with equation (6) should be noted. The value of H in that equation is defined by (3); accordingly if in any region to which it is applied there are magnetic substances the values of the 'current' in (3) must include the molecular currents in the magnetic substance as well as the currents giving rise to the inducing field. The question now arises as to how the value for H thus obtained will be related to the magnetic induction or the magnetic intensity in a magnetic substance if the surface over which H is integrated in (6) lies in that substance.

The molecular currents occupy only an infinitesimal fraction of the whole volume 'filled' by the magnetic substance. Accordingly any surface drawn through the substance will not cut any of those currents, but will lie almost wholly in vacuous regions containing no magnetism. Such a surface will then be similar to an infinitely narrow slit in the magnetic substance and the magnetic intensity perpendicular to it will be simply the quantity which has been defined as the magnetic induction in the substance. We see then that, if the molecular currents are taken into consideration in deducing the value of H from (3), (6) states the familiar relation that the E.M.F. round any circuit is proportional to the rate of decrease of the magnetic induction through it. With the units adopted here the factor of proportionality is $1/c^2$.[1]

---

[1] It is stated in some text-books (for example the first edition of this book) that with the electrostatic system of units the value of μ for a substance in which k is 0 is $1/c^2$. This statement is incorrect; if μ is defined as it is defined here (and this is the usual definition) it is clear that in all systems of units μ must be 1 for all non-magnetic substances; the factor $1/c^2$ must be

58. According to the electronic theory, which has been developed chiefly by Langevin, the molecular circuits of the older writers consist of electrons revolving in closed orbits within the molecule. In a diamagnetic substance each molecule contains many such orbits, which neutralise each other's actions at external points, owing to the fact that the direction of revolution is different in different orbits; while in a paramagnetic substance the neutralisation is not perfect, so that the molecule as a whole has a finite magnetic moment. The diamagnetic susceptibility is due to the changes in the nature of the orbits produced by the magnetic field; the paramagnetic susceptibility is due to the orientation of those molecules which have a finite magnetic moment.

The electronic theory.

For the sake of simplicity let us suppose that the orbits are circular and of radius r and that their planes are all perpendicular to the magnetic intensity H of the inducing field. Let ω be the angular velocity of an electron in its orbit about the direction of H and M the magnetic moment of the magnet to which the revolving electron is equivalent; M will be counted positive when the direction of H is from the south to the north pole of the equivalent magnet; it would be positive if the electron were positively charged and revolved round the direction of H in the positive direction.

Let us first consider the value of M for one of the electronic circuits. It is a consequence of our electromagnetic theory that if a current circuit is equivalent to a certain magnet in respect of one magnetic action, it is equivalent to that magnet in respect of all magnetic actions; accordingly we may choose for the estimation of M any magnetic action which is convenient. The most convenient action for our purpose is the couple exerted on the

introduced by (6). But sometimes μ is defined by means of (6) which is then written in the form

$$\int D ds = -\frac{1}{4\pi}\frac{d}{dt}\int \mu H dS.$$

If this definition of μ is adopted and we continue to call the quantity μH the magnetic induction, we must say that the magnetic induction is not equal to the magnetic intensity in a narrow slit but is $1/c^2$ times that magnetic intensity. Of course if electromagnetic units are adopted, the factor $1/c^2$ disappears in both these cases.

circuit by a magnetic field which lies in the plane of the circuit; if the moment of the equivalent magnet is M this couple is MH.

By (5) if θ is the angle between the direction of H and the radius passing through the electron, the force acting on the electron is $\frac{H\epsilon\omega r\cos\theta}{c^2}$, and the moment of this force about an axis through the centre of the orbit perpendicular to H is $\frac{H\epsilon\omega r\cos\theta \,.\, r\cos\theta}{c^2}$. This force changes its direction when the electron crosses that axis; the average value of the moment of the force in either half of the orbit on one side of the axis is

$$\frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\frac{H\epsilon\omega r^2\cos^2\theta\, d\theta}{c^2} = H\frac{\epsilon\omega r^2}{2c^2},$$

and this is the average couple acting on the electron. But this couple is MH, hence

$$M = \frac{\epsilon\omega r^2}{2c^2} \quad \text{......................}(141).$$

Now let us examine the effect produced on such an orbit by placing it in an inducing magnetic field perpendicular to the plane of the orbit. Consider the forces acting on the electron while the magnetic intensity is being increased from 0 to its final value H_0. First there are the molecular forces which maintain the electron in its orbit when no magnetic intensity acts; if these forces depend only on the radius of the orbit and not on the position of the electron in it, their resultant must be along the radius. Second there is the force $\frac{H\epsilon\omega r}{c^2}$, determined by the instantaneous value of H; this force also acts along the radius, outwards if ϵ and ω are both positive. Third there are the forces arising from the variation of H. In accordance with (6) there is an E.M.F. $E = -\frac{1}{c^2}\frac{d}{dt}(H\,.\,\pi r^2)$ round the orbit in the negative direction; in order to revolve once round the orbit in the positive direction the electron must perform the work ϵE. Now it is certain that the electron revolves so rapidly that dH/dt changes very little during one revolution; we may regard the work as performed against a constant force $\epsilon E/2\pi r$. But the acceleration of the particle along the tangent to the orbit is $\frac{1}{r}\frac{d}{dt}(r^2\omega)$; hence we must have

$$\frac{m_0}{r}\frac{d}{dt}(r^2\omega) = -\frac{\epsilon}{2\pi r c^2}\frac{d}{dt}(H \, . \, \pi r^2) \quad\text{.............(142)}$$

or from (141)

$$\frac{dM}{dt} = -\frac{\epsilon^2}{4m_0 c^4}\frac{d}{dt}(Hr^2) \quad\text{..................(143).}$$

Integrating from the time when H is zero to that when $H = H_0$,

$$\delta M = -\frac{\epsilon^2}{4m_0 c^4} \, . \, H_0 r_0^2 \quad\text{..................(144),}$$

where δM is the whole change in the magnetic moment of the orbit produced by the establishment of the magnetic field, and r_0 is the final value of r. It will be observed that this result (which is confirmed by a more accurate investigation and shown to be applicable to orbits of every form) is independent of the forces acting along the radius. These forces need only be taken into account if we desire to know the change in ω or r separately. This matter will be investigated later; it appears that r is unchanged by the magnetic field, so that the whole change in M is due to a change in ω.

Since M is independent of ω, its sign will be the same for all the mutually compensating orbits in the diamagnetic molecule, even though the electrons in these revolve in opposite directions. Under the influence of the field these orbits no longer compensate each other, and the body will acquire a finite intensity of magnetisation, $I = -\frac{N\epsilon^2 r^2}{4m_0 c^4} H_0$, where N is the number of revolving electrons in unit volume and r the mean radius of the orbit. Since I is negative the body will appear diamagnetic and its susceptibility will be

$$k = -\frac{N\epsilon^2 r^2}{4m_0 c^4} \quad\text{........................(145).}$$

It should be noted that since k depends only on ϵ^2, it would be impossible to draw any conclusions as to the sign of the charge on an electron from the direction of the induced magnetisation.

(145) cannot be tested directly by the insertion of numerical values, since we have no information as to the precise values of N and r; but we can find an upper limit to the susceptibility. For Nm_0, the density of the electrons in the substance, cannot be greater than ρ, the total density of the substance, while r cannot be much greater than 2×10^{-8}, the diameter of the simpler molecules. Hence we must have

$$k/\rho < \frac{1}{4c^4}\left(\frac{\epsilon}{m_0}\right)^2 \times 4 \times 10^{-16}, \text{ i.e. } < \frac{1}{c^2} \times 3 \times 10^{-2}.$$

No body is known which violates this condition; for most substances k/ρ is from $\frac{1}{c^2} \times 10^{-7}$ to $\frac{1}{c^2} \times 10^{-6}$.

Our theory of diamagnetism will be valid only so long as the distortion produced by the magnetic field in the electronic orbit is not so great as to alter the nature of the molecular forces acting on the electron. Accordingly it is important to notice that the change produced in the orbit by the greatest fields attainable experimentally is quite insignificant. If τ is the time of revolution of the electron, we have

$$\delta M/M = \frac{H\epsilon^2 r^2}{4m_0 c^4} \Big/ \frac{\epsilon \pi r^2}{c^2 \tau} = H/c \,.\left(\frac{\epsilon}{m_0}\right).\frac{\tau}{4\pi c} \quad\text{........(146).}$$

τ is not likely to be greater than the free period of the electrons which determine the optical properties of the substance and is, therefore, probably less than 10^{-15} sec. Hence

$$\delta M/M < 1{\cdot}3 \times 10^{-9} \,. H/c.$$

Since for the greatest fields that can be obtained H/c is less than 10^5, the disturbance of the orbit caused by such fields is very small.

Since we have supposed that the electronic orbits which determine the diamagnetic susceptibility are determined wholly by the internal structure of the molecule, it is clear that k should be independent of all influences which do not change the nature of the molecule; in particular it should be independent of the temperature, which affects merely the speed of the molecules and not their nature. k is actually found to be independent of the temperature (if the density remains constant) for all substances except bismuth. Since k is exceptionally large for this substance, it is probable that some other agency is at work. It has been suggested that in this case part of the diamagnetic susceptibility is due to the action of the magnetic field on the motion of the free electrons which give rise to the conductivity of the metal. The explanation is the more plausible because in bismuth the free path of these electrons is exceptionally great.

Paramagnetism.

59. In a paramagnetic substance it is supposed that the electronic orbits in the molecule do not compensate completely each other's magnetic fields, so that the molecule as a whole has a magnetic moment. The first action of the field on such a molecule will be to change slightly the nature of the orbits in the way which has been considered in the previous paragraph, but this change will not affect materially the magnetic moment of the molecule. A further change will now ensue. The molecules, the magnetic axes of which were formerly distributed equally in all directions, will tend to turn so as place those axes along the direction of the field, and, if there are no forces resisting this tendency, equilibrium will be attained only when all the axes lie along the field and the state of magnetic saturation is attained[1]. If M is the moment of each molecule, N the number of molecules in unit volume, the saturated intensity of magnetisation will be $I_0 = MN$.

But this orientation of the molecules is resisted by actions between the molecules. If the substance is a gas the molecules exert no influence on each other except when they collide, but the result of these collisions is to make the energy in each degree of freedom the same for all molecules. Under the action of the field the molecules begin to acquire a kinetic energy of rotation about axes perpendicular to the magnetic axes, and by means of the collisions this energy will be abstracted from this degree of freedom and distributed among the degrees of freedom corresponding to rotations about other axes and to translation; the

---

[1] The molecules of a diamagnetic substance, even when they have acquired a finite magnetic moment owing to the action of the field, do not tend to rotate, because the moment induced in the molecule by the field does not depend on the position of the molecule relative to the field; the energy of the induced magnetism in the field does not depend upon the position of the molecule.

magnetic energy of the molecules will be converted into heat energy of the gas. A state of statistical equilibrium will be reached finally in which the tendency of the molecules to set themselves in one particular position is counteracted by the tendency of the collisions to make them set themselves equally in all directions. When this state is reached no further energy will be taken from the magnetic field and converted into heat.

The nature of this final equilibrium can be calculated by the methods of the dynamical theory of gases. The problem is similar to that of determining the steady distribution of the molecules in a gravitational field (cf. p. 26). It is found that if I is the final intensity of magnetisation of the gas, T its absolute temperature and α the constant of p. 57,

$$\frac{I}{I_0} = \coth a - \frac{1}{a} \quad \text{.....................(147)},$$

where

$$a = \frac{MH}{\alpha T} \quad \text{.............................(148)}.$$

If H/T is small (147) becomes

$$I/I_0 = a/3 \quad \text{........................(149)},$$

so that the susceptibility of the substance for small values of H is

$$k = a/3H \,.\, I_0 = \frac{M^2N}{3\alpha T} \quad \text{.................(150)}.$$

It will be observed that k is inversely proportional to the temperature, a relation which is found experimentally in most cases. Since $\alpha NT = p$, the pressure of the gas, we may also write

$$k = \frac{I_0^2}{3p} \quad \text{.........................(151)}.$$

The most notable example of a paramagnetic gas is oxygen, for which $k = 1{\cdot}43 \times 10^{-7}/c^2$ at atmospheric pressure, so that $I_0 = 0{\cdot}65/c$. If we attribute the magnetic moment of the molecule to the presence of a single electron revolving with a period τ in an orbit of radius r, we have

$$I_0 = N\epsilon \,.\, \frac{\pi r^2}{\tau c^2},$$

or, putting $r = 1{\cdot}5 \times 10^{-8}$, the radius of the oxygen molecule,

$$\tau = \frac{1{\cdot}29 \times 10^{10} \times \pi \times (1{\cdot}5 \times 10^{-8})^2}{0{\cdot}65\, c} = 10^{-15} \text{ sec. about.}$$

The value of τ thus obtained is quite consistent with what we know on other grounds of the nature of electronic orbits.

Ferro-magnetism.

60. If the magnetic substance is solid there will be forces between the molecules other than those due to their collisions; consequently the susceptibility will depend on other quantities than the temperature and we may get the more complicated phenomena of ferro-magnetism. By making plausible assumptions about these intermolecular forces it is found possible to attain an adequate explanation of those phenomena; the consideration of these theories lies outside our scope, but one point is important. We know that when the temperature of a solid is raised sufficiently the molecular forces which determine the mechanical properties of the solid change suddenly when a certain temperature is reached and the solid melts; we should expect a similar sudden change to take place in the molecular forces determining the magnetic properties of the solid at some definite temperature, but, since the molecular forces are doubtless not the same in the two cases, it is not necessary that the sudden change in the susceptibility should take place at the melting point. It is found, indeed, that for all ferro-magnetic substances, there is a definite temperature of transition at which the phenomena of ferro-magnetism disappear and the substance becomes merely paramagnetic. This temperature, which is often called the 'Curie point,' is usually considerably lower than the melting point.

Now we have seen that the magnetisation of a paramagnetic body produces heat in it, and conversely the heating of such a body changes its magnetisation. Such a change requires energy, and we should expect that the amount of heat required to raise the temperature of a body (i.e. its specific heat) would depend upon its paramagnetic properties. Such a dependence is found, and it is found moreover that the specific heat of a ferro-magnetic body changes suddenly at the 'Curie point'; indeed from the study of the magnetic properties of ferro-magnetic substances at lower temperatures, it has been found possible to predict accurately the temperature of the 'Curie point' and the magnitude of the change of the specific heat at that temperature.

Some very interesting results have been obtained by Weiss (to whom our theory of ferro-magnetism is chiefly due) from the study of the magnitude of the paramagnetic susceptibility of various substances. Suppose we consider, not unit volume of the substance, but a gramme-molecule of it, so that in comparing different substances we are dealing with the same number of molecules in each case; then if we write k' for the quantity corresponding to k, we have

$$k' = \frac{M^2 \nu_0^2}{3RT} \qquad \text{.........................(152)},$$

where ν_0 is the number of molecules in 1 gramme-molecule, and $R = \alpha\nu_0$ is the 'gas constant' for a gramme-molecule, a universal constant. Weiss found that for magnetite at temperatures above the Curie point, the curve relating k' and $1/T$ was not as it should be from (152), a straight line passing through the origin, but a series of such lines separated by finite discontinuities in the value of k' at certain temperatures, corresponding to sudden changes in M; the values of M corresponding to these lines were closely approximate to $4B$, $5B$, $6B$, $8B$, $10B$, where $B = 16{\cdot}4 \times 10^{-22}/c$. On studying the values of susceptibility for other paramagnetic substances Weiss found that M always appeared to be a multiple of this same quantity B; hence he concludes that the molecules of different paramagnetic substances differ not in the nature of the uncompensated electronic orbits which they contain, but simply in the number of such circuits, which are always of the same nature. This conclusion, if it is confirmed by further work, is of the utmost importance, for it appears to indicate the existence of a structure common to all atoms, or at least to all paramagnetic atoms. An uncompensated electronic orbit having the moment B is called by Weiss a 'magneton'; (141) shows that it may be represented by an electron revolving in an orbit of radius 10^{-8} cm. with a period of about 10^{-14} sec.

60 A. We must now consider some rather curious and difficult questions connected with the energy of magnetised bodies.

Magnetic energy.

According to our theory of diamagnetism the process of producing the induced magnetic moment in a diamagnetic body consists in exciting currents in the electronic orbits contained in its molecules; we should expect then that the energy of the currents so excited would be equal to the energy spent in exciting them. It appears at first sight that this expectation is obviously correct. If we consider the induced magnetism to be produced by moving the body from a place where $H = 0$ to a place where $H = H_0$, it is easy to prove that the work which must be done in order to produce this displacement is $-\int_0^{H_0} M dH$, where M is the magnetic moment of the body when the magnetic intensity is H. The negative sign must be prefixed to the integral because if M is negative and the body diamagnetic, it is repelled by the magnet producing the field H, so that a positive amount of work must be done on it in order to bring it near that magnet. But $M = kH$, so that the work is

$$-\tfrac{1}{2}kH^2 = \frac{N\epsilon^2 r^2 H^2}{8m_0 c^4}.$$

On the other hand the change in the energy of the currents is the change of the kinetic energy of the electrons in the orbits, that is $\Sigma \tfrac{1}{2} m_0 r^2 \{(\omega + \delta\omega)^2 - \omega^2\}$. Now r is unchanged, while

$$-\frac{\epsilon^2 H_0 r^2}{4m_0 c^4} = \delta M = \frac{\epsilon r^2}{2c^2} . \delta\omega,$$

hence

$$\delta\omega = -\frac{H\epsilon}{2m_0 c^2} \qquad (153)$$

(cf. equation (167), p. 148). But $\Sigma\omega = 0$, since the molecule has no magnetic moment in the absence of the field, so that the change in the energy of the currents is

$$\Sigma \tfrac{1}{2} m_0 (\delta\omega)^2 r^2 = \frac{N\epsilon^2 r^2 H^2}{8m_0 c^4},$$

which is, as we expected, equal to the work done in exciting them.

But now two apparent difficulties arise. Firstly, we appear to have proved that the energy of the diamagnetic substance in the magnetic field is greater than its energy out of the field. But, on the other hand, we know that the introduction of a diamagnetic substance into the neighbourhood of a current circuit exciting a magnetic field diminishes the energy of that field because it diminishes the self-induction of the current. This apparent contradiction is not peculiar to our theory of diamagnetism and is mentioned here only because the discussion of it will help us in treating a second and far more important difficulty which may be stated thus.

We have assumed that the energy necessary to produce the induced currents in the diamagnetic substance is equal to the increase produced in the kinetic energy of the particles carrying those currents. We seem forced to extend the same assumption to other currents, such as those flowing in a conducting wire. Now the energy necessary to make a current i flow in a conducting circuit is $\frac{1}{2}Li^2/c^2 = E_1$ (say), where L is the self-induction of the wire circuit; on the other hand if N is the number of electrons in unit length of the circuit, v their velocity, and l the length of the circuit, the kinetic energy of the electrons is $\frac{1}{2}Nlm_0v^2 = E_2$ (say); since $i = N\epsilon v$, $E_2 = \frac{1}{2}li^2 \,.\, m_0/\epsilon \,.\, \frac{1}{N\epsilon}$. Now L depends on the shape of the circuit, and hence the first expression for the energy depends on that shape, but the second expression depends only on the length of the wire and not on the shape of it. The two expressions cannot possibly agree for all circuits, and it is easy to see that if we substitute for m_0 the value which we have given to it, they will not agree for any circuit. For it appears that L is always greater than $2l$, while $N\epsilon$ for the thinnest obtainable metallic wire cannot be less than 10^3. Hence E_1/E_2 cannot be less than $\frac{2}{9 \times 10^{20}} \times 5{\cdot}3 \times 10^{17} \times 10^3$, i.e. cannot be less than 1·2. For any ordinary circuit the quantity will be very much greater.

I hope that the reader will consider well these contradictions and attempt to discover the fallacy in the arguments which have been used in stating them, before proceeding to the solution of them which will now be given.

In the first case the fallacy is really very simple. In calculating that the work required to produce the induced magnetic moment is $\frac{1}{2}kH^2$, it is assumed that H remains constant while the diamagnetic body is moved into the field; but while the body is being moved into the field the current exciting the inducing field will tend to increase owing to the decrease of the self-induction of its circuit. In order to keep the field due to that current constant, we must subtract energy from its circuit; the energy which must be taken from its circuit in order to keep H constant is greater than the work done on the diamagnetic body, so that the energy taken from the system is greater than that given to it and the whole energy of the system decreases. If the simple form of Weber's theory is adopted, according to which a diamagnetic body is equivalent to a circuit of no resistance and constant self-induction, it is easy to show that the decrease in energy of the whole system, when the diamagnetic body is moved into the field and that field is kept constant, is equal to the work done on the diamagnetic body and therefore equal to the increase in the energy of that part of the system.

Now we can see the solution of the second difficulty. If we considered only one electron in the conducting circuit and estimated the energy which we must communicate *to that electron* in order to give it the velocity v, while the magnetic field due to the rest of the circuit remained unchanged, then we should find that this energy was $\frac{1}{2}m_0v^2$; and if we repeated the operation on each electron in turn we should find that the total energy was $\frac{1}{2}Nlm_0v^2$. But this is not the energy which we measure when we start a current in the circuit; we measure then the energy required to give each electron its actual velocity while the magnetic field due to the rest of the circuit is maintained at its actual value, together with the energy required so to maintain the field at its actual value; the second portion of the energy depends upon the shape of the circuit and is very much greater than the first part in all actual cases.

Let us look at the matter in another way. The difficulties to which attention has been drawn arise from the fact that a moving electron, being equivalent to a current, possesses, like a current, magnetic energy which can be regarded as distributed in the space round it. It is known that, if H is the magnetic intensity at any point in a given volume S, the amount of magnetic energy contained in that volume is $\frac{1}{8\pi c^2}\int_S H^2 dS$. If there is only one moving electron producing the magnetic field, the value of H will be determined at every point by the properties of that electron alone and the magnetic energy may be considered as part of the energy of the moving electron. But if there is a field produced by other currents, the energy in the field will be $\frac{1}{8\pi c^2}\int_S (H_1 + H_2)^2 \, dS$, where H_1 is the value of H determined by the electron under consideration and H_2 the value determined by the other currents. The expression to be integrated is no longer the sum of that which would be present if the electron existed alone and that which would be present if the other currents existed alone, but contains an additional term $2H_1H_2dS$. Thus, in the case of the current flowing in a conducting circuit, the whole magnetic energy in the space surrounding the circuit is not the sum of the energies which would be present in the same space if each electron of the circuit existed alone; there is an additional term depending on the mutual actions of the electrons, and we cannot, in general, regard the energy of the field as made up of parts, each of which is to be attributed to a single electron.

But there is one special instance of importance. If we act on one electron by one set of forces, while with another set we keep constant the field H_2 due to the other electrons, then the work done by both sets when the motion of the first electron is changed so as to produce a change δH_1 in H_1 is $\frac{1}{8\pi c^2}\int_S (\delta H_1^2 + 2H_2\delta H_1) \, dS$. Now it appears that the second term in this expression is the work done by the second set of forces, while the first term is that done by the first set. Accordingly, if we regard only the work done by the first set, the work done in order to produce a given change in the motion of the electron is the same as if the electron existed alone in the field. If we measure the work required to produce a change in the motion of the electron either when there is no other field or when the field due to other sources is kept constant by a totally different source of energy, then we shall always get the same result and, consequently, the same mass for the electron. The reader may see for himself that this condition is fulfilled in all the cases in this volume where we regard the phenomena as determined by the quantity m_0.[1]

61. But what is the proportion of this magnetic energy of the electron to its total kinetic energy? According to (14) the value of H_1 at any point is $ev \sin\theta / r^2$, so that the whole magnetic energy of the electron outside a sphere of radius a drawn round it is

'Electromagnetic mass.'

$$\frac{1}{8\pi c^2}\int_0^\pi \int_a^\infty e^2 v^2 \frac{\sin^2\theta}{r^4} \,.\, 2\pi r^2 \sin\theta \, dr \, d\theta = \tfrac{1}{3}\frac{e^2 v^2}{ac^2}.$$

Thus the energy is proportional to v^2 and the magnetic energy will appear as part of the kinetic energy. The less the value we assign to a the greater this part will be; since the magnetic energy cannot be greater than the whole kinetic energy $\frac{1}{2} m_0 v^2$, a cannot be less than $\frac{2}{3}\frac{e^2}{m_0 c^2}$ or $1{\cdot}8 \times 10^{-13}$ cm. If we assign to a this value and suppose that there is no energy inside the sphere with radius a, the magnetic energy will be equal to the kinetic energy; the electron will have no 'true' kinetic energy and its apparent mass will be solely due to electromagnetic actions.

The true value of a must be related to the distance from the electron at which the law relating the electric displacement and that distance undergoes a change; if we suppose there is no energy inside the sphere the electric displacement must be zero within it. In the case of a molecule carrying a charge we have every reason to believe that a is of the order of 10^{-8} cm. (for it is at about this distance that the forces excited by a molecule change), so that the 'electromagnetic mass' of a molecule, $\frac{2}{3}\frac{e^2}{ac^2}$, is about $1{\cdot}6 \times 10^{-32}$ grm., and vastly less than the whole mass of the molecule, which

---

[1] In many such cases, e.g. optical phenomena or the deflection of kathode rays, we are considering not one electron but a large number of electrons, all of which are moving with the same velocity. In these circumstances, the electrons have no relative velocity and exert on each other no electromagnetic force and each behaves as if the others were not there. In other cases, e.g. diamagnetism, the forces between the electrons are neglected on the ground that they will be too small to be appreciable.

is always greater than 10^{-24} grm. But as a matter of fact there is no experimental evidence whatsoever that the relation between the electric displacement and the distance from an electron is not constant down to distances as small as 10^{-13}; we know of no reason (except that the mass of an electron is finite) for supposing that this relation is not the same down to infinitesimal distances, that is to say, we have no evidence, such as we have in the case of molecules, to make us suppose that the radius of an electron is finite. Accordingly we can make a as small as we please and imagine, if we like, that a has actually the value which would make the magnetic energy of the electron equal to its total kinetic energy. We may adopt this hypothesis, but it must be clearly understood that, though there is no evidence against it, there is equally no evidence for it; in the present state of our knowledge it makes not the smallest difference whether we make it or not; all that we actually know is that an electron behaves as a point charge at the smallest distances which we can investigate directly, and of course these distances are vastly greater than the value we have assigned to a.

It might seem that we could obtain some information as to the value of a from the fact that the total work required to produce a change δH_1 in the field due to an electron is

$$\frac{1}{8\pi c^2}\int \delta (H_1 + H_2)^2\, dS,$$

since we can measure the term $2H_2\delta H_1$ which depends upon the distribution of H_1 round the electron. But it must be remembered that our doubt concerning the distribution of H_1 only extends to distances from the electron which are small compared with molecular distances; it is known that outside those distances an electron behaves as a point charge. Now if we take the value of $\int \delta H_1 dS$ over any sphere surrounding the electron the value will always be zero, since H_1 is distributed symmetrically. Unless, therefore, H_2 varies markedly throughout such a sphere, the value of $\int H_2 \delta H_1 dS$ over the sphere will also be zero. But of course all the magnetic fields that we know anything about directly are sensibly constant over distances small compared with molecular radii, and when we are concerned with such fields the total value of the term will be quite independent of that part of it which corresponds to very small distances from the electron. Whatever distribution for H_1 we assume at such small distances the total value of the term $\int H_2 \delta H_1 dS$ will be the same. Nor can we discover anything from the magnitude of the first term $\int (\delta H_1)^2 dS$, for it is obvious that it makes no difference to our estimate of the work required to change the velocity of an electron whether we regard this term as having the form $\int_a^\infty (\delta H_1)^2 dS$ or the form $\int_{a'}^\infty (\delta H_1)^2 dS + \frac{1}{2}mv^2$, so long as a' and a are less than any distance which we can investigate directly.

It is impossible in any reasonable space to discuss all the difficulties which may occur to the thoughtful student in connection with the distribution of magnetic energy, but I hope that the principles which have been employed in solving those raised in the last two paragraphs will enable him to solve any others that may suggest themselves to him.

REFERENCES.

The electronic theory of magnetism was first placed on a satisfactory basis in a paper by Langevin, *Ann. de Chim. et Phys.* Ser. VIII. 5. 70, 1905; the discussion in this chapter is based entirely upon Langevin's work. Langevin enters upon an elaborate discussion of the distribution of magnetic energy but his treatment of the subject appears to me rather difficult and insufficiently related to experimental methods. Nor can I regard all his conclusions as sound; he attempts to prove from the laws of diamagnetism that the whole mass of an electron is 'electromagnetic'; since exactly the same argument would prove that the whole mass of a molecule is electromagnetic, there must be some fallacy in it; after the discussion in § 60 A the reader may be left to discover this fallacy for himself.

The chief papers by Weiss on the 'magneton' are *Compt. Rend.* 152, 187 and 688, 1911. References to earlier work on the general theory of ferro-magnetism will be found therein.

CHAPTER VI

MAGNETO-OPTICS

General principles.

62. In Chapters II and III we attributed the optical properties of material substances to the presence within them of charged particles which were set into motion by the electric intensity in the light disturbance. The motion of these particles, like that of the electrons giving rise to the phenomena of diamagnetism, will be changed if the body is placed in a magnetic field; accordingly the optical properties of the body will be changed when it is placed in such a field, and these changes are of such importance that it is desirable to consider them in a separate chapter.

Let the direction of the magnetic field be that of the axis of z, and let its intensity be H; let the components of the electric displacement in the light disturbance be P, Q, R outside the body and P_1, Q_1, R_1 within it. Then in addition to the force on the particle due to the electric intensity there will be, in accordance with § 16, a force at right angles to the direction of its motion and to the magnetic field; the equations of the motion of the particle corresponding to (34) will be

$$m\ddot{x} + fx - \frac{eH}{c^2}\dot{y} - 4\pi e P_1 = 0 \quad \text{...........(154)},$$

$$m\ddot{y} + fy + \frac{eH}{c^2}\dot{x} - 4\pi e Q_1 = 0 \quad \text{...........(155)},$$

$$m\ddot{z} + fz - 4\pi e R_1 = 0 \quad \text{....................(156)}.$$

If the electric displacement is parallel to H, the magnetic field has no effect; if it is not parallel to H, the motion of the particle is not, as in § 24, always confined to the plane of the electric displacement. We will consider here only the case when the light is travelling parallel to H, so that $R=0$ and $z=0$. Any harmonic vibration, corresponding to light of a single frequency, in the plane XY can be regarded as composed of two oppositely directed circular vibrations in that plane of suitable amplitude and phase; accordingly if we solve the equations for two oppositely directed circular vibrations we may combine the results to obtain the solution for any other case.

Let a plane wave of circularly polarised light fall on the substance; the electric displacement in it may be represented by $P=A\cos(pt-qz)$, $Q=\pm A\sin(pt-qz)$, where the upper and lower signs denote respectively positive and negative rotations of the displacement about the direction of H. The motions of the particle excited by this light are circular, the rotation being in the same direction as that of the electric displacement[1]. Hence the solution corresponding to (35) is

$$(x, y)=\frac{4\pi e}{f-p^2m\mp\frac{eH}{c^2}p+4\pi Ne^2}(P, Q) \quad \text{......(157)}.$$

Proceeding precisely as in § 24, we find

$$n_\pm^2=1+\frac{Ne^2}{\pi m\left(\nu_0^2-\nu^2\mp\frac{eH\nu}{2\pi c^2m}\right)} \quad \text{............(158)},$$

the suffixes + and − denoting throughout the quantities appropriate to light polarised circularly in the positive or negative direction; ν is the frequency of the incident light, $\nu_0=\frac{1}{2\pi}\sqrt{f/m}$ the frequency of the free vibrations of the particle in the absence of the magnetic field.

The frequency of the free vibrations of the particle in the magnetic field may be determined from the equations resulting

---

[1] Perhaps it is worth while to point out that the direction of the rotation of the particle excited by a given light vibration is independent of the sign of the charge on the particle. What is here called positive circularly polarised light is usually known as 'left-handed circularly polarised light,' for when viewed in the *opposite* direction to that of its propagation, it appears to rotate counter-clockwise.

from putting $P_1 = Q_1 = 0$ in (154), (155). If the vibrations are circular and of frequency ν', they may be represented by $x = a\cos 2\pi\nu' t$, $y = \pm a \sin 2\pi\nu' t$, the upper and lower signs again denoting rotations in the positive and negative directions about the direction of H. We then find

$$\nu_\pm'^2 = \nu_0^2 \mp \frac{eH}{2\pi c^2 m}\nu_\pm' \quad \text{..................(159).}$$

It is to be observed that it is *not* possible to deduce (158) from (45) by substituting ν' for ν_0, because the effect of the magnetic field on the vibration of the particle when executing forced vibrations under the influence of the light is determined by the frequency of that light and not by the frequency of the free vibrations which it would execute in the absence of the light[1].

The Faraday effect.

63. Suppose now that a plane wave of plane polarised light falls on the substance. Such a wave may be regarded as compounded of two opposite circular vibrations of equal amplitude, which travel with equal velocities outside the substance. If the plane of polarisation is that of yz, the electric displacement being perpendicular to it may be represented outside the substance by

$$P = A\left\{\cos p\left(t - \frac{z}{c}\right) + \cos p\left(t - \frac{z}{c}\right)\right\},$$

$$Q = 0 = A\left\{\sin p\left(t - \frac{z}{c}\right) - \sin p\left(t - \frac{z}{c}\right)\right\} \quad \text{...(160).}$$

Inside the substance the opposite circular vibrations travel with different velocities $v_\pm = c/n_\pm$; hence the electric displacement may be represented by

$$P = A\left\{\cos p\left(t - \frac{n_+ z}{c}\right) + \cos p\left(t - \frac{n_- z}{c}\right)\right\},$$

$$Q = A\left\{\sin p\left(t - \frac{n_+ z}{c}\right) - \sin p\left(t - \frac{n_- z}{c}\right)\right\},$$

---

[1] I point this out because the mistake was made in the first edition of this book. The chapter on Magnetism in that edition is almost entirely fallacious, the correct formulae being attained in some cases only by the mutual compensation of errors.

or $$P = 2A \cos p \left\{t - \frac{z}{2c}(n_+ + n_-)\right\} \cos \frac{pz}{2c}(n_- - n_+),$$

$$Q = 2A \cos p \left\{t - \frac{z}{2c}(n_+ + n_-)\right\} \sin \frac{pz}{2c}(n_- - n_+) \ldots (161).$$

This represents for any given value of z a plane polarised wave of which the plane of polarisation makes an angle δ (counted positive when the plane is rotated in the positive direction) with the plane yz, where $\delta = \frac{pz}{2c}(n_- - n_+)$. Hence after passing through a thickness l of the substance the original wave will emerge still plane polarised, but with its plane rotated through an angle δ in the positive direction, where

$$\delta = \frac{\pi l \nu}{c}(n_- - n_+) \quad \ldots\ldots\ldots\ldots\ldots\ldots (162).$$

In the experiments to which we shall apply (162), $\frac{eH\nu}{2\pi c^2 m}$ is always very small compared with $\nu_0^2 - \nu^2$. Hence we may write $n_+ + n_- = 2n$, where n is the refractive index in the absence of a magnetic field, and neglect a term $\left\{\frac{eH\nu}{2\pi c^2 m} \Big/ (\nu_0^2 - \nu^2)\right\}^2$. Therefore

$$\delta = \frac{\pi l \nu}{2cn}(n_-^2 - n_+^2) = -\frac{H}{c} \cdot l \cdot \frac{\nu^2}{2nc^2} \cdot \frac{Ne^3}{\pi m^2 (\nu_0^2 - \nu^2)^2} \ldots (163).$$

According to Chapter II a medium containing such particles as we have considered corresponds to a transparent medium acted on by light the frequency of which is not near to that of the free vibrations of the particles in it; we saw further that in order to explain completely the properties of such a medium it was usually necessary to suppose that there were present particles attached differently to the molecules and having different values of e/m. Hence we should write in place of (163)

$$\delta = -\frac{H}{c} \cdot l \cdot \frac{\nu^2}{2nc^2} \Sigma \frac{Ne^3}{\pi m^2 (\nu_0^2 - \nu^2)^2} \quad \ldots\ldots\ldots (164).$$

Comparing (45) and (164) we see that δ is not in general a function of the refractive index; in order to deduce δ from observations on the refraction we should have to determine the values of N, e, m, ν_0 for each of the different groups of particles, and we have seen that the accurate determination of these quantities for all the groups is impossible even for those substances of which the refraction has been the most completely investigated. But, if we might assume that the value of e/m is the same for all the groups, we might write from (45) and (164)

$$\delta = -\frac{H}{c} \,.\, l \,.\, \frac{\nu}{2c^2} \,.\, \frac{e}{m} \,.\, \frac{dn}{d\nu} \quad \text{...............(165)},$$

or, introducing λ, the wave-length in vacuo of the light, in place of ν,

$$\delta = +\frac{H}{c} \,.\, l \,.\, \frac{\lambda}{2c^2} \,.\, \frac{e}{m} \,.\, \frac{dn}{d\lambda} \quad \text{..................(166)}.$$

A rotation of the plane of polarisation of light travelling through a transparent substance in the direction of the magnetic field was first observed by Faraday, who was unable to explain the action; it has since been observed in all the transparent substances which have been investigated. A formula

$$\delta = \beta \,.\, \frac{H}{c} \,.\, l \,.\, \lambda \frac{dn}{d\lambda},$$

which is identical with (166) if we put $\beta = \frac{e}{m} \,.\, \frac{1}{2c^2}$, was proposed on purely empirical grounds by Becquerel, and agrees fairly well with observation. The following table (taken from Schuster's *Theory of Optics*) gives the value of β for various substances; according to our theory and Becquerel's β should be a universal constant.

Gas Substance	$\beta \times 10^4$	Substance	$\beta \times 10^4$
O_2	1·26	H_2O	1·96
CO_2	1·91	CS_2	1·30

Taking the mean value of β as $1{\cdot}6 \times 10^{-4}$, we find

$$e/m = 1{\cdot}6 \times 10^{-4} \times 18 \times 10^{20} = 2{\cdot}9 \times 10^{17}.$$

This value is not very different from that characteristic of an electron, $5{\cdot}3 \times 10^{17}$; we should expect it to be rather smaller than that deduced on other grounds, because the value of e/m for some of the particles is, as we saw in § 28, a great deal less than that for an electron. We get a much better agreement with experiment if in determining δ we neglect altogether the effect of those particles for which e/m is small. (163) is then accurately true, but in deducing a relation similar to (166) we must consider only that part (n') of the refractive index which is independent of these same particles. Now the particles with the small values of e/m are those which have free periods in the infra-red; accordingly n' is the value of n obtained from the empirical formula (49) by putting $M_r = 0$, and it is easy to see that in (166) we should substitute $\frac{n'}{n}\frac{dn'}{d\lambda}$ for $\frac{dn}{d\lambda}$. It is found that (166) thus altered agrees admirably with experiment.

We may conclude that our general theory of the Faraday effect is adequate and that the rotation of the plane of polarisation is determined by the effect of the magnetic field on the vibration of the 'bound' charged particles, which are for the most part electrons. But we must note that, according to (163), the sign of δ is determined wholly by the sign of the charge on the particle, since all the factors except e^3 are essentially positive. If e is negative δ should be positive, that is to say the plane should be rotated in the direction in which a current would have to flow in order to produce a magnetic field in the direction of H. It is found experimentally that δ is positive for all diamagnetic substances with the single exception of titanium chloride; in paramagnetic substances it is usually negative, but sometimes positive. The meaning of the exceptional cases will be considered shortly.

It may be remarked that the sign of δ changes with that of H, but not with the direction of the propagation of the light. Consequently, if the light travels through the substance in one direction and then is reflected back along the same path, the total rotation of the plane of polarisation will be twice that due to a single passage through the substance. In this respect, as is well known, the magnetic rotation differs from the rotation produced by optically active substances.

Absorption bands.

64. So far we have only considered the phenomena when the frequency of the incident light is far removed from that of any of the free vibrations in the substance. If on the other hand these frequencies become nearly coincident, we have to adopt the treatment of § 29 and to introduce a term $k\dot{x}$ into (154)—(156), giving rise to an absorption of the light. The phenomena now to be observed are far more complicated, and we shall only consider the chief features of them.

Since according to (159) the free period of the bound electron varies with the direction of its rotation relative to the magnetic field, the frequency of the incident light which is equal to that of the free vibration in the same direction is different according to the direction of its circular polarisation. In place of the single 'absorption band' corresponding to the single free period, we shall have two absorption bands of which one absorbs positive and the other negative circularly polarised light. It appears from (159) that if the particle is an electron for which e is negative the frequency corresponding to the absorption band for positive light will be greater than that for negative light.

Such a doubling of the absorption band under the influence of a magnetic field has been observed in a few cases, although the experimental difficulties are very great. The most complete observations have been made in sodium vapour for light approximating in frequency to that of the D lines. The measurements accord in general with those predicted from our theory and it is found that in the case of both these lines the absorption in a magnetic field is greater on the violet side for positive circularly polarised light, as it should be if the D lines represent the free periods of electrons. But some remarkable anomalies have been observed by J. Becquerel in investigations of the doubling of the absorption bands in certain crystals containing didymium and the allied elements and by Wood in the case of other parts of the absorption spectrum of sodium. In these cases the magnitude of the separation of the two absorption bands is not very different from that predicted by the theory on the assumption that the particles concerned in the absorption are electrons, but the direction of the displacement of the bands would indicate in some cases that the particles are positively charged, while in others they are negatively charged. The observations have excited great interest because they have been used as an argument for the existence of 'positive electrons' having approximately the same mass as the negative electrons.

It has been explained already in § 43 that there are the most serious objections to accepting the hypothesis of such positive electrons. If they exist they are either only present in some peculiar bodies, or they are present in all, but only make their presence felt in some. If the latter alternative is adopted, we are no nearer an explanation of the anomalies than before; it is no easier to explain why positive electrons should effect optical phenomena only in a very small minority of cases than to explain why in some cases the magnetic field should seem to produce a change of the wrong sign in the orbit of a negative electron. If the former alternative is accepted we must suppose that the substances which contain positive electrons are fundamentally different in their nature from those which do not; a consideration of the other properties of the anomalous substances offers no support whatever for such a conclusion. The hypothesis of positive electrons always introduces more difficulties than it removes, because the cases in which it is of service are so exceptional. If we suppose that the positive electrons act in these cases, we are always left with the problem of deciding why they do not act in the vastly greater number of normal cases.

A far more plausible explanation of the anomalies connected with the sign of the Faraday effect is available. It appears that all the substances which show anomalies of sign are either paramagnetic or at least contain an element which is paramagnetic in the free state. Now the atoms of paramagnetic substances have a finite magnetic moment and, probably, a magnetic field inside them, which will not generally be in the same direction as the external field in which the substance is placed. We may suppose that when a substance shows a Faraday effect of the 'wrong' sign, the electrons in it are under the influence of such an internal field opposite in direction to the external field. Some support for this hypothesis is found in the fact that the magnitude of the Faraday effect is determined in paramagnetic substances by the field due to the molecules rather than by the external field; for in such substances δ is proportional not to H but to B, and B as we have seen is the resultant of the external field and that due to the molecular magnets.

One further prediction of our theory which is confirmed by experiment should be noted. It will be seen that δ in (163) is determined by $(\nu_0^2-\nu^2)^2$; hence the sign of δ will be the same just outside an absorption band whether ν is slightly greater or slightly less than ν_0, that is to say δ will be of the same sign on both sides of an absorption band. This conclusion, which is confirmed by observation, is important, because it has been used to decide between the theory of the Faraday effect which is given here and an earlier alternative theory which predicted that the sign of δ should be different on opposite sides of the absorption band.

The Kerr effect.

65. In addition to the Faraday effect there is another effect of a magnetic field on light propagation which has so far been detected only in the ferro-magnetic metals. It was found by Kerr that if light is reflected normally from the surface of one of the poles of a powerful magnet (that surface being perpendicular to the magnetic induction) the plane of polarisation is rotated through a small angle; subsequent observers have repeated his experiments and discovered effects on light reflected otherwise from magnetised surfaces which are closely connected with the Kerr effect.

For an explanation of these effects the electron theory is sufficient but not necessary. It was shown by Goldhammer that the main features of the phenomenon could be explained by assuming merely the electromagnetic theory of light in the form given by Maxwell together with such a distribution of magnetic and electric intensities at the magnetised surface as is known to exist. Consequently any theory which is not inconsistent with those assumptions will lead to the same result, and it is unnecessary here to follow out the complicated analysis which is necessary in order to apply the electronic theory; the reflection is, of course, determined by the free electrons in the metal, the motions of which will be influenced by the magnetic field in the way which was described in § 42 when the Hall effect was under consideration. It appears that no conclusion can be drawn from the direction of the rotation of the plane of polarisation as to the sign of the charge on the particles concerned.

According to this view the magnetic effects in a ferro-magnetic metal do not differ in their nature from those in a diamagnetic substance; the Faraday and Kerr effects are very much greater in these metals for a given value of H simply because the magnetic field acting on the charged particles for a given value of H is very much greater in the case of a ferro-magnetic than in that of a diamagnetic substance, for in the former case there is added to H the magnetic intensity due to the magnetic moment of the paramagnetic molecules. But a theory has been proposed which would make the phenomena in ferro-magnetic substances of an entirely different nature. It has been suggested that when the molecules of the substances possess a finite magnetic moment they will be set into vibration by the magnetic field in the light disturbance and will alter the propagation of that disturbance in the same way as the charged particles set into vibration by the electric field in the disturbance alter it. This theory is competent to explain the Kerr effect, but it predicts that in the Faraday effect the sign of δ should be different on different sides of an absorption band and has consequently been rejected. There must be some action of the kind suggested by the theory, but it is probable that, owing to the large mass of the molecules, it is small and altogether masked by the much greater effect due to the electrons.

It should be pointed out here that in framing equations (154)—(156) no account has been taken of the magnetic field due to the light disturbance itself; it has been assumed that the only magnetic intensity acting on the electrons is that of the external field. The assumption is justified by considerations which show that the magnetic intensity in the light disturbance must be very small compared with that of an external field strong enough to produce any measurable Faraday effect. A direct estimate of the magnetic intensity in the most intense light, based on a knowledge of the energy conveyed by such light, leads to such a conclusion; moreover if the magnetic intensity in the light were comparable with that due to the external field the refractive index of the substance (either in or out of the external field) should depend on the intensity of the light; no such dependence has ever been detected.

66. Hitherto in considering optical phenomena we have never inquired how the light which is refracted or reflected has been produced. This question will concern us more closely in a later chapter, but some of the most obvious of the conclusions which will be attained may be conveniently anticipated. According to our theory of light a light disturbance can originate only at a place where the electric displacement is made to vary periodically; the direction of propagation of the light and its plane of polarisation are perpendicular to the electric displacement. If the light travels from the place where it is produced to that where it is observed through a vacuum the relation between the time and the electric displacement will be the same at both places.

The Zeeman effect.

Now such a varying electric displacement may be produced by the motion of an electron relative to a positive charge, the electric displacement at a distant point being represented at any moment both in magnitude and direction by the line joining the two charges. If the motion of the electron is harmonic, the light disturbance will be harmonic; if the electron moves in a straight line the light will be plane polarised; if it moves in a circle the light will be circularly polarised in the direction of the rotation of the electron. A harmonic light disturbance appears in our observations as homogeneous light of a single wave-length such as forms a line in the spectrum. We may assume provisionally that the light corresponding to such a line represents the disturbance emitted from an electron in the source of light which is executing harmonic vibrations, and consequently that the frequency of such light is the frequency with which such an electron executes free vibrations, after it has been disturbed in some manner from its position of equilibrium.

In the absence of a magnetic field the light has the same properties in whatever direction it is emitted and hence the forces on the electron must be independent of the direction of its motion. The equations of motion of its harmonic vibrations must be those obtained by putting $P_1 = Q_1 = R_1 = H = 0$ in (154)—(156)[1]. When

---

[1] This statement is not strictly accurate. The emission of light by the electron involves the loss of energy by it and there should be in the equations some term representing this loss. But if we suppose that the proportion of the

the source of light is placed in a magnetic field H these equations are changed by the introduction of the terms involving H. The solution of these equations shows that the vibrations of the electron in the plane xy, perpendicular to H, must consist of circular vibrations of the electron in one direction or the other, the frequency of these vibrations, $\nu_{\pm}'$, being connected with the frequency ν_0 of the vibrations in the absence of the field by (159). Superimposed on the circular vibrations is a rectilinear vibration along the axis of z with the frequency ν_0. Thus the components of the vibration along the axes of x or y have a frequency either ν_+' or ν_-', that along the axis of z has a frequency ν_0.

Since the nature of the light emitted in any direction is determined by the vibrations of the electrons in a plane perpendicular to that direction, an observer receiving light which is travelling along the direction of the magnetic field should find that light composed of two circularly polarised components of frequencies $\nu_{\pm}'$; the single line in the spectrum which appeared when there was no magnetic field will be split into two lines circularly polarised in opposite directions. If the particle emitting the light is an electron, so that e is negative, the component on the violet side (greater frequency) should show left-handed circular polarisation if the direction of propagation of the light is the same as that of the magnetic field, and right-handed if those directions are opposite. On the other hand an observer receiving light travelling perpendicular to H in a direction which may be identified with the axis of x will observe three lines in place of the single line: (1) a line of frequency ν_0 in the same position as the previous single line, representing the vibration of the electron along the axis of z, and (2) and (3) two lines of frequency $\nu_{\pm}'$ representing vibrations along the axis of y; all three lines will, therefore, appear plane polarised, the first in a plane perpendicular to the magnetic field, the others in a plane containing the field.

Such changes when a source of light is placed in a strong magnetic field were first observed by Zeeman and interpreted

---

total energy of the electron lost in the course of a single vibration is very small the presence of this term will not affect materially the conclusions which we are about to draw as to the frequency of the vibrations.

according to the theory given here by Lorentz, but subsequent investigation has shown that in the case of only one substance, helium, is the change in the spectrum produced by a magnetic field exactly that which is predicted by the theory. In all other cases the change is more complicated, more than two lines appearing when the light is viewed along the field and more than three when it is viewed at right angles to the field. But in certain respects observation agrees with theory in all cases. Firstly, the lines in the light emitted along the field, if polarised at all, are circularly polarised, and in the immense majority of cases the direction of the circular polarisation is that indicated by the hypothesis that e is negative; opposite circular polarisations are always found on opposite sides of the line ν_0. Secondly, the lines in the light emitted perpendicular to the field, if polarised at all, are plane polarised, but in some cases there is no undisplaced line in the position ν_0. Thirdly, there is some quantitative agreement. From (159) we find

$$(\nu_+' - \nu_-')\frac{2\pi c^2}{H} = -e/m \quad\text{.................}(167).$$

If our theory is correct this quantity should be the same for all substances and should have the value $5{\cdot}3 \times 10^{17}$. It is found that if the value of (167) is taken for the most widely separated components of the line which appear in the magnetic field, this value is the same within the limits of experimental error for all substances and agrees well with the value for ϵ/m_0 deduced from experiments on kathode rays. Thus for sodium it is found that

$$\nu_+' - \nu_-' = 5{\cdot}7 \times 10^{10} \; (\nu_0 = 5{\cdot}1 \times 10^{14})$$

when $$H/c = 2{\cdot}24 \times 10^4,$$

whence $$e/m = -4{\cdot}8 \times 10^{17}.$$

The best and most modern value of ϵ/m_0 found by such methods is $5{\cdot}32 \times 10^{17}$; it must be regarded as having equal weight with the values obtained by other methods with which it agrees very closely.

It should be remarked here that the determination of ϵ/m_0 by this method is of very great historical importance. The interpretation by Lorentz of Zeeman's observations was given just after J. J. Thomson's determination of the ratio e/m for the 'corpuscles' constituting the kathode rays; the close agreement of the values obtained by the two methods led to the identification of Lorentz' 'electrons' with Thomson's 'corpuscles,' and showed for the first time what diverse phenomena might be brought within the range of a theory based on the properties of such particles; it may be regarded as the starting-point of the wider electronic theory which is the chief feature of modern physics.

Even when the resolution of the lines is more complex than is predicted by the simple theory there are features indicating that our explanation is right in the main. Runge has shown that the value of $(\nu_{+}' - \nu_{-}') \frac{2\pi c^2}{H}$ for the less widely separated components can always be represented by $a/b \,.\, \epsilon/m_0$, where a and b are small numbers[1]; moreover it is found that the relations between the components are precisely the same for lines in the same spectral series (see Ch. IX). These observations give a clue to the probable nature of the explanation of the discrepancy between our theory and experiment. Lines of the same series must be emitted by the same vibrating system, and hence this system cannot consist of a single electron executing harmonic vibrations; the vibration of the electron must be more complex and the apparent simplicity of a single line in the spectrum must be due to the resolution of the more complex vibration by the spectrometer (see Ch. IX). There appears to be no difficulty in devising vibrating systems which would give most of the effects actually observed; the difficulty is rather to choose between the various possible alternatives. The most difficult matter to explain is the absence of an undisplaced line in light emitted perpendicular to the field; perhaps a suggestion may be drawn from the theory of paramagnetism expounded in the last chapter. If the vibrating system which emits the light really consists of electrons revolving in a plane orbit, then that orbit would have a finite magnetic moment and would tend to set itself with its plane perpendicular to H; in this position there would be no motion of the electron

---

[1] What are 'small' numbers? The relation would be more accurately expressed by saying that the requisite values of a and b are very much smaller than would be expected if the values of the quantity were distributed by chance within the range 0 to ϵ/m_0.

parallel to H and, consequently, no line in the position given by ν_0.

But this suggestion leads to another difficulty. On the one hand we have assumed that light may be emitted by revolving electrons; on the other hand we have attributed to the presence of such revolving electrons the magnetic properties which are displayed by all material bodies under all conditions. Why, then, do not all bodies under all conditions emit light? Or, to put the matter in another way, if revolving electrons emit light they must be losing energy; therefore the revolving electrons in a diamagnetic substance must be losing energy. Why, then, do the revolutions continue? what source of energy is maintaining the revolutions? These are difficult questions which have not been answered completely; we will consider them again in a later chapter.

General summary.

67. Our purpose in this part of the volume has been to explain in terms of the electronic theory the electrical and magnetic properties of material substances, to show why they react as they do under the influence of electrical and magnetic fields. In the course of our discussion we have considered at least one property (thermal conductivity) which the older theories did not regard as electrical, and the question arises whether there are not other such properties, mechanical and chemical properties for example, which may now be brought into relation with electrical theory. The answer is, of course, in the affirmative; we can now suggest some reason for the undoubted connection which exists between the chemical or mechanical properties of a substance and the more distinctively electrical properties which have been considered already. But such suggestions as can be made are somewhat vague and speculative; and since it is desirable to keep those parts of our theory which are firmly established separate from those which are merely tentative, we will leave such suggestions for the present and proceed to the consideration of an entirely different branch of modern electrical theory, but one which is intimately connected with that with which we have been so far concerned.

But before proceeding it will be well to epitomise very briefly the results we have obtained already.

The following are the assumptions characteristic of the electronic theory: (1) That electric charges are distributed discontinuously. This assumption involves the consequence that with each charge is associated a definite mass, but this mass might have been very much smaller than that which we actually attribute to an electron. (2) That the charge on any particle is a multiple of a universal unit. (3) That there are charged particles which are common constituents of all bodies. (4) That the mass of these particles is very much less than that of any atom. (5) That the charge on these particles is $-4{\cdot}7 \times 10^{-10}$.

In so far as our theory agrees with experiment, the assumptions on which it is based are confirmed. We may note, therefore, that our general treatment of refraction involves (1) only; a study of numerical values provides some evidence for (2) and (4). The theory of metallic conduction involves (1) and (3); the explanation of the Wiedemann-Franz law involves (2). The study of the 'supported' current in gases requires (2), (3) and (5); the study of the 'unsupported' current requires all the assumptions and provides the most striking evidence for them. Our treatment of magnetic susceptibility does not really require any of the assumptions of the electronic theory, being merely equivalent to that of Ampère and Weber, but we have actually made use of all of them. The theory of magneto-optics (Faraday and Zeeman effects) requires all the assumptions.

The chief difficulties with which we have met, or the phenomena which appear at first sight to be inconsistent with our theory, are the absolute magnitude of the velocity of gaseous ions, the difference of the sign of the Hall effect in different metals, the complexity of the Zeeman effect and the variation of the sign of the Faraday effect; it is possible that all the difficulties would disappear if we were not so profoundly ignorant of the forces acting on electrons near and inside molecules. But the difficulty concerning the specific heat of conductors and dielectrics is much more serious and cannot be attributed to our ignorance of such matters.

REFERENCES.

Some account of magneto-optics is given in all modern text-books of Optics, such as those of Drude, Wood and Schuster. The treatment of the latter appears to me especially admirable, but the reader should be warned that Schuster adopts different conventions from those used here as to the relative directions of the axes of coordinates ; accordingly his equations often differ in sign from those given here.

A very complete account of the facts and an adequate summary of the theories are to be found in Kayser's *Handbuch der Spektroskopie*, Vol. II. The theory of the phenomena is most fully discussed in Voigt's *Magneto- und Elektro-optik* ; the more complicated developements of the theory are largely due to that author himself. But the book is rather difficult reading to those who are accustomed to think physically rather than mathematically, and the lamentable absence of an index makes it almost useless as a work of reference.

For those who are historically curious it may be mentioned that Zeeman's original experiments are described in the *Proc. Amst. Acad.* 1898 and that Lorentz' explanation of them (though published earlier) is given in the *Rapports au Congrès International* (Paris, 1900), Vol. III.

PART II

CHAPTER VII

RADIATION

68. A SYSTEM A produces some effect upon a system B at a distance from it; in some cases, but not in all, this effect is supposed to be due to a 'radiation' proceeding from A. What is the difference between the two classes of cases? Some years ago this question would probably have been answered by saying that the influence was to be regarded as consisting of a radiation if it were possible, by interposing a suitable body between A and B, to produce on the latter a definite 'shadow' of the obstacle, that is to say, a region, the boundary of which is determined by the outline of the obstacle and within which B is free from the influence of A; the formation of such a shadow would be explained by attributing the effect of A on B to some influence travelling out from A along straight lines and capable of being stopped by the obstacle.

Nature of radiation.

If this definition were applied perfectly strictly it is doubtful whether any influence could be termed radiation, for it is probable that in no case can a perfectly 'dark' shadow be produced; the difference in the nature of the 'shadow' produced by a metal plate according as the influence proceeding from A is an electrostatic attraction or a luminous sensation is one of degree rather than of kind. Moreover nowadays we class as radiations influences which, even if they are capable of producing shadows, do certainly not in all circumstances travel in straight lines, so that the boundary of the shadow might be determined rather by the circumstances of the space between A and B than by the outline of the obstacle.

Consideration will show, I think, that in this case, as in many others, an entirely new definition of radiation has been adopted gradually, but one which gives in practice very nearly the same classification as the old. We regard now as a radiation any influence which consists in the transfer of energy from A to B at a rate which may be independent of the relative motion of A and B and dependent only on their position relative to each other and to neighbouring objects. The two definitions agree in classifying as radiation one extremely important form of influence, namely that due to particles of finite mass projected from A with a definite velocity and undeflected in their course towards B; with regard to the classification of any other form of influence they differ. There is no actual difference of opinion as to what influences are and what are not to be classed as radiation, but it will be found that the classification adopted agrees better with the second than with the first definition.

It is suggested by the second definition that, if we fix our attention on A and B only, there will be several fundamentally important quantities to be determined in connection with any radiation. The first is the velocity, which is the ratio of the distance between A and B to the time which elapses between the occurrence of any change in A and of a related change in B. But when dealing with such radiation as is considered here, it is extremely difficult to determine this velocity, because it is so great compared with that which can be produced in any material body. The velocity of only one kind of radiation, light, has been determined with any accuracy by methods which are beyond dispute; in some other cases a quantity which is termed the velocity has been measured accurately, but it must be remembered that it is not the velocity as we have defined it, and is connected with that velocity only by means of some theory as to the nature of the radiation and its behaviour in various circumstances.

The second quantity is the intensity, which is measured by the quantity of energy communicated by A to B in unit time; since this quantity will vary with the situation of B relative to A, if we want to define an intensity which is determined only by the properties of A (and this is the usual procedure) we must adopt some standard situation. Here again it is extremely difficult to measure the quantity directly with any accuracy and in practice the intensity is always estimated by means of observations on some other effect the magnitude of which is believed to be proportional to the energy communicated. Thus the intensity of a given beam of radiation is often taken to be proportional to the darkening in a photographic plate produced by it in a given time, or the number of ions produced in a given mass of gas, or the charge communicated to the body on which it falls. The use of such methods is justified partly by the fact that, according to our theory of the process by which the radiation produces the effect, these quantities must be proportional to the energy, but more by the fact that the ratio of the intensities of two radiations, if they are strictly of the same kind, is found to be independent of whichever of these methods is employed. It is not legitimate to use such methods in the comparison of the intensities of radiations of different kinds; for such comparisons the only sound method is the direct measurement of the amount of energy conveyed by the radiation, which is most conveniently estimated after it has been completely converted into heat energy. Two radiations are of the 'same kind' if all their characteristic constants except the intensity and direction are the same.

A third quantity characteristic of a given radiation which might be measured by the examination of A and B only is the electric charge conveyed by the radiation. In all cases where the radiation is considered to consist of charged particles the charge received by a body struck by radiation of known intensity has been measured directly; the charge is always proportional to the intensity measured by other methods for radiation of the same kind, and the ratio of the charge to the intensity is a constant characteristic of the kind of radiation concerned.

69. The other characteristic constants of a radiation are determined by the presence of bodies other than A and B. Some of these constants concern the change in the path of the radiation caused by the presence of electric and magnetic fields; the others concern the effect on the radiation of passage through material bodies.

Absorption.

In general the insertion of a material body in the path of the radiation[1] between *A* and *B* causes a change of the intensity at *B*. This change is due partly to an abstraction of part of the energy of the radiation by the body (absorption) and partly to a change in the direction of the radiation (scattering or reflection); there may also be a change in the nature of the radiation. Since all these actions generally result in a diminution of the intensity at *B* they are sometimes spoken of collectively as 'absorption,' and it is in this sense that the word will here be generally employed; when it is necessary to distinguish the scattering from the abstraction of the energy of the radiation the latter will be termed 'true absorption.'

The most important case experimentally of absorption is that of a slab of thickness *d* placed perpendicular to the line[2] *AB*. It is found in many instances that in this case the relation between *d* and the intensity (*I*) of the radiation at *B* is

$$I = I_0 e^{-\lambda d} \quad\text{..........................(168)},$$

where λ is a constant determined only by the nature of the radiation and the substance composing the slab. λ is then termed the 'absorption coefficient' of that substance for that radiation. It should be observed that (168) would hold if each infinitesimal layer *dx* of the slab abstracted from the radiation the same proportion of the radiation incident upon it, and none of the radiation abstracted by a layer nearer *A* subsequently fell on a layer nearer *B*; for in this case we should have simply

$$dI/dx = -\lambda I \quad\text{.....................(169)},$$

of which (168) is the solution. Such conditions would obtain (1) if all the radiation were travelling parallel to *x* and (2) if the slab truly absorbed radiation and scattered none. As a matter of fact these conditions do not obtain in any of the cases in which (168) holds and it is probable that in all those cases a more accurate investigation, if it were possible, would show that (168) is only an approximation to the truth. In other cases (168) does

---

[1] It would be very difficult to give a definition of the path of the radiation applicable to all kinds, but it will appear that it is quite easy to give a separate definition for each kind.

[2] Of course if *A* and *B* have finite dimensions the line *AB* is indeterminate; it is then usually arranged that the surfaces of *A* and *B* are plane and parallel so that *AB* is their common normal.

not hold even within the limits of experimental error and a more complicated relation between I and x must be substituted, often involving more than one constant characteristic of the action between the substance and the radiation.

The scattering of radiation is usually described by stating the relation between the intensities of the original and scattered radiations and the angles which they make with the surface at which they are scattered. In only one case, that of the regular reflection of light or refraction, is this relation at all simple and only in this case is the phenomenon determined entirely by one constant characteristic of the radiation and the scattering body. In some cases the scattered radiation differs very markedly in its properties from the incident radiation; it is then usually termed 'secondary radiation' excited by the incident (primary) radiation. But the distinction between scattered and secondary radiation which is usually made is not based only upon experiment; radiation which is regarded as scattered generally resembles the incident more closely than that which is regarded as secondary, but the difference is merely one of degree. The distinction is really based on the particular theory of radiation which is used to explain the facts; according to one theory a certain radiation might be scattered, according to another it might be secondary. When the radiation is supposed to consist of particles, it is scattered or secondary according as the particles constituting it are or are not supposed to have been present in the incident beam; when the radiation is not supposed to consist of particles, current usage does not seem to make any clear distinction in the use of the terms; we shall term the radiation secondary or scattered according as it is or is not possible to vary the quality of the incident radiation without varying that of the emergent. It should be noted that it is possible to regard any given beam of radiation as partly scattered and partly secondary.

It is in connection with the scattering of radiation that the very important phenomenon of polarisation appears. Since polarisation is usually described only in connection with light incident on smooth surfaces, when scattering takes the very special form of regular reflection and refraction, it is not always realised what are the essential features of the phenomenon. If i is the angle which an incident ray, r the angle which the scattered ray under consideration make with the normal to the scattering surface, θ the angle which the plane containing these rays makes with the same line, then the ratio of the intensity of the scattered to that of the incident ray sometimes depends only upon i, r and θ, the surface being in all cases the same. Thus, if there is only regular reflection, we have the relation that the intensity of the scattered ray is zero unless $i = r$ and $\theta = 0$. But in other cases this ratio depends also on χ, the angle which the plane containing the rays makes with some plane characteristic of the source of radiation, for example one of the planes of symmetry of a crystal through which the radiation has passed, so that for any given values of i, r, θ the ratio is a maximum for some values of χ and a minimum for others; in such cases it is always found that there is only one value of χ (independent of i, r, θ) for which the ratio is a maximum and one for which it is a minimum, these values differing by a right angle. The radiation is then said to be plane polarised and the plane in which the maximum occurs is called the plane of polarisation of the radiation. If the minimum is ever zero and there is no scattered radiation for some values of i, r, θ, χ, then the radiation is said to be completely polarised; if the minimum ratio is always finite, the radiation is said to be partially polarised and is usually regarded as a mixture of polarised with unpolarised radiation.

Theories of radiation.

70. The chief characteristics of a radiation which can be determined directly by experiment have now been enumerated; the business of a theory of radiation is to show how and why these characteristics are related. Of such theories there are two chief types, one supposing that the radiation consists of massive particles emitted with finite velocities, the other that it consists of disturbances propagated through a medium occupying the space between the radiator and receiver in a manner somewhat analogous to that of the propagation of a displacement through a material elastic medium; in the cases we shall have to consider this medium is always supposed to be formed by the Faraday 'lines of force.' Of course it is well known that a long controversy raged as to which of the theories was more suitable as an explanation of light radiation, and since to-day there is a difference of opinion somewhat similar in connection with other radiation, it will be well here to consider very briefly the circumstances which might lead to a decision in favour of one theory rather than the other.

If the fundamental principles of Chapter I are accepted there are two, and, I think, only two, properties of a radiation which can decide perfectly definitely which theory is to be applied to it. Firstly, if the velocity of the radiation is not c when there is no material body between the radiator and receiver, then the radiation cannot consist of a disturbance propagated along lines of force; but there may be some difficulty in deciding whether there is or is not a material body between those systems. Secondly, if the passage of the radiation is necessarily accompanied by the transference of a charge, then the radiation must consist of charged particles, for according to our theory the transference of a charge must involve a transference of either electrons or atoms; but the possibility of charged secondary radiations (Chap. XI) may make it very difficult to establish the condition implied by the word 'necessarily.'

But these methods of deciding between the two classes of theory do not carry us far; in the controversy concerning the nature of light the arguments were based chiefly on the phenomena of interference, diffraction and polarisation[1]. If a radiation exhibits any of these phenomena there is strong *prima facie* evidence that it consists of disturbances and not of particles, for the reason that all such disturbances as we consider must be capable of displaying them, while an explanation of them on the basis of particles needs the introduction of special and somewhat elaborate hypotheses. Thus, if a radiation can be polarised its

---

[1] Prof. Bragg has pointed out to me that it was not until the time of Young and Fresnel that any arguments for or against the rival theories were advanced that would be considered at all adequate to-day. It is difficult to say whether the objections to the wave theory urged by Newton or the objections to the corpuscular theory raised by Huyghens appear the more absurd to those imbued with modern physical ideas. And doubtless our discussions of radiation will appear equally ludicrous to our successors. Differences of opinion in connection with scientific theories really endure only so long as the ideas connected with those theories are not perfectly clearly conceived.

structure cannot be symmetrical about all planes containing the direction of propagation. Since lines of force can transmit only transverse disturbances, the properties of the radiation must vary with the angle which the plane makes with the direction of the disturbance. Diffraction again, i.e. the extension of the disturbance within the 'geometrical' shadow, is characteristic of all disturbances propagated within a continuous medium, and interference must occur when two such disturbances differing in phase fall on the same body. If a particle theory is adopted, it is necessary to suppose that the 'particles' of a polarised radiation have a structure of a very special nature and are not symmetrical about all axes; in order to explain diffraction and interference it would be necessary to make very complicated assumptions about the actions between two particles or a particle and a material body. It might be possible to design suitable assumptions, but any theory based on them would not have the valuable qualities of simplicity and directness.

Other methods of distinguishing between the two classes of theory are best considered in connection with definite types of radiation. We shall now proceed to a description of the types of radiation which are important for electrical theory, of the circumstances in which they arise, and of the prevalent theories as to their nature.

CHAPTER VIII

RAYS FROM RADIOACTIVE SUBSTANCES

71. CERTAIN substances, which are termed radioactive, emit spontaneously radiation capable of affecting a photographic plate and producing ionisation in gases through which it passes. These rays may be divided into three groups which are known as α, β, and γ rays.

α rays.

α rays are now believed to consist of atoms of helium each of which carries a positive charge 2ϵ, projected with a speed which depends on the particular substance from which they are emitted and lies in all the cases which have been investigated so far between $1{\cdot}5 \times 10^9$ and $2{\cdot}3 \times 10^9$ cm./sec. The evidence for this conclusion may be epitomised thus.

The total charge carried by a given beam of α rays may be measured by receiving the rays in suitable circumstances on a body insulated in a vacuum and connected to an instrument for measuring charges; the charge received by the body per second is Ne, where N is the number of particles incident on the body per second and e the charge on each. N may be determined by observations of the phosphorescence excited by the same beam of rays in certain crystals. It is found that this phosphorescence is not continuous but consists of a number of momentary flashes of light, and that the relation between the number of flashes per second and the intensity of the rays is consistent with the view that each flash denotes the incidence of a single particle on the phosphorescent substance. Consequently N is the number of flashes due to the rays per second. Accurate measurements have been made on one source of α rays only; the value of e deduced was $9{\cdot}58 \times 10^{-10}$, which is very nearly twice the value of ϵ deduced by other methods.

The value of e/m for the rays may be found as usual by observations on their deflection in a magnetic field. The direction of this deflection indicates that e is positive and its magnitude that e/m lies between $1{\cdot}6 \times 10^{14}$ and $1{\cdot}4 \times 10^{14}$, the mean for the most accurate observations being $1{\cdot}43 \times 10^{14}$; the differences in the values for rays emitted from different substances lie within the limits of experimental error. On the other hand, since the atomic weight of helium is 4 and the value of e/m for a hydrogen atom carrying a charge ϵ is $2{\cdot}894 \times 10^{14}$, the value for a helium atom carrying a charge 2ϵ should be

$$2 \times 2{\cdot}894 \times 10^{14}/4 = 1{\cdot}45 \times 10^{14}.$$

The conclusion, indicated but not strictly proved by these experiments, that e/m is the same for all α rays whatever their origin, is confirmed by the experiments on their absorption described below, which prove that α rays from different sources differ only in their velocity. The conclusion that the value of m is really that of a helium atom is confirmed in a very remarkable manner by experiments which show that helium, detectable by its spectrum, is produced in a completely closed vessel if α rays are allowed to penetrate the walls.

The velocity of the α rays is determined by the experiments in a magnetic field which also determine e/m, the deduction of both quantities from the observations being made, of course, by means of the theory stated in § 16.

Absorption of α rays.

72. If the source of α rays is a thick layer of a radioactive substance and the intensity of the emergent beam is measured by the ionisation which it produces in a layer of gas so thick that none of the rays can pass through it, then the relation between the intensity of the rays and the thickness of a layer of any homogeneous material placed in the path of the rays is found to be represented with considerable accuracy by the simple formula (168). But in this case the simplicity of the mathematical relationship does not imply a corresponding simplicity in the physical processes, and from this relationship we can draw no conclusions as to the nature of the reaction between the absorbing material and the particle passing through it. For the rays which emerge from the surface of the radioactive substance have traversed different thicknesses of that substance before emerging; if any alteration in the quality of the rays is caused by passage through a material substance they will differ in quality; and, since the rays are emitted from such a surface in all directions, the distances which they travel in a layer of absorbing material placed over such a surface will again be different. In order that we may be sure that we are dealing with a beam of rays originally similar and similarly treated by the absorbing layer, we must use as a source of the rays a layer of material so thin that the rays would not be affected by their passage through it and must limit the emergent rays to a parallel beam by suitable screens, so that they all travel the same distance in any absorbing material placed in their path.

Let us suppose that an experiment is arranged so that these conditions are fulfilled, and let us examine the effect of an absorbing layer of thickness x on the effect produced by the rays in a plane parallel to that from which they are emitted. Three magnitudes connected with different kinds of effect may be investigated with advantage:—I_1, the number of phosphorescent flashes excited per second in a suitable screen, which will be equal to the number of particles penetrating the layer; I_2, the number of ions produced per second in a mass of gas so large that the rays cannot penetrate it, which will be proportional to the total energy possessed by those particles; and I_3, the number of ions produced per second in a very thin layer of gas, which will be proportional to the ionising power of those particles, that is to say, to the rate at which the particles are spending energy in producing ionisation. We shall find that if x is greater than a certain quantity d, determined by the source of the rays and the absorbing material, all three effects are zero; the rays cannot penetrate at all a layer of thickness greater than d, or at least they cannot produce any of their characteristic effects after penetrating such a layer. But as x is decreased from d to 0 the three effects do not vary in proportion. I_1 is constant for all values of x between d and 0, except perhaps for those which are very near d. That is to say, the number of particles penetrating the layer does not vary with the thickness within this range;

either all the particles penetrate a layer or none of them do; the layer treats all the particles in the same way. I_2 increases continuously as x decreases; the total energy of the rays is continuously decreased as they pass through the absorbing layer and consequently, in virtue of the conclusion derived from the study of I_1, the energy of each particle must decrease continuously. If we identify the energy of the rays with the kinetic energy of the particles, we must suppose that the velocity of the particles decreases continuously as they pass through the absorbing material; this conclusion is confirmed by observations of their deflection in a magnetic field which also show that e/m undergoes no change. I_3 at first increases as x decreases and then decreases again; the ionising power of the rays is a maximum for a certain velocity and is less for velocities which are either greater or smaller. It will be observed that I_2 and I_3 must be connected. If dx is the thickness of the layer of gas used in observing I_3, and the magnitude of I_3 is $f(x)\,dx$, then I_2 must be $\int_d^x f(x)\,dx$. The length d is called the 'range' of the rays in the particular absorbing medium; if the medium is atmospheric air, it lies between 2 and 9 cm.

The quantitative relations which have been found between the velocity of the rays and the distance which they have travelled in an absorbing medium lead to interesting conclusions. If v is the velocity of the particles after they have traversed a layer of thickness x, it appears that

$$v^3 = a\,(d-x) \quad \text{.....................}(170),$$

where a is a constant independent of the substance emitting the rays and dependent only on the absorbing material; if this is air under normal conditions a is about $1{\cdot}24 \times 10^{27}$. If E is the energy of a particle, we have

$$E = \tfrac{1}{2}mv^2 = \tfrac{1}{2}ma^{\frac{2}{3}}(d-x)^{\frac{2}{3}} \quad \text{...............}(171),$$

and the rate at which the energy decreases along the path in the absorbing medium is

$$-\frac{dE}{dx} = \frac{\frac{1}{3}ma}{v} = \tfrac{1}{3}ma^{\frac{2}{3}}(d-x)^{-\frac{1}{3}} \quad \text{............}(172).$$

The experiments are especially simple if the substance of which the absorbing properties are being investigated is the gas filling the ionisation chambers used in measuring I_2 and I_3: x is then varied simply by moving the source of rays towards those chambers in a vessel filled throughout with the gas. (172) then states a relation between the amount of energy expended by the rays in passing through the thin layer of gas in which I_3 is measured and the distance of the source from that layer. Now it is assumed that this amount of energy is proportional to the ionisation produced in that layer; if this assumption is correct I_3 should vary as $(d-x)^{-\frac{1}{3}}$. It is found that this relation holds accurately except when x is very nearly equal to d and v is small. It will be seen that, according to (172), I_3 should increase continuously as v decreases, whereas it is found that it has a maximum for a small value of v. Part of the discrepancy may be due to the causes considered in § 74, but part is doubtless due to the fact that the assumption that the particle spends all its energy in ionisation cannot be accurately true; for in order to make a pair of ions the particle must possess a definite amount of energy (p. 105); if the energy is less than this amount a further decrease in energy cannot be accompanied by ionisation. It must also be pointed out that the velocity of the particle cannot be determined directly very near the end of its range and it is doubtful whether (170) is really applicable in this region.

So far the various quantities have been expressed as functions of x, the distance of the particle from the source; if we express them in terms of $y=d-x$, the distance from the end of the range, we have

$$v^3 = ay \text{.........(173)}, \quad \text{and} \quad I_3 = \tfrac{1}{3} m a^{\frac{2}{3}} y^{-\frac{1}{3}} \text{......(174)}.$$

Since a is the same for all rays, we see that all rays, whatever their initial velocity, must have the same speed at the same distance from the end of their range. Experiments show that I_3 is also the same for all rays at the same distance from the end of their range, so that m must be the same for all rays and the conclusion that all α rays, from whatever source they are emitted, are of essentially the same nature receives strong confirmation. Indeed the conclusion is really almost independent of any measurements on the velocity of the rays. Observation alone shows that rays which can still travel a distance d are exactly similar in all their effects whether they are rays which at the time of their emission had a range d or whether they are rays which had a greater range d', of which they have already performed a portion $d' - d$. It would be natural to conclude that the rays differ only in initial velocity.

Stopping power and ionisation.

73. According to (170), the two constants a and d, which have been introduced to represent the influence of an absorbing layer on the α rays, must be connected by the relation $ad = b$, where b is a constant independent of the absorbing substance and equal to v_0^3, where v_0 is the initial velocity of the rays. Hence if we compare the effect on the rays of two substances in which the ranges of the rays are d_1 and d_2, then the effect of a layer x of the first substance on the rays must be precisely the same as the effect of a layer x/s of the second substance, where $s = d_1/d_2$. For if a_1, a_2 are the values of a for the two substances, we must have $a_1 d_1 = a_2 d_2$ and

$$v^3 = v_0^3 - a_1 . x = v_0^3 - a_2 . x/s \qquad (175).$$

s is termed the stopping power of the second substance relative to the first; air is usually chosen as a standard substance, so that the stopping power of any substance is the ratio of the range of the α rays in air to the range in that substance. It is found that the stopping power of a substance, thus defined, is not strictly independent of the velocity of the rays of which the range is measured, so that, though (175) may be true for all homogeneous absorbing substances, we cannot apply it in all cases when the absorbing layer is made up of two or more substances by simply substituting $\Sigma x/s$ for x/s. It appears that the stopping power of light substances generally increases relatively to that of heavy substances as the speed of the rays is decreased; heavy substances are relatively more effective in stopping fast rays. But the variation is not large and any uncertainty as to the exact velocity of the rays for which the stopping power is measured would not affect the conclusions which are stated below.

If we calculate, instead of the thickness of a layer of the substance required to stop the α rays completely, the number of atoms which a ray would have to pass through in penetrating such a layer, we arrive at some interesting results. If N is this number of atoms for any substance and N_0 the corresponding number for air (which may for such purposes be regarded as an element having a diatomic molecule of weight 28·8), the ratio $N_0/N = s_a$ may be called the 'atomic stopping power' of the substance. s_a is connected with s by the relation $s = s_a \frac{W_0 \rho}{W \rho_0}$, where ρ, ρ_0 are the densities and W, W_0 the atomic weights of the substance and air respectively. In the following table the atomic stopping powers of several elements are given, together with the ratio of this quantity to the square root of the atomic weight. The ratio is very nearly constant, except for the lighter atoms.

	s_a	$s_a/\sqrt{W}$		s_a	$s_a/\sqrt{W}$
H	0·24	0·240	Cl	1·78	0·299
C	0·85	·246	Fe	2·29	·307
N	0·94	·251	Ag	3·28	·316
O	1·05	·262	Sn	3·58	·326
Al	1·50	·287	Pt	4·14	·297
S	1·76	·312	Pb	4·27	·298

The stopping power of a compound material depends only on the nature and the number of the atoms which it contains; it is the same as that which would be shown if the various atoms were sorted out and arranged in consecutive layers. This observation is very important, for, since the stopping power is quite independent of the manner in which the atoms are connected in the molecule, we must conclude that the α ray reacts with each atom separately and not with the molecule as a whole.

We have assumed so far that the ionisation produced by the rays in a gas is proportional to the amount of energy which they expend in it. This assumption is probably true when we are dealing with one gas only, but it is certainly not true if we compare different gases; for if we make the rays perform their whole range in the gas the energy which they spend in it must be the same whatever the gas, and yet the ionisation which they produce is not always the same. The following figures compare the total amounts of ionisation which a given beam of rays can produce in several gases.

H_2	100	N_2	96	CO_2	108	C_2H_2	126
O_2	109	N_2O	105	CS_2	137	C_6H_6	129

We must conclude, therefore, either that the energy required to produce one ion varies with the nature of the gas, or that the rays spend some part of their energy, which depends on the nature of the gas, in producing some change other than ionisation. We are at present so ignorant of the exact nature of ionisation that it is difficult to distinguish between the two hypotheses, but perhaps the evidence is rather in favour of the second. Thus, although the stopping power of a given mass of methane (CH_4) is the same as that of the same mass of a mixture of ethane (C_2H_6) and hydrogen containing the same proportions of carbon and hydrogen, the ionisation in the mixture is the greater. The energy must be absorbed by the individual atoms, but the ionisation is determined by the molecule as a whole. It has been suggested that the energy of the rays is not converted directly into energy of ionisation, but that the rays produce some change in the atoms, and that the reactions of the atoms so changed result in the production of ionisation (see Ch. XII).

The differences between the 'total ionisations' of different gases shown in the last table are quite small; the properties of a gas in respect of ionisation are perhaps more accurately represented by the product of its molecular stopping power (defined by the substitution of 'molecule' for 'atom') and total ionisation, for it may be readily seen that this product represents the relative values of the average number of ions which an α ray makes from each molecule which it encounters in passing through the gas; it may be termed the molecular ionisation. The following figures show how greatly the molecular ionisations of different gases may vary:—

Air	1·00	CO	1·00	He	0·211
H_2	0·233	CH_4	1·10	Br	3·9

It should be noted that both the molecular stopping power and the molecular ionisation are quite independent of the temperature of the gas.

Lastly a few figures may be given as to the absolute number of ions which each α ray makes in a gas. It is found that each α ray which has an initial velocity of $2{\cdot}06 \times 10^9$ cm./sec. can produce $2{\cdot}37 \times 10^5$ pairs of ions in air. The initial energy of such a ray is

$$\frac{1}{2}\left(\frac{m}{2\epsilon}\right)v^2 \,.\, 2\epsilon = \epsilon \,.\, \frac{4{\cdot}24 \times 10^{18}}{1{\cdot}43 \times 10^{14}} = 2{\cdot}96 \times 10^4 \epsilon \text{ erg.}$$

If we suppose (as has been said, the supposition is probably not true) that all this energy is expended in producing ions, the energy required to produce an ion in air is

$$\frac{2{\cdot}96 \times 10^4}{2{\cdot}37 \times 10^5} = 0{\cdot}125\epsilon \text{ erg.}$$

This is the energy acquired by an electron in falling through a P.D. of $300 \times 0{\cdot}125 = 37{\cdot}5$ volt.

Again, we may ask what proportion of the molecules which an α ray meets are ionised by it. The answer depends on what we mean by 'meeting'; if we say that a ray meets any molecule which is contained in a circular cylinder the radius of which is equal to the 'radius' of the helium atom and the length of which is equal to its range, then we can make the calculation. The rays which we have just considered have a range in atmospheric air of 7 cm.; the number of molecules in 1 c.c. of such air is $2{\cdot}75 \times 10^{19}$. The dynamical theory of gases makes the 'radius' of the helium atom a little greater than 2×10^{-8} cm.; adopting this value the number of molecules in such a cylinder is almost exactly $2{\cdot}37 \times 10^5$, which is the number of ions made. There is some reason then for saying that an α particle ionises every molecule it meets. If the molecule is complex the ray can probably produce more than one ion from it.

Scattering of α rays.

74. The statement on p. 163 that all the α rays incident on an absorbing layer, the thickness of which is less than the range of the rays, emerge on the further side is not accurately true. There is some 'scattering' of the α rays, though it is very much less than that of most of the rays which we shall have to consider; passage through the absorbing layer does deflect the particles slightly from their original directions, but except when they are very near the end of their range the deflections are generally small. Of the particles incident on a thick plate of gold (which scatters more than any lighter substance) only one in eight thousand is so far deflected that it emerges on the side at which it entered. This scattering, which is of considerable importance in connection with speculations concerning the structure of the atoms, has been investigated by two different methods both of which are beautiful examples of the experimental art.

In Geiger's work a pencil of α rays is sent against a layer of absorbing material and the number emerging in any direction estimated by means of the flashes of phosphorescence which they excite in a screen placed in various positions. His method is well adapted to determining the relation between the intensity and the direction of the scattered radiation, but it cannot decide directly whether the observed deflection is due to a large number of small deflections or whether it arises suddenly at a single encounter with a molecule; this question is of great theoretical importance (Ch. XIII). In C. T. R. Wilson's method, which, as we shall see, is applicable to other important problems, the actual path of the rays through a gas is made visible by the means described on p. 90. Water is condensed round the ions formed by the rays and photographed before the ions have time to move from the positions in which they were formed. The resulting photographs show lines of drops which are almost straight for the greater part of their length, but occasionally are deflected through finite angles; it is unusual for one line to suffer more than one deflection. We may conclude that the large deflections observed by Geiger are generally imposed at a single collision; if they were the result of many collisions the path of the α ray, which is made visible by the drops formed on the ions, would be an irregular curve which would nowhere show sudden changes of curvature. Since we have just seen that the ray doubtless makes many thousands of encounters with molecules in the course of its range, we must conclude that most of these encounters do not result in a deflection. An atom travelling as fast as an α ray can generally pass right through an atom placed in its path without being turned aside.

But, as would be expected, the scattering of the rays increases rapidly as their speed decreases, and near the end of their range, when they possess so little ionising power or power of exciting phosphorescence that it is difficult to observe them, the rays probably lose their original direction entirely and become scattered in all directions. It is probably largely on this account that the 'ionisation' curve, relating their ionising power to the distance which they have travelled in an absorbing medium, becomes anomalous when the distance from the end of the range is very small. Ultimately the rays doubtless become ordinary atoms of helium agitated among the molecules of the absorbing substance with a velocity which depends on the temperature of that substance. And it is for the same reason that the 'positive rays' described in § 54 differ apparently so completely from the α rays, although they also consist of positively charged atoms. These positive rays can also ionise a gas through which they pass, excite phosphorescence and affect a photographic plate; but while the α rays have an energy corresponding to the fall of an electron through about 5 million volts, the positive rays have seldom even 1/100 of that energy. Their range is so small that they cannot pass through the thinnest available films of solid substances and their scattering is so great that they can only preserve their direction in a high vacuum.

Nature of β rays.

75. Just as the α rays are similar in their constitution to the positive rays of Chap. IV, so the β rays are similar to the kathode rays. β rays consist of electrons moving with speeds which are usually considerably greater than can be imposed in a discharge tube and may approach very nearly the speed of light; but velocities as low as 10^{10} cm./sec. have been observed (corresponding to a P.D. of 30,000 volts) which can be exceeded by kathode rays. Henceforward the term β rays will be often used to denote any electrons travelling with a speed so great that they possess the power of ionising gases, whatever their source; kathode rays will be regarded as merely a special form of β rays with relatively low velocities and, except when the contrary is obviously implied, the statements made as to the properties of β rays may be taken as describing also the properties of kathode rays.

The total charge on a beam of β rays may be measured in the same way as that on a beam of α rays, but it is less easy to measure the number of the particles, because the flashes of phosphorescence produced by the rays are less intense. The identification of the β rays with electrons is really based upon measurements of e/m, the most accurate method of determining this quantity being a slight modification of that described on p. 27. A small particle of radioactive material is placed at the centre of two circular parallel plates separated by a fraction of a millimetre; the plates are kept at different potentials, so that there is an electric intensity X perpendicular to them. A photographic film is attached to the inside of a cylinder concentric with the plates, the radius of which is a few centimetres greater than that of the plates; the whole apparatus, highly evacuated, is placed in a uniform magnetic field of intensity H. Only those β rays which travel in straight lines can emerge from between the plates and affect the film, the others strike one or other of the plates; so that, according to (22), if θ is the angle which any ray emerging from the plates makes with H, its velocity v must be $\frac{c^2X}{H\sin\theta}$, and ρ is the radius of curvature of its path after it has emerged from between the plates and is subject to a magnetic field only $\frac{1}{\rho} = \frac{H\sin\theta}{c^2v}.e/m$. Hence the relation between ρ (which can easily be determined by the position of the impress of the rays on the film) and θ should be

$$\frac{1}{\rho} = \sin^2\theta . \frac{H^2}{c^4X} . e/m \quad \text{.................(176).}$$

Now it is found that ~~$\rho/\sin^2\theta$~~ $\rho\sin^2\theta$ is not constant when X and H are constant. Accordingly either e/m varies with θ, or our theory of the motion of a charged particle in electric and magnetic fields is incorrect. This matter will be discussed further in the last chapter; here only the view which is ordinarily taken of the discrepancy will be mentioned.

It is assumed that e is constant and equal to ϵ. It is conceivable that direct evidence for this assumption will be obtained some day by measuring accurately the total charge on the rays and their number. It is assumed also that formula (176) correctly represents the variation of ρ with θ and m; it follows that m must vary with θ. Now this variation of m with θ cannot be attributed to an emission by the radioactive substance of particles of different nature in different directions, unless we suppose that the emission is influenced by the magnetic field, and such a supposition is contrary to all evidence derived from other sources; it must be supposed that the substance is emitting in all directions particles with different values of m. But, according to the theory of § 16, the nature of the particles which emerge from between the plates is determined only by v and not by m; so that, if there were particles having the same value of v but different values of m, we should obtain for a given value of θ several values of ρ, which depends on both m and v. But we do not find several values of ρ; the particles emerging at an angle θ all travel in the same path; the only alternatives left are to suppose that m is a function of v or that the theory is wrong. Now we cannot possibly distinguish between these two hypotheses, for the only way to determine v is to suppose that the theory is right and to put $v = \frac{c^2 X}{H}$. We have called v the 'velocity' of the particles, but we cannot determine this velocity as we determine the velocity of a rifle bullet or of light, because we cannot perform the necessary experiment, and we have no knowledge whatever whether, if we could perform it, the velocity so determined would turn out to be equal to $\frac{c^2 X}{H}$. The course usually adopted is to define the velocity of an electron v as such a quantity that the force acting on an electron in an electric field X and a magnetic field H is

$$F = \epsilon (X + [v \, . \, H]),$$

and to leave quite open the question whether this quantity is the ratio of the distance travelled by the electron to the time of travelling. The equations of § 16 are logical deductions from this definition and consequently, if the definition is adopted, they are necessarily true and we must suppose that m is a function of v. The experiments are consistent with the view that the function is of the form

$$m = m_0 (1 - v^2/c^2)^{-\frac{1}{2}} \quad \text{..................(177)},$$

where m_0 is the quantity which was given on p. 21 as the mass of an electron; m_0 is the mass when the electron is travelling with a velocity which is infinitesimal compared to that of light. Any particle bearing a negative charge ϵ, the mass of which is represented by (177), is termed an electron; the terminology is justified by the fact that a particle which is an electron, so defined, when travelling with one velocity is also an electron when travelling with another velocity (unless it has encountered a material body in the interval). Whenever the velocity of an electron is spoken of, the quantity defined by the relation

$$F = \epsilon (X + [v \,.\, H])$$

is always meant.

Absorption of β rays.

76. In suitable conditions the absorption of the β rays follows very accurately the relation (168). The coefficient of absorption of the rays increases very rapidly as the velocity decreases; the formula $\lambda \propto v^{-4}$ will represent the facts over a considerable range, but it appears that when the rays are all of the same velocity initially (168) does not hold. The range of the faster β rays is very much greater than that of the fastest α rays; the former can penetrate several metres of air while the latter are stopped by some 9 cm. On the other hand if α and β rays of the same velocity are compared, the α rays are much the more penetrating, probably because they have more than 1000 times as much energy to spend.

If ρ is the density of a substance, the quantity λ/ρ for rays having an initial velocity of about 2×10^{10} cm./sec. does not vary in a ratio of more than 3 to 1 for different substances; if the substance is an element, λ/ρ is a periodic function of the atomic weight, the periods following those of Mendeleef's table. The value of λ is determined only by the number and nature of the atoms and not by their state of chemical combination; this conclusion is natural, for an electron being smaller than an atom can hardly react with the molecule as a whole.

For rays having a speed greater than 10^{10} cm./sec. the ionising power varies approximately as v^{-2}, or inversely as the energy, but for the slower rays still more rapidly. Direct measurements of the change of velocity in passing through absorbing media show also that the kinetic energy decreases proportionally to v^{-2} with the length of path, so that, as would be expected, the ionising power is proportional to the rate of loss of energy. It will be remembered that for the α rays the ionising power is inversely proportional to the velocity and not to the energy, so that there would appear to be some difference in the action of the two classes of rays.

But in the case of β rays, as in that of α rays, measurements of λ do not really give us much information as to the reactions between the rays and the material through which they pass. Since both the α rays and β rays consist of charged particles travelling with great velocities, we should expect the nature of this reaction to be the same for both kinds of particles and the 'absorption' of the rays to be a complex phenomenon, compounded of a scattering of the rays with a true absorption due to an abstraction of their energy as they pass through the absorbing substance and ionise the atoms of it. But it is very much more difficult in the case of α rays to determine separately the effect of these two causes of a diminution in the intensity of the rays by passage through material bodies.

The difficulty is partly experimental and arises from the greater difficulty of detecting the presence and action of a single β ray, since the method of phosphorescent flashes is not so easy to apply; and it is partly due to the fact that the ratio of scattering to absorption is very much greater in the case of the β ray. It is probable that no appreciable fraction of a beam of β rays traverses as much as half of its range without suffering a large deflection from its original course; consequently the thickness of a layer which is just sufficient to stop all the rays cannot be determined with any accuracy and, even if it could be determined, would yield no information as to the range of the rays; for, since the rays do not generally pursue a straight path in the substance, the distance which the rays emerging from an absorbing layer have travelled in that layer is not the same for all the rays and is, on the average, very much greater than the thickness of the layer. The ionising power of the rays emerging does not, as in the case of α rays, fall suddenly to zero when a certain thickness is reached, but decreases gradually as the number of the rays deflected from their original direction increases and as the complicated paths pursued by these rays become successively greater than the range.

Nevertheless, considerable progress has been made of late in analysing the absorption of β rays in the same way as that of α rays. If it is assumed that β rays of definite initial velocity also possess a definite range in any given substance, and differ from α rays only in the circumstance that the path having the length of this range is a very complicated curve instead of a straight line, then an ingenious method due to Bragg enables the range of the same rays in different substances to be compared. By certain methods (see § 108) we can cause β rays of known velocity to start in random directions from all points in the interior of a solid body; if the same number of rays are started in the same volume of two substances A and B, it is easy to see that the ratio of the total length of all the paths of the β rays in A to the total length of all the paths in B will be equal to the ratio of the range of the rays in A to the range of the rays in B. If now holes of the same size are made in the interior of A and B, the number of paths crossing a hole will again be proportional to the total length of all the paths in the substance, i.e. proportional to the range of the rays. But this number can be determined by means of the ionisation which the rays produce in a gas contained in the hole, the total mass of which is too small to cause any appreciable absorption of the rays. In this manner Bragg has compared the range of fast β rays in different materials and finds that, as in the case of α rays, the stopping power of an atom increases with its mass, but not so rapidly as the mass; for equal masses heavy atoms are less effective in stopping the rays than light; but it appears that the stopping power is proportional rather to the cube root than to the square root of the atomic weight.

The change in the velocity of the rays when they pass through layers of matter has been measured by direct experiment. It appears that, in place of (175), the relation

$$v^4 = v_0^4 - ax \quad \text{.......................(178)}$$

is obtained. For aluminium a has the value $7{\cdot}32 \times 10^{42}$ cm.^{3}/sec.^{4}; its value for other substances does not appear to be related at all simply to their density or atomic weight. (178) is applicable with considerable accuracy to all but the very fastest rays with a speed near that of light; we have already seen that the behaviour of such rays is anomalous. If E is the energy of the rays, it follows from (178) that

$$-\frac{dE}{dx}=\frac{ma}{4v^2} \quad \text{......................(179)},$$

or

$$-\frac{dE}{dx}=\frac{m^2a}{8E} \quad \text{......................(180)}.$$

Consequently if the energy of the β rays is spent in producing ionisation, we should have for the ionising power of the rays, that is the number of ions which they produce in passing through a given thin layer of a gas,

$$I \propto E^{-1} \propto v^{-2} \quad \text{....................(181)}.$$

(181) appears to be true for the faster rays, but it is not true for the slower; the ionising power of the β rays is a maximum when their velocity is about 9×10^8 cm./sec., corresponding to a fall through about 220 volts. For velocities less than this the ionising power falls rapidly, until, as we saw in § 51, when it is less than about 20 volts, it is zero. For these very slow rays, there is no evidence that (178) holds, since they are not able to penetrate the thinnest obtainable layers of material.

It will be noticed that (180) is different in form from (172), but since the speed of the fastest α rays is less than that of the slowest β rays for which (180) can be tested experimentally, it is not a necessary consequence that the variation of the ionising power of α and β rays with their speed is different. Indeed, it has been recently pointed out that there is good evidence for the view that over a considerable range the ionising power of α rays and β rays of the same speed is always proportional, being about 10 times as great for the α rays. This is a very remarkable result, for of course the energy of an α ray is very much greater than the energy of a β ray of the same speed, and the ratio of the energies of α and β rays of the same speed varies with the speed. We shall return to this matter when we consider the mechanism of ionisation, ~~but it may be noted that the relation cannot~~ hold down ~~to the smallest speeds, for the experiments on ionisation~~ by

~~collision show that at small speeds the ionising power of α rays is less than that of β rays, not only of the same speed, but of~~ the ~~same energy.~~ On the other hand, it appears that there is a real falling off of the ionising power of the α rays below a certain speed, and that the failure of (170) to apply to the lowest speeds is not due simply to the scattering of the rays.

The scattering of β rays.

77. Though the scattering of the β rays has also been the subject of much investigation, no useful purpose would be served by quoting the numerical results obtained. But the question arises again whether the scattering of the rays is due to the occurrence of a very large number of very small deflections or to a much smaller number of larger deflections; the evidence is not quite so clear in this case, but it seems probable that when a β ray is deflected through an angle greater than a right angle the deflection is usually due to a single encounter with a molecule. C. T. R. Wilson's photographs of the paths of the rays made visible by means of water condensed round the ions formed along them, are not quite so conclusive as in the case of α rays, and more reliance has been placed on deductions from the average distribution of the rays scattered from a layer. Unfortunately it is possible to deduce the distribution caused by a large number of deflections only when the angle through which the rays are scattered is very small and this case does not occur with the β rays; the thinnest material layers which can be investigated deflect a considerable proportion of the rays so far that they travel in the direction opposite to that of the incident beam. But it is to be noticed that when such thin layers are investigated the proportion of the rays scattered in any given direction does not vary with the thickness of the layer. This result would follow immediately if it were supposed that the large deflections were caused by single encounters, for the chance that the same particle would suffer two large deflections in such a thin layer would be very small; the only effect of adding another such layer would be to increase the number of rays which suffered even one deflection. If on the other hand the large deflections were caused by many encounters, the rays which had been deflected in the first layer would be deflected again in the second, and it is probable that the proportion deflected through any given angle would increase when the second layer was added.

The scattered β rays are always slower than the incident rays, the difference in similar conditions being greater the less the atomic weight of the scattering substance; but the difference in the absorption coefficient of the scattered and incident rays is not nearly so great as it would be if the former consisted of a mixture of rays having all velocities less than that of the primary in approximately equal proportions; far too few slow rays occur in the scattered beam. We may perhaps find an explanation of this fact by observing that the ionising power of the rays and, therefore, the rate at which they lose energy increases very rapidly as the velocity decreases. The β ray may be regarded as a system which is almost unstable; as soon as its velocity begins to fall the rate at which it falls increases; after it has fallen past a certain point it will vanish completely in a very short distance, so that the ray will not emerge at all from the absorbing substance. But we shall see in Chap. XI that there are other factors to be taken into account when the β rays are scattered. The assumption that we made above that all the energy lost by the β rays is spent in ionisation is certainly not correct; (178) and (180) cannot both be accurately true. Experiments on scattering certainly show that an atom can affect a β ray in two ways; it can cause a small deflection in it or it can cause a large deflection. It is probable that the small deflections are connected with ionisation and with the loss of a definite amount of energy which is given to the ionised atom. Whether the large deflections also are accompanied by a loss of energy, and what form this energy takes, we shall consider later.

There is no evidence that the β rays produce any secondary radiation (see p. 157) which consists of charged particles and has the power of ionising gases. There is a secondary radiation which has the power of ionising gases and is not composed of charged particles, and another which consists of charged particles but does not ionise gases; these radiations will be considered subsequently. It cannot be proved definitely that all the 'scattered' radiation is really scattered and not secondary; there is simply no evidence to the contrary.

78. The third class of radiation emitted by radioactive substances is termed γ radiation. It does not carry a charge and is not deflected by electric and magnetic fields, but there is considerable difference of opinion as to its nature. All that appears quite certain is that it is similar in its nature to Roentgen radiation; the similarity has long been suspected and has been proved recently by causing γ rays to excite a secondary radiation which is especially characteristic of Roentgen. Accordingly the consideration of γ rays will be postponed until we undertake that of Roentgen rays, when we shall make no distinction of kind between the two groups of rays.

γ rays.

79. We must now consider briefly the process by which the radioactive rays are emitted; and for this purpose it will be convenient to describe, not the actual facts, for which the reader may refer to treatises on radioactivity, but an ideal instance which illustrates better the fundamental principles involved.

Radioactive changes.

A chemically homogeneous radioactive substance R, if examined after it has been left untouched for a considerable period, is found to be emitting rays at a rate which depends only on its mass and is quite independent of its temperature and of other physical and chemical circumstances; the rate of emission of the rays appears also to be constant and independent of the time. But if the substance be subjected to suitable chemical treatment it may be divided into several portions R_1, R_2, R_3 ... which have different radioactive properties: some may emit only α rays, some only β and γ rays, and some all three kinds; the largest portion R_1, if tested by the ordinary methods of chemical analysis, appears almost identical with R deprived of the small traces of impurity which were originally present in it; the other portions, which are far smaller in mass, appear to such analysis to consist of these impurities. The total activity of all the portions, that is the number of rays of all sorts emitted in unit time, is still constant and equal to the activity of the original R; but the activity of each portion varies with the time, the relation between the activity and the time being generally somewhat complex; in all cases, if a sufficient interval of time is allowed to elapse all the portions R_2, R_3 ... will have become totally inactive, while R_1 will have become again identical in radioactive properties with the original R.

In the simplest case only two portions R_1, R_2 can be produced in this way. It is then found that the activity I_2 of R_2 is related to the time by the equation

$$dI_2 = -\lambda_2 I_2 \qquad (182),$$

or

$$I_2 = I_2' e^{-\lambda_2 t} \qquad (183),$$

λ_2 is called the coefficient of decay of R_2. Now in the case of the original substance R it is found that the activity is simply proportional to the mass of the substance present, and there is every reason to suppose that the activity of R_2 is also proportional to its mass[1]; in this case the substance R_2 must be disappearing at a rate which is at any moment simply proportional to the mass of it present. On the other hand, since the total activity of R_1 and R_2 is constant, R_1 must be constantly producing R_2, and, unless the rate at which R_2 disappears depends on its association with R_1 (an hypothesis which is contradicted by all the evidence), the rate at which R_1 produces R_2 must be proportional to the mass of R_1. For when R_1 and R_2 were still unseparated, their total activity was constant and proportional to the mass of R_1 and, since the activity of the two substances is not of the same nature, the mass of R_2 present must have been a constant fraction of that of R_1. But R_2 disappears at a rate which is proportional to its mass and must have been produced at the same rate, which must, therefore, have been proportional to the mass R_1.

In the original substance, then, R_1 was producing at a constant rate and R_2 was disappearing at the same rate. The mass of R_2 present was determined by the condition that it should disappear at the same rate as it was formed; when this condition is fulfilled the substance undergoes no change with the time and is said to be in 'radioactive equilibrium.' The question now arises whether R_1 could have produced R_2 without suffering a diminution itself. There is no direct evidence of such a diminution, but the absence of evidence may well be due to the

---

[1] The mass of R_2 is not, of course, the mass of the substance separated out by the chemical treatment, for this substance is known to consist chiefly of traces of impurity which have in themselves no activity.

fact that the mass of R_1 is so much greater than that of R_2 that the subtraction from it of the mass of R_2 produced in any time which we can observe is not appreciable. If we accept this hypothesis we can at once connect the simple case which has just been considered with the more complex case from which we started.

When we can separate the original substance R into many radioactive products we find that the change of the activity of those products with the time is consistent with the view that each of them is decaying in accordance with (182) at a rate which depends only on the nature of the substance and the mass of it present; and that each of these substances, except one, produces as it decays another of the substances. R_1 decays with a coefficient λ_1 producing a mass of R_2 which is proportional to the mass which R_1 has lost; R_2 in its turn decays with a different coefficient λ_2, producing R_3 and so on, until finally R_n is reached which decays but does not produce any substance the presence of which can be detected by radioactive measurements. Some substance is probably produced by R_n, but that substance is not radioactive. It is easy to see that, if the mixture of the substances is left a sufficient time and allowed to attain 'radioactive equilibrium,' the substance with the smallest coefficient of decay will determine the rate at which the activity of the whole mixture diminishes, and that no substance which occurs earlier in the series than this substance and from which this substance is produced will remain. A set of products, such as $R_1 \ldots R_n$, of which each is produced by the disappearance of another is called a 'radioactive series.'

The quantity $T = 1/\lambda$ is called the 'life' of the substance and is the time that the activity of it takes to decay to $1/e$ of its original value. The lives of the various radioactive substances known differ enormously; in some cases they are to be measured in millions of years, in others in fractions of a second. As would be expected, the activity of a substance decreases, in general, as its life increases, so that the total amount of energy which a substance gives up during its life does not vary nearly as much as the life. In fact, in the case of substances which emit α rays, the number of rays emitted by a given mass of a substance during its life is probably the same for all substances, but may be twice as great for some as for the rest. The chief difference lies in the range and velocity of the particles so emitted; the substance with the longer life always emits the slower rays (at any rate if substances in the same radioactive series are compared), and there appears to be a definite relationship between the life and the range d which may be expressed by either of the formulae

$$\log \lambda = a + b \log d \quad \text{..................(184)},$$

or

$$\log \lambda = a' + b' d^c \quad \text{......................(185)}.$$

It cannot be said at present which is the more accurate.

When the substance emits only β rays the energy which it gives up during its life is considerably less; for the number of β particles emitted is about the same as the number of α particles, while their energy is very much less. Indeed at one time it was thought that there might occur in a radioactive series 'rayless' substances, which produced and were produced by radioactive substances but were not radioactive themselves. All these 'rayless' products have subsequently been found to emit β and γ rays which had formerly escaped attention owing to their much smaller energy. It appears that the initial velocity of the β rays, like that of the α rays, is characteristic of the substance from which they are emitted, though the β rays emitted from a single radioactive substance have not all the same velocity but often consist of several groups each of which has a different velocity.

Theories of radioactivity.

80. What is the nature of the change which produces R_2 from R_1 and causes the emission of rays in the process? It is quite certain that the change affects the nature of the atoms of the radioactive element in R_1, resulting in the production of a new element to which the properties of R_2 are due; it is almost certain that the atom of R_2 is always smaller than the atom of R_1 and that the rays are to be regarded as the fragments of the larger atom which are shot off during the change to the smaller. For there is no doubt that the members of a radioactive series differ so greatly that they would be judged by the criteria of chemists to be different elements and, consequently, to be composed of atoms of different structure and not merely of the same atoms differently arranged. And this conclusion is confirmed by the physical evidence. The only properties of atoms which cannot be changed by subjecting them to any change of temperature, pressure, electric and magnetic intensity and other physical conditions are the intrinsic properties of their atoms and not those which depend upon the way atoms are combined. Radioactive substances undoubtedly must differ in the nature of their atoms and, since they are produced one from the other, the process of radioactivity must involve a change in that nature. The view that the rays emitted are the fragments of the atoms thrown off during this change is probable as soon as this conclusion is established; it has been confirmed by experiments which show that, when we can compare the atomic weights of two elements in the same radioactive series, the atomic weight of the later element is less than that of the earlier by the weight of a number of helium atoms equal to the number of the intervening members of the series which give off α rays. That is to say, the change in the atomic weight is what would be expected if an atom in changing into the next atom of the series emits only one α particle (or none if the change is not accompanied by the emission of α rays) and loses at the same time a mass equal to the mass of that particle. Since the mass of an electron is very small compared with that of an α particle and the number of β rays emitted is not much greater than the number of α rays, the change produced in the atomic weight by the emission of β rays is inconsiderable. All the evidence in favour of this theory of the nature of the change produced accompanying radioactivity (which was first proposed by Rutherford and Soddy) cannot be set out here; it is sufficient to say that it has been confirmed from every possible source of evidence.

But it is not so certain what makes the atom of a radioactive substance disintegrate or whence it derives the energy represented by the rays which it emits, an energy which is vastly greater than can be obtained by causing the atom to combine chemically with others. The change must be determined by some cause affecting each atom separately, for if it were due to a reaction between several atoms, the rate of it would not be simply proportional to the mass of the substance present. It is easy to see that the relation (183) for the rate of decay of a substance would hold if the atoms of the substance each 'lived' a period T before changing into the next element of the series, so long as the times at which those atoms were originally produced were distributed at random during the time preceding the observation. For the N atoms present in the substance at any time must all have been born in the preceding period T; the number which were born in any period dt in that period T is $N dt/T$, for this is what is meant by saying that the births were distributed at random; therefore the number born in the first of such periods dt was $N dt/T$; but this is the number which will attain the age T at the instant under consideration and will die then. Hence

$$dN = -N\frac{dt}{T} \text{ or } \frac{dN}{dt} = -\lambda N.$$

Accordingly it has been proposed that some change is continually going on in radioactive atoms which causes no appreciable change in their properties for a time T, but at the end of that time causes some sort of explosion in the atom which leads to the ejection of an α particle together with electrons; and hypotheses have been put forward as to the nature of this change which gradually destroys the stability of the atom (see Ch. XIII). But it would seem a necessary consequence of this view that the rate at which the atoms break up should be different if the times of their births were not distributed at random during a period T preceding the moment of observation. For suppose that we separate a substance R_1 from the associated R_2 at a time t_0, leave it until the time t_1 and then separate from it the R_2 which has been formed and examine the decay of this substance. (Such experiments are possible.) Then none of the atoms in R_2 are older than $t_1 - t_0$; consequently none should break up until a further time $T - t_1 + t_0$ has elapsed, where T is the life of R_2; until that time there should be no emission of rays. On the other hand no fresh atoms of R_2 are produced after t_1 and consequently none of them should survive after a time $T + t_1$; the substance should be radioactive only within the period $T+t_0$ to $T+t_1$. But it is found that from the time t_1 onwards the substance decays perfectly regularly according to the formula (182); if there are N atoms present, Ndt/T will break up during the next instant dt, and it does not matter in the least whether the N atoms have all only just been born or whether they have already lived a period considerably greater than T. The chance that an atom will break up at any instant does not depend upon the time which it has lived already. T must be regarded as only the *average* life of an atom.

If we are to suppose that the sudden breaking up of the atom is determined by processes going on within it, we shall get a better representation of the facts if we imagine that the parts of the atom are continually rearranging themselves in new groupings and that some very small proportion of these groupings are unstable. The chance that one of these unstable groupings occurs at any moment will be independent of the previous history of the atom, if we regard the groupings as determined by 'chance.' But such speculations do not lead us very far; we have no reason on other grounds to suppose that the structure of the atom is not perfectly constant, so long as a radioactive disintegration does not occur, and we can make no suggestion whatever as to the basis of the remarkable relation which has been mentioned between the average life of an atom and the range of the α particle which it emits when it dies.

An alternative theory was put forward in the early days of radioactivity suggesting that the breaking up of the atom was determined by some exterior influence, probably some radiation acting on the atoms from without, and that the energy of the rays emitted was derived from this radiation. All that can be said about this theory is that there is no evidence whatever of the existence of such a radiation, and that the supposition of its existence does not help us to explain any of the numerical relations characteristic of radioactivity. Moreover such a theory would lead us to expect that under some conditions it should be possible to shield the substance from the influence of the radiation and so decrease its activity; no such conditions have been found.

Are all substances radioactive?

81. The lives of the radioactive substances known and the rate at which they emit rays vary in a ratio of at least 10^{12} to 1. The question has often been asked whether there are not other substances which are really radioactive, but are not recognised as such because their lives are so long and the rate at which they emit rays so small; a substance of which the life was as much greater than that of uranium as that of uranium is greater than that of actinium emanation could not be recognised as radioactive by any conceivable experiment. There seems no reason *a priori* why there should be no α rays emitted with a velocity less than $1{\cdot}2 \times 10^9$ cm./sec.; if the relations (184, 185) hold universally, the life of a substance emitting rays with a third of that velocity would be so long that its activity would be quite inappreciable. May not all elements really be radioactive?

Of course we cannot possibly decide by experiment whether there are radioactive substances of which the radioactivity cannot be detected by experiment. All that we can say is that all the substances which are known to emit α rays have an atomic weight greater than that of Bismuth[1] and that the life of an atom does not necessarily increase as its weight decreases; there is no evidence of this kind suggesting that the lighter atoms might have very much longer lives. On the other hand it is highly probable that lead, a substance which shows no appreciable radioactivity, is the last member of a radioactive series; for, if we subtract from the atomic weight of uranium that of the α particles which it and its descendants are known to emit, we arrive at the atomic weight of lead; and again the composition of minerals containing uranium suggests that lead may be the ultimate product of its disintegration. And if lead is a member of a radioactive series it is not very improbable

---

[1] Potassium and rubidium have been shown to emit β rays in a manner very similar to that of some radioactive substances; they are now usually regarded as radioactive. There is no evidence of the production of new elements from them, but the absence of such evidence may be due to the fact that they emit only β rays, the loss of which has hardly any effect on the atomic weight. And it cannot be shown that they have a definite life. The problems presented by these two elements are very interesting, but difficult to treat experimentally because the activity is so small.

that it is radioactive itself. Again, in their general chemical and physical properties, the radioactive elements do not differ markedly from other elements; radioactivity apart, radium can hardly be distinguished from barium. The probabilities are all in favour of a general radioactivity of the elements, but it seems that only observations over millions of years could establish directly the existence of such radioactivity. We might then be able to detect the production of new elements in a substance which did not originally contain them; for instance we might be able to show that lead produces by its own disintegration the small trace of silver with which it is always found to be associated in minerals.

Meanwhile experiments have several times been described which appear to show that the non-radioactive elements can be stimulated to disintegration by the radiation from radioactive elements; it has been announced that copper, acted on by the rays from radium, is converted into lithium. But experiments conducted with due precautions have always failed to confirm these conclusions, which, it may be noted, would not prove that the elements concerned were really radioactive. For it is the most characteristic feature of true radioactivity that the changes involved in it are independent of all external influences.

REFERENCES.

The treatises on Radioactivity of Mme Curie and of Rutherford, to whom the greatest advances in this branch of science are due, contain information concerning all the work of any importance which has been performed up to the time they were written. The third edition of the latter treatise is so recent that no further references need be given.

An admirable discussion of the nature of the radioactive rays is contained in Bragg's *Studies in Radioactivity.* This little volume professes to deal only with the author's work, but, in the case of a rays at least, this includes almost all the discoveries which are interesting for general theory.

CHAPTER IX

LIGHT

82. THE forms of radiation which are not considered to consist in the emission of charged bodies may be divided into two groups; on the one hand Light, on the other X and γ rays. The possibility of a distinction between the two groups on theoretical grounds will be discussed throughout the next two chapters; a distinction valid for all practical purposes may be made on purely experimental grounds by noting that the former but not the latter group can be regularly refracted. Light, thus defined, includes, of course, not only visible radiations but also those which are usually termed infra-red and ultra-violet; it would also include Hertzian rays, but the investigation of that form of radiation throws no light on any of the problems with which this book is concerned and it will be left wholly out of consideration.

Definitions

Light, according to modern theory, consists of disturbances propagated along Faraday lines. The differences in the properties of light which has the same intensity are attributed to differences in the nature of the disturbances of which it consists, that is to say, in the relation between the time and the electric displacement at a point over which the light is passing. If the light is such that its quality (as distinct from its intensity) is unchanged by reflection or refraction under any attainable circumstances, the light is said to be homogeneous; the view is universally accepted that homogeneous light consists of simple harmonic disturbances, such that the relation of the displacement to the time may be expressed by a relation of the form

$$D = D_0 \cos\left(\frac{2\pi t}{T} + \delta\right) \quad \text{..................(186)},$$

valid for all values of t during the period of observation. There is no direct experimental evidence for the view that the disturbances of which homogeneous light consists have this form[1], for it is quite beyond our experimental powers to measure directly the relation between D and t; the most that our measurements can tell us is that the relation is such that D has the same value for all values of t given by $t = t_0 + mT$, where t_0 and T are constants and m any integer, and that D has the same value with the reverse sign when $t = t_0 + (m + \frac{1}{2})T$. These conditions would be satisfied as well if we wrote in place of (186)

$$D = D_0 \cos^3 \left(\frac{2\pi t}{T} + \delta \right) \qquad \text{(187)}.$$

The adoption of (186) rather than (187) seems to have been dictated by a desire for mathematical simplicity and now to be rendered necessary by the assumptions made concerning the nature of the processes of reflection and refraction. I do not wish to suggest that the adoption of any other hypothesis as to the nature of homogeneous light would be preferable, but merely to point out that it would be physically possible.

Heterogeneous light.

83. No actual source of light emits radiations which are even approximately homogeneous; the importance of the conception of homogeneous light for optical theory arises from the possibility of obtaining such light by means of the actions of prisms, gratings and similar instruments on the light originally emitted. These instruments acted on by heterogeneous light travelling in one direction produce rays of light of different quality travelling in different directions; in suitable circumstances, if the light travelling in one direction only is isolated by means of a fine slit, it is found to be approximately homogeneous; the degree of approximation to homogeneous light obtainable by such

---

[1] Some writers maintain strenuously that it is desirable to define homogeneous light as light consisting of disturbances which can be represented by (186). The uncertainty expressed in the text must then be stated by saying that there is no evidence that homogeneous light is unaltered in quality by refraction or reflection. It appears to me desirable to give to such an important word as 'homogeneous' a definition the application of which to any particular light can be tested by experiment, independently of any theory of the nature of light disturbances.

means from any source whatever depends only on the width of the slit and the resolving power of the prism or grating, but with a given slit and grating the degree depends also to some extent on the nature of the source. For the nature of the spectrum into which the original light is resolved depends on the nature of the source; in some cases the spectrum consists of light of varying quality emitted in all directions (continuous spectra), in others of light emitted in certain directions only and not in others (line spectra, among which for the present will be included band spectra). In the last case, of course, no greater degree of approximation to homogeneous light is obtained by narrowing the slit unless the slit becomes narrower than a single line.

From heterogeneous light, then, we may obtain light of a homogeneity limited only by the experimental appliances which we bring to bear on such light. From this fact the conclusion was formerly deduced that the original heterogeneous light consisted of a mixture of homogeneous light of various frequencies and that the function of the instrument producing the spectrum was merely to sort out these frequencies; the correctness of this conclusion has been subjected more recently to much discussion. It seems to me that the question whether a given light disturbance 'consists' of a mixture of homogeneous vibrations is incapable of a definite answer because the question may mean several different things. If we are asked whether a given volume of water 'consists' of four litres of water, the question may mean either 'Could that volume be obtained by mixing four litres together?' or 'Has it been so obtained?' and the answer to these two questions may be quite different. If it be asked whether the original light disturbance could be reproduced by mixing together the rays of homogeneous light into which it can be decomposed, then the answer is probably, yes; if it be asked whether it has been so produced, whether, that is, if the path of the disturbance be traced backward, a point will be found at which it breaks up again into homogeneous disturbances separated from each other, then the answer is almost certainly, no.

Take the second question first. If the light disturbance has been produced by the mingling of homogeneous disturbances which were originally separated, the source of light must consist of as many parts as there are homogeneous trains into which the light can be analysed and must be emitting light with perfect regularity during the whole time over which it is possible to make observations. Now the number of trains, appearing in our experiments to be homogeneous and to have different frequencies, into which the light from a body giving a continuous spectrum can be resolved is certainly greater than 10^5. The evidence, which will be considered later, derived from the study of the production of light, indicates that it is quite impossible that the number of sources of different kinds contained in any body we know should be nearly as great as this, and, moreover, that it is equally impossible that any of these sources should emit homogeneous vibrations over a period of time nearly as long as that of the shortest experiment. It should be noted that if this view were correct there would be a limit to the number of homogeneous trains of different frequency which could be produced, namely the number of sources of different nature emitting the light; so far an increase in the power of resolving instruments has shown no sign of such a limit. On all grounds this view may be dismissed without the smallest hesitation.

Formation of Spectra.

84. The first question requires much more consideration. It might seem that it could be answered at once in the affirmative by pointing out that, if the trains emitted by the resolving instrument in different directions are reflected so as to travel in the same direction, the resulting light appears identical with that falling on the instrument. But when the question is asked, an answer of a different nature is usually required; what we really want to know is how far we can deduce from a study of the homogeneous trains the nature of the original light. The question is usually put in this way. Any light disturbance can be represented by some equation of the form $D = F(t)$, where D is the electric displacement in that part of the disturbance which falls on the instrument at time t. According to Fourier's theorem, some series of the form

$$a_0 + a_1 \cos\left(\frac{2\pi t}{T} + \delta_1\right) + a_2 \cos\left(\frac{4\pi t}{T} + \delta_2\right) + \dots \quad \dots(188)$$

can be found which is equal to $F(t)$ for all values of t between 0 and T. Each of the terms of this series, if it were the whole value of D, would represent a train of homogeneous light. Are the trains of homogeneous light which can be obtained from the original light by means of a grating or prism during the period 0 to T those represented by the terms of this series?

A superficial consideration would show that this question must be answered in the negative, if the answer is to apply to all possible forms of light disturbance and all values of T. Suppose, for example, that T is five minutes. The value of D for values of t between 0 and 5 minutes can be resolved into a series of the form (188) representing homogeneous trains constant in amplitude over a period of 5 minutes, and the quantities a and δ defining that series will depend as much on the values of D during the fifth minute as on those during the first. But it is inconceivable that the nature of the homogeneous trains emitted by the grating during the first minute should depend on the values of D during the fifth, for the part of the disturbance which will reach the grating in the fifth minute has not reached it at all during the first. And it is equally inconceivable that the nature of the homogeneous trains emitted in the fifth minute can depend on the values of D during the first, for the part of the disturbance incident in the first minute has passed far away from the grating before the fifth minute is reached. A resolving instrument, unlike a mathematician viewing the disturbance as a whole, cannot foresee the nature of the disturbance which has not arrived at it and cannot remember indefinitely the nature of that which is passed.

This difficulty appears very serious, but let us consider what are the actual experiments which we could make on the trains of light issuing from a resolving instrument. First we can measure their intensity; if the train investigated was represented for all time by (186), this intensity would be proportional to the square of the amplitude D_0. Now when we make such measurements we are only concerned in a single measurement with the properties of the light during the time requisite to make that measurement; the result is not affected by the properties of the light at other times, and, on the other hand, it is affected equally by the properties of the light at all moments during that time[1]. It can be shown that, if the time occupied by a measurement is T, then, with an ideal measuring instrument, the measured intensity of the trains between frequencies ν and $\nu + d\nu$ is equal to the sum of the squares of the amplitudes of the terms in the Fourier series expressing the original disturbance for the time T which have frequencies in that range. The proof involves, of course, assumptions as to the action of the instrument on the incident light, but these assumptions are very simple and general, and can scarcely be doubted except in one particular which will be mentioned in the next paragraph. There is nothing paradoxical in this result, for, of course, the measuring instrument 'remembers' the disturbance over the time of the measurement just like the mathematician who calculates the Fourier series.

Now suppose that we take two successive measurements of the intensity occupying a time $2T$. Then in general the result of the first measurement will be different from that of the second and both will be different from that of a single measurement which occupied the time $2T$. To this extent it is true that the results of our measurements will vary with the time over which we take them, but again there is nothing paradoxical in the result; if we calculate the total energy received by the instrument during the time $2T$, the result will be the same whether we sum the energies obtained in the two successive measurements or measure it directly by a single measurement. In cases of this kind, we simply say that the source of light is not steady. In practice we concern ourselves almost exclusively with sources which are steady; with such sources the Fourier series equivalent to the disturbance is the same in respect of amplitudes for whatever period, during which the source is steady, the series is calculated.

One qualification needs to be made in these statements. The measured intensity of the train will be the square of the amplitude of the corresponding term in the Fourier series for the time occupied by the measurement, only if that time is long compared with a period T_0 determined by the nature of the resolving instrument; T_0 may be termed roughly the time that

---

[1] Assuming an ideal measuring instrument; corrections may have to be made in the case of actual instruments.

the light takes to pass through it. In all actual experiments this condition is always fulfilled, but the existence of this condition limits the amount of information which we can obtain from intensity measurements even with ideally perfect instruments.

In other experiments, namely those on the visibility of interference bands, we are concerned with the phases of the terms of the Fourier series rather than with the amplitudes; the greatest retardation with which bands can be obtained measures the time T_1 during which the light entering the interference apparatus is regular and the phases of the same Fourier series are unaltered. But it must be remembered that this light is never that emitted directly by a source; it is always a narrow part of the spectrum thrown by a grating or prism; the value of T_1 for this light is not necessarily that of the original light disturbance. This condition is attained only if the T_1 for the original light disturbance is greater than the T_0 of the grating; if it is less, the T_1 of the light separated from the spectrum by the finest possible slit represents the T_0 for the grating rather than the T_1 of the original light, and interference experiments of this kind tell us nothing about the constitution of the original radiation. And in no case can such experiments tell us anything of the relation of the phases of terms of different frequency.

An answer can now be given to the question asked at the beginning of this paragraph. We can say that, if we know the form of the original disturbance and wish to calculate the effect it will produce on any conceivable measuring instrument, we are perfectly justified in substituting for that disturbance during the time of measurement the sets of homogeneous trains represented by the terms of the Fourier series equivalent to the original disturbance during that period. But this problem is not very important. A far more important question is whether from a study of the homogeneous trains into which a resolving instrument splits the original disturbance we can conclude anything about the terms of the Fourier series representing that disturbance and so about the disturbance itself. The answer is that we can conclude something, namely the amplitudes of the terms of the series to which the original disturbance is equivalent during the time of observation, but that we cannot conclude much more.

It would seem, therefore, that the question from which we started in its most obvious sense should be answered in the negative; the resolution by the grating into homogeneous trains is not equivalent to the resolution by the mathematician into Fourier series, for from the former we cannot deduce the latter; but, on the other hand, we must not conclude that the application of Fourier analysis to the problems of resolution by a grating is not justifiable.

Application to experiments.

85. Such appear to be the main results of the theory of the formation of spectra by prisms and gratings. In order to see how it can be applied to the interpretation of our experiments to give us information concerning the nature of light disturbances we must make some preliminary hypotheses concerning that nature.

Light is emitted from bodies in which electrons are colliding with atoms moving with considerable velocities. Mechanical analogies suggest that the disturbances sent out as a consequence of such collisions would consist of having a finite length; they would begin suddenly, continue more or less unchanged for a certain time and then either die away gradually or cease as suddenly as they began. The disturbances would not all be of the same nature, but, if the source of light were steady, the number of any disturbances of any kind sent out during a sufficiently long time would be simply proportional to the length of that time, the periods occupied by the disturbances being distributed at random throughout it. The collection of all the disturbances emitted during such a long time would be representative of those sent out by the source; it might be resolved into a Fourier series over that time and the amplitude of the terms in that series would be the same for all such long times; but the phases of the terms would vary very greatly with the particular time selected and there would be no relation between the phases of two series separated by a time longer than that occupied by a single disturbance. The fact that any source appears steady when we measure the intensity of the trains into which it is resolved by a grating is due to the complete independence of that intensity of the phases of the trains represented by the terms of the Fourier series; it depends only on their amplitudes and these are constant over considerable spaces of time.

But owing to the fact that we cannot discover anything concerning the relative phases of the trains into which the light is resolved by the grating, we cannot deduce directly from our observations anything concerning the nature of the original disturbances. All that we can do is to guess their nature and then see whether a representative collection of such disturbances as we guess would be resolved into a Fourier series agreeing in the amplitude of its terms with the results of our measurements. An infinite number of guesses would turn out to be equally satisfactory and we should have no means of deciding between them, but a still greater infinite number could be definitely disproved.

Many such guesses have been made and in the case of line spectra some of them turn out to be consistent with experiment. But no success has yet been attained in the case of continuous spectra. Some forms which have been suggested for those elementary disturbances of which a collection represents 'white light' will be mentioned in the next chapter, when we proceed to consider that very important form of radiation; in this chapter we shall be concerned only with the speculations which have been made concerning line spectra.

But before proceeding to this matter we must notice an objection which has been raised to the whole theory of the resolution of light disturbances into spectra, on which the interpretation of all experimental results must depend. It affects the theory only so far as it applies to prisms and does not seem applicable to the same extent to gratings.

The resolving action of the prism doubtless depends on the fact that homogeneous trains of different frequency travel through material media with different velocities. In Chapter II we inquired into the reason for these different velocities and concluded (the conclusion is universally accepted) that it is connected with the forced oscillations into which the periodic electric intensity in the homogeneous train throws the electrons contained in the substance. Now since the electrons possess a finite mass the excitation of these oscillations will take a finite time; if for the homogeneous train, containing an endless succession of waves, were substituted a very small portion of the train containing only one or two waves, it is probable that the oscillations would not be excited at all; in popular language, the electron would not have time to get up speed. Accordingly it has been argued that if the light disturbance incident on the prism consisted of only a very small part of such a train, perhaps not even one complete wave, the electrons would not respond, the mechanism by which the different velocities of different frequencies are determined would not be excited and, therefore, the prism would fail to resolve the disturbance. Now we have noted that we are not concerned with what happens to an isolated disturbance, but only with the radiation from steady sources over a considerable period of time. Still if this radiation consisted of a succession of such disturbances as are contemplated, it might satisfy the conditions of a steady source, and the argument given would still show that it would not be resolved into the trains represented by the equivalent Fourier series. The matter is of great importance, but opinion still seems to be divided concerning it. No decision concerning it can be attempted here. But it may be pointed out that there is some evidence that the difficulty is not of much importance when we are dealing with any form of luminous radiation. For the argument seems only to affect the question of resolution by a prism and not that by a grating; if there were light disturbances which were resolved by gratings and not by prisms, we should expect to find differences in the spectra shown by different types of instrument. No such difference as would be expected if the objection raised were valid seems to have been observed.

86. We will summarise briefly some of the facts concerning line spectra and consider their interpretation.

Line Spectra. Breadth of lines.

A line spectrum obtained by an instrument of low resolving power appears to consist of a finite number of beams of homogeneous light of different frequencies, the number of the lines and their frequencies being determined mainly by the chemical constitution of the source of light, but also to some extent by the means by which it is excited. But if one of these beams is examined by an instrument of higher resolving power, it is usually found not to be strictly homogeneous; in many cases each line can be further resolved into a number of lines, which are usually spoken of as the 'components' of the original line; the distinction between 'separate lines' and the 'components of the same line' is convenient in practice, but represents only a difference in degree; the lines which are termed components of the same line merely differ less in frequency than those which are spoken of as separate lines. If the resolution is pushed still further a limit is reached when it is no longer possible to resolve the lines which formerly appeared simple into a finite number of components, but the light of which one of these components consists is still found to be heterogeneous. The component has a finite 'breadth,' that is to say, it can be resolved into a number of trains (this number depending only on the resolving power which can be applied to it) of frequencies varying within a certain small range, and the intensities of these trains vary continuously with their frequency, being a maximum for one definite frequency only, which is regarded as the frequency of the line. The breadth of the line, or the range of the frequencies within which the intensity is finite, varies enormously with the condition of the source; if this is a gas excited by an electrical discharge, the breadth increases with the pressure of the gas and also with its temperature.

On the view that has been suggested that the disturbances constituting light consist of oscillations lasting only for a finite time, the finite breadth of the lines can be readily explained. If these oscillations during the greater part of the time that the disturbance lasts can be represented by (186) and if the period at the two ends of the disturbance during which it is excited and decays is small compared with the period during which these regular vibrations continue, then a beam of light consisting of a large number of such disturbances would be resolved into just such a line as is found in practice; the breadth of the line would really be infinite, but the intensity of the portions which lie at any considerable distance from the maximum (which represents the frequency of the harmonic vibrations) will be so small as to be inappreciable. Further, from the range within which it is appreciable, that is the breadth of the line, can be deduced some information concerning the period within which the regular harmonic vibrations continue undisturbed. In the case of the narrowest lines which have been observed this period must be several hundred thousand times that of a single vibration[1].

It may be supposed that the number of regular vibrations which the atom of a gas can emit is determined by the time that it is free between two collisions with other atoms, or, possibly, with electrons; such collisions would probably lead to a disturbance of the oscillations and a break in the continuity of the regular train. It is, of course, in accordance with this view that the breadth of the line decreases as the pressure of the gas is decreased and the length of the free path increased, but neither the theory nor the measurements are sufficiently accurate to enable this view to be completely tested. It appears however that, if collisions were the only source of breadth, the lines should be narrower than they actually are at the lowest pressures and some other source of the finite breadth of the lines in such conditions must be found.

Such a source is found in the 'Doppler effect,' the theory of which will be considered in the last chapter. If two sources emitting harmonic vibrations of the same frequency are moving one toward and the other away from the observer, the frequency of the first source will appear less than that of the second. Now the ultimate sources of the light emitted by a gas are probably the atoms or molecules of it; in virtue of the temperature of the

---

[1] Many modern treatises on Optics are so careful to explain that the number of interference bands which can be obtained (an inverse measure of the breadth of a line) is determined, in the case of a line separated from a continuous spectrum, only by the resolving powers of the instrument employed to form a spectrum, that readers of them are apt to think that the same is true when the bands are those formed by the line from a line spectrum thrown by an instrument of moderate resolving power. It is perfectly true that the *ultimate* limit to the number of bands which could be obtained by the use of the most powerful instruments is determined only by those instruments, but this is only because we possess instruments so powerful that the homogeneity of the light which can be obtained by their use is greater than that of any spectral line separated by an instrument of moderate power. If we use only an instrument of moderate power and a fairly wide slit to separate out the spectral line, the number of bands obtainable will often be much greater than can be accounted for by the power of the instrument; in this case the number is determined by the nature of the line and not that of the resolving instrument.

gas these molecules will be moving with various velocities in various directions; even if the vibrations emitted by them were all of the same frequency (if observed when the molecules were at rest) they would appear somewhat different; the total light emitted by them will appear as a line of a finite breadth increasing with the velocity of the molecules or the temperature of the gas.

This effect seems sufficient to explain the finite breadth of the lines, even at the lowest possible pressures. Indeed the theory has been used recently to determine the nature of the system emitting the light. For the mean velocity of thermal agitation at a given temperature depends on the mass of the particle which forms a molecule of the gas, and if by observations on the breadth of the lines we can determine this velocity we can also determine the mass. In this way Fabry and Buisson have arrived at the conclusion that the system emitting the most important part of the hydrogen spectrum is the atom of hydrogen and not the molecule consisting of two atoms.

The pressure of the gas from which a line spectrum is emitted may influence that spectrum otherwise than by broadening the lines of it. It is found that, if the pressure of the gas in which an arc discharge is passing is increased, the frequencies of the lines of the spectrum emitted by the arc (which will be characteristic of the material forming the poles and not of the gas) may be changed. In general the wave-lengths of the lines are increased, the whole spectrum being moved towards the red, and the change in wave-length of any line is proportional to the wave-length of that line and to the increase in pressure. The change seems to be determined by the whole pressure of the gas and not merely by the partial pressure of the vapour of the substance emitting the lines. Other numerical relations have been discovered, but they are unimportant for our purpose.

The theory of this pressure shift is probably quite simple. In § 22, Chap. II, when we were calculating the dielectric constant of a substance, we saw that, if the substance were sufficiently dense and the molecules of it sufficiently near together, the force on an electron in a molecule would not be determined only by the properties of the molecule in which it is contained; there would be an additional force arising from the displacement of the electrons in the neighbouring molecules, which would act in the opposite direction to that due to the molecule containing the electron. Now the frequency of the vibration of the electron will be less and its wave-length greater the less the force acting on it, and accordingly the influence of the neighbouring molecules will be to decrease the frequency of the vibration; and the denser the gas, the closer the molecules will be packed and the greater will be the change of frequency due to this cause. This simple theory seems capable of explaining most of the observations on the 'pressure shift.'

Series in Spectra.

87. We see then that the frequencies corresponding to the maximum intensities in the various lines represent the vibrations which the atoms would emit if they were at rest and could be maintained in a continuous state of undisturbed oscillation. It has long been known that these frequencies characteristic of a single element can be represented as terms of one or more quite simple numerical series, the study of which is of great importance as throwing light on the nature of the vibrating systems from which the light is emitted. The facts are so complex that only the simplest of them can be mentioned here.

The lines in the spectrum of potassium might be divided as the result of a careful study of the spectrum in different conditions into six groups. All the lines in any one of these groups are of exactly the same quality; that is to say, they are all strong or all faint, all broad or all narrow, all are resolved into the same number of components in a magnetic field and all are altered in the same way by pressure; lines in different groups do not resemble each other in respect of these qualities. The frequencies (n) of the lines in any one group can be obtained by substituting for m in the following formula the integral numbers from 2 upwards:

$$n = n_0 - \frac{N_0}{(m+\mu)^2} \qquad \text{..................(189).}$$

n_0, μ, N_0 are some numbers characteristic of the series. It will be observed that towards the red end of the spectrum, when m is small, the lines in the group are far apart, but at the violet end, when m is large they become near together and approach ultimately the frequency n_0. As many as 50 lines have been observed in some series, and probably a great many more could be observed if instruments of greater resolving power could be brought to bear.

Moreover there are simple relations between the values of the quantities n_0 and μ for different groups in the spectrum of one element and other simple relations between these quantities for corresponding groups in the spectra of different but similar elements. When, for instance, the spectra of sodium and potassium are compared some of these quantities are seen to be functions of the atomic weights of these elements. And perhaps the most remarkable relation of all is that the quantity N_0 is the same for all groups which can be represented by series of this type in the spectra of all elements.

These facts prove conclusively that the systems vibrating in the frequencies represented by the various lines of the spectrum are not independent; in fact we may be certain that the lines in a single group are all emitted by the same system, which is not emitting simple harmonic vibrations but more complicated vibrations which may be analysed into a finite number of simple vibrations. In acoustics we are familiar with such vibrating systems; a musical instrument never emits a 'pure tone,' but a vibration, which may be analysed into many pure tones, between the frequencies of which there are definite numerical relations; from these relations some information can be deduced concerning the structure of the instrument; by a study of the 'overtones' an open organ pipe could be distinguished from a closed, and a wind from a reed instrument. It is to be hoped that in the future the study of spectral series will discover to us the structure of the atom, but at present we are very far from such a discovery. Various hypotheses have been put forward from time to time but none of them seem to lead to results which can be reconciled with experiment. It is certain that the vibrators are electrons—so much may be deduced with confidence from the study of the Zeeman effect—and that their vibrations must be very complicated, but there our knowledge ends; we do not know even whether the vibrating system is a single electron moving in a very complicated steady field of force, or a system of electrons exerting disturbing forces on each other, or again whether the same electrons compose the systems emitting the different groups of lines.

Band Spectra.

88. Band spectra may be regarded as a special case of line spectra, but it is probable that the distinction between band spectra and the true line spectra is as important theoretically as it is practically. Band spectra differ from line spectra in that they usually do not show the Zeeman effect and that they are not shifted by pressure. These differences might suggest that the vibrators emitting them are not electrons but particles of much greater mass, but against this view is to be placed the fact that a very small change in the conditions under which the light is excited will cause the spectrum to change from one in which only lines are visible to one consisting wholly of bands. Moreover it has been shown recently that some band spectra do show the Zeeman effect, as if they were emitted by electrons. A more plausible explanation may be found in the nature of the vibrations executed by the systems which emit band spectra. In our consideration of the Zeeman effect we assumed that the electrons emitting the light executed simple harmonic vibrations. If the vibrations were more complicated the arguments would not apply, and it is possible that, even if the vibrating systems were electrons, the effect of a magnetic field on them would be too small to be appreciable. But really all that can be said about band spectra is that our ignorance concerning their origin is even more profound than that concerning the origin of line spectra. Simple numerical relations have been traced between the frequencies of the different lines in the same band and between those of the 'heads' of bands in the same spectrum, but, as before, no adequate theoretical explanation of the relations has been suggested.

Are the radiators charged?

89. So much—or so little—can be discovered concerning the nature of the radiators which emit line spectra from a study of the spectra themselves. Such other information as we possess is derived from a study of the conditions in which the spectra are emitted.

Except in a few very exceptional cases, line spectra are emitted only from flames and arc or spark discharges; the spectrum of the same element varies considerably according to which of the methods of excitation is adopted. All three methods are similar in the circumstance that the gas is made luminous by being made a conductor of electricity; moreover the study of the luminous spark discharge indicates that those regions which are most highly luminous are also those in which there is the greatest amount of ionisation and the greatest conductivity. The conclusion is obvious that the conductivity and the luminosity are associated fundamentally[1], and we arrive at once at the view that the light emitted by a gas is connected with the ionisation of its atoms. The vibrating charges may either be the electrons liberated from the atoms, or they may be the electrons, forming part of those atoms, which are disturbed by the ionisation. The existence of such ionised atoms may cause the emission by the same gas of different spectra; for the vibrations associated with an atom are determined by the electric field in its interior and this field is likely to be different according as the atom has or has not lost one of the electrons which belong to it when it is neutral.

Researches both on flame spectra and spark spectra have been conducted with these considerations in view. Lenard showed that different portions of the same spectrum (i.e. different groups of lines such as were considered in § 87) are emitted by different parts of a flame or arc, and further that the parts of the flame emitting different groups of lines were differently affected by placing the flame in an electric field. The 'principal' series of the alkali elements was emitted by a part unaffected by such a field, the 'subsidiary series' by a part which was displaced by such a field as if it carried a positive charge. Thence he concluded that the former series was emitted by uncharged atoms, the latter series by atoms which had lost either one or two electrons. His conclusions have been criticised and it has been

---

[1] Why then are gases not luminous when they are made conducting by the passage of ionising rays through them? Probably only because the concentration of the ions made in such a way is so very much smaller than the concentration in the flame or 'unsupported' discharge that the luminosity cannot be detected. When very intense ionisation is produced, as in the immediate neighbourhood of a strong radioactive preparation, the gas does become luminous.

suggested that his experiments only show differences in the emissive and absorptive power of different parts of the flame.

On the other hand Stark from a study of the mercury spectrum, which has two quite distinct forms, was led to suggest that the whole of one spectrum was emitted by the same atom, but that different spectra were emitted by atoms which had lost different numbers of electrons; that all line spectra were emitted by atoms which had lost at least one electron, while band spectra were emitted by neutral atoms. This suggestion seemed immediately to receive brilliant confirmation. The positive rays in a discharge (p. 111) emit light and they are moving with velocities which may be as high as 10^8 cm./sec.; according to Doppler's principle, the frequency of the light seen by an observer towards whom they travel ought to be greater than that which he would see if they were at rest. Accordingly if the spectrum of the positive rays is examined along the direction in which the rays travel, the wavelengths in the spectrum ought to be shifted towards the violet by an amount determined by the speed of the rays. Such an effect can be detected. If the spectrum of the positive rays of hydrogen (which shows the effect better than any other gas) is so examined, it is found that each line is split into two; one part occupies the normal position, the other part is a somewhat diffuse line on the violet side and separated from the normal line by a finite interval. The velocity determined by the most displaced portion of the line, on the assumption that the resolution is due to the Doppler effect, agrees well with that of the positive rays determined in other ways. The fact that the displaced or 'moving' line is diffuse is to be attributed to the varying velocities of the different positive particles. The finite interval between the moving and the normal or 'resting' line must be due either to the absence of positive rays moving with less than a certain speed or to the incapacity of such rays to emit light. All the lines in any one spectrum appear to be emitted by particles travelling with the same velocities.

At the time these experiments were first made it was believed that all the positive rays consisted of positively charged particles, and the observations seemed to show that all the lines in the hydrogen spectrum, including the principal series, were emitted by such particles. But subsequent investigation has shown that the positive rays contain also neutral and even negative charged particles, and Wien has shown that the light emitted by the neutral particles (which are undeflected in a magnetic field) is quite as intense as that emitted by those positively charged. The question of the charge carried by the particles emitting the line spectrum is therefore still unsettled, but it may be pointed out that the evidence is in no way fatal to Stark's hypothesis. For all the positive rays are positively charged at some time in their career; those which are undeflected or deflected in the negative direction are, as we have seen, those which are neutralised or more than neutralised after passing the kathode. It is not improbable that the disturbance which causes a positively charged particle to emit light is that which causes it to become positive, namely the ejection of an electron from it. The time during which the emission will last as the result of a single disturbance will be short compared to that required for the particle to travel through the field of observation; hence it may be that, even if the light is only emitted when the particle is positively charged, the average intensity of the light which it emits during a given time is determined more by the number of occasions during that time on which it *becomes* positive than by the proportion of the time during which it *is* positive. A particle which is positive the whole time might emit no light at all, while one which was on the whole neutral might emit a great deal. In this way Wien's observation might be reconciled with Stark's theory.

And this view might also account for the finite interval between the resting line and the moving line. For we have seen that a particle must be moving with a certain velocity in order to be ionised on collision with an electron; it would be only particles which possessed a velocity as great as this which would become positive during their passage through the field of observation and so be excited to emit light. Another explanation of this interval, perhaps not really independent of that suggested here, will be mentioned in the next chapter.

Stark has made a further ingenious suggestion as to the origin of band spectra which he still attributes to neutral radiators. Line spectra he considers to be due to the vibrations of those electrons in the atom which are not ejected from it when the atom is ionised[1]; band spectra he considers may be emitted by the electrons which can be separated from the atom when it is ionised and do not form part of its permanent structure, when these electrons return to the atoms from which they have been ejected and restore it once more to the neutral condition. The oscillations of such an electron in returning to its final resting place within the atom would probably differ considerably in their nature from the vibrations of an electron slightly disturbed from a position of equilibrium, and this difference might possibly account for the absence of the Zeeman effect, the pressure shift and the Doppler effect in band spectra. There is little direct evidence for this hypothesis, which seems rather difficult to reconcile with the fluorescent phenomena mentioned below, but it is worthy of a brief mention. It may be pointed out that none of these theories gives any but the vaguest suggestions as to why the spectrum of the same element may in some conditions consist only of lines and in others only of bands.

The more recent speculations as to the nature of light cannot be considered apart from the theoretical investigations of 'white light,' which are of such importance that they require a chapter to themselves. We shall return to the subject at the end of the next chapter. It remains only to consider a few of the properties of light, concerned in its reactions with material media, which require a rather more detailed treatment, in view of their interest for modern theories, than is contained in most text-books of optics. With the contents of such text-books the reader is, of course, supposed to be familiar.

The photo-electric effect.

90. The most important of these properties are connected with what is commonly known as the 'photo-electric effect.' It has been known for some 30 years that a metal surface on which light (especially light of high frequency) is incident tends to acquire a positive charge; it has been proved conclusively that this tendency is due to the emission from the surface of electrons with an initial velocity of a few volts.

---

[1] In Chap. XII the reasons for believing that there are such electrons will be considered.

The many experiments which have been made in the last ten years on this phenomenon appeared at one time to lead to completely inconsistent results as to the relation of the number and the velocity of the electrons so emitted to the nature of the metal and the frequency of the light. All the discrepancies are not yet completely explained, but the following statements may be made with some confidence:

(1) Both the number of the electrons and their initial velocity are determined completely by the nature of the metal and the intensity, frequency, angle of incidence and state of polarisation of the light incident upon it. Apparent discrepancies are due either to a film of impurity on the surface of the metal or to a failure to take into account the Volta difference of potential between the illuminated metal and the other electrode used in measuring the photo-electric current.

(2) There are two distinct kinds of photo-electric effect, the 'selective' and the 'normal.' In the normal effect the number of electrons liberated by a light of given intensity and frequency depends only on the amount of light absorbed by the metal. In the selective effect this number depends also on the angle between the electric displacement in the light and the metal surface; it is zero if this angle is zero.

(3) In the selective effect the number of electrons liberated for a given absorption of light is a maximum for some particular frequency of the light and is less for frequencies either greater or smaller. Little is known of the velocity of the electrons emitted in the selective effect. It has only been discovered with certainty in the alkali metals, and depends on their state of molecular aggregation. No great theoretical significance has been attached to it as yet, and it will be left out of consideration in what follows except when specifically mentioned.

(4) In the normal effect the electrons leave the surface with all velocities between zero and a certain maximum which depends only on the frequency of the light and the nature of the metal. It is probable that all the electrons are emitted from the atoms of the metal with the same speed, but that some of them lose energy before emerging from the surface. The maximum velocity for a given metal increases regularly with the frequency of the light

incident upon it, so long as the frequency is great enough to cause any photo-electric current at all. The very important conclusion that this maximum velocity does not depend on the intensity of the light is reached by all observers.

The majority of observers agree that E, the maximum kinetic energy of the electrons liberated, is related to ν, the frequency of the incident light, by the equation

$$E = h\nu - w_0 \quad\quad (190),$$

where w_0 is a constant characteristic of the metal from which they are liberated and h is a constant which does not vary very greatly for different metals and may possibly be the same for all. The latest experiments give values for h lying between $3{\cdot}5 \times 10^{-27}$ erg sec. (bismuth) and $5{\cdot}8 \times 10^{-27}$ (platinum). If we write $w_0 = \epsilon V_0$, where V_0 is the P.D. through which the electron would have to fall in order to acquire the energy w_0, then V_0 lies between 1·7 volts (sodium) and 3·9 volts (platinum). In general it appears that V_0 is less for the more electro-positive element.

It will be observed that if ν is less than w_0/h, (190) breaks down; and it appears in fact that there is for each metal a definite frequency ν_0, given by $h\nu_0 = w_0$, which is the smallest which will produce any photo-electric current at all. Thus light which has a frequency less than $5{\cdot}15 \times 10^{14}$ (green) will excite no photo-electric current in sodium, and that which has a frequency less than $10{\cdot}4 \times 10^{14}$ (ultra-violet) none in platinum. It may also be remarked that the maximum value of the velocity which these experiments suggest to be possible (i.e. with $w_0 = 0$) for the light of highest frequency obtainable ($\nu = 3 \times 10^{15}$) would be that corresponding to a fall of the electron through 11 volts.

(5) There is a direct conflict of evidence, which has not yet been resolved, as to the relation between the intensity of the light[1] and the rate at which electrons are liberated, other conditions being constant. Some observers maintain that the number of electrons liberated from a given surface in unit time by light of given frequency and direction is simply proportional to the intensity of that light. Others maintain that the number

---

[1] The intensity of the light is defined in all such experiments to be inversely proportional to the square of the distance from a point source.

decreases more rapidly than the intensity. An examination shows that the former worked with the alkali metals and visible light (the conditions were such as to exclude the selective effect), while the latter used other metals and ultra-violet light. Possibly there is a real difference in the relation under the two conditions, but it is more likely that some disturbing phenomenon has been overlooked. It is high time that this important matter was definitely settled.

(6) It is difficult to make any valuable comparison between the rate at which electrons are liberated from different metals by lights of different frequencies because the sensitiveness of every metal varies with the frequency of the light in a different manner owing to differences in ν_0, but, as might be expected, in common experimental conditions it is found that the rate of emission increases with the maximum velocity of the electrons emitted, when either the frequency of the light or the nature of the metal is varied; but little work has been done on the precise relation between these quantities. When the nature of the metal and the frequency of the light are kept constant, the angle of incidence of the light alone being varied, it appears that the photo-electric current is proportional to the amount of light absorbed.

(7) All phenomena connected with the normal photo-electric effect appear to be independent of the temperature within the range 0° to 150°. At higher temperatures experiment is difficult.

(8) Chemical compounds and non-metallic elements in general show far smaller photo-electric effects than metals; in fact with most inorganic compounds no photo-electric action has been observed. It has been stated that among the salts of potassium the photo-electric activity is less the more electro-negative the element with which the potassium is combined; the quantity w_0 is doubtless very much greater for these substances than for metals.

91. Many of the features of the photo-electric effect are readily explicable on the view which we have taken already of the nature of light and of the properties of electrons in metals. We have already seen (Chapter III) that it is necessary that the free electrons in metals possess a certain amount of kinetic energy before they can pass

Theory of the effect.

out through the surface of the metal and also that the action of light upon such electrons is to set them in motion and increase their kinetic energy. If we identify w_0 in (190) with the work which an electron must do in order to pass through the surface of the metal and suppose that light of frequency ν can communicate an amount of energy $h\nu$ to an electron, then we should expect that it would be possible to liberate electrons from metals by the action of light on their surfaces and that the relation between the energy of the electrons after their liberation and the frequency of the light would be given by (190). From thermionic measurements (p. 83) the value of w_0 for Pt was found to be a little more than 5 volts, whereas the photo-electric experiments make it only 3·9 volts, but the uncertainty of the measurements is such that it is quite possible that the values given by the two methods are really identical.

If the electrons liberated by the action of light are really the free electrons of the metal, the kinetic energy which they must absorb from the light in order to be able to emerge cannot be truly independent of the temperature, for that which they already possess is proportional to the absolute temperature. But we saw on p. 58 that the energy of a free electron at ordinary temperatures is less than $\frac{1}{20}$ of a volt and is very small compared to that which is necessary to liberate electrons from the metal. It would not become comparable with that energy until the temperature of thermionic emission was reached and at that stage the complications arising from this action would render observations on the photo-electric effect impossible. In any practicable conditions it is to be expected that the number and velocity of the electrons liberated by light should appear independent of the temperature.

It is, of course, a natural consequence of this view of the photo-electric effect that it should not be shown, or at least be shown in a far smaller degree, by chemical compounds and non-metals, which do not possess metallic conductivity and in which, therefore, there are not free electrons. In such bodies the electrons will absorb energy from the light, but they will not emerge from the body until that energy is sufficient to enable the electron to break away from the atom in which it is contained; this energy is that which has been previously recognised as the energy necessary to produce ionisation of the atom, and the minimum estimate which we have obtained of its magnitude is about 10 volts (p. 105). We have just noted that if (190) is universally applicable it is only the most extreme ultra-violet light which can be observed that can give to an electron such an amount of energy and, as a matter of fact, it is found that it is only extreme ultra-violet light which is capable of producing ionisation in air and other gases[1]. If light of lower frequency than that for which $h\nu_0/\epsilon = 10$ volts can liberate electrons from solid non-conductors it must be because the energy to ionise the atom of a solid is less than that required to ionise the atom of a gas.

So far we have considered only the normal photo-electric effect; the mechanism of the selective effect is less clear, but it is almost certainly due to 'resonance.' We saw in Chapter II that if the period of the light incident on an atom coincided with the free period of the electrons in it, those electrons would be set into vibrations of very large amplitude and energy. It is possible that the vibrations might become so large that the electron would break free from the forces restraining it and emerge from the atom; it has been suggested, indeed, and the view is very plausible, that the absorption of light in such cases is due to the ejection of the electron, the energy of which would then be converted by collisions into heat energy of the other molecules and would be lost to the radiation. Such an ejection of the electrons from their atoms is possibly the cause of the selective photo-electric effect, and the frequency of the light for which it is a maximum is probably that of the free vibration of some electron contained in the substance; though it should be mentioned that this frequency does not coincide with any of those revealed by observations on the spectra of the substances concerned. The fact that the selective effect does not occur unless the electric intensity in the light has a component perpendicular to the surface of the substance would be due to the fact that

---

[1] There has been considerable discussion in the past as to whether ionisation could be produced in air by ultra-violet light. It seems likely that the apparent ionisation produced by such light as can be transmitted by quartz is only due to a photo-electric action on solid particles contained in the gas, and that true ionisation can only be produced by the 'Schumann rays' of which the frequency is not less than $1{\cdot}8 \times 10^{15}$.

unless the electron were ejected along that perpendicular it would not emerge at all; its velocity would not be as great as that necessary to enable a free electron to emerge from the surface, and if it collided with another atom it would be at once reduced to the condition of the free electrons in the substance[1].

But though we can explain in this manner some of the phenomena of photo-electric emission without revising in any way our hypotheses concerning the nature of light or the structure of atoms, we can certainly not explain all of them. In the first place no reason whatever can be suggested as yet for the relation (190) so far as it concerns the frequency of the light. It appears from that relation that light of frequency ν communicates to an electron an amount of energy $h\nu$ quite independently of the intensity or direction of the light or the nature of the metal. It would certainly have been expected that the electric intensity in the wave-front and, therefore, the forces exerted by the wave on an electron would increase with the intensity of the light, and, in general, we should have expected the effect of the light to decrease rather than increase with increase in frequency (cf. p. 35). The approximate proportionality of the number of electrons emitted to the intensity of the light might have been expected until we knew that the velocity of them was independent of that quantity; no hypothesis that has been suggested so far can provide any reason for this distinction. It is on account of these difficulties that the photo-electric effect is of such immense importance for the theories of the nature of light which we shall consider presently.

Phosphorescence.

92. It is probable that in the future developement of the theory of light radiation the phenomena of phosphorescence and fluorescence will be found of equal importance with the photo-electric effect; but though at the present time it is difficult to draw any precise theoretical conclusions from the results of the great amount of experimental

---

[1] It is certainly not quite clear why in the normal effect there is not a similar influence of the direction of the electric intensity in the light. For, if the assumption of p. 59 that a free electron is always reduced by a collision to the same velocity is correct, it would be only when an electron was accelerated in such a direction that it emerged without collision that it would have a greater chance

work of which they have been the subject in recent times, it would be undesirable to pass them by without a brief notice.

Phosphorescence and fluorescence both consist in the emission by a body on which radiation is falling or has fallen of light different in quality from the incident radiation. The former term is used when the emission continues an appreciable time after the action of the incident radiation has ceased, the latter term when the emission appears to continue only while the exciting radiation acts; there are cases in which it is difficult to decide which term is the more appropriate and it is uncertain whether further investigation will lead to a maintenance of the distinction. It appears that phosphorescence is only displayed by solid bodies, fluorescence chiefly by liquids and gases.

The most notable phosphorescent bodies are the sulphides of the alkaline earth metals, calcium, strontium and barium, containing minute traces of impurity. When perfectly pure the sulphides seem to show no phosphorescence; it is excited only when small quantities (often only 1 part in a million) of copper, manganese, bismuth and some other metals are added. The spectrum of the phosphorescent light consists of one or more continuous bands having maxima at different wave-lengths, which are determined only by the nature of the phosphorescent substance; the positions of the maxima depend chiefly on the metal present as impurity, but are also influenced by the alkaline earth sulphide in which it is contained. Lenard, to whom most of our knowledge on this subject is due, has stated the remarkable law that the wave-length in vacuo of such a maximum determined by a given 'impurity' is proportional to the square root of the dielectric constant of the medium containing it. If the relation $K = n^2$ were accurately true, this proposition would mean that the wave-length in the *substance* corresponding to such a maximum is always the same.

The intensity and duration of the phosphorescent light, but not its quality, depend chiefly on the nature of the exciting

---

of emerging than any other electron. It is possible that the energy w_0 necessary to make an electron emerge is that for which this assumption ceases to be even approximately true, and is such that an electron possessing it can collide with an atom without losing any appreciable portion of its velocity.

radiation. The light which is capable of exciting a given phosphorescent band for a given duration also consists of one or more bands in the spectrum within each of which lies a definite wave-length associated with a maximum of exciting power. All the wave-lengths corresponding to maximum exciting power are shorter than those of the maxima of the phosphorescent bands which they excite; but in some cases the band containing light possessing exciting power overlaps to some extent that containing the light which it excites. In general, therefore, the exciting light is of shorter wave-length and greater frequency than that of the excited. This proposition, which was formerly supposed to be applicable to all cases of phosphorescence and fluorescence, is known as Stokes' Law; it is now known to be only true in certain classes of cases.

The intensity and duration of the emission of a phosphorescent band, but not its position, are profoundly influenced by the temperature. In general, the duration of the emission is greatly diminished by rise in temperature, but the total amount of light which a body can emit after a given exposure to radiation appears to be approximately constant. A body illuminated at a very low temperature will exhibit in that state no sign of phosphorescence, but if its temperature is suddenly increased greatly, it will become intensely phosphorescent for a very short time.

Lenard has attempted to explain these observations by supposing that the immediate action of the exciting light is to cause a photo-electric liberation of electrons from the atoms of the phosphorescent substance, and that the subsequent emission of light represents the return of these electrons to the atoms from which they have been ejected. This return will require an appreciable time because the substances have a very small electrical conductivity and oppose a great resistance to the motion of electrons; the return, however, would be accelerated by any influence which, like rise in temperature, increased that conductivity. This theory, combined with that of Stark (§ 89), would explain immediately why the light consists of bands and not lines, but Lenard considers it more probable that the light is emitted by the permanent electrons in the atom, set into vibration by the restoration of the negative charge. The original ejection of the electron doubtless is similar to the 'selective' photo-electric effect; it is a 'resonance' effect due to the coincidence of the frequency of the exciting light with that of the free period of the electron.

It cannot be said that this theory covers all the facts. It explains the similarity of the phosphorescence excited in some such substances by light and by kathode rays (which also consist of electrons striking the atoms), but it makes only the vaguest suggestion as to the extraordinary influence exerted by extremely small changes in the composition of the substance. Its bearing upon Stokes' Law will be considered later; the other quantitative relations which have been discovered it leaves wholly unexplained. However it is the first attempt to bring the phenomena of phosphorescence within the reach of modern optical theory, and is likely to be the basis of future research.

Fluorescence.

93. Fluorescent bodies have been divided into two classes, those which do and those which do not follow Stokes' Law that the exciting light is of greater frequency than the excited light; this classification appears to have theoretical importance. Bodies which obey Stokes' Law have no sharp absorption line in the visible part of the spectrum; they absorb to some extent the light which causes them to fluoresce and it is doubtless the energy of this light which is transformed into that of the fluorescent light, but the regions in which this absorption takes place are broad bands utterly unlike the sharp lines the presence of which characterises the bodies which do not obey Stokes' Law. The fluorescent light which they emit seems to form a spectrum of continuous bands which have not been resolved into separate lines, and there is usually a distinct gap between the lowest frequency which will excite fluorescence and the highest frequency of the fluorescent light which it excites. The light emitted by these bodies when fluorescing is unpolarised, even if the exciting light is polarised. The most familiar example of this class is quinine sulphate.

In such bodies the relation between the exciting and the excited light is precisely similar to that in the phosphorescent bodies which have just been considered, except for the fact that the emission ceases with the excitation. It is probable that the mechanism involved in the two processes is the same and Lenard's theory is as applicable to one as to the other; we have only to imagine that in the fluorescent substances the electrons liberated by the light cannot remain an appreciable time separated from the atoms, but pass straight from the atom which they leave into another where they come to rest.

In the case of bodies which do not obey Stokes' Law fluorescence is excited only if the exciting light coincides in frequency with one of the sharp absorption bands of the substance. The most remarkable instances of such bodies are the vapours of sodium and iodine which have been subjected to a very searching investigation by R. W. Wood. The fluorescent light varies very greatly in its nature with the frequency of the exciting light; its spectrum has been shown to consist of sharp lines, of which one always coincides in frequency with the exciting light and, therefore, with one line in the absorption spectrum of the substance; but the remaining lines usually differ from those of the absorption spectrum of the substance and from those in the spectrum which it emits when excited in other ways. The vapours can be made to emit an enormous variety of fluorescent spectra by exciting them with light coinciding in frequency with different absorption bands and in no case do the fluorescent spectra excited by two different frequencies seem to be identical. Interesting relations, similar but not identical with those mentioned in § 87, have been discovered between the lines of the fluorescent spectra. In some cases the fluorescent light has been shown to be polarised, even when the exciting light is unpolarised, the electric intensity displacement being perpendicular to the plane containing the directions of propagation of the exciting and excited light.

In such bodies the exciting light does not cause the ejection of electrons from the atom, but causes them to execute resonant vibrations within it. A fairly complete analogy to the phenomenon is to be found in the analysis of the sound emitted by a musical instrument by means of Helmholtz resonators. The vibrations composing such sound are not simply harmonic, but may be regarded as the resultant of several harmonic vibrations of different frequency, and the same is true of the sound emitted by the resonators. If the frequency of the fundamental vibration of one of the resonators is the same as that of one of the harmonic components of the exciting sound, that resonator will be excited; but, unless the resonator is constituted in exactly the same manner as the musical instrument, the frequency of the vibrations other than the fundamental which it emits when excited will not coincide with that of any of the harmonic vibrations composing the exciting sound. The nature of the sound determines which resonators are excited, but the nature of the resonators and not the nature of the exciting sound determines what vibrations other than those by which they are excited the resonators will emit.

The application of this analogy presents no difficulty, for we have already seen that it is probable that the atom contains systems which execute vibrations which are not simply harmonic but can be resolved into a finite number of harmonic components. The systems which are excited to emit fluorescent light are not the same as those which are concerned when the substance is rendered self-luminous, but the difficulty raised by the difference is no greater than that which already attaches to the explanation of the difference of the arc from the spark spectrum. Fluorescent phenomena only prove once more what an exceedingly complicated system the atom must be and how far we are from any complete theory of its structure. Perhaps the difference between the fluorescent and the self-luminous spectra is based on the fact that the atoms when they are fluorescing contain their full complement of electrons, whereas by the more ordinary methods of excitation we can only excite the atom to vibrate by dragging an electron out of it, so that we never observe the vibrations of the neutral atom. But there are objections to this view as there are to every other which has been proposed.

One other remarkable observation on this kind of fluorescence should be noted. It has been found that the nature of the fluorescent light emitted by iodine vapour (and probably other similar substances) is greatly influenced by the mixture of the vapour with other substances. The presence of any other gas in sufficient quantity causes the spectrum to change from a line to a band or a continuous spectrum; at the same time the intensity of the fluorescent light is greatly reduced. The power of gases to produce these changes is found to increase with the electronegative character of the gas; a trace of chlorine or oxygen destroys the fluorescence entirely, while a considerable quantity of the inert gases produces scarcely any effect. This fact undoubtedly suggests that the difference between a line spectrum and a band spectrum is to be found in the influence on the vibrating atom of neighbouring atoms; the foreign gas might damp in some way the vibrations of the resonators, and it would be those atoms which attract electrons most strongly (see Ch. XIII) which would produce the most powerful damping effect. But there are other phenomena, such as the persistence of band spectra of some substances even at the lowest pressures, which render this view difficult.

REFERENCES.

The questions concerning the resolution of natural light by means of gratings and prisms are fully discussed in most modern treatises on Optics, e.g. those of Schuster and Wood. They have been further considered here because it is hoped that yet another discussion may help students to understand this extremely difficult problem. The following papers by Rayleigh are very important:—*Phil. Mag.* 28. 460, 1889; 34. 407, 1892; 10. 401, 1905; 11. 123, 1906. Larmor discusses the objection raised in § 85 in *Phil. Mag.* 10. 574, 1905.

The main facts concerning our knowledge of line spectra are admirably given in Baly's *Spectroscopy*. Michelson's *Light waves and their uses* is also very interesting in this connection.

One of the best examples of attempts to explain spectra series on mechanical principles is to be found in a paper by Ritz, *Compt. Rend.* 144. 634, 1907.

The theories of Lenard and Stark concerning the origin of spectra are to be found in Lenard, *Ann. d. Phys.* 17. 197, 1905, and Stark, *Ann. d. Phys.* 16. 490, 1905; 21. 401, 1906; *Phys. Zeit.* 6. 892, 1905; 9. 85, 1908. The many speculations of Stark on this and other subjects are also set forth in his recent book *Die Principien des Atomdynamik*. Wien's work is contained in the papers on Positive Rays mentioned at the end of Chapter IV.

The earlier work on the photo-electric effect is given in Thomson's *Conduction of Electricity through Gases.* The later literature is extremely voluminous, the writings of Elster and Geitel, Ladenburg and v. Baeyer being of especial importance. The latest important paper at the time of writing is by Richardson and Compton, *Phil. Mag.* 24. 575, 1912 ; in it a summary of previous work is given. Pohl and Pringsheim first distinguished the normal and selective effects, *Deutsch. Phys. Gesell.* 12. 16. 682 and 12. 17. 697, 1910.

The chief papers by Lenard and his pupils on Phosphorescence are *Ann. d. Phys.* 15. 425, 1904 ; 28. 476, 1909 ; 31. 641, 1910 ; they are all rather difficult reading. Wood's work on Fluorescent gases is described in his text-book.

CHAPTER X

COMPLETE[1] RADIATION AND THE STRUCTURE OF LIGHT

Thermal radiation.

94. WHEN any solid or liquid body is heated to a sufficiently high temperature it emits light. There is no certain evidence that a gas can be made luminous by heat alone, even when the temperature to which it is raised is that at which all solids and liquids glow brilliantly. The light emitted by the hot bodies can always be resolved into a continuous spectrum, but the intensity is not constant throughout that spectrum, nor does it generally increase continuously from one end of the spectrum to the other. The positions of the maxima of intensity (for there may be more than one) depend partly on the nature of the body and partly on the temperature to which it is raised; in general, as the temperature is raised, the intensity of the light in all parts of the spectrum is increased, but the intensity of the parts of higher frequency is increased relatively to those of lower frequency, so that the maxima move towards the violet.

Concerning this thermal radiation questions might be asked similar to those discussed in the last chapter; we might inquire what is the nature of the radiators emitting the radiation and what is the nature of the disturbances which they emit. Two answers to this question have been proposed, which are not necessarily mutually exclusive and resemble the theories proposed by Stark for the origin of line and band spectra respectively. According to one theory the radiators are systems characterised by

---

[1] Complete radiation is also called 'black' and 'full' radiation; the former term is very objectionable, since it assumes from the outset the identity of 'complete' and 'equilibrium' radiation.

natural frequencies, such as emit line spectra, the vibrations of which are heavily damped owing to the disturbing influences of neighbouring molecules. That continuous spectra can be produced in such circumstances seems to be proved by the experiments on the fluorescence of iodine vapour (p. 220), and the fact that the selective emission of metals when they are hot resembles their selective absorption when they are cold seems to indicate that systems having definite natural frequencies must take a part in the radiation. The other theory is based on the observation that when bodies are emitting thermal radiation they are nearly always metallic conductors; gases, which appear never to acquire metallic conductivity, show no thermal radiation. Now metallic conductivity is attributed to the presence of free electrons possessing kinetic energy of thermal origin; it is suggested that thermal radiation represents the disturbances emitted by these free electrons when they collide with the molecules of the substance. In the next chapter we shall be further concerned with the nature of the disturbances which would be emitted in consequence of such collisions; here we will only note that they would take the form of 'pulses,' lasting for a time comparable with that of collision[1].

Neither of these theories has been very successful in explaining the facts. The theory of damped vibrations has been worked out in considerable detail; of course the exact nature of the resulting disturbances would depend upon the form of the damping, but if any simple assumptions are made on this point, it appears that the spectrum into which a large collection of damped vibrations would be analysed by a grating would not be characterised by a relation between the intensity and the frequency of the

---

[1] The term 'pulse' occurs frequently in all treatments of radiation, but it is rather difficult to define precisely. A disturbance in which the electric displacement had only one maximum would certainly be termed a pulse and distinguished from a 'vibration' consisting of a series of maxima and minima; the pulses mentioned here would probably be of this type. But apparently disturbances characterised by more than one maximum are also called pulses. It appears that the term is used somewhat vaguely to denote a disturbance such that, if the relation of the electric displacement to the time were represented graphically, the resulting curve would not resemble at all closely a portion, including more than one complete period, of the curve representing the simple harmonic vibration (186).

trains such as is found in the cases which are of most experimental importance. The other theory depends even more upon the nature of the assumptions that are made concerning the collisions. The problem of deciding what would be the spectrum into which a set of the pulses would be analysed has been attacked with great power by Lorentz, who shows that, with any plausible assumptions as to the nature of the collisions, a relation between intensity and frequency can be deduced that is in accordance with experiment so far as the part of the spectrum consisting of long wave-lengths is concerned; conclusions as to the short wave-lengths depend entirely on assumptions made as to collisions. His result would appear to favour the second theory, but we shall see that the same result can be attained from much more general considerations to which we will now proceed. Moreover these considerations raise difficulties so formidable and so fundamental that it may be doubted whether the questions have been put in a form in which they are capable of an answer.

Complete radiation.

95. If only the observations already mentioned had been made thermal radiation would be of little importance for modern theory. But it has been found that under certain conditions the thermal radiation emitted by a body does not depend on the material of which the body is composed, and it is the study of this radiation, common to all substances, with which for the rest of the chapter we shall be concerned. In this matter the distinction between the newer and the old science made in the first chapter is invalid; here modern theory is interested in the phenomena in which all bodies agree and not in those in which they differ.

The necessary conditions are easily attained. If a hollow box, in the side of which there is a hole of an area small compared to that of the whole box, is maintained at a uniform temperature and the radiation issuing from the hole is examined, then it is found that the spectrum of this radiation is entirely independent of the material of which the box is made and dependent only on its temperature. If $E_\lambda d\lambda$ is the intensity[1] of the radiation

---

[1] The intensity E_λ is defined in any treatise on radiation in a manner which makes it independent of the size of the hole or the distance of the hole from the observer and other geometrical conditions.

emitted which has a wave-length lying between λ and $\lambda + d\lambda$, then E_λ is a function of λ and T, the absolute temperature of the box, only; it does not depend upon the material of which the box is composed.

Three formulae representing this function are of importance. The first is that of Rayleigh:

$$E_\lambda = ckT\lambda^{-4} \quad \text{.....................(191)},$$

where k is a constant. This formula agrees with experiment only if the quantity λT is sufficiently large; that is, for a given temperature it is valid only for the longer wave-lengths and for a given wave-length only for the higher temperatures. The second is that of Wien:

$$E_\lambda = c^2 h\lambda^{-5} . e^{-\frac{ch}{k\lambda T}} \quad \text{..................(192)},$$

h being another constant. This formula is valid only if λT is sufficiently small. The third is that of Planck:

$$E_\lambda = c^2 h\lambda^{-5} \frac{1}{e^{\frac{ch}{k\lambda T}} - 1} \quad \text{..................(193)}.$$

This formula agrees with experiment for all values of λ and T; it may be noticed that it becomes identical with (191) for large values of λT and with (192) for small values of that product. It is useful to transform these formulae by the introduction of the frequency ν in place of λ. If E_ν is the intensity of the radiation which has a frequency between ν and $\nu + d\nu$, then, remembering that $d\lambda = -\frac{c}{\nu^2} d\nu$, we see that the three formulae become

$$E_\nu = c^{-2} k\nu^2 T \text{ (Rayleigh) ....(194)},$$

$$E_\nu = c^{-2} h\nu^3 . e^{-\frac{h\nu}{kT}} \text{ (Wien) ....(195)},$$

$$E_\nu = c^{-2} h\nu^3 . \frac{1}{e^{\frac{h\nu}{kT}} - 1} \text{ (Planck) ....(196)}.$$

An important feature of (193) is that it shows that E_λ will be a maximum for some finite value of λ, which will, of course, be determined by the equation $\frac{dE_\lambda}{d\lambda} = 0$. If λ_m is the value of λ for which this equation is satisfied, it will be seen that

$$\lambda_m T = \frac{ch}{k\beta} = b \text{ (say)} \quad \text{.................(197)},$$

where β is the root of the equation

$$e^{-\beta} + \frac{\beta}{5} - 1 = 0 \quad \text{....................(198)},$$

which gives $\beta = 4{\cdot}9651$. It is to be noted that as the temperature rises, the wave-length for which the intensity of the radiation is a maximum becomes shorter. The experimentally determined value of b is 0·294 cm. degree.

The total intensity of radiations of all frequencies (which might be measured by receiving the radiation on a thermopile without resolving it by a grating) is given by

$$E = \int_0^\infty E_\nu d\nu \quad \text{....................(199)},$$

where we must insert for E_ν the value (196) valid for all frequencies. The integration cannot be effected in finite terms, but if the exponential term is expanded in a series and it is remembered that $\int_0^\infty x^n e^{-x} dx = n\,!$, it will be found that

$$E = ST^4 \quad \text{.......................(200)},$$

where
$$S = \frac{6k^4}{c^2h^3}\left(1 + \frac{1}{2^4} + \frac{1}{3^4} + \dots\right).$$

S is found to have the value $8{\cdot}427 \times 10^{-6}$ erg/cm.^{2} degree^4 sec.

(200), which states that the amount of energy radiated per second from each square centimetre of such a hole as we have considered is proportional to the fourth power of the absolute temperature of the box, has long been known under the name of Stefan's Law. From the measured values of S and b the following values of h and k can be deduced:—

$$h = 6{\cdot}548 \times 10^{-27} \text{ erg sec.}; \quad k = 1{\cdot}346 \times 10^{-16} \text{ erg/degree.}$$

Radiation such as we are considering is termed the 'complete radiation' corresponding to the temperature T.

96. The possibility of the existence of a thermal radiation determined only by temperature and not by the peculiar properties of any material body is a consequence of the theory of such radiation founded by Prevost and elaborated by Kirchhoff. This theory is so familiar that it need only be sketched here. Three fundamental assumptions are made:—(1) That a body in such a condition that it can only undergo thermal changes emits a radiation determined wholly by its constitution and its temperature; and that such a body, if radiation of constant quality falls on it, absorbs a proportion of that radiation determined by the constitution and temperature of the body and by the quality of the radiation, but not by its intensity. The first part of this proposition is really a definition of thermal radiation; the second is very important but not always explicitly stated. (2) That any set of bodies enclosed in a boundary impervious to all forms of energy will ultimately attain equal temperatures. (3) That if radiation of any quality starting from A along AP arrives at B along QB, then radiation of the same quality starting along BQ will arrive at A along PA. It is always convenient to assume—the assumption does not limit the practical application of the theory—that all bodies absorb to some extent radiations of all qualities. From these assumptions it can be proved that, if E is the emissive power of any body at a temperature T, A its absorptive power, E/A is a function only of the temperature and the frequency of the radiation concerned and is equal to I, the intensity of the radiation of that quality in the medium surrounding the body when the final condition contemplated in (2) has been reached[1]. If $A = 1$, $E = I$ and the emissive power of the body is equal to the intensity of the radiation in the space surrounding it; a body for which $A = 1$ is termed 'black.'

Equilibrium radiation.

The radiation of which the intensity is I, traversing all parts of an impervious enclosure in thermal equilibrium, may be called 'equilibrium radiation.' It is clear that it resembles 'complete radiation,' for its total intensity is determined wholly by the temperature, and the intensity of any one of the homogeneous beams

---

[1] For a precise definition of E, A, I the reader must refer to treatises on thermal radiation.

into which it could be resolved is determined only by the frequency of that beam and the temperature. Moreover it appears likely that the conditions in which the complete radiation is obtained are also those which would give rise to this equilibrium radiation. For, on the one hand, we can regard the hole in the box as a black body, because radiation entering the hole would be neither reflected nor refracted at the hole, but wholly absorbed within after scattering or reflection at the walls; or, on the other hand, we may regard the box as forming such an impervious enclosure as has been contemplated and the hole as a means of extracting from the box a small sample of the equilibrium radiation within without disturbing the equilibrium. Accordingly until recently it was not disputed that the complete radiation, the properties of which have been described, was exactly the same thing as the radiation which would be, according to Kirchhoff's theory, in equilibrium with any set of bodies whatever kept in an impervious enclosure at a constant temperature.

The question now arises whether we can deduce without making further assumptions any of the quantitative relations concerning complete radiation which are stated in (191) to (200). An ingenious argument due to Wien enables us to make some progress in this direction. Wien contemplates a cylinder with perfectly reflecting walls and movable piston, kept at a constant temperature and containing a particle of some substance of infinitesimal mass. The presence of this particle ensures that the quality of the radiation in the cylinder shall always be that of radiation in thermal equilibrium with material bodies at the temperature of the cylinder. In virtue of the pressure exerted by radiation (§ 111) work must be done to push the piston in, and in virtue of the Doppler effect the frequency of the radiation will be changed by reflection at the moving piston; by calculating this work and this change of frequency while the cylinder is carried through a completely reversible cycle, Wien calculated from thermodynamical principles a relation between the frequency and the intensity of the various components of the complete radiation. He arrived at the result that for this radiation

$$E_\lambda = c^2 \lambda^{-5} F(\lambda T) \quad \text{......................(201)},$$

where F is some function of which the form cannot be determined without further assumptions. This proposition is known as Wien's Displacement Law; it was of great importance in the historical developement of the theory, but it becomes unnecessary as soon as the further assumptions which are about to be discussed are introduced. It should be observed that all three formulae (191, 192, 193) are special cases of (200), so that the identification of complete radiation with equilibrium radiation is confirmed.

Rayleigh's theory.

97. To proceed further it is absolutely necessary to introduce some hypothesis as to the nature of the mechanism by which the equilibrium is attained between the bodies in the enclosure and the complete radiation traversing it. The simplest and most plausible of the hypotheses which have been suggested for the purpose is the proposition generally known as the 'equipartition of energy.' This proposition has been the subject of much dispute, but it is now agreed that it is a necessary consequence of the fundamental principles of generalised dynamics as stated in the Hamiltonian equation of least action; these principles are more general than those of ordinary dynamics (which they include as a special case) and would have been regarded a few years ago as the most secure basis of physical science. It involves the conception of a 'degree of freedom,' which is of some difficulty, but may be defined, if scrupulous accuracy is not required, as any quantity involved in the expression for the kinetic energy of the system in respect of which a physical system can vary, while all other quantities determining the energy of the system remain unchanged; the number of degrees of freedom of a system is the number of terms, each involving one independent coordinate only, which are necessary to express its kinetic energy as a function of the coordinates. The proposition of the equipartition of energy states that, when a system is in equilibrium, the amount of energy associated with any degree of freedom is the same, and the terms in such an expression for the kinetic energy are all equal.

Now consider the impervious enclosure in thermal equilibrium. The Hamiltonian expression for the kinetic energy of this system consists of two groups of terms. The first group consists of the terms expressing the kinetic energy of the molecules of the bodies in the enclosure, the second group consists of terms expressing the energy of the complete radiation. The number of terms in the first group is equal to three times the number of molecules, since each molecule has three degrees of freedom, and is finite; the number of terms in the second group is proportional to the number of possible different forms of radiation. If we suppose a grating contained in the enclosure, the radiation will at some point be resolved into a continuous spectrum and, so far as we know, the intensity of the beam forming any infinitesimal part of this spectrum is independent of the intensity of each other such beam, that is to say, it is physically possible to change the intensity of one such beam without changing that of any other. Accordingly the number of independent forms of radiation is equal to the number of infinitesimal parts into which the spectrum can be divided. But this number is infinite. Accordingly if the energy associated with each term is the same, the energy in the second group of terms will be infinite compared to that in the first group. If the energy of the whole system is finite (as it is) the first group will have no energy at all. Hence this argument would lead to the conclusion that, when equilibrium was attained, the kinetic energy of the molecules would be zero and the bodies would be at the zero of temperature, while all the energy was contained in the radiation. This conclusion appears utterly inconsistent with experiment, for the bodies in the enclosure remain in many cases at a very high temperature.

But let us go further and inquire into the distribution of the energy among the beams of radiation of different frequency, assuming for the moment that the whole energy in the enclosure is infinite. The problem is simplified without loss of generality if we suppose the enclosure to consist of a cube, the side of which is l. It can then be proved that the only forms of radiation which can exist permanently in the enclosure are those for which

$$\frac{4\nu^2 l^2}{c^2} = p^2 + q^2 + r^2 \quad \text{..................(202)},$$

where p, q, r are any integers. Such radiation alone can form the 'standing waves' which must represent all the radiation in a state of complete equilibrium. If we want to know how many such forms there are having frequencies between ν and $\nu + d\nu$, we must find how many sets of integers p, q, r there are such that

$$\frac{c}{2l}(p^2 + q^2 + r^2)^{\frac{1}{2}}$$

lies between ν and $\nu + d\nu$. It appears that the number of such sets is $\frac{4\pi\nu^2 l^3}{c^3}d\nu$, assuming, as of course we may, that $\frac{\nu l}{c}$ or $\frac{l}{\lambda}$ is a very large number. In order to obtain the number of independent beams of radiation having frequencies between ν and $\nu + d\nu$, and the number of terms corresponding to such beams in the expression for the kinetic energy, we must double this number, for waves which are polarised in perpendicular planes must be counted independent. The kinetic energy associated with each of these terms is the same, according to the principle of equipartition, and is equal to that associated with one of the terms of the first group, that is one-third of the kinetic energy of a single molecule. But this energy according to p. 57 is $\frac{1}{3}\alpha T$. Moreover the radiation, unlike the molecules, possesses a potential energy equal to its kinetic energy; hence the total energy in the cube contained in the radiation between frequencies ν and $\nu + d\nu$ is $\frac{16\pi\nu^2 l^3 . \alpha T}{3c^3} . d\nu$.

Now it can easily be shown that U_ν, the energy, travelling with velocity c, contained in unit volume of the enclosure, is connected with E_ν, the intensity of the radiation proceeding from the hole in it, by the relation

$$E_\nu = \frac{c}{8\pi} U_\nu \quad \text{..........................(203).}$$

Hence
$$E_\nu = \tfrac{2}{3}\alpha c^{-2}\nu^2 T \quad \text{......................(204).}$$

If (204) and (194) are compared, it will be seen that they are identical if $\frac{2}{3}\alpha = k$. Now (p. 58) $\alpha = 4{\cdot}3 \times 10^{-7}\,\epsilon$, and if we put $\epsilon = 4{\cdot}7 \times 10^{-10}$, the most probable value, we find $\frac{2}{3}\alpha = 1{\cdot}35 \times 10^{-16}$, a most remarkable agreement with the experimental value[1]

$$1{\cdot}346 \times 10^{-16}.$$

[1] Observe that, conversely, if we assume $k = \frac{2}{3}\alpha$, we obtain an independent determination of ϵ, for $\frac{\alpha}{\epsilon}$ is known from observations on gases. The value obtained

Accordingly we find ourselves in this strange position. The arguments which have been given lead in general to a result utterly irreconcilable with experiment, and yet, if we apply them only to conditions in which λT is large (for it is only to such conditions that (194) can be applied), they lead to a very striking agreement with experiment and enable us to predict from observations on the compressibility of hydrogen how much radiation of a given frequency will be emitted from a hole in a box made of platinum. Why is the agreement limited to this range? Which of the assumptions we have made is true within this range but false outside it?

Jeans' theory.

98. On this point there has been much discussion; there are two views worthy of consideration. The first is that of Jeans, who proposes to deny the identity of complete radiation with equilibrium radiation. He points out that, while the principle of equipartition defines the condition of equilibrium which a system will ultimately attain, it does not set any limit to the time which a system will take to acquire that condition. He gives reasons for thinking that, while the system would reach equilibrium in respect of the radiations of lower frequency in a small fraction of a second, it might take years to attain equilibrium in respect of the radiations of higher frequency; the time would be determined by the relation between the period of the radiation and the average time between two collisions of the same molecule; if the latter time were much greater than the former, an immense number of collisions would have to take place before the radiations of higher frequency acquired the amount of energy which they would have when the final condition of equipartition was reached. Now no enclosure which is used experimentally is really impervious to radiation, if only because a hole has to be made in its wall in order that the radiation inside may be observed; if the rate at which radiation of a given frequency escaped from the enclosure

---

in this way is $\epsilon = 4{\cdot}69 \times 10^{-10}$ and is one of the most reliable. The only assumption necessary is that complete radiation and equilibrium radiation are identical. The latest measurements of the constants S and b lead to somewhat greater values of k and of ϵ.

were comparable with the rate at which it acquired energy by the processes tending to equipartition, then that radiation would never possess its full amount of energy; the fact that the energy associated with a radiation of high frequency determined experimentally is very much less than that predicted by (204) is due, according to this view, to continual escape of energy from the enclosure, which prevents equilibrium being ever attained. Moreover the difficulty connected with the prediction that the radiation should possess all the energy in the enclosure and the molecules none is overcome, for if we integrate (204) for those values of ν to which it applies (namely those for which λT is sufficiently large) we shall find that the total amount of energy associated with these wave-lengths is very small compared with that associated with the molecules. It is true that if the conditions were such that equilibrium could be reached, the energy in the radiations of smaller wave-length would become infinitely great compared with that in the molecules, but as they never acquire their full share of energy the total energy of the radiation remains small compared with that of the molecules. This view does not require the rejection of any of the assumptions of the theory of equilibrium radiation, but only the denial that the conditions contemplated by that theory are ever realised in experiments on complete radiation.

The main objection to Jeans' view is that it makes it difficult to explain the independence observed experimentally of the complete radiation on the nature of the enclosure. We should expect that the rate at which energy is transferred to a radiation of given wave-length and the rate at which that energy escapes would depend on the nature of the enclosure; unless these two rates were always proportional (and it is hard to believe that such a relation could be universal) the energy contained in radiation of any frequency, which depends on both rates, should vary with that nature. But it is likely that plausible additional hypotheses could be found to overcome this difficulty and that the eminently reasonable basis of Jeans' theory would have been accepted, if an alternative theory of exceptional boldness and brilliance had not been proposed by Planck.

Before considering this theory it may be noted that there is one other assumption on which the arguments of § 97 are based that might be denied. It has been assumed that each of the forms of radiation, possessing different frequencies, which can exist in the enclosure represents independent degrees of freedom. But it must be remembered that we have concluded (p. 192) that each of these frequencies does not represent the vibrations of a corresponding radiator in the source; it merely represents a term in the Fourier series into which a large number of complicated disturbances can be resolved. If—to take an extreme case—these disturbances were all similar and emitted at regular intervals, the relative intensities of the terms of the series into which they could be resolved would be wholly determined by that nature and interval. It would not be possible for one term of the series to have a certain intensity unless other terms had at the same time an intensity proportional to it. There would be a definite relation between the intensities of the radiations of different frequency into which the disturbances could be resolved, and these radiations would not represent independent degrees of freedom; they would represent such independent degrees only if the original disturbances could be of any nature or separated by any intervals. The possibility of escaping in this direction from the difficulties which have been met does not appear to have been considered.

99. Planck's theory of complete radiation is based on a denial of the principle of the equipartition of energy. It asserts that complete radiation is identical with equilibrium radiation, and that the reason why the radiations of higher frequency contain much less energy than that predicted by the arguments of § 97 is to be sought in the nature of light and of the radiators which emit it; the action of the radiators is not determined by the Hamiltonian principle of least action from which the principle of equipartition is deduced. Now the rejection of the principle of least action underlies not only dynamics, but also the theory of electromagnetism sketched in the first chapter; if we reject it, we must frame a new set of hypotheses covering all electrical science before we can proceed to mathematical argument at all. In order to see exactly how the new hypotheses must differ from the old, it will be convenient

Planck's theory.

to give the outline of Planck's theory as it was originally stated, though it is now agreed that this form of it cannot be regarded as logically sound.

Planck's theory contains three groups of assumptions. The first concerns the mechanism which establishes equilibrium between the hot bodies in the enclosure and the complete radiation. This mechanism is supposed to consist of 'oscillators,' consisting of a positively and a negatively charged part which can execute relatively to each other linear harmonic vibrations; the frequency of these vibrations is different for different oscillators, and in any material body there are oscillators having every frequency for which the validity of (193) has been observed. The radiation which these oscillators emit when they vibrate and the vibrations which they execute when acted on by the periodic electric intensity in a beam of light can be calculated by means of the fundamental equations of Chapter I.

These assumptions are used to determine the relation between L_ν, the mean energy possessed by an oscillator of frequency ν, and E_p, the intensity of radiation of frequency p contained in the enclosure when equilibrium is attained. It is easily seen that, since this radiation causes the oscillator to vibrate, and since in virtue of its vibrations it emits radiation, there will be a condition of equilibrium in which the energy taken by the oscillator from the incident radiation is equal to that of the radiation which it emits. It appears on detailed calculation that L_ν is determined only by the value of E_p when $p=\nu$, and that

$$L_\nu = \frac{c^2}{\nu^2} E_\nu \quad \text{......................(205)}.$$

The assumptions of the second group are thermodynamical. It is assumed that the principles of thermodynamics can be applied (as in Kirchhoff's theory) to the oscillators and radiation in the enclosure. It is a consequence of these principles that if S is the entropy of any part of the system, which is isolated from external forces, U its energy, then equilibrium is attained when the value of S is such that

$$\frac{dS}{dU} = \frac{1}{T} \quad \text{......................(206)}.$$

Further, Boltzmann's equation for the entropy is assumed,

$$S = k \log W + C \quad \text{.....................(207)},$$

where k and C are universal constants and W is the number of 'complexions' of the system when it has a given energy U. By the number of complexions is meant the number of possible different states of the system which are necessarily indistinguishable by experiment; thus in the case of a gas at a given temperature, pressure and volume, we might be able to distinguish experimentally states in which the number of molecules possessing a given velocity or position was different, but we could not possibly distinguish states in which that number remained the same while the individuality of the molecules possessing this velocity and position changed. With the precise value of C we are not concerned; the value of k can be found by applying (207) to a perfect gas, which leads to the conclusion that $k = \frac{2}{3}\alpha$; we have just seen that the k of (191—196) has the same value.

Suppose that there are in the enclosure N oscillators of frequency ν, each having a mean energy L_ν, so that the total energy U is NL_ν. The number of complexions will be the number of ways indistinguishable by experiment in which this energy can be divided up among the oscillators. Suppose that we imagine the total energy split up into a large number P of equal parts δ, so that $P = \frac{NL_\nu}{\delta}$. Then W is the number of ways P cards could be dealt to N players, the cards being supposed to be indistinguishable from each other, but the players distinguishable[1]. The number of such ways can be shown to be

$$W = \frac{(N+P-1)!}{(N-1)!\,P!} \quad \text{..................(208)}.$$

Now N and P are very large numbers, so that we make use of Stirling's theorem, according to which

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} \quad \text{.........................(209)},$$

---

[1] Of course we cannot really tell which oscillator has a given amount of energy, but it is conceivable that we could, because the oscillators differ in position and are therefore distinguishable, while the energy elements are, *ex hypothesi*, equal and absolutely indistinguishable. In counting complexions we are not concerned with the states which can actually be distinguished, but the states which, according to the theory, could conceivably be distinguished.

and we may neglect 1 in comparison with N or P. Hence we find

$$S = k \log W + C = kN \left\{ \left(1 + \frac{L_\nu}{\delta}\right) \log \left(1 + \frac{L_\nu}{\delta}\right) - \frac{L_\nu}{\delta} \log \frac{L_\nu}{\delta} \right\} + C \quad \text{.........(210)}$$

or

$$\frac{1}{T} = \frac{dS}{d(NL_\nu)} = \frac{k}{\delta} \log \left(\frac{\delta}{L_\nu} + 1\right) \quad \text{...........(211)},$$

whence

$$L_\nu = \frac{\delta}{e^{\frac{\delta}{kT}} - 1} \quad \text{....................(212)}.$$

Hence from (205)

$$E_\nu = \frac{\delta \nu^2}{c^2} \cdot \frac{1}{e^{\frac{\delta}{kT}} - 1} \quad \text{.................(213)}.$$

In calculating (205) according to ordinary electromagnetic theory it was assumed, of course, that the energy of any oscillator can vary continuously, i.e. by infinitesimal increments; in (212) δ is the smallest increment by which the energy of an oscillator can vary and so (212) will only be true in the limit when δ tends to 0. In that limit we have

$$L_\nu = kT \quad \text{........................(214)};$$

$$E_\nu = c^{-2}\nu^2 kT \quad \text{....................(215)},$$

and we arrive once more at the consequences of the Rayleigh-Jeans theory[1].

But now Planck makes his third and characteristic assumption, which is that δ is not infinitesimal but that

$$\delta = h\nu \quad \text{.........................(216)}.$$

If this substitution is made (213) becomes identical with (196), the experimentally established relation for complete radiation. If this assumption can be accepted, Planck's theory will have succeeded

---

[1] There appears to be a difference, however. We assumed before that the energy of a system having one degree of freedom, such as a linear oscillator, was $\frac{1}{2}aT = \frac{1}{2}kT$; now we find it kT. The difference is only apparent; before we were calculating only the kinetic energy of the system, now we are calculating the sum of the kinetic and potential energies. It may be noted that we arrive at the same result because we have made the same assumptions. We assumed the applicability of the principle of least action in establishing (205) and we assumed that the radiations of different frequency were independent when we assumed that the value of S could be found from (206) for the group of oscillators of a given frequency without taking the others into account.

in explaining the properties of that radiation. The assumption means that the energy of an oscillator varies not continuously, but by finite jumps and that, since the radiation changes its energy by exchange with an oscillator, the energy of the radiation can only vary by the same jumps. The magnitude of the jumps by which the energy of radiation of any frequency varies is proportional to that frequency. The physical basis of Planck's explanation of why the radiation of high frequency contains less energy than that which is allotted to it by the principle of equipartition may be stated roughly thus:—Radiation of a given frequency can receive energy only in finite increments δ; the magnitude of these increments is greater for the radiation of higher frequency. The chance that the system will be in a condition such that an amount of energy δ can be transferred suddenly and discontinuously from an oscillator to the radiation decreases very rapidly as δ increases. Consequently, though, when a radiation of high frequency receives energy, it receives more than a radiation of low frequency, the chance that it will receive any energy at all is so much less than in the case of a radiation which can take energy in smaller increments that the average energy of the latter radiation over a long time is greater than that of the former.

Planck's theory is open to two very obvious objections. The first is that it contains two mutually contradictory assumptions and therefore, according to logical principles, could be used to prove any proposition, true or false. In deducing (205) it is assumed that the energy of an oscillator varies continuously; in stating (216) it is assumed that it varies discontinuously. Perhaps the objection is not so serious as it appears at first sight, for examination shows that all that is assumed in proving (205) is that the *average* energy of a large number of oscillators varies continuously, while (216) is applied to a single oscillator; all that seems necessary to the theory is that the ordinary electromagnetic principles should be valid when applied to the average of a large number of similar systems. But it may be noted that, for the higher frequencies, the finite increments by which the energy of an oscillator is supposed to vary are large compared with the average energy of the oscillator; if $\nu = 5 \times 10^{14}$ (yellow light) and $T = 1500$ (white heat), then δ is about 16 times the average energy of the oscillator. Planck has endeavoured to reduce the inconsistency yet further by showing that the discontinuous changes in energy need only occur in the emission of radiation and not in its absorption, but according to his theory there must be some discontinuity in the transfer of energy, which is absolutely inconsistent with the application of the principle of least action in any of its forms to systems in which such transfers are possible.

In any case the objection need not overthrow the theory. It is not logically inconsistent to assume that (205) and (216) are both true, but only that (216) and the principle of least action are both true. Before we can prove anything we must make some assumptions and hitherto that principle has been thought the most certain of assumptions; but if we choose to throw it over, assume (205) in its place and try to build up theoretical physics afresh, there is no logical objection to the proceeding; the only question is whether the new building will be more in harmony with experiment than the old.

The second objection to Planck's theory seems to me much more serious. The theory is based entirely on the conception of the radiation being emitted from a large number of oscillators, all vibrating harmonically but each characterised by a different frequency; there must be as many different kinds of oscillators as there are homogeneous trains into which the spectrum can be resolved. It has been stated already that such a view of the nature of light emission appears contrary to all our evidence as to the structure of an atom; there is no evidence that any atomic oscillator emits simple harmonic vibrations and there is no evidence for the possibility of the simultaneous presence of so many different oscillators of different kinds. Perhaps the theory might be developed to deal with oscillators of which the vibrations were not simply harmonic by regarding these oscillators as composed of others of which the vibrations were harmonic; but it seems probable that the radiations of a single oscillator in some bodies would have to be regarded as made up of an infinite number of harmonic constituents and it would appear then that for such systems δ would have to be infinite.

I think it would be now generally agreed that, in view of these difficulties, Planck's theory would not have been regarded as a very serious attempt to solve the problem of complete radiation, unless evidence for the essential validity of its fundamental conceptions had been derived from experiments altogether independent of such radiation. The later developements of the theory, chiefly due to Einstein, have made these conceptions as important for modern physics as those which underlie the electronic theory; we will consider briefly these developements and return later to the question of how the difficulties may be overcome.

Consequences of Planck's theory.

100. Let us assume that the mean energy of a linear oscillator at temperature T is

$$L_\nu = \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1} \quad \text{.......................(217)}.$$

We can arrive at this result by assuming either (212) and (216) or (205) and (196), while the identity of complete with equilibrium radiation is assumed[1]. If the oscillator is not linear but can vibrate in any direction, its mean energy will be $3L_\nu$. Let us suppose that each atom of a solid element contains one of these oscillators, and that all have the same natural frequency ν. Then the energy required to raise the temperature of these oscillators will represent one part of the specific heat of the element; the magnitude of this part will be

$$C = N \frac{d}{dT}(3L_\nu) \quad \text{..................(218)},$$

where N is the number of atoms in one gramme. If we take the 'atomic heat,' or the heat capacity of a gramme-atom, we must substitute ν_0 (p. 58) for N. Dulong and Petit's law states that the atomic heat of a solid element is

$$C_0 = 3k\nu_0 = 2{\cdot}49 \times 10^8 \text{ ergs} = 5{\cdot}95 \text{ cals.};$$

this 'normal' value can be deduced by assuming the principle of equipartition and supposing that each atom has three degrees of

---

[1] Several attempts have been made recently to deduce the chief results of Planck's theory without making explicit use of (216). (196) is taken as a result known empirically for equilibrium radiation and the Hamiltonian principle of least action is assumed. None of the real difficulties are avoided by this process; the difficulty is to see how, if that principle is true, (196) can be true for equilibrium radiation.

freedom and possesses potential as well as kinetic energy, for then the energy of each atom at temperature T will be $6 \times \frac{1}{2}kT$. But Dulong and Petit's law is not universally true. Einstein first suggested that the deviations from it could be explained by supposing that the whole heat energy of the solid is the energy of the oscillators in the atoms and, consequently, that C is the whole atomic heat.

We have
$$C/C_0 = \frac{e^{\frac{h\nu}{kT}}\left(\frac{h\nu}{kT}\right)^2}{\left(e^{\frac{h\nu}{kT}} - 1\right)^2} \qquad \text{..................(219).}$$

The right-hand side is a function of $\frac{h\nu}{kT}$; when this quantity is very great $C/C_0 = 0$, when it is very small $C/C_0 = 1$. Thus if $\frac{h\nu}{kT} < 1$, $C/C_0 > 0{\cdot}92$ and if $\frac{h\nu}{kT} > 10$, $C/C_0 < 0{\cdot}004$. In the first case the specific heat of the solid will be practically equal to the normal value; in the second case it will be practically zero. If the constitution of the oscillators and, therefore, ν are independent of T, all elements will have the normal value for their atomic heats at sufficiently high temperatures and all will have zero atomic heat at sufficiently low temperatures. It has long been known that the atomic heat of even those elements which have abnormally low values at ordinary temperatures approaches the normal value as the temperature is raised; the recent researches of Nernst and his pupils show that the atomic heat of all substances falls rapidly with the temperature and is almost zero at the lowest temperatures which can be reached. The exact variation of the atomic heat with the temperature is not generally given precisely by (219), but in that equation it has been assumed that there is only one kind of oscillator and that its vibrations are independent of T. In § 117 we shall see reason to believe that there are a large number of oscillators with different frequencies present, all these frequencies being less than a limit which is characteristic of the substance; this limit is approximately equal to the single frequency introduced by the simple theory. If this assumption is made the predictions of this theory of specific heat agree admirably with experiment.

It will be observed that by means of (219) we can calculate from an abnormal value of the atomic heat at a single temperature the value of ν, assuming that there is only one kind of oscillator present. At ordinary temperatures ($T = 300$) the condition for an atomic heat which lies between the normal value and zero is $6{\cdot}5 \times 10^{13} > \nu > 6{\cdot}5 \times 10^{12}$; both these values of ν correspond to the infra-red region of the spectrum. All elements which have abnormally low specific heats at ordinary temperatures must have a value for ν in the infra-red. The following table gives the values calculated for ν from the ratio C/C_0 at ordinary temperatures.

TABLE VIII.

Element	C/C_0	ν
S and P	0·91	$7{\cdot}1 \times 10^{12}$
Fl	0·84	9·1 ,,
O	0·67	14·3 ,,
Si	0·64	15·0 ,,
B	0·45	20·0 ,,
H	0·39	23·1 ,,
C	0·30	25·0 ,,

Now in Chapter II we saw that the frequency corresponding to an absorption band was equal to the natural frequency of some vibrator in the atom; hence the elements named in the table ought to have absorption bands in the positions indicated by the values of ν. Observations on the elements themselves are not available to test this view, but if we examine such compounds as have been investigated, we find the following result:

Substance	ν for absorption bands	ν for elements
CaFl	12·5, $9{\cdot}5 \times 10^{12}$	$9{\cdot}1 \times 10^{12}$
$CaCO_3$	44·8, 26·3, 10·2 ,,	14·3, 25·0 ,,
SiO_2	35·3, 33·3, 14·5 ,,	14·3, 15·0 ,,

The second column gives the observed absorption bands, the third the values of ν from Table VIII corresponding to the atoms contained in the compounds. The agreement is perhaps not very good, but the fact is to be remarked that no compound is known which has an absorption range between the limits

$$6{\cdot}5 \times 10^{13} > \nu > 6{\cdot}5 \times 10^{12},$$

corresponding to an abnormal value of the atomic heat calculated on Einstein's theory, which does not contain an element which has an abnormally low value of the atomic heat.

Indeed it is rather difficult to explain why the agreement should be even as good as it is. For in Chapter II we concluded that all the absorption bands in the infra-red represented the free vibrations of atoms and not of electrons inside atoms, and it is surprising to find that we can predict anything at all about the vibration of an atom in a compound from the vibrations of the same atom when it forms part of the solid element; we should have imagined that the forces under which the atom would vibrate would be entirely different in the two cases. It may be noted that the electrons in transparent bodies, having all frequencies in the ultra-violet, contribute nothing to the specific heat; the difficulty which has been so often raised, as to why a body possessing as many degrees of freedom as are indicated by its optical properties has not a higher specific heat, is overcome; those degrees of freedom which are most important optically correspond to frequencies so high that they never obtain an appreciable amount of energy.

The conception that the heat energy of a solid body consists of the energy of the oscillators in its atoms and that the mean energy of these oscillators is $3L_\nu$ has been applied also to the phenomena of electrical conductivity and magnetic susceptibility. We will notice only the former. In Chapter III we assumed that the mean energy of a free electron was αT or $\frac{3}{2}kT$, which we then thought was the mean kinetic energy of a molecule, and we found that, in order to explain the proportionality of the resistance and the absolute temperature, we had to conclude that the mean free path of an electron, l, was proportional to $T^{-\frac{1}{2}}$. Recent work has shown that at very low temperatures the resistance of pure metals falls more rapidly than T, so that the resistance of mercury is inappreciable at the temperature of liquid helium which is some degrees above the absolute zero. It appears that the relation between σ and T can be explained if we continue to assume that $l \propto T^{-\frac{1}{2}}$ and that the kinetic energy of a free electron is equal to that of a molecule, but that the kinetic energy of a molecule is $\frac{3}{2}L_\nu$ (the other half of its energy being potential), where ν is the natural frequency of the mercury molecule as determined by observations on the 'remaining rays' (p. 44) at ordinary temperatures. The explanation of the matter will not be complete until some adequate reason is adduced for believing that l is proportional to $T^{-\frac{1}{2}}$, but still the new hypothesis does throw some light on the difficulty concerning the specific heat of conductors which was mentioned in § 45. The external source of heat communicates energy only to the oscillators. If we could suppose that the energy of an oscillator consists of the energy of a vibrating electron, the similarity in respect of specific heat of insulators and conductors might be explained; for the only difference between the two cases would be that in the insulator the electron vibrated within the atom and could not move under an external electric field, while in the conductor it moved outside the atom and so could drift from one to the next under such a field. There are several objections to such a view; one is based on the fact that optical experiments show that infra-red oscillators (and these alone have any appreciable heat-energy) do not consist of electrons; others will probably occur to the reader; but it is probably in the view that the energy of the free electrons is not derived immediately from the source of heat but only through the agency of the oscillators that the solution of the difficulties will be found. At present it is wise to attempt none but the vaguest suggestions.

The structure of light.

101. The considerations of the last paragraph are based only on the assumption that Planck's theory gives correctly the mean energy of an oscillator; they do not involve necessarily the fundamental equation (216) by the use of which that result was actually deduced. Now we will make use of that equation and suppose that the energy of an oscillator cannot change by amounts less than $\delta = h\nu$, so that the oscillator will not be able to emit light at all unless it is given an amount of energy not less than δ. There are two sets of experiments which have been mentioned already which can be explained by the use of this supposition. The first are those which show that the velocity of the particles giving the 'moving line' in the positive ray spectrum (p. 206) is always above a certain limit. For, if the energy in that line is derived from the kinetic energy of the particles, that kinetic energy must be greater than $h\nu$, where $h\nu$ is the frequency of the line. On p. 207 we attempted to explain this observation on the grounds that the particle must possess sufficient energy to be ionised by collision before it could emit light; the explanation now offered is not necessarily inconsistent with the earlier idea, but may provide a hint as to the physical basis of the apparent discontinuity in the changes of the energy of an oscillator (see below)[1]. The second set is that summed up in Stokes' Law (p. 216). We have attributed the phosphorescence or fluorescence of bodies which obey that law to the impact on their molecules of electrons liberated by the photo-electric action. Now if ν is the frequency of the exciting light, experiments show that the energy of these electrons is never greater than $h\nu$, where h is very nearly the same as the quantity denoted in this chapter by that symbol. But if the oscillators cannot take up energy from more than one of the electrons at a time, the energy which they gain cannot be greater than $h\nu$ and, in consequence of the hypothesis we are now considering, no oscillator will be stimulated to emit light if its frequency is greater than ν. Only those oscillators will be stimulated by the exciting light of which the frequency is less than that of the light. This is Stokes' Law. In the case of substances which do not obey that law, the excitation is probably not caused by the photo-electric effect, so that the considerations which have just been advanced do not apply.

But of course the most interesting matter which we have to consider in this connection is the approximate identity of the h in (190) with the h in (216); it has been left till now because it leads to a new hypothesis which is suggested by, but not actually involved in, Planck's theory. If (190) is true, the kinetic energy of the emergent photo-electric electrons is determined by two factors[2]; first, the quantity w_0 which depends on the substance

---

[1] It may be noted that the kinetic energy of the slowest hydrogen particles which emit the moving hydrogen line is about 80 volts, a quantity which is probably considerably greater than that required to ionise a hydrogen atom by collision.

[2] It will be assumed that the two h's are actually identical; the experiments alone are not sufficient to establish the fact with certainty, but they are sufficient to make considerations based on that identity highly interesting. It must be remembered that, apart from the theory here considered, there would be no reason

from which the electrons are liberated and, if the electrons are liberated from oscillators, may depend on ν', the frequency of those oscillators, and, second, the quantity $h\nu$ which depends only on ν, the frequency of the incident light. The form of (190) suggests that the condition of the electron absorbing the light determines only how much of the energy taken from the incident light is transformed into kinetic energy of the emergent electrons; while the amount of energy taken depends only on the frequency of the incident light. Now hitherto we have regarded the quantity $\delta = h\nu$ as determined by the properties of the oscillators; their energy could only change by steps of this magnitude, and, if the energy of the radiation of corresponding frequency also changed by steps of the same magnitude, it was only because in the circumstances considered this radiation could draw energy from no source but the oscillators of the same frequency. But now it is strongly suggested that this view ought to be inverted, that it is the radiant energy which can only be absorbed in steps of magnitude $h\nu$ and that the energy of the oscillators can only change by the same steps, because in the circumstances they can only draw energy from radiation of the same frequency. When we were considering a state of equilibrium between the oscillators and the radiation, there was no means of distinguishing between the two views or any combination of them, but when we come to consider states which are not those of equilibrium it is very important to decide which we are going to adopt.

The view that the finite discontinuous changes in the energy are determined by the properties of the oscillators has one great advantage, namely, that it might be possible to explain the fact without any very fundamental changes in our theories. If the emitting and absorbing system were an electron within the atom, then it is probable that changes in its energy while it remained within the atom would be inappreciable by any observations; it would only be when the atom gained energy by the entrance of an electron, or lost it by the exit of one, that appreciable changes in its nature would occur. Now the energy involved in either of

---

for supposing that the two quantities were even of the same order of magnitude. It will also be assumed that the electrons are all liberated with the same velocity, differences being introduced by their subsequent history.

these two changes would be perfectly definite and it would probably increase with the frequency of the vibrations of the electrons; for that frequency increases with the magnitude of the elastic forces holding the electron within the atom and, therefore, with the energy required to drag it out. The energy of the oscillator might really vary continuously, but it would only be when its energy varied by a certain finite amount that changes which would produce any effect on our observations would occur.

Such an hypothesis concerning the action of oscillators may explain some of the facts, but it cannot explain all. If it is to be applied to the emission of complete radiation it requires the assumption that all absorption and emission of this radiation takes place in oscillators of definite natural frequency. Now, though this assumption is involved in the theory which led to our discussion, our developements have carried us further and further from it. The theory of specific heat sketched in § 100 leads to the view that the heat energy of a body is distributed among oscillators, not of all possible frequencies, but of a comparatively small number of definite frequencies all of which are less than a certain limit. Unless we are to conclude that the remarkable agreement between experiment and theory which was found is misleading, it is absolutely necessary to assume that the part of the complete radiation emitted by such bodies which has frequencies higher than this limit does not come from oscillators of the same natural frequency as the radiation. Now if each form of radiation does not come from an oscillator of the same frequency, the hypothesis which has been suggested as to the constitution of oscillators offers no explanation as to how the radiation comes to be emitted in finite amounts $h\nu$, and yet it is the assumption that the radiation is so emitted which is really fundamental in Planck's theory of radiation and leads to an agreement with experiment. In the calculations of § 99 the oscillators might be left out of consideration completely; we might apply (207) to the energy of the radiation U_ν and not to that of the oscillators L_ν. Assuming that U_ν is always a multiple of $h\nu$, and remembering the estimate given in § 97 for the number of frequencies which the radiation can have between the limits

ν and $\nu + d\nu$, we might consider the number of ways in which the P energy elements could be distributed among the N trains of radiation in unit volume of the enclosure and should arrive again at the value for E_ν given by (215). From E_ν we might now deduce L_ν as in (205) and obtain the same result for the mean value of an oscillator, but the value now would be determined by the discontinuity of the changes of energy in the radiation and not by the constitution of the oscillator itself. We might still proceed to the conclusions of § 100, if we assumed that in measurements of specific heat the heat energy was transferred by means of radiation, and yet avoid the inconsistency of assuming that the number of groups of oscillators with different frequencies in the substance was infinite and then concluding that it was small and finite.

Light-quanta.

102. On this ground alone it appears desirable to suppose that (216) applies to the radiation apart from any considerations about oscillators (though it may apply to oscillators also apart from any considerations about radiation) and that radiant energy of frequency ν, from whatever source it is emitted and by whatever recipient it is absorbed, is always composed of bundles of finite amount $h\nu$, so that if a recipient takes any energy from the radiation it must always take a complete bundle. This view has been put forward by Einstein, who terms a bundle of magnitude $h\nu$ a 'light-quantum'; the hypothesis is that an elementary recipient of radiant energy can take from light a whole quantum at one time and can never take less than a quantum; there may be cases in which it can take several quanta at the same time. The view, of course, requires a complete reversal of our conceptions of the structure of light. According to the view which has prevailed hitherto, the light from a source however small spreads out uniformly round the source and a recipient of energy can take continuously that part of the energy which falls on it; if it subtends at the source a solid angle ω it may take from the source during a given time a fraction $\omega/4\pi$ of the energy emitted by the source in that time but no more. According to Einstein's view it can always take up a quantum $h\nu$, so long as that quantum has been emitted from the source and a lower limit to the amount of energy which it can take up is not set by the angle ω.

There is one remarkable feature connected with the photo-electric effect, to which reference has not yet been made, which shows that the older view is quite inadequate and that some reconstruction of our ideas is necessary, while it is in complete agreement with Einstein's theory. A photo-electric effect can certainly be observed when the energy falling per second on 1 sq. cm. of the substance is much less than 1 erg (which is equivalent to a standard candle at a distance of 2 metres), and the energy of each electron liberated by it can be certainly greater than 10^{-12} erg (corresponding to 0·6 volt). On the ordinary theory of light an electron cannot absorb more energy than falls on the molecule in which it is contained; but the area covered by the section of a single molecule is certainly less than 10^{-15} cm.^{2} It appears then that no electron could acquire the energy with which it actually emerges unless light had acted for 10^3 secs. or about $\frac{1}{4}$ hr.; until this time has elapsed from the moment when the light was turned on, there should be no photo-electric effect. As a matter of fact the effect appears to start absolutely simultaneously with the action of the light.

But if the light energy is done up in bundles or quanta, and an electron can take up a whole quantum instantaneously, then there can be a photo-electric effect resulting in the emission of a single electron as soon as the *total energy* of the light emitted is equal to one quantum. Now the source contemplated is emitting light at the rate of 5000 ergs a second, so that only 2×10^{-16} sec. need elapse before a quantum is emitted. This time is, of course, wholly inappreciable.

The theory of light-quanta also explains at once the relation (190), if it be assumed that in the photo-electric effect an electron cannot take more than one quantum at a time; for the energy given to an electron will always be $h\nu$ and will depend only on the frequency of the incident light, while only the fraction of this energy appearing as kinetic energy will be determined by the absorbing substance. But too much weight must not be attached to this argument, in view of the comparatively slight evidence for the universal validity of (190). Again the hypothesis explains why the velocity of the photo-electric electrons is independent of the intensity of the incident light. The magnitude of the quanta does not depend upon the intensity of the light, which is determined only by the number of them falling on the substance in unit time; so long as this number is so small that an electron never has an opportunity to take up more than one quantum before it is ejected from the substance or loses its energy by collision with other systems, the amount of energy it can receive will be wholly independent of the intensity of the light.

In the next chapter we shall meet with other and still more convincing proofs that the older theory of the structure of light radiation is inadequate and that many facts can be explained by means of Einstein's hypothesis. Attempts have been made by Planck and by Sommerfeld to prove that this hypothesis is unnecessary and that the older theory, combined with the assumption of discontinuous energy changes in oscillators, is sufficient, but to me at least these attempts do not appear successful. It is quite possible that an oscillator, when it receives energy from a source other than radiation, such as the impacts of electrons, changes discontinuously, but it seems also necessary to suppose that the energy of light, even when it is not acting on oscillators of its own frequency, changes discontinuously. But of course Einstein's hypothesis presents difficulties, which must now be considered; not the least of these is to devise any theory, similar to that proposed above for oscillators, whereby a connection may be traced between the discontinuous energy changes in radiation and our general theory of electrical phenomena. One attempt to overcome this difficulty has been made.

The essential feature of the older theories is that the processes of the emission and absorption of radiation by two small systems *A* and *B* are represented as unsymmetrical. Even if *A* and *B* are precisely similar it is regarded as impossible that *A* should absorb all the radiation emitted by *B* or *vice versa*; if *A* is absorbing as much radiation as it emits, the radiation which it absorbs does not come from a system similar to itself but from a very great number of such systems distributed all round it. If we reject this view and imagine that it is possible for all the radiation emitted by *A* to fall on *B* and be absorbed by it, the difficulties as to the time required to acquire a given amount of energy are overcome. It has been pointed out by J. J. Thomson that this supposed dissymmetry is by no means a natural consequence of our fundamental hypothesis that radiation consists of disturbances propagated along Faraday lines. Once we accept the conception of these lines, it is at least as natural to suppose that a single elementary radiator emits disturbances along one line only as that it emits them along a very large number of lines distributed round it in all directions; it is a most obvious consequence of the view that the charge on a body is contained on a finite number of indivisible electrons that the medium round that body should be regarded as composed of indivisible lines each of which is associated with one electron. But if the radiation from a single element is sent along one Faraday line then, of course, all that radiation could be absorbed by another single system through which that line passed. According to this suggestion a disturbance travelling along a single Faraday line would form a bundle of energy indivisible except in so far as it possessed a finite length; it might be difficult to explain why a recipient on the line always absorbed the whole train of waves and not only a part of it, but difficulties connected with the time necessary to absorb the energy would vanish; for the time required for the longest train to pass over a receiving system would always be short compared with any that could be observed.

The speculation is notable as the only attempt that has been made to visualise the mechanism by which the energy of radiation is divisible into quanta of finite amount which can be absorbed almost instantaneously by a system on which they fall. It requires a revision of the whole foundation of electromagnetism as applied to electrons or other elementary systems, since the inverse square law for the attraction would not hold for such systems. But the equations of Chapter I might still hold for systems containing large numbers of electrons (such as alone can be observed directly), for the number of Faraday lines connected with such systems would be so large that they might be considered as distributed uniformly in the space surrounding the system. However, it does not overcome two great difficulties connected with the conception of quanta of radiant energy, on which a few words must be said.

In the first place we must ask what are the quanta composing radiation which does not consist of a train of homogeneous light of a single frequency ν, and, if quanta are indivisible, how does it come about that such radiation is resolved by a grating into homogeneous trains characterised by different quanta? No answer can be attempted to this question. It can only be pointed out that, if the quantum theory of radiation is true, all the assumptions on which the resolution of light by a grating was explained become dubitable or demonstrably false; we are once more left in complete ignorance as to how that resolution is effected and as to the relation between the nature of the resolved and the incident light. We may after all have to return to the Newtonian idea and say that the incident light really does 'consist' and has been made up of the homogeneous trains into which it is resolved. The objections which were urged against that view vanish if we suppose, as according to the quantum theory we must suppose, that our theory of the mechanism by which radiators emit radiation is wrong. Homogeneous trains existing in white light need not certainly be supposed to come from radiators vibrating harmonically in the corresponding frequencies; even the most irregular motions might, for all we know, give rise directly to trains of regular vibrations.

In the second place there are difficulties connected with interference and diffraction[1]. If quanta are indivisible, two beams which are brought to interference, whatever their source, contain different quanta. Why then do beams from the same small source interfere and those from different sources not interfere? The ordinary explanation based on irregular phase differences is inadequate unless we suppose that there are phase relations between the quanta emitted from the same source which do not hold between those from different sources. A more serious difficulty is

---

[1] Of course the first difficulty really concerns interference also, for gratings and prisms act by producing interference, and Thomson's theory could explain resolution by a grating if these difficulties could be overcome. It should be observed that on that theory diffraction is not required to explain the rectilinear propagation of light; for the light would travel along the Faraday lines radiating from the source and would only be stopped if the lines fell on an absorbing body. The lines, outside the field giving rise to the disturbance, would be straight, so that an explanation of geometrical shadows follows immediately.

that numerical calculation shows that in some cases when interference and diffraction are observed with very feeble lights the number of quanta in unit volume of the radiation is so small that it would be very improbable that two quanta would fall on the same point at the same time; and if they do not so fall it is difficult to give any explanation of their interference. Diffraction patterns have been photographed when the intensity of the light was such that there could not have been more than one quantum in each cubic decimetre of space. If a single quantum occupies any appreciable fraction of this space how can it be absorbed by a single electron; if it does not, how can it interfere with another quantum?

The problem of solving these and similar difficulties without abandoning the fundamental conceptions of the Planck-Einstein theory is the most important of modern physics. Of the truth of the essential features of that theory there can be no doubt; the universal constant h which it introduces is as important as the charge or the mass of an electron; but a satisfactory explanation of its physical nature has not yet been attained.

103. In order that the developement of these ideas should not be interrupted a few subsidiary questions have been raised but have not been discussed. It will be convenient to notice them briefly here.

(1) What is the relation between the thermal radiation characteristic of a body and complete radiation?—The former may be regarded as a complete radiation on which is superimposed radiation determined by the natural frequency of the oscillators peculiar to the body. As in Jeans' theory, if the body is in such a condition that the energy it radiates is immediately removed from it and does not fall on it again, those frequencies to which energy is transferred most rapidly will have more than their normal share of energy. But if the energy does not escape and true equilibrium is attained, these frequencies are also the more absorbed and their excess of energy is removed.

(2) What light does the theory of complete radiation throw on the nature of the disturbances constituting white light?—Of course, if the light-quantum theory is correct, all conclusions

as to the nature of those disturbances based on observations on the resolved spectrum are unreliable, until a new theory of that resolution is developed. But, even if the facts could be reconciled with the older theory of light, it appears impossible that, even if we could determine the nature of the disturbances which fall on the grating, we could conclude therefrom anything about the nature of the radiators from which they were originally emitted. For in all cases of thermal radiation, the amount of the radiation and, probably, the nature of the disturbances of which it consists, are determined as much by absorption as by emission. A disturbance originally emitted from one radiator may be absorbed in another and its original form utterly distorted.

(3) How would the light-quantum theory explain cases of fluorescence which do not obey Stokes' Law?—Probably by the absorption by the fluorescent radiators of more than one quantum from the original light. When the fluorescence is due to resonance it is not difficult to see that such absorption might occur, for the system might be able to store energy for a considerable time.

(4) What bearing have the conclusions of this chapter on those of the last?—All that was said about resolution of spectra is dubitable unless the problems raised in the last paragraph are solved. Conclusions concerning the nature of the systems emitting line spectra are rendered doubtful only if the nature of a disturbance gives no clue to that of a radiator; these spectra lie almost entirely in the visible and ultra-violet regions; they therefore possess no appreciable heat energy and nothing can be ascertained concerning them from studies of the condition of thermal equilibrium.

REFERENCES.

The general theory of equilibrium radiation is expounded in most text-books on Heat; the account at the beginning of Planck's *Theorie der Wärmestrahlung* (Barth, 1906) is much the best that I know.

For Rayleigh's theory see *Phil. Mag.* 49. 539, 1900, and *Nature*, 72. 54, 1905. Jeans' theory in its final form is given in *Phil. Mag.* 17. 229, 1909. Planck's theory is contained in his book just mentioned; the later form of it in *Deutsch. Phys. Gesell. Ver.* 13. 3, 138, 1911.

Einstein's developements of the theory are to be found in *Ann. d. Phys.* 17. 132, 1905; 20. 199, 1906; 22. 180, 1907. All Einstein's papers are especially lucid and interesting and should certainly be read.

The most important features of Nernst's work on specific heats are summarised in *Ann. d. Phys.* 36. 395, 1911. Other researches on the properties of matter at low temperatures are described in various Communications from the Physical Laboratory at Leyden during the last few years, chiefly by Kammerlingh Onnes.

A very large number of papers have been published recently discussing various aspects of the Planck-Einstein theory; a consideration of the alternatives discussed in § 101 is contained in a paper by Ehrenfest, *Ann. d. Phys.* 36. 91, 1911. Two very interesting summaries leading to discussions by various authorities should certainly be consulted:—Einstein, *Phys. Zeit.* 10. 817, 1909, and Sommerfeld, *Phys. Zeit.* 12. 1057, 1911.

J. J. Thomson's speculations on the structure of light were first published in his book *Electricity and Matter* (Constable, 1904) and developed in *Phil. Mag.* 19. 301, 1910.

A most interesting volume containing discussions of the subjects treated in this chapter by the most eminent authorities has just been published under the title *La Théorie du Rayonnement*, edited by Langevin and de Broglie (Paris, Gauthier-Villars: 1912). Its appearance makes the other references given here almost unnecessary.

CHAPTER XI

X-RAYS AND γ-RAYS

104. X-RAYS (which are also known as Roentgen[1] rays) are produced when kathode rays strike a solid body; they are usually obtained from discharge tubes of special form, containing gas at a very low pressure, so arranged that almost all the kathode rays leaving the kathode fall on a small area of a solid body, called the anti-kathode, placed in their path. γ-rays are emitted by any radioactive source of β-rays. These two groups of rays are now universally considered to be of the same nature, partly on account of the general similarity of their properties and partly on account of theories as to their structure; the evidence for this view will appear as we proceed. They differ from α- and β-rays by being undeflected in electric and magnetic fields; they differ from light by being incapable of regular reflection and refraction in non-crystalline media (see § 114). All X-rays and γ-rays produce ionisation in gases and darken a photographic plate.

General properties.

Like all other rays, X-rays and γ-rays are diminished in intensity by passing through material bodies, and this absorption depends, of course, both on the nature of the body and on the circumstances in which the rays are produced. In a given sheet of matter the absorption of γ-rays is in general very much less than that of X-rays, but the difference between γ-rays and X-rays in the matter of absorption is not as great as that between the different kinds of X-rays emitted under different conditions.

---

[1] I prefer the term X-rays because it is shorter and because it is that given by the discoverer. Nowadays perhaps it would be best to use X-rays as a general term to include both Roentgen- and γ-rays.

Rays which have a great penetration and are little absorbed are termed 'hard,' those which are easily absorbed 'soft'; a group of rays which is harder than another in respect of the penetration of one kind of material may be softer than that group in respect of the penetration of another kind of material; if rays are to be classified according to their hardness the substance in which they are absorbed must be stated. Aluminium is the substance chosen for the classification usually employed.

In general experiments on absorption alone do not yield much information as to the nature of the rays, owing to the complications arising from scattered and secondary radiations. But it is found that under certain conditions X-rays can be produced which are absorbed in any substance in accordance with the simple equation (168). Rays so absorbed are termed 'homogeneous.' It appears that all homogeneous rays for which the value of λ in the same absorbing medium is the same are alike in respect of all their qualities and can differ only in intensity; but two groups of heterogeneous rays, absorbed to the same extent in a given sheet of material, may differ in their absorption in other sheets of the same material or in their absorption in other materials or in other qualities. Homogeneous rays, then, but not heterogeneous rays, can be identified by means of their value of λ for aluminium, just as β-rays are identified by their velocity, or light by its wavelength. A group of heterogeneous rays can always be regarded as composed of a mixture of homogeneous rays of different qualities in suitable proportions.

The absorption of a given beam of homogeneous rays by a given mass of any material is independent of the thickness of the layer over which that mass is distributed. Accordingly it is convenient to introduce a new constant λ_m, determined by the nature of the rays and the absorbing medium, defined by

$$I = I_0 e^{-\lambda_m m} \quad\text{.......................}(220),$$

where m is the mass of a portion of the layer having unit surface. It is easily seen that, if ρ is the density of the absorbing material,

$$\lambda_m = \lambda/\rho \quad\text{.......................}(221);$$

λ_m measures the absorption of the material per unit mass, λ per unit thickness.

When X-rays or γ-rays fall on any material body, they may excite a scattered radiation and two kinds of secondary radiation[1]; one of these kinds consists of β-rays (i.e. electrons moving with velocities sufficient to enable them to ionise gases), the other kind of X- or γ-rays. In accordance with § 69, the scattered radiation is distinguished from the second kind of secondary by the fact that the quality of the former is determined wholly by and is nearly identical with the quality of the incident rays, while the quality of the latter is not determined wholly by that of the incident rays. The total energy of the scattered or secondary rays is always less than that of the primary.

X-rays may be polarised; the experimental evidence is not sufficient to determine whether the same is true of γ-rays. The X-rays produced by kathode rays in a discharge tube are never completely polarised; they are usually partially polarised, the plane of polarisation (defined as in § 69) being perpendicular to the plane containing the X-rays and the kathode rays. The polarisation of the softer rays emitted by such a tube is more complete than that of the harder. On the other hand the scattered rays are completely polarised, the plane of polarisation being that containing the incident and the scattered ray. The secondary X-rays are unpolarised.

We will now consider briefly the chief properties of the scattered and secondary radiations, the study of which has led lately to a great advance in our theory of the nature of the rays.

Scattered radiation.

105. If a beam of X-rays is allowed to fall on a plate of some material and the ratio of the intensity of the X-rays emergent from the plate to the intensity of the incident rays is measured, it is found that this ratio depends on the thickness of the plate, on the material of which it is composed, on the hardness of the incident rays and on the angle between the emergent and the incident rays. If the plate is so thin that it absorbs no considerable fraction of the incident rays,

---

[1] It has not yet been established experimentally that all bodies struck by X-rays can emit secondary X-rays. Scattered X-rays and secondary β-rays are excited by all X- and γ-rays in all bodies; secondary X-rays by some rays in some (probably all) bodies.

the ratio ceases to be dependent on the precise thickness; a thick plate gives a different ratio from a thin only because it absorbs the softer portion of the heterogeneous incident rays and so alters their quality; accordingly, if the plate is always thin, this factor may be left out of consideration.

If both the hardness of the incident rays and the angle between them and the emergent rays which are being measured are kept constant, the effect of the material alone may be investigated. If the rays are such as are sent out by an X-ray tube of average hardness ($\lambda_m =$ about 10 for aluminium), it is found that so long as the material on which they fall contains no element of which the atomic weight is greater than 40, the intensity of the emergent rays is nearly proportional to the mass of the material in which they are excited, whatever its chemical nature; the hardness of the emergent rays is almost exactly the same as that of the incident and varies continuously with that hardness[1]. Such radiation is 'scattered' X-radiation. It is completely or almost completely polarised, as may be shown, according to § 69, if the radiation is allowed to fall on another material body and the radiation excited by it in that body examined; its plane of polarisation is that containing the original incident beam and the polarised scattered beam.

Now, keeping the material and the hardness of the incident beam the same, the variation of the scattered to the incident beam with the angle between them can be investigated. Since the incident beam is partially polarised, this ratio will vary with the plane containing the scattered and incident beams, but if the mean value of the ratio is taken for all positions of this plane its value will be very nearly that which would be given by an unpolarised incident beam. If R_e, R_i, R_p are respectively the values of the ratio for the scattered radiation emerging in the direction in which the incident beam travels, in the opposite direction to which it travels and at right angles to this direction[2],

---

[1] Some experiments indicate that the scattered rays are slightly softer than the incident.

[2] It is convenient to adopt Bragg's nomenclature according to which scattered or secondary radiations leaving the scattering material on the side at which the incident radiation emerges and travelling in the same direction as that radiation are called 'emergence radiation,' while those which leave the scattering material

then it is found that $R_e > R_i > R_p$. The ratio R_e/R_i varies with the hardness of the rays; it is very much greater for γ-rays than for X-rays; the evidence as to whether it is greater for hard X-rays than for soft is conflicting, but the difference between the X-rays and the γ-rays suggests that it is greater for the harder rays. The value of R_i/R_p also varies with the hardness of the incident rays; for the very softest rays it appears to be 2; it is less for harder rays.

The hardness of the scattered radiation depends entirely on that of the incident; if it did not we should not call it scattered; but it appears not always to be of exactly the same hardness as the incident. In the case of the γ-rays the scattered radiation is certainly somewhat softer than the incident and the incidence is softer than the emergence; in the case of X-rays there is no certain evidence of a change of hardness in scattering. It is probable that at least part of this apparent change in hardness is due to the heterogeneity of the incident rays. In a given layer of material soft rays are scattered more than hard; accordingly in the first layers through which the rays pass the softer portion will be scattered leaving only the harder portion to be scattered in the last layers; moreover in passing through the scattering layer from the place where they are produced to that where they emerge the softer portions are, of course, absorbed more than the hard; both these causes tend to make the incidence radiation appear softer than the emergence. There are also complications arising from the secondary radiations; it cannot be said that there is any evidence that the scattered radiation from an infinitely thin layer which emits no secondary differs in hardness from the incident radiation.

The total intensity of the scattered radiation is, of course, less than that of the incident; its energy appears to be derived purely from that of the incident. The amount scattered increases with the amount truly absorbed, but not always in the same proportion, so that no definite statement can be made as to the fraction of

---

on the side on which the incident rays are incident and travelling in the opposite direction to those rays are called 'incidence radiation.' But the reader must be careful to distinguish incident from incidence radiation. The former is primary, the latter scattered or secondary.

the energy of a given beam of rays which is converted into scattered radiation when it passes through unit mass of material. In the case of rays of average hardness from an X-ray tube about one-tenth of the whole amount of energy abstracted from the rays in passing through a thin layer of material which emits no secondary X-radiation appears to be converted into scattered radiation.

Secondary X-radiation.

106. If, while the other factors are kept constant, the atomic weight of the material constituting the radiator is increased, the ratio of the emergent to the incident radiation increases very markedly after an atomic weight of 40 is passed. When chromium (52) is reached, the emergent radiation is about 300 times as great as it was for the light elements. At the same time the emergent radiation has undergone a change in character; it is now very much softer than the incident radiation and its hardness is independent of that of the incident; it is almost completely unpolarised and its intensity is almost independent of the angle between the emergent and incident rays. The fact that the qualification 'almost' has to be inserted in these statements is due to the fact that this new form of emergent radiation, which has all the characteristics of secondary radiation, is mixed with a small proportion of scattered radiation, such as emerges from radiators of lower atomic weight. The secondary radiation from an element is homogeneous and its hardness depends only on the nature of the element; the nature of the secondary radiation from a compound is determined only by the elements contained in that material and not by their state of chemical combination. There is a secondary radiation of definite hardness emitted from any radiator containing chromium and another emitted from any radiator containing zinc; chromium oxide or zinc oxide emits the homogeneous radiations characteristic of chromium and zinc respectively, together with a certain amount of scattered radiation due to the presence of the oxygen; a mixture of chromium and zinc would emit a heterogeneous radiation which might be regarded as made up of the radiations characteristic of chromium and zinc. Accordingly we may speak of zinc rays or tin rays, meaning thereby the homogeneous secondary radiations emitted from zinc and tin, and these expressions will denote X-radiation of perfectly definite quality. The following table due to Barkla, who has been foremost in such investigations, gives the value of λ_m in aluminium for the secondary radiations emitted from radiators containing some of the elements from calcium to bismuth. It will be observed that, in general, the characteristic radiation of the element of higher atomic weight is harder.

TABLE IX.

Element	Ca	Cr	Fe	Cu	Br	Mo	Ag
Atomic Weight	40	52	56	63·5	80	96	108
λ_m in Al	435	136	88·5	47·7	16·4	4·7	2·5 700

Element	Sn	I	Ba	Ce	W	Au	Bi
Atomic Weight	119	127	137	140	184	197	208
λ_m in Al	1·57 ?	0·92 306	0·8 224	0·6 ?	? 33	? 25	? 19

The ratio of the intensity of secondary radiation to that of the incident radiation is a maximum when the incident rays are just harder than the secondary radiation characteristic of the substance on which they fall; it is zero if the incident rays are softer than this radiation and small if they are much harder. This relation is most easily demonstrated by allowing the homogeneous radiation from one element to fall on another and observing whether it excites in it the homogeneous radiation characteristic of it; it is found that the radiation from an element *A* will excite the homogeneous radiation from another element *B* only if the atomic weight of *A* is greater than that of *B*. Zinc rays excite copper, iron or chromium rays in radiators containing those metals, but copper, iron or chromium rays do not excite zinc rays when they fall on zinc, nor do zinc rays excite selenium or tin rays; the intensity of the secondary radiation excited by a given beam of zinc rays is greater when they fall on copper than when they fall on chromium.

It will be seen that for some elements two values of λ_m are given. In these cases the secondary radiation is not homogeneous. If a very thin screen is used, the value of λ_m calculated from (220) for iodine is 306, but as the thickness of the screen is increased this value decreases; after a certain thickness is attained, the value of λ_m calculated for any additional thickness becomes constant and equal to 0·92. These facts are readily explained if it be supposed that the radiation is made up of two homogeneous radiations of widely different hardness; the very thin screen absorbs practically only the softer radiation and so gives the value of λ_m for that radiation, while after the thickness is sufficient to absorb all the softer radiation, an additional thickness gives an absorption determined only by the properties of the harder constituent.

An examination of the table undoubtedly suggests that the emission of more than one homogeneous radiation is not peculiar to some elements only; it would seem probable that there are two series of homogeneous radiations, in each of which the hardness of the radiation increases with the atomic weight of the element, while the radiation in one series is always much harder than that in the other for the same element. The fact that only one radiation has been discovered in most elements can be readily attributed to experimental limitations. It is impossible to detect with any certainty a very soft radiation because it is absorbed before it can enter any measuring instrument, partly in the substance which emits it, partly in the air or the apparatus surrounding it. It is quite possible that elements of lower atomic weight than silver emit radiation belonging to the softer series, but since the value of λ_m for this series in the case of such elements would be greater than 700, it would be completely absorbed in a few centimetres of air and would probably escape detection.

Again it would be very difficult to observe the very hard radiation which would be emitted by the heaviest elements, if these elements possess a radiation belonging to the harder series. For we have noted that in order to excite a given secondary radiation we must always use a primary which is harder than it. Accordingly in order to obtain the very hard radiation which we should expect to be emitted by bismuth, we should have to use very hard primary rays; but the hardness of the primary X-rays which we can obtain is limited by the potential difference which we can apply to the X-ray tube (see below) and it is not possible to produce primary rays which are considerably harder than the cerium rays for which $\lambda_m = 0{\cdot}6$; if there are very hard bismuth rays, they would be harder than such rays and no X-rays which we can produce would excite them.

But can we not use the γ-rays, which are harder than X-rays, to produce these radiations? For γ-rays λ_m may be as low as 0·04, but here again there are difficulties, for we cannot obtain γ-rays as intense as X-rays; also it should be noted that as the amount of secondary radiation produced is a maximum if the primary rays are only slightly harder than the secondary, the γ-rays may be too hard to produce much secondary radiation. The production of cerium rays by the incidence of γ-rays on cerium compounds has been observed and affords perhaps the most certain experimental evidence for the essential similarity of γ- and X-rays, but the production of secondary rays harder than those of cerium does not seem to have been recorded. Certain anomalies which have been noted in experiments on the 'scattering' of γ-rays by lead may be due to the fact that the emergent rays in these cases are not simply scattered but are also partly secondary.

It may be concluded that it is highly probable that all elements emit two homogeneous radiations, one of which is very much harder than the other. The series of hard radiations (denoted by the upper row in Table IX) is termed by Barkla the *K* series; the soft radiation (the lower row) is termed the *L* series. There is some evidence for the existence of a series harder than the *K* series, for it appears that under the action of very hard primary rays the elements of low atomic weight (less than 40) give an emergent radiation which is not purely scattered, but contains a part which is secondary radiation characteristic of the element. The radiations belonging to the *K* and *L* series characteristic of these elements would be too soft to be detected; but if there is a *J* series, harder than the *K* series, the radiations belonging to it characteristic of the heavier elements would be too hard to be excited by any primary rays at our disposal, and it would only appear in the case of the lighter elements.

Whiddington has pointed out that there is a simple relation between the *K* and *L* series which may be stated as follows:—

The radiation belonging to the L series emitted by an element of atomic weight W is identical with that in the K series emitted by an element of atomic weight W', where

$$W' = \tfrac{1}{2}(W - 50) \quad \ldots\ldots\ldots\ldots\ldots\ldots(222).$$

There is a very obvious analogy between the emission of these characteristic secondary radiations under the action of X-rays and the fluorescence of various substances under the action of light. In both cases we observe the emission of a true secondary radiation, characteristic of the radiator and not of the exciting rays, and the members of the K and L series for the same element may be analogous to the different lines in the fluorescent spectrum. Further, if hard rays are analogous to light of high frequency, the rule that the secondary rays are always softer than the primary would correspond to Stokes' Law[1]. On the ground of this analogy Barkla terms the characteristic secondary radiations 'fluorescent radiations,' but while the nature of the X-rays is still uncertain, it is unwise to describe the facts in terms which connote one definite theory rather than another.

Secondary β-rays.

107. When X-rays or γ-rays fall on any material body they cause the emission from its surface of electrons moving with considerable velocities, which form rays essentially of the same nature as β- or kathode rays. The relation between the velocity of these electrons and the exciting rays or the material in which they are excited may be stated in the following propositions:—

The velocities of the electrons are not all the same but probably vary continuously from a certain maximum to zero; nothing appears to be known concerning the proportion of electrons having a velocity which is a given fraction of the maximum, but probably the greater number have velocities near the maximum.

The maximum velocity is determined only by the quality of the exciting rays. It does not depend on their intensity, nor upon the material upon which they act. If the exciting rays are those characteristic of an element of atomic weight W belonging to the

---

[1] It should be noted however that Stokes' Law does not apply to bodies which emit a fluorescent *line* spectrum.

K series, the relation between v_s, the maximum velocity of the secondary β-rays, and W is

$$v_s = kW \quad \text{......................(223)},$$

where k is very nearly 10^8 cm./sec.

If the exciting rays belong to the L series, the same relation, in accordance with (222), is given by

$$v_s = \tfrac{1}{2}k(W - 50) \text{....................(224)}.$$

There is then a secondary β-radiation which is characteristic of a given element just as there is a secondary X-radiation, and we may speak of tin β-rays, just as we spoke of tin X-rays, meaning β-rays of a perfectly definite velocity, which is given by substituting for W in (223) the atomic weight of tin. It must be carefully remembered, however, that, while the tin X-rays are the rays emitted from a radiator of tin, the tin β-rays are not those emitted from a radiator of tin, but the fastest of those emitted from any radiator whatsoever when the tin X-rays are incident upon that radiator.

It will be observed that (223) provides a third method of defining the quality of a given beam of homogeneous X-rays. We may define it not only by the values of λ_m and W, the atomic weight of the element of which the rays are characteristic, but also by v_s. The last method is theoretically much the most satisfactory, for it does not depend on the properties of any material body.

The number of electrons emitted per second by a beam of X-rays or γ-rays of given intensity increases very greatly when the hardness of those rays is increased from that just insufficient to excite the characteristic secondary radiation in the radiator to that just sufficient to excite that radiation. But it must be remembered that the maximum velocity of the electrons does not change especially rapidly at this point; it changes by the same amount over the same range of hardness of the primary rays whether or no that range includes the radiation characteristic of the radiator.

Even when the ratio of the secondary β-rays to the incident rays is a maximum, the number of electrons emitted from a given mass of material by the most powerful rays we can produce is very much less than (not more than one-millionth of) the number of atoms over which the rays have passed. The rays never cause nearly every atom subject to their influence to emit a secondary β-ray.

If the radiator has the form of a thin sheet through which the exciting rays pass normally the number of emergence secondary β-rays is much greater than that of the incidence. The ratio of the number of the emergence to the incidence rays increases as the atomic weight of the radiator decreases and as the hardness of the exciting rays increases; it may vary from 1 to 40 under actual experimental conditions. But the value of v_s appears to be the same for incidence and emergence rays.

So far as is known, the velocity of the secondary β-rays, like all the quantities determining the action of X-rays and γ-rays on matter, is independent of the temperature.

There is a remarkable relation between the velocity of the secondary β-rays excited by X-rays from a discharge tube and that of the kathode rays which are falling on the anti-kathode of that tube and so causing the emission of those X-rays. It has long been known that the hardness of the rays emitted by an X-ray tube increases with the velocity of the kathode rays in that tube[1]. Whiddington has shown that, if v_p is the least velocity of the kathode rays in the discharge tube which, falling on any anti-kathode, will give rise to X-rays hard enough to excite the characteristic X-radiation of an element of atomic weight W (K series), then the relation between v_p and W is again (223) and that, within the limits of experimental error, the constant k has the same value as in (223). Now the softest radiation which will excite the characteristic radiation is that which is just harder than that characteristic radiation; accordingly the remarkable fact appears that the slowest kathode rays which will excite a given X-radiation are the fastest rays which that X-radiation can excite when it falls on another body. To put the matter another way,

---

[1] It was first discovered that the hardness of the rays increased as the gas pressure in the tube was decreased, the apparatus driving the discharge through it remaining the same. It is now known that this variation with the pressure is due to the fact that when the pressure is high a larger current flows through the tube so that the potential difference between the electrodes (which determines the speed of the kathode rays) is less.

the fastest kathode rays in the tube emitting the X-rays have the same speed as the fastest secondary β-rays excited by those X-rays.

It is probable that a similar relation holds in the case of γ-rays, the β-rays which are emitted at the same time as the γ-rays corresponding to the kathode rays exciting the X-rays. The speed of the fastest secondary β-rays excited by the γ-rays appears to be the same as that of the β-rays emitted by the source of the γ-rays. The evidence in this case is not quite so complete partly on account of experimental difficulties and partly because it is impossible to procure homogeneous γ-rays.

Two other similarities between the emission of secondary β-rays by X-rays and the emission of X-rays by kathode rays must be noted. If those kathode rays fall on an anti-kathode so thin as to cause little absorption in the X-rays excited, the amount of X-rays emerging in the direction in which the kathode rays travel is greater than that emitted in the opposite direction. The relation between the direction and the intensity of the X-rays excited by kathode rays is the same as that between the direction and the intensity of the β-rays excited by X-rays.

Again the amount of X-rays excited in a given anti-kathode by a given number of kathode rays increases very rapidly when the velocity of those rays becomes just great enough to excite the radiation characteristic of that anti-kathode; just as X-rays excite an abnormally large number of β-rays of velocity v_s when v_s is the value characteristic of the element on which they act, so β-rays of velocity v_s excite an abnormally large number of X-rays when the same condition is fulfilled.

Absorption and ionisation.

108. The existence of these secondary β-rays raises a very interesting question. It has been said that X-rays and γ-rays produce ionisation in gases and cause the emission of secondary and scattered X- or γ-rays. But now we know that these rays produce secondary β-rays, and we also know that such β-rays can produce ionisation in a gas and, when they strike on any material body (as in the case of the anti-kathode of a discharge tube), cause the emission of X-rays. Are then the effects which we have attributed to the X-rays a direct effect of these rays or do they merely represent the action of the β-rays excited by these rays?

If it had been known that the X-rays produced β-rays before it was known that they produced ionisation and secondary X-rays, it would probably never have been doubted that the effects apparently due to the X-rays were immediately due to the β-rays which they produce. But as the order of discovery was reversed, and as some people find difficulty in abandoning a belief even when the grounds on which they originally founded it have been removed, there has been some discussion of the question, to which a few words must be devoted.

Let us consider the ionisation first. The question might be settled definitely by determining the amount of β-rays excited by a beam of X-rays and the amount of ionisation produced by those β-rays and examining whether the amount of ionisation so produced is equal to the amount produced by the X-rays. Bragg's experiments in this direction are completely in accordance with the view that all the ionisation is due to the β-rays excited by the X-rays in the walls of the ionisation chamber and in the gas itself[1]. But the experiments are difficult and not completely conclusive. The β-rays excited by the hardest X-rays which can be obtained have a velocity so low that they are completely absorbed in a few millimetres of air, and the determination of their absorption coefficient in the gas and the walls of the vessel, which enters into the calculation, is a matter of great difficulty. The experiments can apparently be reconciled with the opposite view that there is some ionisation produced directly by the X-rays.

Support for the view that the ionisation is all due to the secondary β-rays may be derived from a study of the ionisation of different gases. The ratio of the amounts of ionisation produced in two gases by the same beam of rays (the total absorption of the rays by the gases being small) depends on the nature of the rays; it is different for the same two gases according as the rays are α-rays or β-rays. But it is exactly the same for β-rays and γ-rays, so that it is probable that the actual ionising agent is the same

---

[1] The secondary β-rays excited in a gas have never been observed directly, but there is plenty of indirect evidence that such rays are excited in gases as in solids.

whether the rays directly incident on the ionisation chamber are β-rays or γ-rays. On the other hand experiments in which the ionisation chamber is placed in a strong magnetic field have been used to support the contrary view; it has been argued that if the ionisation is due to β-rays from the walls (the effect of the β-rays excited in the gas may be neglected if the gas is light and the rays hard), it should be reduced to zero if the chamber is placed in a magnetic field strong enough to 'curl up' the paths of the β-rays and prevent them traversing the gas. But this argument may be answered by pointing out that there is no evidence that there is such a field; the magnetic field alters the direction of the paths of the β-rays, but there is no reason for thinking that it can reduce the path of the rays in the gas to zero.

But the question has been definitely settled by the beautiful experiments of C. T. R. Wilson in which the tracks of ionising rays through a gas are made visible (p. 170). When the gas is ionised by X-rays the ions are not found to be distributed more or less uniformly through the gas, as they would be if they were produced directly by the X-rays, but are concentrated along a large number of fine irregularly curved lines, which resemble exactly the paths which such soft β-rays as are excited would be expected to follow. The space between the lines is wholly free from ions. The lines start suddenly where the β-ray is produced in the gas and the irregular changes of direction are doubtless due to the scattering of those β-rays, which would be peculiarly intense in the case of such slow rays. The length of the lines agrees well with the value, determined from experiments on absorption, of the distance which these β-rays are able to travel in the gas.

The ionisation produced by X-rays is, then, due entirely to the secondary β-rays which they excite. It is much more difficult to determine whether the secondary X-radiation is due to the impact of the β-rays on the substance in which they are excited, or whether it is due to a direct action of the original X-rays on the atoms of the substance. There is no conclusive experimental evidence in favour of the first view, but there is no evidence of any kind against it. The secondary β-rays must produce some characteristic secondary X-rays, for they are precisely similar to the kathode rays which produce such X-rays when they strike the same substance in a discharge tube; there is no reason for believing that they do not produce all. Perhaps the best evidence for this view is to be based on the difference between the scattered and the secondary X-radiation in the matter of polarisation; the former is polarised in a way that depends on the incident radiation and therefore, whatever view is taken of the nature of that polarisation, probably represents the direct action of the incident radiation. The secondary radiation is unpolarised and it is accordingly likely that it is not produced directly by the primary, but through some intermediate process which destroys the relation between directed quantities in the primary and the secondary; such an intermediate process would be provided by the liberation of the secondary β-rays.

Apart from any questions of theory the facts that have been described enable order to be introduced into the very complicated results of measurements of the absorption of the X-rays. Suppose that a narrow pencil of X-rays is allowed to fall on a thin layer of absorbing material and the ratio of the energy of that part of the emergent rays which is included in the same pencil as the incident rays to the energy of those incident rays is measured. The energy of the emergent pencil will be less than that of the incident because energy has been subtracted from the incident rays by scattering and by the emission of secondary β- and X-rays; the energy of the secondary β-rays is truly absorbed in the layer by producing ionisation in it (unless, as is probable, a considerable part of it is converted into secondary X-radiation), while the energy of the scattered and secondary X-radiation is distributed almost uniformly round the radiator, so that only an inappreciable fraction of it enters the measuring chamber. The fraction of the incident energy which appears to be 'absorbed' is that which is transformed into these three forms of radiation, and the absorption coefficient would be more accurately regarded as a transformation coefficient of the energy from its original to its derived form. The rate of transformation into scattered radiation by unit mass of the substance increases regularly with the atomic weight of the substance and the softness of the rays, but we have seen that the rate of transformation into secondary β- and X-rays is a maximum when the incident radiation is just harder than the secondary X-radiation characteristic of the substance. Accordingly, we should expect that, while the absorption would generally increase with the atomic weight of the absorber and decrease with the hardness of the rays, a sudden increase in the absorption should occur when this condition is fulfilled. Some of the experimental facts are given in Table X, in which the first column gives the kind of rays (i.e. the elements of which they form the characteristic rays of the *K* series), the remaining columns the values of λ_m for these rays in the absorbing material stated at the head of the column.

TABLE X.

	Absorber					
Rays	C	Al	Fe	Cu	Zn	Ag
Cr	15·3	136	103·8	143	170·5	580·5
Fe	10·1	88·5	66·1	95·1	112·5	381
Cu	5·22	47·7	268	53·0	60·9	214
Zn	4·26	39·4	221	55·5	50·1	175
Se	2·04	18·9	116·3	149·8	174·6	87·5
Ag	·41	2·5	17·4	24·3	27·1	13·3

A vertical column read downwards gives the absorption of a single material for rays of increasing hardness (measured in aluminium), i.e. rays characterised by increasing values of v_s; a horizontal row read from left to right gives the absorption of the same rays in materials of increasing atomic weight. It will be observed that the numbers in the columns decrease regularly except at the point where the rays become capable of exciting the characteristic radiation in the absorber; then they increase rapidly, so that the harder rays are sometimes absorbed more than the softer. In a horizontal row the figures increase more regularly, but there is a remarkable minimum when the absorber is such that the rays just can not excite the characteristic radiation in it, followed by a rapid increase; the rate of transformation into secondary β-rays appears to be especially small just before the sudden increase in its value. It may be observed that if the ratio of the figures in successive columns for the various rays is taken it is found that this ratio is constant so long as the rays are not able to excite the characteristic radiation in the absorber. Thus for all rays which fulfil this condition the value of λ_m in aluminium is about 9 times that in carbon. We shall refer to this result later.

109. The origin of γ-rays. Throughout this chapter less is said of the γ-rays than of the X-rays, chiefly because they differ from very hard X-rays only by being more difficult to investigate experimentally. But a few words must be said about a very interesting theory of the origin of the γ-rays which has been based by Rutherford on the consideration of the characteristic secondary X-radiation.

Apparently all radioactive bodies which emit β-rays also emit γ-rays, but there is no simple relation between the quantity or the hardness of the γ-rays emitted and the velocity of the β-rays. It is found that the β-rays of those substances which emit a large quantity of γ-rays are not homogeneous in velocity, but can be divided into a considerable number of groups (often more than 20) each of which is homogeneous. Now, since it appears that when a radioactive atom breaks up it emits never more than two β-rays and usually only one, it is unlikely that these β-rays were originally ejected with different velocities: it is far more probable that they were originally homogeneous and have subsequently suffered different losses of energy. Rutherford suggests that they have lost energy by exciting the characteristic radiation of the atoms from which they were ejected and that it is the radiation which forms the γ-rays; for it has been shown by direct experiment that β-rays excite γ-rays in solid substances on which they strike, just as kathode rays excite X-rays.

Evidence for this view may be derived from the fact that the hardness of the γ-rays is very approximately that predicted for the characteristic radiation of such radioactive elements by extrapolating Barkla's results to elements of great atomic weight. Further, at least in one case, the difference between the energy of the fastest group of β-rays and that of any other can be accurately expressed by $pE_1 + qE_2$, where p and q are integers, and E_1 and E_2 are not very different from the values obtained from (223) and (224) for the energy which must be spent by a β-ray in order to excite the X-ray of series K or L characteristic of the radioactive atom. This relation is immediately explained if the groups of slower rays are those which have lost energy by exciting p X-rays of series K and q of series L. The theory is very convincing and, though it does not yet explain accurately all the facts, it is certainly substantially correct. It is of great theoretical importance, for it would show that a β-ray does not always lose all its energy, but sometimes only part, when it excites a γ-ray.

It has also been announced that γ-rays have been excited by the impact of α-rays and not only β-rays, on other substances. This result, if it is substantiated, is again of very great theoretical importance.

The Stokes-Wiechert theory.

110. The facts which have been enumerated so far can be stated quite satisfactorily without any theory of the nature of X-rays and γ-rays; the consideration of other important facts can only be based on some theory. We shall now consider the most important of these theories.

The first plausible suggestion as to the nature of X-rays was made simultaneously by Stokes and Wiechert; it was developed further by J. J. Thomson and in very great detail by Sommerfeld, who still maintains that in the form he has given to it it is capable of a satisfactory explanation of the facts. The theory states that X-rays differ from light only in the nature of the disturbance of which they consist and that the Maxwellian theory of light is fundamentally correct. X-rays, like light, are supposed to consist of transverse disturbances propagated along Faraday lines; the disturbances are generated by the impacts of the electrons constituting the kathode rays on the atoms of the anti-kathode which they strike.

Hitherto we have merely noted the fact that, according to Faraday's theory, any change in the motion of a charged body must result in disturbances propagated along the lines of force attached to it, and have made no attempt to calculate a relation between the change of motion and the consequent disturbance. Such a relation can be obtained from the fundamental equations of Chapter I; a change in the value of v will determine from (9) a change in H, which will again determine from (6) a change in D; the distribution of the vectors D and H which represent the state of the Faraday lines will be different when the charge is accelerated from that which obtains when it is unaccelerated, and the acceleration will result in a distortion of those lines. But the calculations based on these equations are difficult and they raise some very important and perplexing questions which will be discussed in the last chapter. A treatment of the matter has been given by J. J. Thomson, which, though not rigid, brings out more clearly the physical principles involved.

In Fig. 5 suppose that an electron is moving along the line XO with a velocity w, which is small compared with that of light. When it reaches the point O let it be subjected to a retardation

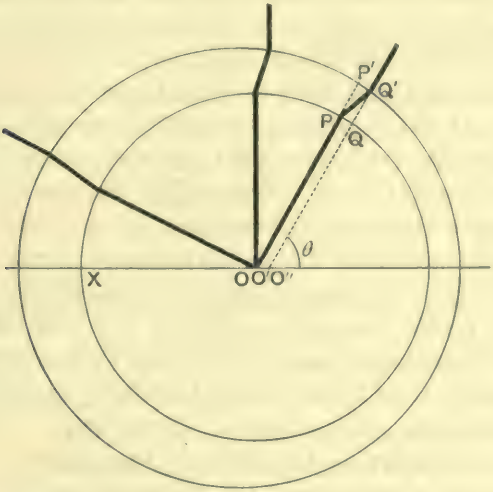


Fig. 5.

which brings it to rest at the point O'. The Faraday lines are flexible and they possess mass when they move at right angles to their length; consequently the parts of them remote from the electron will not stop immediately the electron is stopped, but will continue to move in that direction. But since their ends are attached to the electron, this motion will cause a distortion of the parts near the electron and they will be no longer in equilibrium under the forces acting on them; new forces will arise which will ultimately bring them into positions in which they are in equilibrium with regard to the electron after it has been stopped; a disturbance will travel out along the lines bringing them into this new position and this disturbance, like all others travelling along Faraday lines, will travel with the velocity of light.

Consider then the distribution of the lines at a time t after the electron has been brought to rest at O'. If a sphere of radius $r = ct$ is drawn about O', the disturbance will have reached all the lines within this sphere, and these parts will have assumed the final position of equilibrium, being uniformly distributed about O'. On the other hand, if τ is the time between the electron reaching O and being reduced to rest at O', the parts of the lines outside a sphere of radius $c(t+\tau)$ drawn about O, will be in the positions which they would have occupied if the electron had not been stopped but had continued to move uniformly; for an observer in the region outside this sphere could not be aware that anything had happened at O, for news of what happened at O could only reach him[1] with the velocity c. But if the electron had continued to move uniformly, the lines would have been uniformly distributed about O'', where $OO'' = wt$. Accordingly the distribution of the Faraday lines will resemble that shown by the thick lines in the figure. The disturbance due to the stoppage of the electron will be contained between the two spheres; the precise form of the lines in this region cannot be predicted, but it is clear that they will have a component tangential to the surface of the sphere; our calculations would not be affected if we assumed any other form for them than that shown in the figure.

Let us now calculate the value of the electric displacement and the magnetic intensity in the region covered by the disturbance. If OO' is very small compared with PP' (i.e. $\frac{1}{2}w\tau$ small compared to $c\tau$) we may regard the spheres as concentric and the region of disturbance as a spherical shell of thickness $\delta = PP' = c\tau$ drawn about O'. The electric displacement in any direction is

---

[1] We might be inclined at first sight to argue that, since the news of what happens at O is conveyed along the Faraday lines, the sphere of radius $c(t+\tau)$, beyond which the news has not reached, ought to be drawn about O' or O'' rather than about O. It will be seen in the last chapter why this deduction, which seems a natural consequence of the conception of Faraday lines, has not been preferred to that drawn from the principle that an observer distant r from O cannot be aware of what happened at O before a time r/c has elapsed. If the sphere had been drawn about O'' this principle would have been abandoned.

equal to the number of lines passing through unit area perpendicular to that direction. The lines which pass through that portion of a plane QQ' contained in the spherical shell and of a length q perpendicular to the paper are clearly those which within the sphere are contained between two planes perpendicular to the paper passing through $O'P$ and $O'Q$. If the angle between these last planes is $d\theta$, we must have $r d\theta = wt \sin\theta$, since

$$PQ = OO'' \sin\theta \quad \text{and} \quad OO'' = wt;$$

but $r = ct$ so that $d\theta = \frac{w}{c}\sin\theta$. But the laws of electrostatics show that the number of lines contained in the angle $d\theta$ is $\frac{\epsilon}{4\pi r^2} qr d\theta$ and the area of the plane through which they pass is δq; consequently D_p, the electric displacement perpendicular to QQ' and tangential to the shell, is given by

$$D_p = \left(\frac{\epsilon q}{4\pi r} \cdot \frac{w}{c} \sin\theta\right) \Big/ \delta q$$

$$= \frac{\epsilon w \sin\theta}{4\pi r c \delta} \qquad \text{(225).}$$

The electric displacement, D_s, along QQ' normal to the shell, is, of course, $\frac{\epsilon}{4\pi r^2}$. Hence $D_p/D_s = \frac{rw\sin\theta}{c\delta}$. If we consider only distances from the electron which are large compared with δ, the thickness of the shell, D_s is negligible compared with D_p, and we may regard the disturbance as consisting of an electric displacement $D = D_p$, tangential to the shell, travelling outwards with the velocity c. But these lines travelling at right angles to their length will produce a magnetic intensity H, perpendicular to the plane of the paper, where

$$H = 4\pi Dc = \frac{\epsilon w \sin\theta}{r\delta} \qquad \text{(226).}$$

The energy in unit volume of the shell is $2\pi D^2 + \frac{1}{8\pi c^2} H^2$; hence it can be shown immediately that the energy of the shell contained within a cone of solid angle $d\omega$, the axis of which makes an angle θ with OO'', is

$$S_\theta \, d\omega = \frac{\epsilon^2 w^2 \sin^2\theta}{4\pi c^2 \delta} \, d\omega \qquad \text{(227),}$$

and the energy contained in the whole shell is

$$S = \frac{2}{3}\frac{\epsilon^2 w^2}{c^2 \delta} \quad \text{......................(228).}$$

According to the Stokes-Wiechert theory, X-rays consist of such 'pulses,' in the form of spherical shells in which the electric and magnetic intensities are tangential, travelling out with the velocity of light from the place where the electrons of the kathode rays are stopped; but before we consider the evidence for this theory, we must develope further the formulae which have been obtained.

111. Since f is the retardation which brings the electron with velocity w to rest in time τ, we must have $w = f\tau$; hence, since $\delta = c\tau$, we may write (225), (226), (227), (228) respectively

Radiation from an electron.

$$D = \frac{\epsilon f \sin\theta}{4\pi r c^2} \quad \text{......................(229),}$$

$$H = \frac{\epsilon f \sin\theta}{rc} \quad \text{......................(230),}$$

$$S_\theta = \frac{\epsilon^2 f^2}{4\pi c^3}\sin^2\theta \, . \, \tau \quad \text{................(231),}$$

$$S = \frac{2}{3}\frac{\epsilon^2 f^2}{c^3}\tau \quad \text{......................(232).}$$

The amount of energy S is radiated in a time τ, so that the energy radiated per second is given by

$$S_0 = \frac{2}{3}\frac{\epsilon^2 f^2}{c^3} \quad \text{......................(233).}$$

Now the question arises whether this relation will hold when the electron is not brought to rest, but only has its velocity decreased by an infinitesimal amount dw in an infinitesimal time dt, so that $\frac{dw}{dt}$ is still equal to f. If we attempt to answer this question by considering, as we have done, the motion of the Faraday lines, we meet a grave difficulty. If the electron is not brought to rest at O' but continues its motion, reaching O_1 at the time t when the observation is made at P, the arguments which have been given would seem to show that the whole system of lines would have moved some distance to the right; there would still be a disturbance at P, but it would be propagated along the line O_1P and not OP, and the time at which it would reach P would be O_1P/c and not OP/c. It would seem therefore that our calculations would be altered. These considerations undoubtedly introduce a difficulty into our conception of Faraday lines which will be considered in the last chapter; but more rigid methods of calculation show that we may take (233) as giving generally the rate at which energy is radiated by an electron which is subject to any acceleration f. It should be noted that the energy radiated by an electron depends only on the magnitude and not on the direction of f; an electron will radiate in just the same way when it is started or accelerated as when it is stopped or retarded.

The formulae which have been developed so far are all based on the assumption that the velocity of the accelerated electron is always small compared with that of light. This condition is not fulfilled in many of the cases we shall have to consider. If w is comparable with c our theory will be incorrect for two reasons; firstly OO' will not be small compared with PP', so that the region in which the disturbance is contained can no longer be considered a spherical shell, but will have the form shown in the figure, being thinner on the side towards which the electron is moving; secondly we have seen reason (p. 173) to believe that our fundamental electromagnetic equations are not applicable in their present form to electrons moving with such velocities. We are not yet in a position to consider the second source of error, so that the result which is in accordance with modern theory will be simply quoted. It is that the energy emitted in a cone of angle $d\omega$, the axis of which makes an angle θ with the motion of the electron, in the time during which the acceleration f acts is $S_\theta d\omega$, where

$$S_\theta = \frac{\epsilon^2 f}{16\pi c^2}\frac{\sin^2\theta}{\cos\theta}\left(\frac{1}{(1-\gamma\cos\theta)^4}-1\right) \quad \text{......(234)}$$

and $\gamma = w/c$. It will be seen that, since $\delta = c\tau$, $w = f\tau$, (234) becomes identical with (227) when $\gamma = 0$. The whole energy emitted during the same time is

$$S = \frac{2\epsilon^2 f}{3c^2}\cdot\frac{\gamma}{(1-\gamma)^{\frac{1}{2}}} \quad \text{..................(235).}$$

As an illustration of these results we may apply them to the radiation of light, assuming that in this case w/c is so small that we can use (229–233). If the electron emitting the light is vibrating harmonically with a frequency ν, so that its motion is given by

$$x = A \sin 2\pi\nu t,$$

then $f = -4\pi^2\nu^2 A \sin 2\pi\nu t$ and the energy emitted by it during a single period, occupying a time $1/\nu$, is $\dfrac{16\pi^4\nu^3\epsilon^2 A^2}{3c^3}$. The whole energy of the vibrating electron is $\frac{1}{2}m_0 . 4\pi^2\nu^2 A^2$, so that the fraction of it radiated in a single period is

$$\frac{8\pi^2}{3c^3} . \frac{\epsilon^2}{m_0}\nu = 10^{-30} \times 5{\cdot}3 \times 10^{17} \times 4{\cdot}7 \times 10^{-10}\nu = 2{\cdot}5 \times 10^{-22}\nu.$$

In the case of the radiations which we are able to recognise as light ν is never greater than 4×10^{16}, so that the fraction of the energy radiated in a single period is very small and the electron will be able to execute a large number of vibrations before its energy is sensibly decreased. In other words, the damping of the vibrations by the emission of energy radiation is very small and our assumption that an electron is able to emit long trains of practically undamped vibrations is justified.

On the other hand, if the retardation to which the electron is subjected is very much more intense than in the case just considered, a considerable fraction of its energy may be emitted in the form of radiation and not given up to the system producing the retarding forces. Returning to (228), we see that the whole of the kinetic energy of the electron will be converted into radiation if δ, the distance in which it is reduced to rest, is such that

$$\frac{2}{3}\frac{\epsilon^2 w^2}{c^2\delta} = \tfrac{1}{2}m_0 w^2 \quad \text{.....................(236).}$$

Using the known values of ϵ/m_0 and ϵ, it appears that in this case $\delta = 4 \times 10^{-13}$ cm. But what will happen if δ is less than this distance; the electron cannot radiate more energy than it possesses? In § 61 we came to the conclusion that a, the 'radius of an electron,' that is the radius of the sphere outside which an electron behaves as a point charge, cannot be less than that given by the relation

$$\frac{2}{3}\frac{\epsilon^2}{ac^2} = m_0 \quad \text{.......................(237)}.$$

If we assume that a has the value given by (237) and substitute for m_0 in (236) the quantity $\frac{2}{3}\frac{\epsilon^2}{ac^2}$, we see that (236) states that the electron will radiate all its energy if it is stopped in a distance equal to twice its radius. This statement must not be interpreted too literally, for we do not know exactly what the radius of an electron means; we have no definite knowledge that the electron behaves as a point charge outside a certain sphere and as an uncharged body within it. But at least we may conclude that, for distances comparable with a, calculations based on the assumption that the electron behaves as a point charge are invalid and that, therefore, there is no reason to fear that (228) will ever lead to the conclusion that an electron suddenly stopped will radiate more energy than it possesses.

The loss of energy by a retarded electron owing to radiation will, of course, affect its motion in a given field of force; we are not strictly justified in assuming that the acceleration of an electron in a field of electric intensity X is always $\epsilon X/m_0$, whatever the value of X; but so long as the value of X is such that $\frac{m_0 w^2}{2\epsilon X}$, the distance in which, according to the ordinary theory, the electron would be brought to rest, is large compared with a, the radius of an electron, calculations of the motion of the electron which neglect the loss of energy by radiation will be practically correct. All the cases in which we have considered hitherto the motion of an electron fulfil this condition.

When the radiation from one electron falls on another at a distance r in a direction θ this electron will be subject to an acceleration

$$f_1 = \frac{4\pi\epsilon D}{m_0} = \frac{\epsilon^2 f \sin\theta}{m_0 r c^2} \quad \text{..................(238)}.$$

As a consequence of this acceleration, the second electron will emit radiation in its turn; the quantities determining this radiation (which will be distinguished by the suffix 1) may be found by substituting in (229–233) the values of f_1, θ_1, r_1; θ_1 is the angle between the direction of the excited radiation and the electric displacement in the exciting radiation. Since f_1 is always proportional to f, the relation between D_1 and t in the excited radiation will be the same as that between D and t in the exciting radiation; consequently the two radiations will have the same quality and the excited radiation may be termed 'scattered.' An example of such scattered radiation is to be found when light falls on a turbid medium and causes accelerations in the electrons contained in its particles; if the particles are separated by distances small compared with the wave-length of the light, the scattering becomes regular refraction, but the phenomenon is complicated by the fact that the electrons are subject to forces other than those due to the radiation, namely those determining their free periods.

The scattered radiation will be polarised, for it can be seen that the electric displacements in the incident and scattered beams must lie in the same plane; they are also perpendicular respectively to the direction of those rays. Accordingly, if the plane of polarisation is identified with that perpendicular to the electric displacement, it contains the incident and scattered beams. The intensity of the scattered radiation about a given incident beam is proportional to $\sin^2\theta_1$. If the scattered ray has the same direction as the incident, $\theta_1 = \frac{\pi}{2}$ and $\sin^2\theta_1 = 1$; if the scattered ray is perpendicular to the incident, θ_1 will vary from 0 to $\pi/2$ with the plane of polarisation of the incident beam. If this beam is unpolarised the average value of $\sin^2\theta_1$ will be $\frac{1}{2}$.

The velocity w_1 which the electron acquires under the influence of the radiation falling upon it is given by $w_1 = \int f_1 dt$. If the original radiation is that obtained by stopping completely an electron moving originally with a velocity w, we have

$$w_1 = \frac{\epsilon^2 \sin\theta}{m_0 r c^2}\int_0^\tau f\,dt = \frac{\epsilon^2 \sin\theta}{m_0 r c^2}\,w \quad \text{............(239).}$$

Hence if E is the initial kinetic energy of the electron which emitted the radiation, E_1 the final kinetic energy of the electron on which that radiation acts,

$$\frac{E_1}{E} = \frac{w_1^2}{w^2} = \frac{\epsilon^4 \sin^2\theta}{m_0^2 r^2 c^4} \quad \text{..................(240).}$$

Hitherto we have neglected to observe that the radiation consists of a magnetic intensity as well as of an electric displacement. As soon as the electron has acquired energy from the electric displacement the direction of its motion will be deflected by the magnetic intensity; but, since the force due to this action is always perpendicular to the direction of its motion, the energy of the electron will not be increased. A consideration of the direction of the electric displacement and the magnetic intensity will show that the force due to the magnetic intensity acts in the direction of the propagation of the radiation, so that the motion of the electron has a component in the direction of the propagation. From (18–20) it can be easily deduced that, if the radiation acts for a time t small compared with $1/\omega = c^2/H \,.\, m_0/\epsilon$, the ratio of the component of the velocity in the direction of the radiation to that in the direction of the electric displacement is $\frac{1}{2}\omega t$.

This tendency of the electrons when they are absorbing radiation to move in the direction of the propagation of that radiation represents the effect of what is usually termed the 'pressure of radiation.' The existence of such a pressure can be predicted from considerations much more general than those advanced here, which show that the pressure is a necessary consequence of the absorption of any form of radiation which conveys energy. It appears also that our fundamental conception of radiation as propagated along Faraday lines of force leads to the same conclusion; for a disturbance propagated along such lines may be regarded, as we have seen, as a set of lines superimposed on the undisturbed lines, travelling out along them with the velocity of the disturbance. Now these lines possess mass and, consequently, they must carry momentum as well as energy; when they are absorbed and disappear the body which absorbs them must receive their momentum as well as their energy and so tend to move in the direction of the propagation of the light.

112. We have now developed sufficiently the theory of the radiation from an accelerated charged body; let us compare with experiment the Stokes-Wiechert hypothesis that the X-rays consist of such radiation emitted by the kathode rays when they are stopped at the anti-kathode. Before we undertake this comparison, a further question must be decided. A beam of X-rays is characterised not only by its intensity, but also by its hardness and polarisation; what are the circumstances in the disturbance which determine these characteristics?

Comparison with experiment.

In the case of polarisation the question is easily answered; the plane of polarisation of the X-rays, like that of light, must be considered to be perpendicular to the electric displacement in the disturbance. Accordingly in the simple case shown in Fig. 5 the plane of polarisation ought to be perpendicular to that containing the motion of the electron and the X-ray which it emits. It has been stated that the X-rays emitted by kathode rays are partially polarised in this manner; that they are not completely polarised is doubtless due to the fact that the direction of the motion of the electrons is not always that of the kathode stream. β-rays or kathode rays are not stopped immediately they fall on a solid body; they are rather scattered. The acceleration to which an electron is subjected in being deflected will, of course, cause it to emit radiation, but, even at the first deflection, this acceleration will not be exactly in the direction of its original motion and after it has been deflected several times it will bear no relation to that direction[1].

It is always assumed that the hardness of an X-ray is determined by δ, the thickness of the pulse of which it consists, and that the thinner the pulse the harder the ray. This assumption is obviously based upon the theory of light in which the wave-length is a quantity of the same nature as the thickness of a pulse. If X-rays are like light at all, they must obviously be like light of

---

[1] It will be observed that it is left undecided whether the plane of polarisation is determined by the direction of the acceleration or of the original motion of the electron. Since the electric displacement is in the same direction throughout the pulse only when these directions coincide, it is only in this case that the plane of polarisation can be defined.

very short wave-length rather than very long, if only because there are no considerable gaps in our knowledge of light of long wave-length; we know all about light of which the wave-length lies between a small fraction of a millimetre and several kilometres, and no light in this region resembles X-rays in the least. On the other hand there is a definite inferior limit to the wave-length of light of which we know anything. Now hard X-rays are less like light than soft X-rays; they can penetrate a greater thickness of opaque matter; so that if soft X-rays differ from light in being characterised by a shorter length, hard rays must be characterised by one still shorter. At the present time other arguments which will appear incidentally to our discussion could be used to support this assumption, but the arguments given appear to be those on which it was originally put forward and accepted.

Accordingly, if the velocity of the electron is comparable with that of light, Fig. 5 shows that the hardness of the rays emitted by it should vary with the direction of emission and that the hardest rays should be those sent in the direction of its velocity, for the distance between the spheres is least in this direction. Such experimental evidence as there is appears to be in accordance with this conclusion.

A direct test of this theory of X-rays could be carried out if experiments were made on the variation of the energy of the rays with θ, w, and f. Unfortunately we have no direct knowledge of f, so that experiment and theory cannot be compared without making more or less arbitrary assumptions concerning this quantity. All that can be said at the present time is that no experiments on the average distribution of energy round kathode rays striking an anti-kathode with known velocities are certainly in conflict with (234); in many cases the assumptions concerning f which are necessary to make experiment and theory agree are so plausible that the theory may be said to be confirmed by the experiments.

If it is assumed that f is in the direction of w, then (234) states a relation between S_θ and θ which is independent of the magnitude of f and depends only on γ. It appears that S_θ should be zero when $\theta = 0$ and be a maximum for some value of θ (determined by γ) less than $\pi/2$. As γ increases the value of θ for this maximum decreases and a greater proportion of the rays is sent forwards, i.e. in the direction of the motion of the electron, and at the same time the sharpness of the maximum increases, until, when $\gamma = 0{\cdot}99$, practically all the energy of the rays is contained between $\theta = 2^\circ$ and $\theta = 10^\circ$. This prediction agrees qualitatively with experiment; it would probably be accurately true if we could measure only the polarised part of the beam, which is that given out when f and w are in the same direction.

In order to obtain a relation between S, the whole energy of the radiation which is emitted, and E, the original energy of the accelerated electron, the assumption has usually been made that l, the distance over which the acceleration lasts, is independent of the velocity; it then appears that

$$S/E = \frac{2\epsilon^2\gamma}{3c^2 l m_0 \sqrt{1-\gamma^2}} \quad\text{..................(241).}$$

Accordingly the ratio of the energy of the X-rays to that of the kathode rays which excite them should increase with the speed of the latter. Such an increase is indicated by experiment, but (241) cannot be accurately true, for we have noted that the amount of X-rays emitted by a stream of kathode rays increases very rapidly when the energy of those rays becomes just sufficient to cause the emission of the radiation characteristic of the anti-kathode. A better agreement of (241) with experiment is found, if that formula is applied to the emission of γ-rays and it is assumed that these rays are produced by the sudden starting of the β-rays from the atoms of the radioactive substance, just as the X-rays are produced by the sudden stopping of the kathode rays. If for γ-rays we use the experimentally determined values of S/E to calculate l from (241), it is found that l lies between

$$3 \times 10^{-11}\text{ cm. } (\gamma = 0{\cdot}78) \quad\text{and}\quad 10^{-12}\text{ cm. } (\gamma = 0{\cdot}95).$$

All that we can say of this result is that it is not necessarily inconsistent with the theory; we should expect, as we find, that l would turn out to be less than the diameter of an atom, but perhaps we should have expected it not to be so very greatly less. The basis of the assumption that l is constant was the idea that it must be the distance over which an atom exerts appreciable forces.

Another interesting assumption, suggested by Planck's theory of radiation, has been considered by Sommerfeld. The quantity h which, according to that theory, is constant for all forms of radiation has the dimensions (energy) × (time); Sommerfeld suggests that in the case of X- or γ-radiation the quantity which is constant for all velocities of the electron is the time integral of its energy during the process of acceleration, i.e. $\int_0^\tau E\,dt$. If we call this integral h, we find

$$S/E = \frac{2\epsilon^2\gamma^2}{3ch\,(1-\gamma^2)\,(1-\sqrt{1-\gamma^2})} \quad \text{.........(242)},$$

and the experiments on γ-rays indicate that h lies between

$$2{\cdot}4 \times 10^{-27} \quad \text{and} \quad 1{\cdot}4 \times 10^{-28}.$$

The order of these quantities is that of the h involved in Planck's theory ($6{\cdot}5 \times 10^{-27}$). If we use this value of h in (242), the calculated values of S/E do not differ from the experimental values of this ratio by more than can be attributed to the uncertainty of the observations.

δ, the thickness of the pulse, is determined by l and w. According to our simple theory $\delta = c\tau$, $l = \frac{1}{2} w\tau$, so that

$$\delta = \frac{2lc}{w} = \frac{2l}{\gamma},$$

but if γ is not very small, δ is not constant over the whole pulse and our calculations need revision in this and other respects. It is found that if δ_m is the mean thickness of the pulse

$$\delta_m = l\,\frac{1+\sqrt{1-\gamma^2}}{\gamma} \quad \text{..................(243)}.$$

From (243) and the values of l obtained above, we find that δ_m lies between 3×10^{-9} for hard X-rays and 5×10^{-10} for hard γ-rays. All these results are obtained on the assumption that the β-ray is completely stopped when it emits X-rays, but they would not be greatly changed if it is merely deflected through a large angle in place of being stopped. The heterogeneity of the rays from an X-ray tube is doubtless due to the fact that the deflection suffered by the β-ray is not always the same.

Let us now turn from the production of the X-rays to their effects upon material bodies.

The scattered radiation is doubtless to be identified with that which was termed 'scattered' on p. 282. The plane of polarisation of this radiation is that predicted by the theory. The theory also predicts that, if γ is small, R_e (p. 259) should be equal to R_i and equal to $2R_p$. The value predicted for R_i/R_p agrees well with experiment in the case of the softer rays; this is one of the few cases where a numerical relation predicted by any theory of the nature of X-rays can be tested directly, and the agreement is doubtless a strong argument for the validity of the theory. But the value does not agree for the harder rays, nor is the value predicted for R_e/R_i correct. It has been suggested that the discrepancy is due to the presence of a radiation having the same hardness as the primary, but not truly scattered, being produced by the impact of the secondary β-rays on the substance in which they are liberated. Since there are more emergence than incidence β-rays and since β-rays produce more emergence than incidence X-rays, this view would account for the greater amount of emergence radiation apparently scattered.

The scattering of X-rays is then, according to this theory, essentially similar to the scattering of light in a turbid medium; its presence would show that media which appear clear and homogeneous to light appear turbid to X-rays (cf. p. 299). Similarly the secondary X-radiation would be considered to be analogous to fluorescence; the fact that it is unpolarised suggests that it is analogous to the fluorescence which is excited by the impact of the photo-electric electrons rather than to that which is due to resonance; indeed we have already seen reasons for believing that it is due to the impact of the secondary β-rays. Thomson applies Lenard's theory of phosphorescence to the phenomenon and suggests that the secondary X-radiation is emitted when the secondary β-rays return to the atoms from which they have been ejected.

But it is in connection with the ejection of these electrons that the greatest difficulties of the theory we are considering are encountered. We have concluded that in most cases the energy which will be radiated by the β-ray in the form of X-rays is only a small proportion of the total energy of that ray, and that the fraction of this energy which is absorbed by an electron over which the rays pass is given by (239). According to (239) the fraction should change with the distance between the radiating and the absorbing electron and vary with the angle θ; indeed a comparison of (231) and (239) will show that the fraction depends on these two quantities in the same way as the square root of the energy that is radiated per unit of time through unit area perpendicular to the direction of propagation. On the other hand the experimental evidence shows that the energy of the secondary β-ray is independent of the intensity of the primary ray and simply equal to the energy of the β-ray giving rise to that primary ray; in fact it shows that E_1/E is always 1. Now the value of E_1/E given by (240) is always very much less than 1; in experimental conditions r is never less than 10, and $\sin^2\theta$ can never be greater than 1, so that the greatest possible value of the right-hand side of (239) is

$$\epsilon^2 . \left(\frac{\epsilon}{m_0}\right)^2 . \frac{10^{-2}}{c^4} = 8 \times 10^{-28}.$$

If we are to suppose that the secondary β-ray receives its energy by the action of the electric intensity in the X-rays in the manner contemplated in § 111 the only possibility of reconciling experiment and theory lies in supposing that the energy of the secondary β-ray is derived, not from a single pulse emitted by one original kathode ray, but from a succession of such pulses emitted by a large number of kathode rays. If we could imagine that the electron which is ultimately ejected and forms the secondary β-ray can store the energy which it receives from N successive pulses, then the value of E_1/E when it is ultimately ejected would be $N . \frac{\epsilon^4 \sin^2\theta}{m_0{}^2 r^2 c^4}$ and, if N were large enough, might be equal to 1.

An hypothesis of this kind involves, of course, assumptions as to the nature of the forces other than those due to the rays which act on the electron. There are very grave difficulties in the way of devising any mechanism which would give rise to forces of a suitable nature and, even if these could be overcome, there would be other objections to the explanation proposed. The forces which enable the electron to store the energy must be determined by the atoms of the substance in which it is contained, and it would seem that the amount of energy which an electron can store before it is ejected must depend on the nature of the substance as much as on the nature of the rays. We might be able to invent a means by which the electron could store energy, but it is almost impossible to invent one which would be such that the amount of energy which the electron could store would be wholly dependent on the nature of the radiation, but independent of its intensity and of the nature of the atoms. An even more serious objection can be based on the magnitude which would have to be attributed to the quantity N, the least number of pulses which would have to act on the electron before it could acquire the requisite amount of energy. N cannot be greater than the number of kathode rays which have struck the anti-kathode and consequently $N\epsilon$ cannot be greater than the quantity of electricity which has been carried between the kathode and the anti-kathode by the kathode rays. Measurements of the current conveyed by these rays show that the value of $N\epsilon$ for one second is never greater than 10^8, so that for this period N is not greater than 2×10^{17}. Hence in order that $1/(8 \times 10^{-28})$ pulses should be emitted, the current would have to flow for 6×10^9 seconds or about 200 years! According to the theory suggested, no secondary β-ray could possibly be emitted with the energy which it is actually found to possess until the X-ray tube had been in action for this time, even if an electron stored all the energy which falls upon it. Of course the secondary β-rays are really emitted so soon after the current in the tube is started that the interval has never been measured.

Another explanation of the difficulty has been considered; it has been suggested that the energy of the secondary β-ray is not derived from the X-ray at all, but from the atom from which it is liberated, and that the action of the X-rays is merely to stimulate the atom to emit an electron by some process similar to that involved in radioactivity. The hypothesis of such an action ('trigger' action it is often called) has been introduced to explain in turn all the difficulties connected with radiation, but nobody appears to support it now. The first objection mentioned in the last paragraph applies also to this hypothesis; a graver objection is that the total amount of energy contained in all the secondary β-rays liberated by a beam of X-rays is not greater than that which, according to the theory which has been discussed, is contained in the X-rays and is, moreover, the same on whatever material the rays act. If we regard the total energy of all the secondary rays liberated when a given beam of X-rays is completely absorbed as equal to the energy of that beam of X-rays, it appears that this energy agrees well with that calculated by the Stokes-Wiechert theory from the nature of the kathode rays giving rise to the X-rays; difficulties are raised only when we consider the distribution of this energy and inquire why it appears to be concentrated on a very small proportion of the electrons contained in the material on which the rays act, giving to this small proportion an energy far greater than is to be expected and to the rest none at all[1].

In view of these discrepancies it is not worth while to develope further the theory of the motion of an electron under the influence of the forces in the pulse constituting the X-rays, but it may be noted that some light is thrown on the difference between the emergence and incidence secondary β-radiation by the considerations of the pressure of radiation given on p. 283. We saw that the ratio of the momentum of the electron in the emergence direction to its total momentum should be $\frac{1}{2}\frac{H}{c^2}\cdot\frac{\epsilon}{m_0}t$. Remembering that the time during which the pulse acts on an electron is δ/c, where δ is the thickness of the pulse, and taking the value of H from (226), we conclude that this ratio is proportional to w, the velocity of the kathode rays exciting the X-rays. Accordingly the ratio should increase with this velocity and the ratio of the emergence to the incidence radiation increase with the hardness of the X-rays. This relation is actually found by experiment, but

---

[1] These difficulties are not overcome, even if we adopt the older view that the rays produce the ionisation, by which their energy is often measured, directly and not through the agency of the secondary β-rays. The X-rays ionise only a very small proportion of the atoms in a gas through which they pass and the energy which they give to each atom when they ionise it is still very much greater than that predicted by (239).

neither the measurements nor the theory are sufficiently definite to make a quantitative comparison possible.

113. The considerations which have been advanced suggest very strongly that an explanation of the discrepancy between theory and experiment is to be found at least as much in an error in the older view of the nature of X-rays as in our ignorance of the forces acting on an electron before it emerges as a secondary β-ray. It was these considerations which led Bragg to suggest his corpuscular theory of X-rays.

Bragg's theory.

The difficulties which are raised by the theory which has just been considered are all connected with the assumption that the energy of an X-ray pulse is distributed over a finite volume depending on the distance from its place of origin[1] and is not, like that of an α- or β-ray, concentrated in a very small volume independent of the distance from the origin. Bragg suggests that this assumption is false, that the energy of an X-ray, like that of the β-ray from which it arises, is concentrated on a corpuscle; these corpuscles start where the β-ray is stopped and, so long as they remain X-ray corpuscles, travel through material bodies without loss of energy and through electric or magnetic fields without deflection; but when they are passing through material bodies they are liable to be reconverted into β-rays, which once more may later be reconverted into X-ray corpuscles. The fraction of the X-rays which is converted into β-rays in passing through a given layer of material and the fraction of the β-rays which is retransformed into X-rays depend on the nature of the matter and the quality of the rays.

Stated in this form Bragg's theory is little more (and nothing less) than an adequate description of the experimental facts mentioned in § 108; the only assumptions made are practically certain, namely that the X-rays produce all their ionisation and secondary X-radiation through the agency of the secondary β-rays

---

[1] If the original kathode ray is moving with a speed near that of light the energy of its radiation is not spread uniformly over a sphere, but even in the case of the fastest β-rays the cone in which it is concentrated occupies not less than $\frac{1}{100}$ of the whole space round the origin. The considerations presented here apply also to the hardest γ-rays.

and that the energy of these β-rays is derived from the X-rays. In this form the theory is most certainly true and most highly important; it is true and it is important that when X-rays pass through material bodies there is a certain rate of transformation of β-rays into X-rays and *vice versa* and that this rate determines the phenomena of ionisation and absorption. But in this form it does not solve the main problem; it does not explain how the transformation between the β-ray and the X-ray is effected. For this purpose some hypothesis as to the nature of the X-ray corpuscle must be made.

Bragg's original suggestion was that the X-ray corpuscle consisted of a 'neutral pair,' formed by the association of the electron forming the β-ray with a positive charge of equal magnitude which it picked up when it was converted into an X-ray and lost again when that X-ray was converted into a β-ray. In the process of picking up the positive charge no energy was lost; accordingly, if the whole energy of the X-ray corpuscle, like that of the β-ray, is kinetic, the mass of the positive particle must be infinitesimal compared with that of the electron. Such a neutral pair would be undeflected in electric and magnetic fields; it is possible that, being uncharged, it would produce no ionisation; but it might be deflected in passing through a material and so give rise to the scattered radiation.

This theory is, of course, greatly superior to the Stokes-Wiechert theory in explaining how a secondary β-ray can receive the whole energy of an original β-ray, and why the ionisation due to the X-rays is produced entirely by the secondary β-rays. These propositions were suggested by Bragg on the ground of his theory before they had been established experimentally. But in many, if not all, other respects the 'neutral pair' is not superior to the pulse theory. It explains well why the emergence radiation is always greater than the incidence, for the corpuscle would possess the momentum as well as the energy of the original β-ray and so tend to move in the same direction, but it appears now that this difference is quite in accordance with the older theory. In respect of the relation between the rate of transformation of the rays and the nature of the matter on which they act there is little to choose between the theories; additional hypotheses as to the constitution of the matter have to be introduced in either case and these can be suited to one theory as well as to the other.

But there are very serious objections to Bragg's theory. The first of these—at the present time it is not fatal—is that it makes it necessary to suppose that, if any of the energy of a β-ray is converted into X-rays, the whole of that energy is converted. There is no direct evidence against this view, but it is becoming increasingly difficult; if Rutherford's theory of the origin of γ-rays is confirmed by further measurements, it will have to be rejected. Again, we have seen that β-rays in their passage through matter are subject to two kinds of deflection; they suffer many small deflections leading to a general curvature of path and a few large deflections. On the older theory it would be natural to associate the former with loss of energy by ionisation, the latter with loss of energy by the emission of X-rays; on Bragg's theory a β-ray which has given rise to an X-ray must disappear completely.

Another objection may be based on the difficulty of explaining by this theory how the X-rays can be polarised. Bragg has suggested that the neutral pair revolves in an orbit of which the plane determines the plane of polarisation, but it is not easy to see why this plane should remain perpendicular to the direction of propagation when the rays are scattered. And this is only a special instance of the essential weakness of the 'neutral pair' theory, which is its failure to explain the indubitable general resemblance between X-rays and light[1]. The main argument for accepting the theory at the time when it was proposed is now one of the chief reasons for rejecting it.

The most fatal objection to the theory would be the undoubted proof of the excitation of γ-rays by α-rays. The production of

---

[1] To me it appears a serious objection to Bragg's theory that it necessarily fails entirely in performing one of the chief duties of theories, that of correlating diverse phenomena, though it performs admirably the other of those duties, that of visualising the processes which it explains. No other physical phenomena give us the smallest reason to suspect the existence of positively charged particles of which the mass is small compared with that of an electron, so that the theory introduces a mechanism which is foreign to all other theories. It will be remembered that the 'positive electrons,' the existence of which was discussed in §§ 43, 66, have a mass comparable with that of an electron; even if we could

γ-rays by any particles but electrons can hardly be reconciled with the 'neutral pair' theory in its present form, while it is in complete accordance with the Stokes-Wiechert theory, which predicts that all accelerated charges should emit electromagnetic radiation.

114. X-rays and light. The neutral pair theory was originally proposed because of the inadequacy of the Stokes-Wiechert theory. The latter theory involves two assumptions, that the X-rays are similar to light, and that the Maxwellian theory of light is correct. At the time that Bragg proposed his theory, it was far more plausible to deny the first of these assumptions than the second; but the reverse is now true. We have seen in the last chapter that the Maxwellian theory is difficult to reconcile with observations on light; it is no more difficult to reconcile with observations on X-rays; an attempt to save the theory by denying that X-rays are like light is useless; the theory is past saving. And meanwhile the evidence of the similarity of X-rays and light has steadily accumulated; let us examine this evidence further.

It is, of course, obvious that the emission of secondary β-rays is very similar to the photo-electric effect. In each case we have electrons emitted with a speed that is determined by the quality but not the intensity of the incident rays[1]; in each case we find the difficulty that the energy of the electrons emitted is very much greater than can be accounted for by the Maxwellian theory. Further the occurrence of a sharp maximum in the emission of secondary β-rays when the nature of the substance and the hardness of the X-rays are suitably related is analogous to the 'selective' photo-electric effect. The difference between the two processes is that in the case of the X-rays the velocity of the electrons is

---

accept that hypothesis, the position of Bragg's theory would be unaltered. Of course Bragg's theory cannot be applied to light, because it is known that light travels with a constant velocity c, whereas the neutral pairs travel with a velocity dependent on their energy.

[1] Experiments have been described which seemed to prove that the emergence photo-electric effect was, like the emergence secondary β-ray effect, greater than the incidence. But the most recent work shows that these experiments are capable of a quite different explanation (see Partzsch and Hallwachs, *Sitz. Ber. Bayer. Akad.* 64. 147, 1912).

independent of the material from which they are emitted. But this difference is a natural consequence of the difference which must be supposed to exist between the two kinds of disturbances by any theory which assimilates X-rays and light; the X-rays must correspond to light of very short wave-length, and the quantity which corresponds to ν must be very much greater for X-rays. If ν is so great that in (190) $h\nu$ is always very large compared with w_0, the velocities of the electrons emitted from substances with different values of w_0 will be indistinguishable.

There are two kinds of experiment to which no reference has yet been made which, if they could be carried out in a manner open to no doubt, would go far to deciding whether X-rays are or are not similar to light, namely experiments on the velocity of the X-rays and experiments on their interference. If it could be shown that the X-rays do not travel with the velocity c, then there can be no doubt of their difference from light, while if they do travel with that velocity there can be little doubt of their essential similarity to light.

Many attempts, notably by Blondlot and Marx, have been made to measure the velocity of the rays, always by comparing it with that of Hertzian waves. In Marx' work, such waves, as they travel along wires, are made to charge alternatively positively and negatively a metal plate on which fall X-rays excited by the alternating current which is also the source of the Hertzian waves. Let us suppose that the distance of the plate from the X-ray bulb is the same, whether it is measured in a straight line or along the wires. The X-rays are only excited when the electrode designed to act as the kathode is actually negative; during the other half of the alternation of the potential difference no rays are excited. If the rays so excited arrive at the metal plate when it is positive, no secondary β-rays will leave it, while if they arrive when it is negative such rays will be emitted. If the X-rays travel with the same velocity as the Hertzian waves, the potential of the plate when the rays arrive at it will be the same whatever the distance of the plate from the bulb; but if one set of rays travel faster than the other, then for some distances the rays will arrive when the plate is positive and for some when it is negative, and the emission of the secondary rays will depend on the distance. From experiments based on this principle, Marx has concluded that the X-rays and Hertzian waves (which, of course, travel with the velocity c) have the same velocity; but though his experiments were conducted with great care and elaboration there appears to be some doubt as to the interpretation of his results, and the conclusion that the X-rays travel with the velocity of light cannot be regarded as incontrovertible.

It is impossible to perform with X-rays any experiments analogous to the classical optical experiments on interference, because up to the present time it has been impossible to divide a beam of X-rays into two parts subsequently brought together again by any process of regular reflection or refraction; but attempts have been made to observe a diffraction of X-rays when they pass through a narrow slit. An extension of the action of the rays into the geometrical shadow has undoubtedly been observed, but there are complications arising from the scattered and secondary radiations to be taken into account. However Sommerfeld concludes from the observations of Walter and Pohl that the distribution of the intensity of the rays about such a narrow slit is completely in accordance with the theory discussed in § 111, if it be assumed that the thickness of the pulse is not greater than 4×10^{-9} cm. It is to be remarked that the value for this thickness calculated from the ratio of the intensity of the X-rays to that of the kathode rays exciting them is of the same order of magnitude. This is important evidence in favour of the similarity of X-rays and light, but it is not absolutely conclusive, while the limits of error in the interpretation of the 'diffraction pattern' are so great.

Recently—almost too recently for adequate discussion in this volume—some very interesting experiments have been described which appear to show true interference by X-rays. If a narrow pencil of the rays is passed through a plate of a cubical crystal, cut perpendicular to one of the chief axes of symmetry, it is found that the emergent rays do not consist of a single pencil in the direction of the incident rays, surrounded by a halo of scattered rays, but form a large number of sharply defined pencils which produce on a screen placed behind the crystal a complicated but regular pattern of small spots. Laue suggested that the crystal acted towards the X-rays as a three-dimensional diffraction grating, the lines and spaces being represented by the regularly arranged atoms in the crystal and the spaces between them. Light of wave-length λ falling on such a system of regularly arranged atoms and exciting in them secondary disturbances would emerge perpendicular to planes such that the normals to them through each atom were multiples of λ, so long as λ was not large compared with the distance between the atoms. Since the distance between the atoms in the case of a crystal is about 10^{-8} cm., the crystal would not act as a diffraction grating towards ordinary light, but X-rays with a wave-length of about 10^{-9} cm. might be diffracted by the crystal and resolved into a spectrum. Laue suggested that the pencils of X-rays emergent from the crystal represented the line spectrum of the X-rays falling on it[1].

But W. L. Bragg has pointed out that the position of the emergent pencils can be more readily explained in a manner somewhat different, but quite as interesting; it appears that the pencils can all be regarded as the result of the regular reflection of the incident pencil in various planes in the crystal; these planes are those which contain the greatest number of atoms in unit area. On this view the optical analogy to a crystal acting on X-rays is not a diffraction grating, but a pile of very thin parallel plates of glass; such a pile reflects at an angle θ light of which the wave-length is $d \cos \theta$, where d is the distance between the plates. Following up this suggestion he has obtained evidence of the regular reflection of X-rays at the surface of crystals of mica. This result raises the question which has not been discussed hitherto why, if the X-rays are like light, they are not always regularly reflected and refracted. The explanation often given is that these rays, consisting of mere pulses and not trains of vibrations, cannot

---

[1] Here and elsewhere mention has been made of the 'wave-length of the X-rays' although it is not believed that such rays are simple harmonic vibrations characterised by a single length. The thickness of the pulse is really meant, but the use of the term wave-length is justified because it appears that if a large random collection of pulses of the same thickness were analysed into a Fourier series in the manner sketched in § 84, then the amplitude of the term having a wave-length equal to that of the thickness of the pulse would be much greater than that of any other term.

set the atoms of the reflecting substance into vibrations and so call into play the mechanism on which reflection depends (see p. 198); but as we have every reason to believe that white light, which is regularly reflected, also consists of a succession of pulses, this explanation does not seem entirely adequate. A reference to § 24 will show that the theory of the refractive index which we have developed suggests that disturbances, even if they consist of regular vibrations, would not be refracted if their wave-length were as small as that which we associate with X-rays. If in (45) ν is greater than 3×10^{18}, corresponding to $\lambda = 10^{-8}$, $n^2 - 1$ will be very small, unless, either ν is very nearly equal to ν_0, or $N\epsilon^2/\pi m_0$ is of the order of 10^{36}. We shall see in § 124 that $N\epsilon$ is probably not greater than $10^{15} d$, where d is the density of the substance, so that $N\epsilon^2/\pi m_0$ will not be greater than 2×10^{33}; accordingly there should be no appreciable refraction of such rays, except when ν is nearly equal to ν_0 and in this case refraction would be masked by absorption. The reflecting power of a substance is determined by the value of $n^2 - 1$, so that absence of reflection would accompany absence of refraction. But if there are in a crystal certain planes containing an abnormally large number of atoms per unit area, the effective value of $N\epsilon^2/\pi m_0$ would be much larger for those planes and they would act as reflecting surfaces. Another reason for the absence of *regular* reflection and refraction might be found in the coarseness of the grain of all material media. A substance will not act as a smooth surface or as a clear medium unless the distance between its grains is small compared with the wave-length of the light; the distance between the grains (or atoms) in such media as we know is not less than 10^{-8} cm., which is small compared with the wave-length of any light, but not compared with that of X-rays.

Waves or corpuscles?

115. It appears then that all the evidence which is available tends to prove that X-rays are essentially similar to light, and that the apparently great difference in their properties may be attributed to a difference in the wave-length. Even the difference in the penetrating power of the rays for various media is readily explicable in this manner. For the absorption, like the refractivity, of a medium for radiation of wave-length λ is roughly proportional to the number of electrons in a cube of volume λ^3 which are resonant with that radiation; so that, even if the number of electrons in unit volume resonant with X-rays is greater than the number resonant with light (see § 124), the absorption which they cause may still be very much less. The question is no longer whether the X-rays are like light, but what theory of light we are to adopt in order to explain the properties of both light and X-rays. If we adopt the Maxwellian theory, we encounter the difficulty of explaining how a single electron acquires from the radiation the energy which experiment shows that it acquires; some hypothesis as to the forces acting on the electron must be devised, but it is difficult to see how such an hypothesis can solve the problems raised in the last chapter or in this; is it possible, consistently with the doctrine of energy, to explain how an electron can absorb an amount of energy which is very great compared with that contained in the volume occupied by a molecule? If we adopt any corpuscular theory we encounter all the difficulties which led to the rejection of the similar Newtonian theory; we have to explain how radiations of different quality are distinguished by different lengths or times (as shown by interference experiments) as well as by different energies (as shown by the photo-electric effect), how these radiations all travel *in vacuo* with the same velocity and how they can be polarised. Some attempt at such an explanation is offered by the Planck-Einstein theory which was discussed in the last chapter, and it must be noted that, though the evidence is less decisive, that theory appears as applicable to X-rays as to light. The figures given on p. 287 show that, if we take for the value of ν or λ that indicated by experiments on diffraction, the quantity $h\nu$, where h is the constant characteristic of Planck's theory, is approximately at least equal to the average energy radiated by a single β-ray in the form of X-rays or to that received by a β-ray from X-rays. Moreover the arguments in favour of Einstein's view that the quantity h is determined by the structure of the radiation rather than by that of the systems on which it acts are even more convincing in the case of X-rays than in the case of light. However the theory will not be generally considered complete until it explains the physical meaning of h and visualises the processes involved as satisfactorily as the alternative theories; Thomson's suggestion (p. 251) might explain why the energy appears to be divided into quanta of finite magnitude, but it does not explain why there is the relation which doubtless exists between the magnitude of a quantum and the frequency of the light[1]. Of the importance of the Planck-Einstein theory there cannot be the smallest doubt, but at present it is rather an illuminating statement of the experimental facts which have to be explained by any theory than an attempt at explaining them.

The greatest difficulty that has to be met by any theory which, on account of considerations of the transference of energy, would substitute for the Maxwellian theory of spherical waves any form of corpuscular theory is that connected with interference and diffraction. There are two forms of this difficulty. The first is that mentioned on p. 252; if the energy is concentrated in very small volumes, how does diffraction take place when these volumes are so sparsely distributed? The second is that, if the energy travelling in different beams is carried by different systems, it is hard to explain why these systems can act on one another and interfere if they come originally from the same source and not if they come from different sources. I do not think that any satisfactory solution of this difficulty has been offered so far, but interesting attempts have been made to decide between the wave and the corpuscular hypotheses by investigating in a new manner whether the energy carried from a given source in two different directions is or is not carried by different systems.

The idea underlying these attempts is due to Schweidler. Suppose that a source of rays is emitting on the average N rays per second, but that the moments of emission are not regularly spaced but distributed at random over the period considered. Then in any period t the number of rays emitted is not exactly Nt, but $Nt+q$ where q is a quantity which will be different for different periods t. If we perform the experiment many times,

---

[1] It must be remembered there is no evidence whatever for any other magnitude for the quanta than $h\nu$; if light is not done up in quanta of this magnitude then there is no evidence whatever that it is done up in quanta at all. This remark is necessary in view of some of the explanations of the difficulties of radiation which have been offered.

always observing over the same period t, we shall obtain each time a different value of q; of course the arithmetic mean of these values for a sufficiently large number of observations will be 0, for q is sometimes positive and sometimes negative, but the mean of the values of q^2 will not be 0; it is easy to show that, if the moments of emission are really distributed at random, $\overline{q^2}$, the mean value of the square of the deviation from the mean Nt, is Nt[1].

Now suppose that we observe the number of rays emitted by the same source in two different directions, and measure for each period t the difference between these numbers; this difference will be $q_1 - q_2$, where q_1 and q_2 are the deviations from the mean for the rays emitted in the two directions. If the rays emitted in the two directions are quite independent, so that it is possible for the source to emit rays in one direction without emitting them in the other, then q_1 and q_2 will be quite independent and it is easy to see that

$$\overline{(q_1 - q_2)^2} = 2\overline{q_1^2} = 2\overline{q_2^2}.$$

But if when the source emits a ray in one direction it also necessarily emits one in the other then, of course, q_1 will always be equal to q_2 and $\overline{(q_1 - q_2)^2}$ will be 0.

A radioactive body is a source of rays emitted at moments distributed at random. If the rays are α- or β-rays, those which are emitted in different directions are doubtless independent in the sense which has just been described; and experiments have been carried out on α-rays which show that the value of $\overline{(q_1 - q_2)^2}$ is, as it should be, $2Nt$. But what if the rays are γ-rays? If the spherical wave theory is true, then the rays are not independent;

---

[1] In the same way, suppose we toss a penny a very large number of times and divide this number into a large number of series each of which contains 200 tosses; then the mean value of the number of heads in one series is 100, but the actual value will be $100 + q$ where q is different for each series. The mean value of q^2 for all the series will be 100. It is to be observed that while the absolute value of $(\overline{q^2})^{\frac{1}{2}}$ increases with the number of tosses in a series, the ratio of this value to that number decreases with the number of tosses. If we toss a penny a million times, the probability that we shall get a deviation of 100 from the mean number of heads is much greater than if we toss it only 100 times, but the probability that we shall get a deviation which is $\frac{1}{100}$ of the whole number of tosses is much less.

if the source emits a γ-ray in one direction it also emits one in the other and $\overline{(q_1 - q_2)^2}$ should be 0. But if the corpuscular theory is true, the rays are independent, just as are the α-rays, and $(q_1 - q_2)^2$ should be equal to $2Nt$. The experiment has been tried; the actual measurements are very difficult and their interpretation is rendered somewhat uncertain by factors which have been left out of account in this simple description; one source of uncertainty is to be found in the fact that we cannot measure the number of γ-rays, but only the number of β-rays to which they give rise, and in any measuring instrument the number of β-rays liberated by a given number of γ-rays may show fluctuations about the mean. The results are not conclusive, but, so far as they go, they appear to favour the wave theory as against the corpuscular theory; the numbers of γ-rays emitted by the same source in two different directions do not appear to be perfectly independent. Attempts have also been made to apply the same ideas to the case of the emission of light and X-rays, but here the experimental difficulties are insuperable.

These results, if they are reliable, are extremely important, but they do not raise any new difficulty; they only show that certain ways of overcoming the old objections based on optical interference are impossible. The present position appears to be that the wave theory, as supported by Sommerfeld, and the corpuscular theory, supported by Einstein, are each capable of explaining a great many facts connected with radiation; but that, in general, the facts which are explicable by one theory are not explicable by the other. The corpuscular theory alone can explain how the energy of radiation is transferred from one place to another, while the wave theory alone can explain why the transference along one path is not independent of that along another. It almost seems that the energy itself is transferred by corpuscles, while the power of absorbing energy and making it perceptible to experiment is transferred by spherical waves, but it is impossible to suggest any theory capable of explaining this distinction. The problems of radiation are not yet solved.

References.

There is a good general account of the properties of X-rays and γ-rays in Pohl's *Die Physik der Röntgenstrahlen* (Vieweg, Brunswick: 1912); other summaries which have been published have been written deliberately to support one theory against another. The chief facts about the actions of the rays in their passage through material bodies are to be found in two summaries by Barkla in *Jahrb. d. Radioakt.* 5. 246, 1908, and *Phil. Mag.* 22. 396, 1911, and in Bragg's book *Studies in Radioactivity.* Marx' measurements of the velocity of the rays are described in *Ann. d. Phys.* 33. 1305, 1910, and criticised by Franck and Pohl in *Ann. d. Phys.* 34. 936, 1911. Sommerfeld discusses the diffraction of the rays, giving references to the experimental work, in *Ann. d. Phys.* 38. 473, 1912. The experiments on crystals mentioned on p. 297 are described by Laue, Friedrich and Knipping, *Sitz. Ber. Bayer. Akad.* July 1912 and W. L. Bragg, *Proc. Camb. Phil. Soc.* 17. 43, 1912. For the experiments mentioned in § 115 see v. Schweidler, *Phys. Zeit.* 11. 225, 1910; E. Meyer, *Ann. d. Phys.* 37. 700, 1912; Campbell, *Proc. Camb. Phil. Soc.* 15. 310 and 513, 1910.

The theory of the radiation from an accelerated electron based on ordinary electromagnetic theory is given in Abraham and Foppl's *Theorie der Elektrizität* and in Lorentz' *Theory of Electrons.* There is a most interesting summary of its application to X-rays by Sommerfeld in *Sitz. Ber. Bayer. Akad.* Jan. 7, 1911; this paper is perhaps the best statement of the case for this theory of the rays. Thomson's form of the theory is given in his *Conduction of Electricity through Gases*, but of course the account of the facts given there is now quite obsolete. Bragg's theory is given in *Jahrb. d. Radioakt.* 7. 348, 1910 and also in his book mentioned above. The application of the Planck-Einstein theory of radiation to X-rays is discussed in most of the papers and in the volume on *La Théorie du Rayonnement* mentioned at the end of Chapter X, and also in Stark's *Prinzipien der Atomdynamik.*

PART III

CHAPTER XII

THE PROPERTIES OF MATTER

116. IN the second Part of this volume, when we were developing the electronic theory of electrical and magnetic phenomena, we had occasion many times to make assumptions concerning the relation of the electrons to the substance in which they are contained. We had to suppose that in some bodies the electrons are bound within the atom by elastic forces, in others that they move freely throughout the substance or revolve in orbits of definite magnitude; also that, in certain circumstances, external agencies may detach an electron from an atom to which it is usually bound and cause it to become free. But so far no attempt has been made to correlate these various assumptions, to discover, for example, a connection between the properties of the bound and the free electrons of the same substance and to invent hypotheses to account for such connections. This problem will be considered in this Part; we shall try to explain the connection which experiment shows to exist between two or more different properties of the same substance.

The properties of matter.

In order to distinguish the more from the less speculative portions of the attempt, we shall in the present chapter take a general survey of the properties of matter, discover which of them experiment shows to be intimately connected and investigate whether we can explain the connection to any extent without making assumptions other than those which have been introduced in our theory of those properties. In the next chapter we shall consider the attempts that have been made to coordinate all these properties one with the other by means of a more detailed theory as to the structure of the atom.

The connections that exist between the various properties of the same substance may be brought to light by classifying those properties in two ways. First we may classify them according to the degree in which they depend upon the relations of the atoms in the substance and not only on the properties of the atoms themselves. We thus arrive at the following classification:—

I. Properties which depend neither on the nature of the atoms nor on the relations between them:—The gravitational constant, the constants which have been denoted hitherto by ϵ, m_0, h, k, c and, perhaps, the constant N_0 of formula (189). These are the universal constants of physics.

II. Properties which depend only on the number and nature of the atoms in the substance, but not on their state of chemical combination or on the physical state of the substance:—Mass, the radioactive constants (decay constant and those defining the number and speed of the rays emitted), the stopping power for α-rays and (probably) for β-rays, the constants defining the behaviour towards X- and γ-rays (transformation constants and v_s or λ_m, defining the quality of the characteristic radiation).

III. Properties which are not determined only by the number and nature of the atoms in the substance but depend also on their state of chemical combination and physical aggregation:—Chemical properties, mechanical properties (elastic constants, melting and boiling points, surface tension and so on), thermal conductivity, electrical conductivity, optical properties (refractive index, position of absorption bands, emission spectrum and so on), photoelectric properties (the constant w_0 of (190)), the power of being ionised, paramagnetic and diamagnetic susceptibility, specific heat.

It is not pretended, of course, that this list is exhaustive, but it includes all the properties which are important for our purpose. In class III, the properties are named roughly in the order of the degree in which they depend upon the relations of the atoms. Thus the elasticity or the electrical conductivity of a given chemical compound appears to be perfectly independent of the elasticity or conductivity of the elements of which it is composed, while its specific heat can be deduced with a high degree of approximation from that of its constituent elements.

The second classification is applicable only to the elements themselves and not to their compounds. We may classify the properties of a compound according as it is or is not a periodic function of the atomic weight. Thus, as is well known, the chemical properties of an element and its density are periodic functions of the atomic weight, a fact which is expressed in Mendeleef's table; a quantity defining one of these properties does not increase or decrease regularly with the atomic weight, but increases and decreases alternately in periods which are approximately the same for all these periodic properties. On the other hand, the quantity v_s defining the properties of the characteristic X-radiation, though a function of the atomic weight, is not a periodic function; v_s (for either the K or the L series) increases regularly with the atomic weight. Now the remarkable fact appears that the second classification leads to very nearly the same result as the first. Properties which are independent of the chemical combination of the atoms are functions of the atomic weight which are not periodic, while properties which depend on the chemical combination of the atoms and the physical state of the substance are also those which, when they are properties of the elements themselves, are periodic functions of the atomic weight[1]. These considerations undoubtedly suggest that when we make hypotheses as to the structure of the atom, we shall have to make that structure consist of two distinct parts, one which is changed by chemical combination and is a periodic function of the atomic weight, while the other is not changed by

---

[1] Perhaps the identification of the two classifications cannot be maintained in every detail. Thus λ_m for β-rays of given speed falls in class II, but appears to be a periodic function of the atomic weight; but it must be remembered that the absorption coefficient for any kind of rays is a very complex quantity, determined partly by true absorption and partly by scattering, which does not correspond to a simple physical property. Again all the properties in class III cannot at present be expressed as periodic functions of the atomic weight; in the case of diamagnetism, for example, it can only be said that on the whole elements of small atomic volume are paramagnetic, those of large atomic volume diamagnetic. But in every case (except specific heat, which is a doubtful member of class III) elements of the same chemical group resemble each other in respect of properties in class III more than they do elements in other groups.

chemical combination and is not a periodic function of the atomic weight.

The next part of our task is to discover whether, without making any definite assumptions as to the structure of the atoms, we can explain the connection which exists between properties in the same group. We have already made some advances in this direction in showing that there is a necessary connection between the thermal and electrical conductivities and between the refractive index and the dielectric constant. We will proceed first to consider the other properties of class III.

117. In § 28 we saw that the free vibrations of a substance which determine its absorption bands in the infra-red are those of the atoms executed under the forces by which they are held in their places. If we regard the molecules of a solid element as monatomic, the molecules are identical with the atoms and the forces which hold the atoms in their places are those which hold the molecules. Now the elastic forces which are called into play when the body is deformed by compression or shearing are doubtless also those which hold the molecules in their places, and the question arises whether the elastic forces resisting mechanical strain have anything to do with those which determine the optical vibrations of the solid element. An exceptionally bold speculation by Madelung and Sutherland suggests that the forces are exactly the same in the two cases and that from the mechanical properties of such a body we can calculate its infra-red frequencies.

Optical and mechanical properties.

A solid body can transmit mechanical vibrations of two kinds, compressional and transversal. The former travel through the body with a velocity a_1, the latter with a velocity a_2, where

$$a_1^2 = (\kappa + \tfrac{4}{3}n)/\rho \ldots\ldots(244), \qquad a_2^2 = n/\rho \ldots\ldots(245),$$

and κ is the modulus of bulk elasticity, n the rigidity, ρ the density. If the body has the form of a cube and it is maintained by external forces in a state of vibration, a condition will ultimately be reached in which it is traversed by steady standing waves of both types; the wave-length of those waves which lie along the edges of the cube will be an integral sub-multiple of

l, the length of an edge; there will be standing waves of which the wave-length is l, others of which the wave-length is $l/2$ and so on. Some of these vibrations will be perceptible to our senses as sound vibrations emitted by the body, some may have too low a frequency to be heard and some will certainly have a frequency too high to be heard. The frequency of a vibration of which the wave-length is λ will, of course, be a_1/λ or a_2/λ according as the wave belongs to one type or the other. Now it seems that there must be an upper limit to the frequency of these vibrations, conditioned by the fact that the body is not infinitely fine-grained; it seems reasonable to suppose that there can be no vibration of which the wave-length is less than twice the distance between two atoms, for it is difficult to see how a smaller wave-length could be propagated. Madelung's suggestion is that there is indeed this lower limit to the wave-length of the vibrations which can exist in the body, that the vibrations of the highest frequencies are perceptible to our senses not as sound vibrations but as light vibrations, and that the highest possible frequency is the highest frequency in the infra-red shown by the absorption spectrum of the body. The suggestion really only involves the assumption that the atoms are charged, so that their vibrations can give rise to electromagnetic disturbances, and that the forces which determine the mechanical properties of the body act only over ranges comparable with the distance between two atoms.

If this theory is true the highest infra-red frequency of a body (and we can only expect to be able to detect optically those of highest frequency) should be $a_1/2d$, where d is the distance between the atoms. Now a_1 can be calculated from elastic measurements by means of (244) and, if N is the number of atoms in unit volume, d must be equal to $N^{\frac{1}{3}}$ and can be determined from the density and the atomic weight. Accordingly the predicted and the measured frequency of the infra-red vibrations can be compared, but the comparison is difficult because so little is known directly of the infra-red vibrations of the solid elements[1]. However so far as a comparison can be made

---

[1] Very few elements at ordinary temperatures are solid and transparent. The theory cannot be applied immediately to compounds, for in such bodies the

it is not unsatisfactory; the theory indicates that the infra-red frequency should be of the order 3×10^{12}, i.e. in the region where the infra-red absorption bands of many substances are known to occur.

But the theory of specific heat given in § 100 affords another way of obtaining information concerning the infra-red vibrations of a substance and a better way of testing this view. The essence of that theory is that each vibrating system in the body which has a frequency ν contributes to the specific heat a quantity $\frac{d}{dT} L_\nu$, where L_ν is the function of ν and T given by (217). Hitherto we have regarded all the vibrating systems as parts of the molecules of the substance, but the considerations which have been pointed out suggest that we must take into consideration the vibrations of the substance as a whole; these vibrations will not have one definite frequency, but as many frequencies as there are sets of standing waves of different wave-lengths which can exist in it. Now the problem of determining how many of these sets of standing waves there are which have frequencies between ν and $\nu + d\nu$ is exactly the same as that which was considered on p. 231, when we were dealing with equilibrium radiation. To discover the number of waves of different frequencies travelling with velocity a in unit volume of the substance we have only to substitute a for the velocity of light[1]; if the waves are transversal each frequency must be counted twice, because vibrations polarised in different planes must be considered as different vibrations. Hence, if $Z_\nu d\nu$ is the number of vibrations of frequency ν in unit volume of the solid substance,

$$Z_\nu = 4\pi\nu^2 \left(\frac{1}{a_1^3} + \frac{2}{a_2^3}\right) \qquad (246).$$

---

infra-red vibrations may be due to atoms vibrating within the molecule and subject to forces very different from those which resist efforts to move the molecules from their places. However the theory seems true for compounds.

[1] The proof given here of (246) is not rigid. It is assumed that the transversal and compressional vibrations are independent, whereas one kind cannot exist without the other. No completely rigid proof has yet been given; the difficulty lies in the fact that it is not permissible to apply calculations which assume a continuous medium to one which is, *ex hypothesi*, discontinuous.

But now the upper limit to ν is not, as it was in the case of light, infinity, but some definite frequency ν_m, which is that of the waves which have a wave-length $2d$; a better view (which does not lead to a very different result) is that since there are only N atoms present, each of which has only three degrees of freedom, the total number of vibrations cannot be greater than $3N$. Therefore ν_m is determined by the equation

$$3N = \int_0^{\nu_m} Z_\nu d\nu \quad \text{.....................(247)},$$

whence

$$Z_\nu = 9N \frac{\nu^2}{{\nu_m}^3} \cancel{d\nu} \quad \text{.....................(248)}.$$

Hence the specific heat of unit volume of the substance is

$$C = \frac{d}{dT} \int_0^{\nu_m} Z_\nu L_\nu d\nu \quad \text{..............................(249)}$$

$$= 3NRT \left\{ \frac{12}{p^3} \int_0^p \frac{x^3 dx}{e^x - 1} - \frac{3p}{e^{p-1}} \right\} \quad \text{...............(250)},$$

where p is written as an abbreviation for $\frac{\lambda \nu_m}{kT}$.

(244)—(250) express the specific heat of the substance as a function of the quantities κ, n, ρ, N, determined by its properties, T, the absolute temperature, and the universal constants h, k. So far as the variations of C/C_0 with T are concerned, the formula agrees admirably with experiment; there is also a very remarkable agreement of the observed values of C/C_0 for different elements with the values predicted from their elastic constants; the outstanding discrepancies can certainly be attributed to uncertainties as to the value of these constants, which vary considerably with the nature of the treatment which the substance has undergone[1].

The importance of these results for our present purpose is that they show that there is a connection between the elastic constants of an element and its specific heat and explain this connection by relating both these properties to the possession of natural frequencies in the infra-red portion of the spectrum. The connection can be extended to include other of the mechanical

---

[1] Moreover it is assumed that the elastic constants are independent of the temperature, an assumption which is certainly not true.

properties of the element; for empirical relations have been found to hold between the elastic constants and such properties as the melting point, the latent heat of fusion and the atomic volume, and some attempt at explaining the existence of these relations can be made by regarding all these properties as determined ultimately by the forces which control the atoms in their vibrations. Thus the melting point may be regarded as the temperature at which the energy of the oscillations of the atom becomes so great that they exceed the limits of its sphere of influence, the latent heat of fusion as the difference between the energy of an atom when it is an oscillator of frequency ν and when it is a free atom at the same temperature. A great number of relations of this nature have been suggested and confirmed by experiment, but a detailed description of them would lead us too far. The conclusion which it is important to draw is that there is no need for separate hypotheses as to the structure of the atom in order to explain each of the various mechanical properties, the specific heat and those optical properties which depend upon frequencies in the infra-red; if an hypothesis could be devised which would predict for the forces acting between the atoms of a solid element values from which one of these properties might be deduced satisfactorily, then that hypothesis would explain equally all these properties. The problem of atomic structure is greatly simplified by these considerations[1].

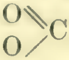
118. In Chapter II we saw that every body which shows normal dispersion in any part of the visible spectrum (that is, every transparent body and probably every body) must possess natural frequencies in the ultra-violet, and that these ultra-violet frequencies represent the vibrations of electrons within the atom or molecule. From a

Optical and chemical properties.

---

[1] It is perhaps remarkable that, if the specific heat of a substance is so closely related to its mechanical and optical properties, the former property does not behave more like the latter in respect of the variations on which the classifications of § 116 were founded. However it is probable that the specific heat of an element does not appear to be so distinctly a periodic function of the atomic weight as its

complete knowledge of the dispersion of the substance we could find out the frequency of these vibrations and also the number of electrons in each atom which execute them, but there are great practical difficulties to determining these quantities with accuracy, because we do not generally know how many different natural frequencies there are; our calculations have usually to be based on the assumption that there is only one ultra-violet frequency, and this assumption is almost certainly untrue. Nevertheless some useful information can be obtained.

Let us first consider the number of electrons in each molecule which have natural frequencies in the ultra-violet. If it is assumed that all these electrons have the same natural frequency, so that there is only one absorption band in the ultra-violet, this quantity can be calculated by the principles of § 28. The results obtained by measurements on substances which have no absorption band in the visible spectrum show that the quantity p is additive, that is to say, if we attribute to each atom in the molecule a certain value of p characteristic of that atom, then, in general, the value of p for the whole molecule will be the sum of the values of p for the atoms which it contains. Hence we may regard each atom as contributing a definite number of electrons which have natural frequencies in the ultra-violet. However the number contributed by any atom is not quite independent of the state of chemical combination of that atom; the number contributed by an oxygen atom is different according as it is combined in the form $C=O$ or in the form  $\begin{matrix} O \\ O \end{matrix}\!\!>\!C$, and the number contributed by a carbon atom is different according as it forms part of an 'open chain' or a benzene ring; in general it may be stated that if atoms are joined by a 'double bond' the number of electrons which they contribute is less than if they are

---

boiling point mainly because the variation of the value of C/C_0 for different elements at ordinary temperatures is so slight that the periodicity has not been detected. But it is certainly strange that the specific heat of compounds should not be as completely independent of those of the atoms in the compounds as is the boiling point. As has been said, the theory of this paragraph may not be strictly applicable to compounds, because such bodies may have natural vibrations within a molecule which are quite unconnected with those forces between the molecules which alone are called into play by mechanical forces.

joined by a single bond. The following table gives a few values of p for certain atoms and groups of atoms[1]:

H	1·07	O (in form $\mathrm{C}=\mathrm{O}$)	1·34	Br	3·95
C	1·63	O $\left(\text{in form } \mathrm{C}\genfrac{}{}{0pt}{}{/\!\!/\mathrm{O}}{\diagdown\mathrm{O}}\right)$	2·35	I	2·49
N	2·36	Cl	4·17	F	1·48

In the case of H, O, N, C the value deduced for p from a study of the compounds of those elements agrees well with that deduced from a study of the dispersion of the elements themselves. But in the case of the halogen elements, the dispersion of the elements themselves indicates that only one electron per atom is involved in the absorption band.

From the fact that the values of p calculated by this method for a given atom depend to some extent upon the state of chemical combination of that atom it must not be concluded at once that the number of electrons associated with it really depends on that state. For p is only the number of electrons of which the natural frequencies are so near the part of the spectrum which can be investigated that they affect the dispersion in that region, and we have seen that there are probably a large number of other electrons which have frequencies so great that they do not affect the dispersion. Perhaps the distinction between the two classes is not perfectly sharp and an increase in the frequency of an electron due to chemical combination might cause it to be transferred from the first to the second class, so that the value of p was decreased. Estimates of p only tell us the number of electrons having natural frequencies in an ill-defined region in the ultra-violet, but, as we shall see, valuable considerations may be based on them.

When an absorption band lies in the visible spectrum or in the part of the ultra-violet which can be investigated with ease, a study of it based on equations (78), (79) leads to a more direct knowledge of the value of p for the electrons involved in that

---

[1] p is calculated on the assumption that ϵ/m_0 has the value $4\cdot5\times10^{17}$ and not the usual value $5\cdot3\times10^{17}$ (see footnote, p. 46). The abnormal value of ϵ/m_0 is doubtless connected with the fact that the value of $p\,.\,\epsilon/m_0$ appears to decrease with increase of temperature.

band. It is found that absorption bands (of solids and liquids at least) can be divided into two classes; first, narrow absorption lines in which the absorption is very intense and accompanied by metallic reflection of the light 'absorbed'; second, much broader absorption bands in which the absorption is much less intense and unaccompanied by any special reflection. To the first class belong the lines which give to many artificial dyes their intense colour; to the second the bands to which the colour of such substances as copper sulphate is due. In the first class it is found that the value of p for the molecule is about 1; the absorption band represents the natural frequency of one electron in the molecule[1]. In the second class $p . e/m$ turns out to be of the order of 10^{15}, instead of 10^{18}, so that if we regard the vibrating system as an electron, there can only be one electron concerned in about 1000 molecules. On the other hand, if we regard the vibrating system as an atom, we should often have to suppose that the atom carried a charge of more than 100ϵ. Either of these hypotheses is very difficult; we have no reason in many of these cases to suppose that all the molecules are not in the same condition[2], nor is there any reason to suppose on any other ground that an atom can carry a large multiple of the electronic charge.

It should be noted that this proof that a single absorption band never represents the vibrations of more than a very small number of electrons in each molecule shows that the estimates of the number of electrons having natural frequencies in the ultra-violet based on the assumption that these frequencies are all the same are unreliable. But if the absorption frequencies are not very far apart the results of the calculation will not err in order of magnitude.

So much for the number of electrons giving rise to an absorption band; a few words must be said on the influence of the chemical constitution of the body on the position of those

---

[1] Sometimes the number is more nearly two, sometimes it is only a half. In the last case it must be supposed that polymerisation occurs in the solid state.

[2] There are cases in which chemists have attempted to explain the presence of absorption bands by supposing that molecules of two kinds are present, but it seems probable that the bands in these cases belong to the first class rather than the second.

bands. The technical importance of this relation for dyeing and other industries has led to the accumulation of an enormous number of facts concerning it, but unfortunately they are not of such a nature as to throw much light on the physical problem. To the dyer it is of immense importance whether or no an absorption band falls within the visible spectrum; to the physicist it is of no importance at all. Attention has been directed mainly to the influences which cause the bands to fall within the visible spectrum, and those which cause the band to move from one part to another of the ultra-violet have been neglected; various theories have been put forward to account for the presence of bands in the visible spectrum which take no account of the fact that there are always absorption bands in the ultra-violet. But it has been proved that there are definite chemical groups (called 'chromophores') which tend to move the absorption bands from the ultra-violet to the visible spectrum, that is to lower the frequency of those bands. These groups are in almost all cases such as contain 'double bonds,' and it may be said in general that they are such as to cause the molecule to be chemically unstable; for in general 'unsaturated compounds' are less stable chemically than 'saturated.' The general conclusion, if it can be substantiated, that the less stable chemical compound is characterised by lower natural frequencies of the electrons contained in it is of great importance. It must be remembered—for the point is usually ignored by chemists—that in all cases which have been investigated the intense absorption bands which cause the colour of dyes represent the vibrations of electrons and not of entire atoms or groups of atoms. On the other hand the frequency of the vibration of these electrons undoubtedly depends on the whole constitution of the molecule rather than on that of the individual atoms contained in it. It is found that in solid substances a rise of temperature tends to lower the frequency of the absorption bands.

The natural frequency of the electrons in a substance is indicated, of course, not only by its dispersion and absorption spectrum, but also by its emission spectrum. Such facts as are available concerning the emission spectrum of an element and its chemical properties have been briefly stated in § 87; it appears that there is a great similarity between the spectra of elements of the same chemical group, and the quantities which define such spectra have been shown to be functions of the atomic weight of the element and consequently also of such properties as its boiling point and atomic volume. Little is known of the precise relation between the spectra of elements of different groups beyond the fact that the constant N_0 is very nearly the same for all. Of the relation between the emission spectrum of a compound and its chemical constitution still less is known; our ignorance arises partly from the fact that the processes which excite emission spectra often cause a change in the chemical constitution. However a large number of facts are available in the case of organic compounds as to the relation between their composition and the spectra which they emit under the influence of kathode rays or when they are made to fluoresce. Here again undue attention has been paid to that part of the emission spectrum which is visible; it is possible that if the range of observation were extended the results of it would appear less chaotic. In the present state of knowledge it is impossible to draw from such evidence any valuable conclusions affecting the structure of the atom or the forces which bind the atoms into a molecule. The relations which can be traced between optical and chemical properties are certainly less definite than those described in the last paragraph between optical and mechanical properties, but in the next chapter we shall see that they suggest interesting conclusions.

A few remarks may be made about the constant N_0. If it is indeed a universal constant, it should be on our theory a function of the universal constants ϵ, m_0, c, h, k[1]. The only function of these constants which has the dimensions of a frequency is m_0c^2/h, and the value of this expression is $3 \cdot 8 \times 10^9 c$, while N_0 is about $10^5 c$. Since these numbers are so different it is rather improbable that N_0 is simply a function of the universal

---

[1] It is of course possible that the gravitational constant might also enter into the function; its use renders the problem indeterminate. The reader must be warned against the mistake of supposing that such a function as m_0c^3/ϵ^2 has the dimensions of a time. The value of this quantity is not determined only by the choice of the unit of time, but also by the choice of the unit of k in equation (1).

constants, for the argument from dimensions usually leads to a value differing from the true value only by a comparatively small factor[1]. The most recent work seems to show that N_0 is not quite the same for all elements; it may depend on some quantity, such as the radius of the atom, which is nearly but not quite the same for all elements.

Free electrons.

119. It appears that in certain circumstances the electrons contained in an atom can be freed from the forces which hold them and emerge as free electrons. In some cases the emergence appears to be due to the vanishing of the restraining forces and to require no external work (as in the case of the free electrons in metals), in others it is due to the communication of energy to the electron by external forces (the photo-electric effect, secondary β rays, and the ionisation of gases). Can we connect the power of a substance to give off electrons with any of its other properties?

In the case of the free electrons of conductors the connection is obvious. The best conductors[2] are all metals in the chemical sense; they are all highly electro-positive elements which are capable of appearing in suitable solutions as positively charged ions. A positive electrolytic ion is doubtless an atom which has lost electrons, and the conclusion is clear that the properties of an atom which make it give off free electrons in the solid state are also those which make it lose electrons in aqueous solutions; since these last properties determine the chemical behaviour of the element, there is a direct connection between the conductivity and the chemical behaviour. The connection would be still clearer if it appeared that the number of free electrons per atom in the solid metal were equal to the number which the atom loses when it becomes an electrolytic ion, that is, equal to its valency; determinations of the number of free electrons in a metal are very difficult because of the uncertainties attaching to the theory

---

[1] There are, of course, quite obvious reasons why the factor always turns out to be small; it is always very difficult to measure accurately a quantity which is a very large multiple of the standard which is employed to measure it.

[2] The conductivity of a conductor is not, of course, simply proportional to the number of free electrons which it contains, but the good conductors as a class do differ from the bad by containing many more free electrons.

of Chapter III, and it cannot be said that there is any evidence for such a relation; indeed, in spite of the results given in Table IV (p. 64), it is not certain that there is ever more than one free electron per atom.

The study of the secondary β rays (of which we must regard the photo-electrically liberated electrons as a special case) does not yield much information as to the capacity of an atom to give off electrons; for, first, the number of electrons given off from any piece of matter is always a very small proportion of the number of atoms in it and is wholly independent of the number of electrons which each atom could give off, and, second, the energy of the emergent electrons, so far as it depends on the material from which they come, is probably determined by the forces at the surface of the material rather than by those which restrain the electron within the atom. We can only note that w_0 is very much greater for compounds than for metals because in the former class, but not in the latter, work has to be done to detach an electron from the atom as well as to drag it from the surface of the material, and that, since in compounds the value of w_0 seems to depend in a definite way on the chemical character, the forces restraining the electron in the atom are probably those giving rise to the chemical combination.

120. **Ionisation of gases.** From the study of ionisation we might hope to obtain more information as to the processes which determine the loss of electrons by an atom. There are two quantities which might be measured and possibly connected with the other properties of the atom, namely the number of electrons which an atom can lose and the work required to detach those electrons.

When a gas is ionised by α or β rays of considerable speed (of course we regard the ionisation due to X and γ rays as a special case of ionisation by β rays) there is no clear evidence that more than one electron can be detached from an atom; it has been maintained that in some cases doubly-charged positive ions appear, from which two electrons have been detached, but the existence of these ions is uncertain and at the most they are present in very small proportion. On the other hand in the conditions which give rise to the positive rays (§ 54), when the ionisation is effected by β rays or by positive particles of comparatively low speed, there is no doubt that multiply-charged atoms, which have lost several electrons, are produced. The facts which result from the researches of J. J. Thomson are complicated and difficult to arrange in any definite scheme, but the following conclusions can be stated. All atoms and many molecules can appear as singly-charged positive ions which have lost one electron; in most cases a complex molecule is, in part at least, broken up into its constituent atoms or into simpler molecules, but there is no evidence to determine whether this splitting up is a direct or indirect consequence of ionisation. There appears to be some connection between the valency of an atom (i.e. the number of electrons which it could lose if it formed an electrolytic ion) and the number of electrons which it actually loses when it becomes a positive ray. Thus the hydrogen atom usually loses one electron, while the carbon atom usually loses four. But the relation is not simple; the oxygen and nitrogen atoms usually lose two electrons, while the inert gases (argon and helium) often lose three; the value of e/m for some particles is most easily explained by supposing that they consist of aggregations of atoms the existence of which there is no reason to suspect on other grounds, such as a doubly-charged molecule H_3. The greatest number of electrons which is known to be lost by one atom is eight, in the case of mercury. Thomson believes that there are two kinds of ionisation, one resulting in the production of singly-charged particles and probably due to the action of β rays, the other resulting in the loss of the maximum number of electrons possible and probably due to the action of the positive particles. But this conclusion is somewhat speculative.

The formation in the same conditions of negatively charged particles, which doubtless represent atoms or molecules which have gained an electron, is certainly connected with the chemical properties of the atom. The electro-negative atoms which gain electrons in electrolytic solutions are those which appear most readily negatively charged in the 'positive' rays; the chief exception to this rule is hydrogen, which seems to form negative particles as easily as positive.

If we could measure the energy required to ionise an atom, that is to detach an electron from it, we should doubtless obtain very valuable information as to the constitution of the atom. But the problem is very difficult and, in order that the nature of the difficulties may be realised, it will be well to consider a rough theory of how an atom might be ionised by a moving charged particle.

When the atom is not ionised the electrons within it are doubtless restrained in their positions or orbits (for they may be in motion) chiefly by the attraction of the positively charged portion of the atom. The passage of a rapidly moving charged particle through such a system will be similar to the passage of a comet through the solar system; the comet is somewhat deflected from its course and, if its mass is not too small, produces perturbations in the orbits of the planets to which it communicates some part of its kinetic energy. The general problem of determining the path of the comet and the changes of energy in any circumstances cannot be solved completely, but the case when the comet affects and is affected by only one planet may be treated simply. We may suppose that the planet is not acted on by any other forces than those due to the comet, the effect of the forces restraining it in its orbit being practically equivalent to an increase in its mass. Let e_1, m_1, e_2, m_2 be the charges and masses of the comet and planet, v their relative velocity when the comet is still far away, b the distance of the planet from the line representing this relative velocity or the distance to which the comet and planet would approach if the former were not deflected, 2θ the angle through which the comet is deflected, $T = \frac{1}{2} m_1 v^2$ the original kinetic energy of the comet, and δT the loss of energy which it suffers. Then, if the force between the comet and planet is due to their charges and equal to $e_1 e_2 / r^2$, we find

$$\sin^2\theta = \frac{1}{1 + \frac{b^2 v^4}{e_1^2 e_2^2}\left(\frac{m_1 m_2}{m_1 + m_2}\right)^2} \quad \text{............(251)};$$

$$\delta T / T = \frac{4 m_1 m_2}{(m_1 + m_2)^2} \sin^2\theta \quad \text{...............(252)}.$$

If θ is small and the comet little deflected, and if we put $m_1 = m_2$, $e_1 = e_2$, we obtain

$$\delta T/T = \sin^2\theta \qquad \text{.....................(253)},$$

and

$$\delta T = \frac{e^4}{b^2} \cdot \frac{1}{T} \qquad \text{.......................(254)}.$$

These conditions probably represent approximately the interaction between a swift β-ray and the electrons in the atom through which it passes. Since such a ray can form many thousand ions, the amount of its energy which it gives up at one encounter with an electron in an atom must be small; from (253) we see accordingly that θ must be small and we may conclude, in accordance with the experimental facts, that the ionisation caused by such a rapid ray need not be accompanied by a marked deflection of it. Further we see from (254) that the loss of energy by the ray at a single collision is proportional to $1/T$; since the number of collisions and the quantity b are probably independent of v, we arrive at the experimental law (181) that the rate at which the ray spends energy along its path is inversely proportional to its energy. Numerical calculation will show that a β-ray having a velocity of 10^{10} cm./sec. (i.e. the slowest ray emitted by a radioactive substance) will not be deflected through a right angle unless $b < 10^{-12}$ cm. and will not give up even a thousandth part of its energy (which is probably about the amount required to ionise) unless $b < 10^{-11}$ cm. The corresponding values of b will, of course, be still less if the velocity of the ray is greater.

When the 'comet' is an α particle similar calculations are impossible because the particle is so much larger than the electrons in the atom that it doubtless does not react with them as a whole; the forces exerted by such a particle in an electron are probably determined by the forces inside the particle rather than by those outside it, and of these forces we have no knowledge whatever.

On this theory we should expect that the average amount of energy which a ray would expend in producing a certain number of ions would be considerably greater than the amount required to detach the electrons; for the ray may waste energy either by giving to an electron a velocity too small to carry it outside the atom, so that it does not appear as an ion, or by giving it a greater velocity so that the electron leaves the atom with a finite speed, though only one ion is formed. In the last case, part of this 'waste' of energy may only be temporary, for if the ejected electron possess more than a certain amount of energy it would be able to ionise a fresh atom; but energy would be wasted by giving to electrons a velocity greater than that needed to carry them clear of the atom and yet insufficient to enable them to ionise another atom.

Some information as to this waste of energy might be obtained by investigating the velocity with which the electrons leave the ionised atoms. It is scarcely possible to conduct such investigations when the substance ionised is a gas, but it has been found that when ionising rays fall on solid substances they cause the emission from them of electrons, possessing velocities too small to ionise gases. When this emission was first discovered the electrons thus liberated by α-rays were termed 'δ-rays' and those liberated by β-rays 'secondary kathode or β-rays.' They may be confidently identified with the electrons liberated when a gas containing the same atoms is ionised; but the first name may still conveniently be applied to them. It appears that the velocity with which these δ-rays leave the solid substance is quite independent of the nature and velocity of the ionising rays, so long as this velocity is considerably greater than that required for ionisation, and that there is no evidence whatever that it varies with the nature of the substance ionised. The velocity is not the same for all the rays; about half of them have a velocity less than 2 volts, about 0·9 a velocity less than 10, and all of them a velocity less than 30 volts. We cannot immediately identify these velocities with those with which the electrons leave the atoms, for some of the electrons probably come from layers below the surface and may have collided with other atoms on their way. But the layer from which they come is very thin, certainly thinner than the thinnest films of solid material which can be prepared, and we should expect a considerable proportion of them to possess the velocity with which they were emitted. It is very curious, therefore, to find that, even when the ionising rays are electrons having a velocity of several thousand volts, none of the

δ-rays seem to possess a velocity between 30 volts and a value near that of the ionising rays, this latter value being that of the 'scattered' rays; on our theory of the action of a β-ray on the electrons in an atom we should expect to find all velocities between that of the incident rays and zero.

On the other hand the considerations which have been advanced might explain the fact recorded on p. 168 that, while the energy spent by a ray in passing through a given layer of matter depends only on the number and nature of the atoms in that layer, the amount of ionisation produced depends on their chemical combination. For the energy spent would probably depend almost entirely on the number of electrons in the layer, while the fraction of that energy actually used in liberating electrons would depend on the forces restraining those electrons in the molecule. But if it is true that the velocity with which the electrons are ejected is independent of the nature of the material this explanation is not so plausible. It seems preferable to suppose that ionisation is not a direct effect of the action of the charged particles in the rays on the electrons in the atom ionised and that the energy lost by the ray is spent in penetrating to the interior of the atom, the ejection of the electron from it being a secondary result of the presence of the charged particle within the atom. An α particle doubtless requires more energy than a β particle in order to penetrate within the atom, and so an explanation might be found of the fact that at very low speeds an α particle is much less efficient in producing ionisation than a β particle possessing less energy. If we could suppose that the chance of ionisation being produced by the presence of a charged particle within the atom depends simply on the time during which that particle is within, we might explain the remarkable fact (p. 177) that the ratio of the ionising powers of α- and β-rays of equal velocity is constant for all values of that velocity above a certain limit; for rays of equal velocity would be those which remain equal times within the atom.

It appears then that measurements of the ratio of the amount of energy spent by the rays in a gas to the number of ions produced do not necessarily give us much information as to the work necessary to detach an electron from an atom; they can at most give us a maximum estimate of that work[1]. Hence the numbers given on p. 169 for the number of ions produced in air by an α-ray only show that the work required to detach an electron cannot be greater than 37·5 volts. Another, and probably closer, maximum estimate is given by the least energy of a ray which will produce any ionisation at all; the experiments on ionisation by collision give this energy as between 21 volts (nitrogen) and 11 volts (helium); more direct experiments by Lenard give the value 11 volts for many substances. But we are not so much interested in the absolute value of this quantity as in its variation for different substances; the uncertainties as to the exact nature of ionisation make it impossible to draw from such measurements any conclusions as to this variation. The least energy required to produce ionisation may vary, not because the work required to detach the electron varies, but because the proportion of the whole energy lost which is spent in performing that work varies. It is impossible at present to deduce from such observations any valuable conclusions as to the structure of the atom.

It may be mentioned, however, that the estimates just given agree in order of magnitude with those which we should expect on general grounds. The energy required to remove a charge $-\epsilon$ to an infinite distance from a sphere of radius a bearing a charge $+\epsilon$ is $\epsilon . \epsilon/a$. Outside a sphere of radius about 2×10^{-8} cm. an atom behaves as a point charge at the centre of the sphere; using this value for a we find that ϵ/a is about 7 volts.

Magnetic properties.

121. Of the properties of matter placed in class III there remain for consideration only the paramagnetic and diamagnetic susceptibilities. And of these little can be said. The fact that the paramagnetism of iron (for example) depends so greatly on its state of chemical combination indicates once more that the electrons concerned in it are those concerned in chemical action, but no correlation appears to have been traced

---

[1] Assuming, that is, that the work required to detach the electron is derived from the ionising rays. An alternative theory, similar to that mentioned on p. 290, according to which this energy is derived not from the rays but from the atom, has been proposed to solve this and every other difficulty connected with ionisation. The objection to this theory is that there is no evidence that the energy of ionisation ever greater than that lost by the rays.

between the paramagnetic susceptibility and other properties. It is uncertain whether the diamagnetic susceptibility really belongs to class II or to class III. But a reference should be made to the difficulty raised on p. 150, that, if the magnetic properties are due to the presence of revolving electrons, these electrons should be sending out radiation and losing energy, whereas no kind of radiation peculiar to magnetic bodies has been discovered, and there is no evidence that a supply of energy is required to maintain their magnetism. We have seen in Chapter XI that the whole theory of the emission of radiation by accelerated charged systems is uncertain, so that it is useless to enter on detailed speculations for the solution of the difficulty; but it may be noted that according to (229) it is not necessary that the radiation from a large number of electrons should be the sum of the radiations which each would emit if it were alone. If there are two electrons very close together characterised by equal and opposite values of f, the value of D and H at distant points due to the pair will be zero and there will be no radiation; an illustration of this principle is to be found in the absence of radiation from a circuit round which a steady current is flowing, when the tendency of one electron to radiate is counteracted by that of the others; we may regard one electron as absorbing the radiation emitted by another, so that no energy on the whole is lost. Accordingly if a magnetic substance contains a large enough number of electrons, accelerated at any instant in different directions, the radiations due to the whole system would be zero. Again, according to the Planck-Einstein theory we might suppose that the electrons never have the opportunity to acquire the energy $h\nu$, which they must possess in order to radiate, because the frequency of the light which they would emit if they radiated at all is so great.

Radioactivity.

122. We must now turn our attention to the properties of class II and endeavour again to discover relations between them which will throw light on the structure of the system to which they are due.

Since mass is a property belonging to class II (that is since the mass is 'conserved'), we must regard the mass of any constituents of which we regard the atom as composed as independent of their association. It is perhaps just worth while making this remark in view of the anomalies concerning the mass of electrons which were noted on p. 173, but it is not of any importance for our discussions. Nor is it possible to deduce any information of importance from the second property of this class, radioactivity. Beyond the fact that, with the possible exception of potassium and rubidium, the radioactive elements are those which have atomic weights greater than 208, not even an empirical relation is known between the radioactive properties of an element and its other properties. Nor have we at present even a vague outline of a theory of those properties. When the property was first discovered and it was imagined that the probability of an atom disintegrating depended on its age, it was suggested that the instability which finally led to disintegration was due to the slow radiation of electrons revolving in the atom (such as those to which we have attributed the property of diamagnetism), involving a gradual loss in their energy and, finally, a change in their orbits. But since it appears certain that radioactive atoms 'die,' not of old age, but by accident, this theory is obviously unacceptable.

Scattering of α- and β-rays.

123. Much more information can be obtained from a study of the action of material bodies on α- and β-rays which pass through them. In all cases there is a scattering of the rays, some of the particles being deflected from their original direction. The view is generally accepted that this scattering arises from the forces exerted on those particles by other particles contained in the atoms through which they pass, and that the analogy of the passage of a comet through the solar system, which was introduced in connection with the ionisation of gases, is applicable to this action also. If it is assumed that the force between the particle in the rays (carrying a charge e_1) and that in the atom (carrying a charge e_2) is e_1e_2/r^2, where r is the distance between the particles, (251) will give the value of 2θ, the deflection of the ray by a single encounter. By considering the probability that a ray will pass within a distance b from the deflecting particle in the atom, the probability that it will suffer a given deflection at a single encounter is easily calculated. If the particle makes more than one encounter, the problem of determining the probable deflection from assumptions concerning the charges and masses of the particles becomes more complicated, but so long as the total deflection is small the calculation can be carried out; it is found, of course, that the ratio of the probability that a particle will suffer a given deflection θ_1 to the probability that it will suffer a deflection θ_2 is different according as these deflections are produced at one encounter or at several; consequently the comparison of the number of particles deflected through different angles can be used to determine whether the deflections are produced at one or at several encounters.

In the case of α-rays passing through layers of material which represent only a small portion of their range there is no doubt that the deflections produced are due to single encounters. Moreover, even when the angle of deflection is more than a right angle, the α particle loses only a very small proportion of its energy at a single encounter; therefore, from (252), the mass of the deflecting particle must be very great compared to that of the deflected particle and in (251) we may write $m_1/m_2 = 0$[1]. The probability of a given deflection now becomes a function of e/m for the deflected particle and e_2, the charge on the deflecting particle. The former quantity is known and hence the latter may be calculated from the observed value of the probability. In this way it is found that e_2 for the deflecting particles in gold is about 100ϵ, and that generally the value of e_2 for any substance is $N\epsilon$, where N is a number intermediate between the atomic weight and half that number. The remarkable conclusion appears that the deflecting particles are certainly not electrons, but such as carry very much greater charges, the magnitude of this charge being determined by the atomic weight of the substance. No evidence is obtained by such experiments as to the sign of these charges.

In the case of the β-rays it is less certain whether the scattering actually observed is simple (i.e. the result of single encounters) or compound (i.e. the result of many encounters).

---

[1] m_2 is the 'effective mass' of the deflecting particle and will be great if that particle is restrained from moving under the reaction of the deflection by powerful forces, even if the actual mass is not great.

Probably both types of scattering are important in producing the actual deflections. Hence it is not possible to deduce from the observations certain evidence as to the nature of the particles which are producing the scattering, but it appears quite consistent with the experimental data to suppose that in this case also the scattering is largely, if not entirely, due to encounters with the particles carrying large multiples of ϵ, the existence of which is proved by the observations on α-rays. It appears certain that all the scattering cannot be due to encounters with electrons only, but if there is any considerable amount of compound scattering involved in the deflections which are observed, this scattering may well be due to encounters with electrons.

The results which have been obtained by experiments on the stopping power of various materials for α- and β-rays are less easy to interpret. The loss of energy suffered by such rays in passing through matter may be due to the same process as causes the deflection; if the ratio m_2/m_1, where m_2 is the effective mass of the deflecting particle, is not infinite, there will be a loss of energy given by (252). But the most recent calculations show that, while the deflection of the α-rays is due to interaction with particles of large mass, their absorption is probably due to interaction with electrons. On this hypothesis many of the quantitative relations of § 73 can be explained and values obtained for the total number of electrons in an atom which agree well with those indicated by other arguments.

When the loss of energy of β-rays in passing through matter is measured and relation (178) obtained, the electrons concerned have probably lost their energy by encounters with other electrons; for the mass of the multiply-charged centres is doubtless so great compared with that of an electron that an electron would not lose any appreciable portion of its energy by an encounter with them, unless it were deflected through a large angle and so disappeared from the emergent rays. On the assumption that the energy is lost in this case by collision with the electrons contained in the matter, and that the effective mass of these electrons is not very different from m_0, a minimum estimate of the number of electrons in the matter can be obtained by (254). For if there are N electrons per unit volume of the matter the number in a layer x

for which the value of b will lie between b and $b+db$ is $2\pi b\,db\,.\,Nx$; consequently the loss of energy of a ray in passing through this layer is given by

$$T\,.\,\delta T = Nx\int_{b_1}^{b_2}\frac{2\pi\epsilon^4}{b}\,db = 2\pi N\epsilon^4\log\frac{b_2}{b_1}\,.\,x \;\ldots\ldots(255).$$

The widest possible limits for the integral are set by the fact that the rays are not deflected through more than a right angle and that b can never be greater than the distance between two atoms; hence we may take $b_1 = 10^{-12}$ cm., $b_2 = 10^{-8}$ cm., and obtain $T\,.\,\delta T = 2{\cdot}8\times 10^{-36}Nx$. Now from (180)

$$T\,.\,\delta T = \frac{m_0{}^2 a}{8}\,x \ldots\ldots\ldots\ldots\ldots\ldots\ldots(256),$$

so that

$$N = 7\times 10^{-20}a \ldots\ldots\ldots\ldots\ldots\ldots\ldots(257).$$

For aluminium and gold a is $7{\cdot}32\times 10^{42}$ and $2{\cdot}54\times 10^{43}$ respectively. Inserting these values in (257) we find that aluminium cannot contain less than 9 electrons for each atom, nor gold less than 30. These estimates are very rough and little reliance can be placed on them, but they are consistent with those obtained from other sources of evidence and seem to show that the whole number of electrons in the atom is considerably greater than the number which was estimated from a consideration of any of the properties of matter in class III. It should be remarked that if it is assumed that all the scattering of β-rays is due to encounters with electrons, the estimated number of electrons present for each atom of the material substance appears to be not very different from the atomic weight. This conclusion is, of course, in accordance with that which has just been stated.

Action on X-rays.

124. All considerations as to the structure of the atoms based on experiments on X-rays are doubtful while a satisfactory theory of those rays is still lacking; but if the theory of the scattering of the rays mentioned on p. 282 can be accepted, measurements of that scattering lead to a very direct estimate of the number of electrons concerned in it. It is easy to see from (238) that E_1, the total energy of the radiation scattered by a layer dx of matter containing N electrons per unit volume is $\frac{8\pi\epsilon^4}{3m_0^2c^4} . N . E$, where E is the energy of the primary radiation crossing that layer. It should be observed that the ratio E_1/E should be independent of the quality of the rays and of everything except the properties of the scattering particles[1]. Experiment shows that this condition is certainly not fulfilled universally, E_1/E being less for hard γ-rays than for soft X-rays, but it is apparently fulfilled over a smaller range of quality. If the value for E_1/E is measured when the rays lie within this range of quality and the scattering material is air, the experiments indicate that the number of electrons present per atom is approximately equal to the atomic weight; the scattering of elements differing largely from those of air in atomic weight has not been investigated sufficiently for any conclusion to be drawn as to the general validity of this relation. If it were generally valid the scattering of a piece of matter should depend solely upon its mass and not on the nature of the atoms contained in it; it is possible that this may be true in the case of all the lighter elements, but it probably does not hold for the heavier; more evidence on this point is required.

The remaining features in the behaviour of substances towards X-rays could probably be expressed in terms of the transformation coefficients of X-rays into β-rays and of β-rays into X-rays. The value of these constants is not known as yet, but a rough measure of the former quantity is probably given by the absorption coefficient λ_m. A study of the value of λ_m for different substances and rays indicates that the ratio of λ_m for a substance A to λ_m for a substance B is independent of the hardness of the rays to which λ_m applies, so long as the rays are not hard enough to excite the characteristic radiation in either A or B. Thus, the absorption of aluminium for any kind of rays is always about nine times as great as that of carbon for the same rays, even though the absolute value of the absorption of aluminium or of carbon

---

[1] It is unlikely that any particles other than electrons are concerned in the scattering, for it will be observed that the ratio decreases with the value of e/m for the particles. Even the multiply-charged particles concerned in the scattering of α-rays cannot have values for e/m nearly as large as that for electrons.

may vary in the ratio of 8 to 1 for different kinds of rays. The simplest way to explain this relation would be to imagine that the systems which are concerned in absorbing the rays (when the rays are not hard enough to excite the characteristic radiation) are the same in all atoms, but that some atoms contain more of these systems than others; aluminium, for example, would have to be supposed to contain nine times as many of them as carbon. But in the present state of our knowledge this speculation cannot be developed further; the systems might be electrons controlled by some special forces, the same in all atoms, but such suggestions are too vague to be useful.

On the other hand the excitation of characteristic radiations of different qualities in different substances undoubtedly indicates the action of different systems, and the analogy between the X-rays and light further suggests that these systems are 'resonators' of some kind, characterised by natural periods or at least by some quantity of the dimensions of a time. But here again speculation is useless until it can be based on a more complete and satisfactory theory of the nature of the rays; before we can make any suggestion as to the constitution of such resonators we must have some physical theory as to how the energy of radiation appears to be always transformed in quanta which depend on the time characteristic of the resonator.

Gravitation.

125. The properties of class I, the universal constants, are scarcely to be regarded as explicable by any theory of matter; they represent rather the conceptions in terms of which any theory of matter must explain the remaining properties. But consideration of dimensions will show that it might be possible to regard one of those constants as a function of others and so to explain it in terms of those others. Attempts have been made in this way to explain the universal property of gravitation and to regard it as a special case of the electrical actions the theory of which underlies all our considerations; these attempts require a brief notice.

The attempts to explain gravitation as a special case of electrical action are all similar to that due originally to Mosotti. It is supposed that gravitational attraction represents a residual electrostatic attraction; it has been supposed, for example, that the attraction of two equal and opposite charges is not equal to, but slightly greater than, the repulsion of two equal and similar charges, so that two neutral bodies, each containing equal and opposite charges, show on the whole a residual attraction. Any such theory explains why the gravitational attraction varies inversely as the square of the distance and if, as the calculations just given would appear to show, the positive and negative charges in an atom are proportional to its mass, the gravitational attraction would also be proportional to the masses of the bodies. In order to explain the fact, amply proved by astronomical observations, that the attraction between two moving bodies is determined simply by their instantaneous positions and not by their positions at some previous time, so that the attraction appears to be propagated with an infinite velocity, some assumption must be made concerning the way in which the residual electrostatic attraction varies with the relative velocity of the bodies. Lorentz has shown that suitable assumptions can be made which are generally consistent with the principles of Chap. I. One of the main difficulties which has always beset electric theories of gravitation is thus removed, but it cannot be said that there is any evidence for such theories; there is simply no evidence against them. It is quite impossible to test any theory of gravitation while no variation of the gravitational attraction between two bodies with the circumstances in which they are placed is found. Experiments have been described lately in which a difference in the weight of a body according as it is charged or uncharged has been noted and it does not appear that any flaw in the experiments has been detected; but the observations are necessarily extremely delicate and some further confirmation by other workers must be awaited before the results can be accepted with confidence. If those results are confirmed, they would, of course, give great support to some electric theory of gravitation.

We have now completed our general survey of the properties of matter and will consider briefly the speculations as to the structure of the atom which have been made to explain some of them.

REFERENCES.

The following papers give important additional information about some of the points to which only a brief reference is made in the text.

§ 117. Lindemann, *Phys. Zeit.* 11. 699, 1910; Debye, *Ann. d. Phys.* 39. 789, 1912; Nernst, *Zeit. f. Elektrochemie*, 17. 817, 1911.

§ 118. The papers by Erfle and by Koenisberger and Kilchling mentioned in the references at the end of Chap. II. A book by Ley on Colour and Constitution in organic compounds (Leipzig, 1911) gives an enormous number of facts concerning this question.

§§ 120, 123. J. J. Thomson, *Phil. Mag.* 23. 449, 1912; Ramsauer, *Jahrb. d. Elektronik*, 9. 515, 1912; Campbell, *ibid.* 419; Bohr, *Phil. Mag.* 25. 10, 1913; Rutherford's *Radioactive Substances* (1913).

§ 125. Zenneck, *Encyklo. d. Math. Wiss.* Vol. 5.

CHAPTER XIII

THE STRUCTURE OF THE ATOM

The internal structure.

126. FROM the considerations of the last chapter some information has been obtained as to the electrons contained in atoms. It appears probable that the whole number of electrons in any atom is not very different from that expressing its atomic weight, when that of hydrogen is taken as unity, and that the number increases regularly with the atomic weight and is not a periodic function of it. But in every atom there are a few electrons, which determine the properties of class III, differing greatly from the rest in the manner in which they are attached to the atom; the natural frequencies of these electrons are much less than those of the remainder and they are more easily detached from the atom. It is these electrons which determine the dispersion of the substance in the ultra-violet or visible spectrum and it is these which leave an electropositive atom when it becomes a positive electrolytic ion. For reasons which will appear more definitely later, these electrons will be termed the 'valency electrons' of the atom, the remainder being termed 'fixed electrons.'

If, as we have assumed throughout[1], the electrons behave as point charges, an atom cannot consist wholly of electrons. It

---

[1] It must be remembered that there is not, and there cannot be, any direct evidence of the nature of the forces between two electrons. It may be found in the future that a more satisfactory theory of the properties of matter can be attained by attributing to those forces a form different from that which we have assumed. And in this manner it might be possible to do without the conception of positive electricity, as a structure different from electrons. It might be possible to suppose that a complicated collection of electrons would behave towards another electron like a positively charged body. This possibility must be kept in mind when hypotheses concerning the nature of the atom are framed, but it cannot be taken into consideration here after the whole electronic theory has been based on the assumption that an electron behaves like a point charge of ordinary electrostatic theory.

must contain also a positive charge equal and opposite to the sum of the negative charges on all the electrons, which neutralises the effect of those charges at distant points and binds the electrons together. Two chief hypotheses have been proposed for the distribution of this positive charge. The first suggests that it is distributed uniformly throughout a sphere the radius of which represents the 'radius of the atom'; the second suggests that it is concentrated on a single particle round which the electrons are arranged; on this theory the radius of the atom is a vague quantity representing roughly the greatest distance of any electron from the positive particle. These two hypotheses are not, of course, the only two which have any intrinsic probability, but they are the only two which are sufficiently simple to enable any calculations whatever to be made as to the properties of the atom. At points outside the radius of the atom the forces due to the positive charge will be the same whichever hypothesis is adopted, and the forces due to the electrons will differ only so far as the two hypotheses lead to different distributions of the electrons within the atom. But even with such simple hypotheses as to the distribution of the positive electricity, a deduction of the positions which the electrons will take up under the attraction of the positive charge and their mutual repulsions is quite impossible; hence no consideration of the forces exerted by an atom outside its radius can lead to a distinction between the two hypotheses.

But when α- or β-rays penetrate solid bodies, the fact that they can travel considerable distances without deflection proves that they must pass within the radii of the atoms, and the forces to which they are subject must be determined largely by those within the radii. The conclusions given in § 123 undoubtedly suggest that the particles bearing the charge $N\epsilon$, where N is not very different from the atomic weight, are the positively charged part of the atom, and that the view that the positive charge is concentrated on a single central particle rather than distributed throughout a finite sphere is the more correct. The only possible alternative would be to suppose that all the electrons in the atom were collected at one point in the atom so that they could react as a single particle, a view which it is almost impossible to reconcile with the evidence from other sources.

It is suggested then that an atom of atomic weight N consists of about N electrons grouped about a positive charge $N\epsilon$. It is well to point out at once that this structure is much too simple to account for nearly all the properties of the atom. Consider for example the spectrum of hydrogen. If hydrogen contains only two or three electrons and a single positive charge, the number of degrees of freedom of the system is certainly very much less than the number of lines which are known to exist in the spectrum of hydrogen, and since a system cannot have more natural frequencies than it has degrees of freedom, it is utterly impossible on such a simple theory to account for the complexity of the spectrum. Consider again a radioactive atom. When such an atom disintegrates it emits a helium atom, which contains not only electrons but also a positively charged part. If the helium atom is actually part of the radioactive atom, we must suppose that its positively charged part was in some way differentiated from the rest of the original atom; the 'particle' forming the positive part of the original atom must have had some structure. Arguments such as these might be multiplied indefinitely; the structure of the atom must be more complicated than is suggested by either of the two hypotheses which have been mentioned and no useful purpose would be served by tracing in detail their consequences.

Perhaps from other sources some suggestions as to the more complicated structure may be obtained. Thus the fact that helium is a product of the disintegration of all the known radioactive elements suggests that the atom of helium is in some way a constituent of the atoms of those elements[1], and if, as is possible, all atoms are really radioactive, the structure of the helium atom would be involved in that of all atoms. Several attempts, similar to that of Prout, have been made to show that the atomic weights of the actual elements can all be regarded as sums of multiples of those of a small number of hypothetical 'proto-elements,' of which helium and hydrogen are usually taken as two; it is suggested that the actual elements are all built up

---

[1] The helium atoms making up a radioactive atom would have to be supposed to interpenetrate each other, for the atomic volume of the heavy elements is little, if at all, greater than that of helium.

out of these proto-elements. Other arguments for the presence in all atoms of some common structure more complicated than that of an electron have been based on the existence of the number N_0, characteristic of the spectra of all elements, and of the magneton, characteristic of their susceptibility. Such speculations are interesting and, in the future, may lead to important results, but they do not at present enable us in any case to explain the relations which are found to hold between the several properties of the same element.

Thomson's theory.

127. However, an attempt has been made by J. J. Thomson to explain some of the most striking relationships between the different properties of the various atoms by assuming that these atoms contain no common structure other than electrons. It has already been pointed out that, if even we assume the simplest possible form for the distribution of the positive charge in the atom, the problem of determining accurately the arrangement of the electrons round it is quite insoluble except when the number of those electrons is not greater than 4. But it is very probable that whatever the distribution of the positive charge, the arrangement of the electrons has always certain features; the electrons will be arranged on spherical shells round the centre of the atom, the number of the shells being determined by the number of electrons present. If the number of electrons is increased (the positive charge being always equal and opposite to that on the electrons), the electrons added will, in general, take their place in the pre-existing shells, but if the number is increased still further a stage must be reached at which a new shell is formed. If the problem is simplified very greatly by supposing that all the forces between the charges are confined to one plane, the distribution of the electrons in a sphere of uniformly distributed positive electricity can be calculated[1]. The shells now become rings and the following numbers show how many electrons there will be in each ring when n, the total number of electrons, is not greater than 31:

---

[1] The main features of the distribution would not be changed if the positive electricity were concentrated at one centre instead of spread over the whole sphere.

1	2	3	4	5										
5	6	7	8	8	8	9	10	10	10	11				
1	1	1	1	2	3	3	3	4	5	5				
11	11	11	12	12	12	13	13	13	13	13	14	14	15	15
5	6	7	7	8	8	8	8	9	10	10	10	10	10	11
1	1	1	1	1	2	2	3	3	3	4	4	5	5	5

Now suppose that there is some property of the system which depends on the presence of a ring containing two electrons; this property occurs when $n=2$; as n is increased it disappears and reappears only when $n=10$ and a new ring has been formed; it disappears again and reappears when a third ring has been formed and $n=22$. According to Thomson's theory, the order of increasing atomic weight of the elements is the order of increasing n; if, as seems probable, the actual distribution of the electrons in the atom shows the feature which has been indicated in the simplified case, then, since we may suppose that the properties of an element are determined by the distribution of the electrons in it, these properties will be periodic functions of the atomic weight. As the atomic weight is increased, any given property will appear and disappear at intervals which will be approximately the same whatever property is chosen; such a periodicity is, of course, characteristic of the actual elements.

This theory might explain the similarity of the emission spectra of allied elements, for such elements would be those the atoms of which contain common arrangements of electrons. It has also been used by Thomson to explain the simpler facts of chemical combination. For some of these arrangements of electrons will be more stable than others[1]; let us suppose that the most stable arrangements are those which are obtained when a new ring has just been formed, that is, those corresponding to $n=6$ and $n=17$. It is possible that if one of the less stable arrangements were given the opportunity of taking up or parting with electrons (the positive charge remaining unaltered) it would tend to do so in such a way as to pass towards one of the more stable arrangements. An atom which tended to become more stable by losing electrons would behave as an electropositive atom

---

[1] It is quite possible that some of the arrangements given are actually unstable; the calculation of the stability of such systems is quite beyond our powers.

in the presence of one which tended to become more stable by taking up electrons. Further, the same atom might become more stable either by losing electrons and so tending towards the condition in which the last ring had just been formed, or by taking up more electrons and forming a new ring. That is to say, it would tend to become positive or negative according to the conditions in which it were placed, and there would be a relation between the number of electrons which it lost in becoming positive and stable and that which it gained in becoming negative and stable; it would have a positive and a negative valency related in some such way as is required by Abegg's theory of valency (see below).

Of course all these considerations are highly speculative, but they are interesting as representing the only considerable attempt that has been made to explain the periodic classification by Thomson's conceptions of the electronic theory. It should be noted that the theory, while it does not deny the existence of valency electrons bound very loosely to the atoms, does not make the sharp distinction between valency and fixed electrons which experiment seems to suggest. It is impossible at present to be sure whether this feature is a strong or a weak point of the theory.

Valency electrons.

128. If we simply accept the fact that atoms contain valency electrons which can be readily detached from them, without inquiring how the forces holding these electrons arise, we can explain satisfactorily many of the phenomena of chemical combination. In considering these phenomena the theory of valency due to Abegg is helpful.

For the question arises how many valency electrons an atom possesses. In the case of an atom which forms a positive ion the answer is obvious; if the positive charge which the atom acquires when it becomes an electrolytic ion is due to the loss of its valency electrons, the number of those electrons must be equal to its valency. But in the case of an electronegative atom, the negative charge which it gains is due to the gain of electrons and tells us nothing concerning the number of electrons which it possessed before. However, all elements sometimes act as electropositive elements. Abegg calls the valency of an atom when it acts electropositively its 'positive valency,' its value when it acts electronegatively its 'negative valency,' and points out that the arithmetical sum of the negative and positive valencies of an element which is usually electronegative is 8. Thus chlorine may have a negative valency 1 (in HCl) and a positive valency 7 (in Cl_2O_7); sulphur may have a negative valency 2 (in H_2S) and a positive valency 6 (in SO_3); nitrogen a negative valency 3 (in H_3N) and a positive valency 5 (in N_2O_5); carbon a negative valency 4 (in H_4C) and a positive valency 4 (in CO_2). Accordingly Drude suggested that, in the case of an electronegative as well as in that of an electropositive element, the number of valency electrons in an atom is equal to its positive valency. Consequently the number of valency electrons should be greatest in the elements with the smallest negative valency. Some support for this suggestion is to be found in the numbers on p. 314, which show that the halogen elements (which should on this theory have 7 valency electrons) do possess on the whole a greater number of valency electrons than the electropositive elements.

In what follows we shall mean by an 'atom' the positively charged structure devoid of valency electrons; the positive valency of an atom is the number of electrons necessary to neutralise the positive charge; the negative valency the number of electrons, additional to this number, which can attach themselves to the atom.

129. Stark's theory. Stark imagines that a chemical combination between two atoms represents not a direct attraction of one atom for the other, but a simultaneous attraction of both atoms for the same electron which thus forms a bond between the atoms. The simplest form of chemical combination may be represented diagrammatically by Fig. 6, in which the large circles represent atoms, the small an electron and the dotted lines the lines of force of the electron falling partly on one atom and partly on the other. Further he supposes that a mono- or di-valent electropositive atom in the neutralised condition may be represented by *a* or *b*, Fig. 7, the valency electrons being at an appreciable distance from the atom. The greater the positive charge on the atom the greater is the force with which it is likely to attract the electrons which are necessary to neutralise it; we

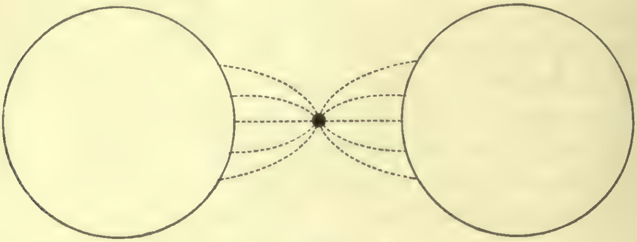


Fig. 6.

should expect therefore that it would be more difficult to detach an electron from an atom of greater positive valency, and that

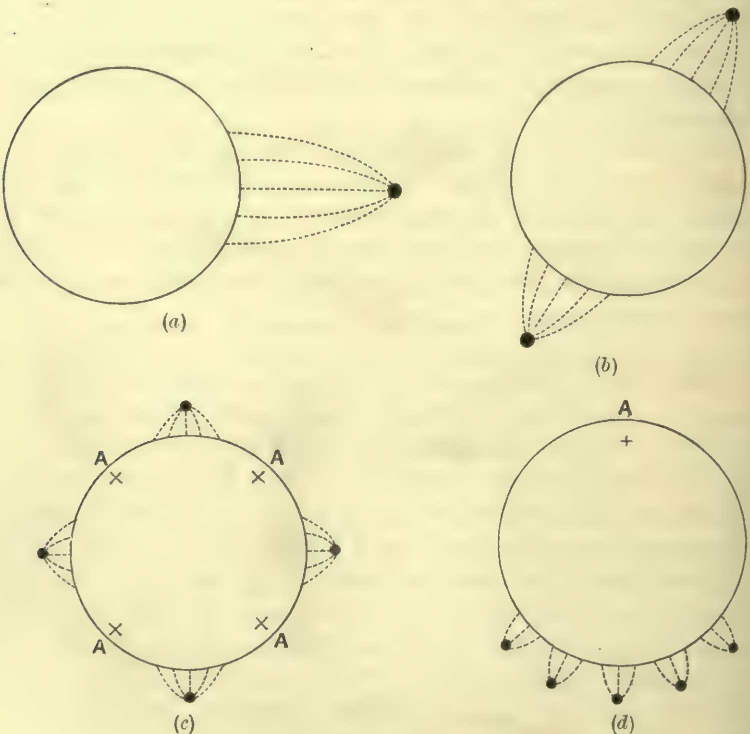


Fig. 7.

the atom with the greater positive valency would be less electropositive. If an atom has a large positive charge it is possible that, even when sufficient electrons are attached to it to neutralise that charge on the whole, there will still be portions of the atom which will react as positively charged to an electron brought near to them; such 'positive patches' are marked *A* in *c* and *d*, Fig. 7, which represent possible forms for an atom of high positive valency. An electron brought near to one of these patches may be attracted to the atom with a force comparable with that which attaches it to a neutralised atom of small positive valency; so that some of the lines of force on such an electron would leave the atom of low positive valency and end on the positive patches of the atom of high positive valency. Thus, if four of the atoms represented by *a* were brought near the atom represented by *c* we might get the arrangement shown in Fig. 8; the atom *c* would behave as an atom of negative valency 4. This view perhaps suggests why an atom of high positive valency behaves also as an atom of negative valency; the diagrams given suggest that the number of positive patches and, therefore, the number of negative valencies, should increase with the positive valency. Since as a matter of fact the number of negative valencies decreases as the number of

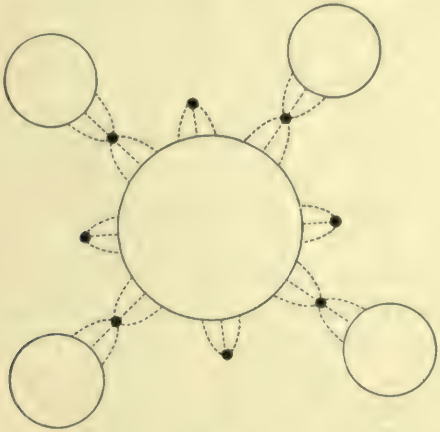


Fig. 8.

positive valencies increases, we may suppose that the valency electrons are not distributed uniformly round the atom but concentrated in one part as indicated in *d*, Fig. 7, which shows an atom with five positive valencies and yet only one positive patch *A*. We must confess, however, that such an arrangement would not have been expected; but if we accept the fact it is clear that we should expect that in the element with the lower negative valency and the fewer positive patches the attractive power of each of those patches would be greater than when there are many positive patches and the distribution of the electrons round the neutralised atom is more uniform.

Since a chemical compound is electrically neutral, the whole number of valency electrons in the compound will be the sum of those which are required to neutralise each of the atoms contained in it, but the valency electrons can hardly be regarded definitely as belonging to either of the atoms which they join. The electrons which join the atoms are those which, when the atoms were in the free state, were attached to the atoms which had the smaller attraction for electrons and were, therefore, more electropositive[1]. But in the compound these electrons may well be attached more firmly to the more electronegative atom, so that if by any means we resolve the compound represented in Fig. 6, the electron which binds the compound may remain attached to the electronegative element rather than to the electropositive element and we shall obtain the electropositive atom positively charged and the electronegative atom negatively charged. Such a resolution takes place when the substance is dissolved and forms an electrolyte.

The energy of chemical combination represents the change in the potential energy of the valency electrons connecting the atoms which takes place when they transfer some of their lines of force from the electropositive to the electronegative atom. It will be the greater the less the attraction of the former atom for an electron and the greater the attraction of the latter. Thus there

---

[1] Of course it must be remembered that the atom which is the more electropositive is that which has the smaller positive charge and requires the smaller number of valency electrons to neutralise its charge; its electropositive valency is represented by the smaller number.

is a great energy change involved in the combination of sodium with chlorine, not so much because the sodium atom as a whole has a great attraction for the chlorine atom as a whole, as because the attraction of the chlorine atom for electrons is so much greater than that of the sodium atom, that the chlorine atom, even when on the whole it is electrically neutral, still possesses a greater attraction for electrons than the sodium atom when it is positively charged. The strongly electropositive elements appear highly reactive, not because they exert powerful forces, but because they exert such very feeble forces on electrons[1].

Applications of the theory.

130. This theory gives a very definite meaning to the 'bonds' of ordinary chemical notation. A bond represents an electron which is attracted notably by both of the atoms which it joins; it is readily seen that such bonds have the definite spatial relations which are necessary to explain the facts of stereochemistry, for, at least in the case of electronegative elements, the positive patches where the lines of force of the electrons end are definite regions on the surfaces of the atoms; moreover the number of bonds which can be attached to an atom is the number representing its positive or negative valency according as the atoms with which it is combined are negative or positive. The main facts which are represented in the conception of valency follow directly from our theory.

Perhaps also the theory provides some help in solving certain of the difficulties connected with the theory of valency. For instance it shows immediately that we should expect an element to form molecules consisting of two atoms; for if two similar atoms are brought together there is clearly no reason why the valency electrons should attach themselves to one rather than to the other; they are likely to attach themselves to both and so form a bond between them[2]. Again it explains very clearly the difference

---

[1] Nothing has been said about the inert gases which have no valency. These may be regarded as composed of atoms which have no attraction for electrons at all, or, more probably, as those which, when they are neutralised by their full complement of valency electrons, have no residual 'positive patches.'

[2] It appears that a valency electron is never attached at one time to more than two atoms, at least in such a manner as to give rise to chemical combination. Hence monovalent elements should not form molecules consisting of more than two atoms.

between a 'saturated' compound and an 'unsaturated.' A saturated compound is one in which each valency electron attached to the saturated atom is attached also to another atom, different for each valency electron. An unsaturated compound is one in which two or more of the valency electrons attached to the unsaturated atom are also attached to one other atom. A 'free' valency[1] occurs when one valency electron is attached to one atom only and does not join two atoms. A simple representation of this view is obtained if we draw the bonds with a 'hinge' in them, representing the electron the attractions of which give rise to the bond. *a*, *b*, *c*, Fig. 9, show an atom saturated, unsaturated and with free valencies. It should be observed that the double bond of the unsaturated atom is characterised by a definite plane, as is required by the existence of certain stereo-isomers.

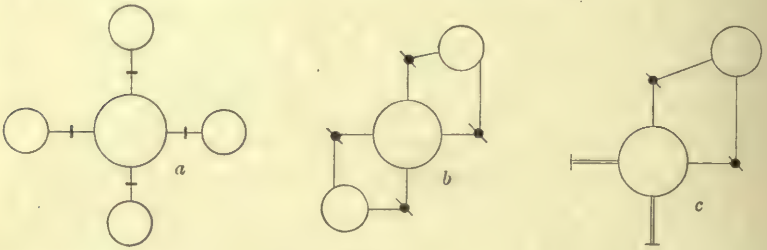


Fig. 9.

This theory cannot, of course, give any definite explanation of why the behaviour in respect of valency of some atoms is anomalous; to give such an explanation it is necessary to have a theory of the internal structure of the atom which ultimately determines its attraction for electrons. But such elements can be represented in our scheme. Iron, for example, which is sometimes divalent and sometimes trivalent, must be supposed to have three places to which an electron can attach itself, one of which differs from the other two in its power of attraction. Hydrogen, which is sometimes strongly electropositive and sometimes strongly electronegative, must be supposed to require only one valency

---

[1] It is for chemists to decide whether it is necessary to suppose that there are free valencies. Our theory only indicates that they are possible, not that they are probable or necessary.

electron to neutralise it, but at the same time to have one well-marked positive patch. It may be remembered that experiments on the positive rays show that hydrogen is peculiar in being almost equally likely to acquire a positive or negative charge. But a consideration of all the facts of chemistry according to this theory would lead us much too far and would probably not lead to any very useful conclusions.

Stark's theory of chemical combination is very well adapted to explaining the connection between optical and chemical properties. The optical properties for frequencies greater than the infra-red are determined by the forces holding the valency electrons. Since we regard these forces as also those which determine chemical combination, it is natural that there should be a very intimate connection between optical properties and chemical constitution. Indeed it is a little difficult to understand why the connection is not even more intimate than it is; a bond joining two definite elements represents an electron attached to both those elements, and we should expect that whenever that bond occurred we should find an electron with a certain definite frequency. As a matter of fact, the frequency associated with a definite bond depends to some extent also on the other bonds present in the molecule; we must suppose that the electron is attached not only to the atoms which it joins but also to some extent to other atoms. There is nothing difficult in such a view, except in so far as it suggests that the valency of an atom should not be quite so definite a property of that atom, uninfluenced by other atoms, as it appears to be. Again we may note that the forces holding an electron are likely to be less intense the less stable the chemical combination; hence on the whole we should expect the less stable combination to give rise to the lower frequencies, a relation which appears to be found in experiment.

Indeed it may be noted that Stark's theory combined with the Planck-Einstein theory of radiation enables us to calculate something about the frequency of a given bond from a knowledge of its chemical properties. For suppose that the heat of formation of the bond in one molecule is E; then it would seem that the electron forming that bond could never have an energy greater than E, for if it acquired a greater energy the bond would be broken. But if it cannot acquire an energy greater than E, it cannot emit radiation of a frequency greater than ν, where $E=h\nu$. But these considerations do not lead to very accurate results. The heat of formation of KCl is about $4{\cdot}5 \times 10^{12}$ ergs per gramme molecule, and we find that ν should not be greater than 10^{15}, which lies just in the ultra-violet; experiment shows that the compound certainly possesses absorption bands for which ν is greater than 2×10^{15}. The discrepancy is probably due to the fact that the heat of formation does not accurately represent the greatest possible energy of the electron in the bond.

According to Stark it is the band spectra of substances which represent the vibrations of the electrons concerned in chemical combination. It has been suggested already that the absence of the Zeeman effect in such spectra may be attributed to the fact that these vibrations are not simply harmonic; this suggestion accords well with the fact that the forces acting on an electron between two atoms would be exceedingly complex and not likely to be simply proportional to the displacement. The line spectrum, according to Stark, is due to the electrons which remain when the atom has lost its valency electrons; if this view is correct the electrons giving rise to the line spectrum must be those in the interior of the atom which are not directly concerned in chemical combination. There are certain difficulties in this view, such as the intimate connection between the spectrum of an element and its chemical nature and the small change in the physical conditions which will change a line spectrum into a band spectrum. It would also be possible to suppose that the line spectrum is emitted by the electron when it represents a free valency and is attached only to one atom; of course all the valency electrons of an uncombined atom represent free valencies. When it is attached to only one atom the forces acting on the electron are probably simpler and the vibrations more nearly periodic; a very small change in the physical conditions might cause the electron to throw some of its lines of force on another atom and so give rise to the band spectrum. A decision between the two views could be obtained if it could be ascertained definitely whether or no an atom is always charged positively when it emits its line spectrum.

Stark's theory also gives some indication as to why the metallic properties, the possession of free electrons and great tensile strength, should be associated with highly electropositive elements. Such elements are those the atoms of which have a very slight attraction for electrons; it is reasonable to suppose that, when they are neutralised by valency electrons, these electrons are not so close to the positive nucleus forming the atom as in the case of less highly electropositive elements. Now when the atoms are closely packed together in the solid state the force due to one atom will tend to counteract that due to the others, and at points which are equidistant from two atoms the electric field due to both of them will vanish. If the valency electron is so far from its atom that it lies in the region in which the electric field vanishes it will be attracted equally by each of the atoms surrounding it and will be subject to no resultant forces; it will be able to move freely through the substance. On the other hand the force which one atom exerts on another is likely to be greater if the valency electron is not very near the atom to which it is attached, so that the system of atom and electron forms an electric doublet of considerable moment. Sutherland has attempted to work out a complete theory of the mechanical properties of substances on the view that the forces between the molecules are those between electric doublets, and he finds, of course, that the tensile strength should be greater the greater the moment of the doublets. If the doublets are those formed by the valency electrons and the positive charges to which they are attached, the intimate connection between chemical and mechanical properties is explained at once. Indeed the 'molecular forces' of a substance are to be regarded as differing only in degree and not in kind from its chemical forces.

This brief sketch of the speculations which have been made as to the structure of the atom must suffice; a full discussion of all of them would occupy far too much space. It cannot be pretended that we have at present any but a most rudimentary theory of the atom; none of the suggestions which have been made as to its structure lead in any case to the possibility of calculating in detail one property of an atom from a knowledge of the others; no quantitative relations can be deduced. Until such calculation is possible all the hypotheses must be regarded as somewhat vague and tentative, but the considerations which have been given seem to indicate that the fundamental ideas which underlie all these speculations are not very far from the truth.

REFERENCES.

Thomson's theory of the atom is given in his *Corpuscular Theory of Matter* (Constable, London, 1907), Stark's in his *Prinzipien der Atomdynamik* (Hirzel, Leipzig, 1910). Other isolated suggestions are to be found in many of the papers mentioned at the end of Chapter XII. It should be pointed out that it is not certain that either of these authors would accept in every detail the statement of their theories which is given here.

CHAPTER XIV

THE PROPERTIES OF MOVING SYSTEMS

Statement of the problem.

131. THIS last chapter will be devoted to a consideration of certain problems which in the past have been the subject of much discussion and difference of opinion among physicists. These problems have little direct connection with those which have concerned us hitherto and our study of modern electrical theory would be logically complete with only the barest reference to them. But since, on the other hand, they are of great interest and are ordinarily accounted of very great difficulty, some treatment of them appears to be desirable. Many of the difficulties which are encountered by students when they first turn their attention to these problems are due to the errors of their preconceived ideas, and readers should therefore consider in the first place, not whether the statements made are in accordance with the opinions which they already entertain on such subjects, but only whether those statements are intelligible, coherent and consistent with each other.

The general problem which we have to consider may be stated thus:—Any set of bodies concerned in any experiment may be divided into two groups. The first group, S, includes the observer who is making the experiment and the various instruments by which he measures the properties concerned in the experiment; in all actual experiments all the bodies concerned in this group are at rest relatively to each other. The second group R contains the bodies on which the experiment is made; in some cases these bodies are at rest relatively to each other and to the first group; in others they are at rest relatively to each other but in motion relatively to the first group; in others again they are in motion relatively to each other and therefore, in some instances at least, in motion relative to the first group. The problem is to find a relation between the result of the experiment in the first case and the result in either of the other cases, the bodies concerned in the experiment being the same in all instances.

It would appear at first sight that a general deduction applicable to all kinds of experiment must be impossible and that the form of the desired relation must differ with the nature of the experiment. This expectation is perfectly correct; but the instances which we shall consider form a special class. We shall consider at present only those instances in which the second experiment differs from the first in the fact that all the material bodies in the group *R* which, in the first experiment, were at rest relatively to the group *S* are, in the second experiment, moving relatively to that group with a uniform velocity *v*. In these instances we shall say that in the second experiment the group *R* is moving with a uniform velocity *v* relatively to the group *S*, even though there are bodies which in the first experiment were not at rest relatively to *S* and are not in the second experiment moving relatively to *S* with a velocity *v*.

By means of a simple device we can now divide the problem into two parts, such that a solution of one part can be given which is valid for all cases, while a solution of the other is obtainable very easily in a large number of important cases. We imagine that the group *R* includes an observer *S′* who is also making experiments on the group *R*, and is also moving relatively to *S* with the velocity *v*, being at rest relatively to *R*. If we can discover how the observations made by *S′* on *R* when *R* is moving relatively to *S* are related (1) to those made by *S* when *R* is moving relatively to *S*, and (2) to those made by *S* when *R* is at rest relatively to *S*, we can obviously discover how the observations (1) and (2) are related.

Principle of Moving Axes.

132. A general solution of the first part of the problem has long been employed to solve certain problems in dynamics. The principle of this solution, which will here be termed the Principle of Moving Axes, may be stated in the following way; a consideration of its basis and meaning will be postponed for the present.

Suppose that S and S' are furnished with measuring instruments which, when compared while S and S' are relatively at rest, are found to be identical; the most important of these instruments will be those by which the distances of material points from the coordinate axes and the intervals between moments of time are measured. Let the position of a point relative to his axes and the time of an event measured by S' be denoted by (x', y', z', t') and the corresponding quantities measured by S be denoted by (x, y, z, t). Suppose further that the axes of coordinates employed by S' are parallel to those employed by S, that S takes his axis of x in the direction of the velocity v of R relative to S and that the origins of the axes of S and S' are such that they coincide when $t = t' = 0$. Then as the result of his observations S' will arrive at certain laws, which will take the form of statements that a certain event (such as the presence of a material particle capable of identification) will take place at a point (x', y', z') at a time t' if these quantities are such that some relation $f'(x', y', z', t') = 0$ is fulfilled. Then the Principle of Moving Axes states that S, as a result of his observations, will arrive at the law that the same event will take place at a point (x, y, z) and a time t, if these quantities are such that $f'(x - vt, y, z, t) = 0$. This equation may also be written $f(x, y, z, v, t)$, where f is some function which will, in general, be different from f'.

This rule enables us to deduce from a knowledge of the laws observed by S' a knowledge of those observed by S by subjecting the former to a simple algebraic transformation in which $x - vt$ is substituted for x', y for y', z for z' and t for t'. A simple example may make the matter clearer. Suppose that one of the laws discovered by S' is that a certain particle is moving with a uniform velocity u' parallel to the axis of x'. This law may be put in the form that the presence of this particle will occur whenever $x' - u't' = 0$; then S will arrive at the law that the presence of the same particle will occur whenever $(x - vt) - u't = 0$, that is whenever $x - (u' + v)t = 0$. In other words S will find that the particle is moving parallel to x with a uniform velocity $u = u' + v$.

The Principle of Moving Axes, then, solves the first part of the problem. The second part, that of finding a relation between the laws observed by S' when R is moving relatively to S and those observed by S (or S') when R is at rest relatively to S, is not capable of a general solution. But there is one special solution which is applicable to a large class of cases. It is found in some cases that the results of the experiments made by S' on R are the same whether R is or is not moving relatively to S with a uniform velocity; R is then said to be physically independent of S.

In these cases we have solved our original problem of finding a relation between the laws observed by S when R was at rest relatively to him and those which he observes when R is moving relatively to him. If R is physically independent of S, the laws observed by S' are identical with those observed by S when R is at rest relatively to him; so that when we know the laws observed by S when R is at rest relatively to him, we know those observed by S' and hence, by means of the transformation, those observed by S when R is moving relatively to him.

The Principle of Relativity.

133. However, it has appeared recently that the results of this solution of the problem are not always correct. The doctrine, called by its originator, Einstein, the Principle of Relativity, states that this method of attacking the problem is of universal applicability, but that the algebraic transformation which is necessary to convert the laws observed by S' into those observed by S is not that which has just been given. In place of that transformation we must use the following:

$$\text{For}\quad \left.\begin{array}{llll} x' & \text{we must substitute} & & \beta(x-vt) \\ y' & \quad ,, & ,, & y \\ z' & \quad ,, & ,, & z \\ t' & \quad ,, & ,, & \beta\left(t-\dfrac{vx}{c^2}\right) \end{array}\right\} \quad \ldots\ldots\ldots(258),$$

where $\beta=\left(1-\dfrac{v^2}{c^2}\right)^{-\frac{1}{2}}$, and c is, as always, the velocity of light. The discussion of the meaning and basis of this transformation will again be postponed, until we have deduced some of its consequences and shown that they are in accordance with experiment.

Let us take first the simple example already treated. With our new transformation the law observed by S is that the particle will be present whenever

$$\beta(x-vt)-u'.\beta\left(t-\frac{vx}{c^2}\right)=0,$$

that is, whenever

$$x-\frac{u'+v}{1+\frac{u'v}{c^2}}.t=0.$$

The velocity of the particle relative to the axes of S will be parallel to the axis of x and will be given by

$$u=\frac{u'+v}{1+\frac{u'v}{c^2}} \quad \text{.....................(259).}$$

This result is not capable of immediate experimental verification, because we cannot determine directly the velocities of material particles which are moving relatively to us with a speed comparable with that of light. But it should be observed that if v is very small compared with c, this transformation becomes identical with that adopted before; we could only hope to determine which of the transformations ought to be adopted if we study the properties of systems moving relatively to us with velocities comparable with that of light. Let us now turn to the problems which are of classical importance.

The optics of moving systems.

134. Suppose that R includes a source of light which, when R is at rest relative to S, appears to have a frequency ν' and to emit a parallel beam travelling in a direction of which the direction cosines are (l', m', n'). Experiment shows that this frequency and this direction are different according as the source of light is or is not at rest relatively to S; the change of frequency and of direction with the motion of the source are called respectively the Doppler effect and aberration[1]. Experiment shows also that in these cases the source of light is physically independent of S in the sense, already given,

---

[1] See Appendix II.

that to an observer relative to whom the source is always at rest the source appears to have exactly the same properties whether it is or is not moving relatively to S. Or, again, the source of light may be at rest relative to S, but between S and the source there may be a material medium through which the light travels and this medium may be in some cases and not in others at rest relative to S; it is found that the velocity with which the light appears to S to travel through the medium differs with the motion of the medium relative to S; this change of velocity is known as the Fizeau effect. In this instance also, although there is no direct experimental evidence, it is believed that the medium may be considered physically independent of S. The oldest problem concerning moving systems is that of finding a general means by which the changes known as the Doppler effect, aberration and the Fizeau effect, may be predicted from a knowledge of the properties of the source and of the medium when they are at rest relative to S and a knowledge of their velocity v relative to S.

Let us now apply the methods of the Principle of Relativity and consider first the Doppler effect and aberration. Since the source is physically independent of the observer and his instruments, the light will appear to S' to be a parallel beam having the frequency ν' and the direction (l', m', n'); in other words the law observed by S' will be that the light vector will have the same value whenever (x', y', z', t') are such that

$$\sin 2\pi\nu' \left(t' + \frac{l'x' + m'y' + n'z'}{c}\right)$$

has the same value. Consequently the law observed by S will be that the light vector will have the same value whenever (x, y, z, t) are such that

$$\sin 2\pi\nu' \left\{\beta\left(t - \frac{vx}{c^2}\right) + \frac{l'\beta(x - vt) + m'y + n'z}{c}\right\}$$

has the same value, that is, whenever

$$\sin 2\pi\nu \left(t + \frac{lx + my + nz}{c}\right)$$

has the same value, where

$$\nu = \nu' . \beta \left(1 - l' \frac{v}{c}\right), \quad l = \frac{l' - \frac{v}{c}}{1 - l' \frac{v}{c}}, \quad (m, n) = \frac{(m', n')}{\beta \left(1 - l' \frac{v}{c}\right)} \text{ (260).}$$

Consequently the light will appear to S to have a frequency ν and to travel in a direction l, m, n. There will be changes in frequency and in direction of the light, according as S is or is not moving relatively to R; the predicted values of the changes agree with those which are actually observed. The ordinary formulae relating ν with ν' and l, m, n with l', m', n' differ from those just obtained by the absence of the factor β; but in all cases this term is so nearly 1 that our experiments cannot decide whether or no it ought to be inserted.

We might treat in the same manner the case of light travelling through a medium which is moving relatively to the observer, by substituting V, the velocity of the light in the medium, for c in the above expression for the light vector. But it is simpler to make use of the result (259) which we have already obtained. Suppose that to S', who is at rest relatively to the medium, the velocity of the light through the medium appears to be V' and to be parallel to the axis of x'; then to S the light will appear to travel parallel to the axis of x with a velocity

$$V = \frac{V' + v}{1 + \frac{vV'}{c^2}}.$$

If v is so small that we may neglect $\frac{v^2}{c^2}$ in comparison with 1, we may write

$$V = V' + v\left(1 - \frac{V'^2}{c^2}\right);$$

now $\frac{c}{V'}$ is n, the refractive index of the medium when it is at rest relatively to the observer; hence we arrive at the ordinary formula for the Fizeau effect

$$V = V' + v\,(1 - n^{-2}) \text{ ....(261).}$$

It is to be observed that, if the refractive index of the medium is

1 and the velocity of the light in it c, the velocity of the light measured by S will also be c; the velocity of a beam of light which appears to one observer to have a velocity c will appear to have the same velocity to any other observer moving relative to the first with any uniform velocity.

135. In § 76 we noted that the motion of an electron in an electromagnetic field with a velocity comparable with that of light shows certain anomalies which are not completely explicable by the general principles enunciated in Chapter I. The question arises whether these anomalies can be removed by the method of treating moving systems which has just been described. At first sight we may be inclined to argue thus. The motion of a particle in an electromagnetic field relative to which it is at rest is given by

Application to electromagnetism.

$$m\frac{d^2x'}{dt'^2}=eX \quad \text{....................(262).}$$

Therefore, if it is moving relative to that field with a velocity v, its motion must be given by

$$m\frac{d^2\beta(x-vt)}{d\left\{\beta\left(t-\frac{vx}{c^2}\right)\right\}^2}=eX.$$

But this argument would be erroneous. If the electron, together with an observer S' always at rest relative to the electron, form the system R, R is not physically independent of S, if S includes the instruments by which the electromagnetic field is produced. These instruments exert forces on R which depend on the velocity of R relative to S, and the acceleration of the electron relative to S' will change according as R is or is not moving relatively to S. The transformation which has been described gives in all cases the laws observed by S in terms of those observed by S', but it leads directly to a solution of the problem only when R is physically independent of S, for it is only in this case that we can identify the laws discovered by S' with those which would be discovered by S if R were at rest relative to S.

To deal with this problem we must introduce a new assumption and one that is as fundamental in the methods of the Principle of Relativity as the transformation (258). This assumption is that the laws stated by the equations (2), (7), (9) of Chapter I are found to be true by all observers who are moving with any uniform velocity (including zero) relatively to the instruments producing the electromagnetic field. Suppose then that the system R includes instruments which are producing an electromagnetic field, and that the quantities characteristic of this field, as measured by S', are an electric intensity X', Y', Z', a magnetic intensity L', M', N', a density of electric charge ρ', and a velocity u_x', u_y', u_z' of the particles bearing this charge. All these quantities vary with x', y', z', t', and in virtue of (2), (7), (9), equations must hold between them of which the following are typical:—

$$\left.\begin{aligned} \left(4\pi\rho' u_x' + \frac{dX'}{dt'}\right) = \frac{dN'}{dy'} - \frac{dM'}{dz'}; \quad \frac{1}{c^2}\frac{dL'}{dt'} = \frac{dY'}{dz'} - \frac{dZ'}{dy'} \\ 4\pi\rho' = \frac{dX'}{dx'} + \frac{dY'}{dy'} + \frac{dZ'}{dz'} \end{aligned}\right\} \quad (263).$$

Now let S perform measurements similar to those by which S' arrived at the values of these quantities and let him find as a result of these measurements that the electric intensity is X, Y, Z, the magnetic intensity L, M, N, the density of charge ρ and the velocity of the particles carrying it u_x, u_y, u_z. Then, if the laws of the electromagnetic field are unaltered even when the observer is moving relatively to the instruments producing the field, S must find relations of which the following are typical:—

$$\left.\begin{aligned} \left(4\pi\rho u_x + \frac{dX}{dt}\right) = \frac{dN}{dy} - \frac{dM}{dz}; \quad \frac{1}{c^2}\frac{dL}{dt} = \frac{dY}{dz} - \frac{dZ}{dy} \\ 4\pi\rho = \frac{dX}{dx} + \frac{dY}{dy} + \frac{dZ}{dz} \end{aligned}\right\} \quad (264).$$

But if we transform equations (263) according to (258) and rearrange the terms, they become equations of which the following are typical:—

$$\left.\begin{aligned} \left\{4\pi\beta\rho'(u_x' + v) + \frac{dX'}{dt}\right\} = \beta\left\{\frac{d(N' + vY')}{dy} - \frac{d(M' - vZ')}{dz}\right\} \\ \frac{1}{c^2}\frac{dL'}{dt} = \beta\left\{\frac{d\left(Y' + \frac{v}{c^2}N'\right)}{dz} - \frac{d\left(Z' - \frac{v}{c^2}M'\right)}{dy}\right\} \end{aligned}\right\}$$

$$\dots\dots\dots(265).$$

These equations state the relations between the quantities X', Y', Z', L', M', N', as these relations are measured by S. But what is the physical significance to S of these quantities? We must not assume immediately that X' represents the electric intensity to S as well as to S', for when S says the electric intensity has a certain value he is stating the occurrence, not of a single event, but of a complicated connected series of events the places and times associated with which are subject to certain relations. And these relations will differ, as we have seen, according as they are measured by S' or by S, so that the events which are described by S' in the statement that the electric intensity has a certain value will not be described by S in a similar statement. Indeed we can see that $(X'Y'Z')$ does not represent to S the electric intensity, for we know that (264) and (265) must be consistent since they are merely different ways of describing the experiments made by S; and they are not consistent if we put $X = X'$, $Y = Y'$, etc. On the other hand, they are consistent with each other if we put

$$\left.\begin{array}{ll} X = X', & L = L' \\ Y = \beta\left(Y' + \frac{v}{c^2}N'\right), & M = \beta\left(M' - vZ'\right) \\ Z = \beta\left(Z' - \frac{v}{c^2}M'\right), & N = \beta\left(N' + vY'\right) \\ \rho = \beta\left(1 + \frac{vu_x'}{c^2}\right)\rho' & \\ u_x = \dfrac{u_x' + v}{1 + \dfrac{u_x' v}{c^2}}, & (u_y, u_z) = \dfrac{(u_y', u_z')}{\beta\left(1 + \dfrac{u_x' v}{c^2}\right)} \end{array}\right\} \quad \ldots(266).$$

These equations give the values of the quantities characteristic of an electromagnetic field measured by S (moving relatively to the instruments producing the field) in terms of those measured by S' (at rest relatively to those instruments). If, on the other hand, it is the observer S who is at rest relatively to the instruments, we may obtain the values of the field observed by S' in terms of those observed by S by solving (266) for the dashed quantities; the same solution will be obtained (as may be seen from general considerations) by interchanging dashed and undashed quantities, while the sign of v is reversed.

Before proceeding, two points may be noted. First, it will be seen that the relation between u_x and u_x' is the same as that obtained in (259); accordingly we may conclude that the assumption which we have made concerning the universal validity of the electromagnetic equations is not inconsistent with the Principle of Relativity. Second, we may ask, if the charge on a given particle at rest relative to S' is e', what is the value of e, the charge on that particle as it appears to S. If dx', dy', dz' is the volume of the particle as it appears to S', we have

$$e' = \rho' dx' dy' dz';$$

and similarly $e = \rho dx dy dz$; if we make our measurements of the limits of the particle simultaneously (i.e. if t is constant),

$$dx' dy' dz' = \beta dx dy dz;$$

hence $e' = e$. The charge on a particle will appear the same, whether or no it is moving relative to the observer.

Motion of a charged particle.

136. We may now proceed to investigate the laws of the motion of a charged particle which is moving with a great velocity relatively to the instruments producing a field. Suppose now that these instruments are at rest relatively to S. The particle will be accelerated relatively to S, but if its acceleration is small, we may consider that at each instant it is moving relatively to S with a uniform velocity v, and that the observer S' is at rest relatively to the particle. The electromagnetic field (X', Y', Z', L', M', N') observed by S' will be related to that observed by S (X, Y, Z, L, M, N) by (266). If we make the additional assumption that a system consisting of a charged particle and instruments producing an electromagnetic field is physically independent of the observer, we may remove the instruments at rest relatively to S and suppose the field observed by S to be produced by certain other instruments at rest relatively to S'. Then the equations of the motion of the particle relatively to S' are, according to (262),

$$m\frac{d^2x'}{dt'^2} = eX', \quad m\frac{d^2y'}{dt'^2} = eY', \quad m\frac{d^2z'}{dt'^2} = eZ' \quad \ldots(267).$$

In order to obtain the equations of the motion of the particle relatively to S, in terms of the field observed by S, we must transform (267), by substituting for x', y', z', t' according to (258) and for (X', Y', Z') according to (266). We thus obtain

$$m\beta^3 \frac{d^2x}{dt^2} = eX, \quad m\beta \frac{d^2y}{dt^2} = e\left(Y - \frac{v}{c^2}N\right), \quad m\beta \frac{d^2z}{dt^2} = e\left(Z + \frac{v}{c^2}M\right) \quad \text{.........(268).}$$

These equations become more symmetrical if we remove the restriction that the axis of x is parallel to v and imagine the equations referred to any system of rectangular coordinates at rest relatively to S, having the same origin as those we have used so far. Then we shall find that, if $q = (\dot{x}^2 + \dot{y}^2 + \dot{z}^2)^{\frac{1}{2}}$ is the velocity of the particle, the equations become

$$\frac{d}{dt}(\beta m\dot{x}) = K_x, \quad \frac{d}{dt}(\beta m\dot{y}) = K_y, \quad \frac{d}{dt}(\beta m\dot{z}) = K_z \text{...(269),}$$

where

$$\left.\begin{aligned} K_x &= e\left(X + \frac{\dot{y}}{c^2}N - \frac{\dot{z}}{c^2}M\right) \\ K_y &= e\left(Y + \frac{\dot{z}}{c^2}L - \frac{\dot{x}}{c^2}N\right) \\ K_z &= e\left(Z + \frac{\dot{x}}{c^2}M - \frac{\dot{y}}{c^2}L\right) \end{aligned}\right\} \text{..............(270),}$$

and β is now written for $\left(1 - \frac{q^2}{c^2}\right)^{-\frac{1}{2}}$.

Equations (269) are simply the general equations which reduce to (268), when the axes to which they are referred are turned so that the velocity of the particle lies along the axis of x.

Comparing (269) with the equation $\frac{d}{dt}(m\dot{x}) = P_x$ (where P_x is the 'force' on the particle parallel to x), which represents the fundamental proposition of Newtonian mechanics, we see that it is suggested that the charged particle moves like a Newtonian particle having a mass βm and acted on by a force (K_x, K_y, K_z). But before we can be sure that we can adopt this suggestion, we must see whether the principles of electromagnetism allow us to attribute to the quantity (K_x, K_y, K_z) *all* the properties of a Newtonian force. A thorough investigation shows that they do allow us so to do; here we will only investigate one special question.

According to Newtonian mechanics a body acted on by a force (K_x, K_y, K_z) must acquire in a time dt an energy

$$(K_x\dot{x} + K_y\dot{y} + K_z\dot{z})\,dt.$$

Hence, if the quantity which we have called (K_x, K_y, K_z) is a force, we must have

$$-\frac{dE}{dt} = K_x\dot{x} + K_y\dot{y} + K_z\dot{z} = e\,(X\dot{x} + Y\dot{y} + Z\dot{z}) \quad \ldots(271),$$

where E is the energy of the electromagnetic field from which the particle receives its energy. Now

$$E = \int_S \left\{\frac{1}{8\pi}(X^2 + Y^2 + Z^2) + \frac{1}{8\pi c^2}(L^2 + M^2 + N^2)\right\} dS \ldots(272),$$

where S is an element of volume in the field, and it may be easily shown from (264) that, if ρ is zero except at the point occupied by the moving particle, $-\frac{dE}{dt}$ has the value required by (271). In this respect therefore the quantity may be regarded as a Newtonian force. It is interesting and important to observe that the value of the force acting on a charged particle, moving with a velocity v relative to the instruments which are producing an electromagnetic field, which is given by (270) is exactly that which is given by (5). In Chapter I we treated the properties of a moving charge as altogether distinct from those of a charge at rest relative to the observer; we now see that, if we adopt the Principle of Relativity, some of the properties of the moving charge may be deduced directly from those of the resting charge; a further inquiry shows that the same statement is true of all those properties. We see in fact that a magnetic field is not something altogether different in its nature from an electrostatic field; whether a given field appears electrostatic or magnetic depends entirely upon the state of motion of the observer relative to the instruments which are producing it.

And we also see that the relation which has been obtained

$$m' = \beta m \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(273),$$

where m' is the mass of a body moving relatively to the observer with a velocity v, and m its mass when that velocity is infinitely small, is precisely that which was quoted in (177) as being consistent with experiment. The apparent anomalies of the motion of fast electrons have been shown to be a necessary consequence of the fundamental electromagnetic equations for a system at rest together with the Principle of Relativity. This agreement of our theory with experiment is perhaps the most notable evidence for the validity of that theory.

The energy of the particle, E, is given by

$$E = \int (K_x \dot{x} + K_y \dot{y} + K_z \dot{z})\, dt.$$

Inserting in this expression the values of K_x, K_y, K_z obtained from (269) we find

$$E = \beta m c^2 + \text{constant} \qquad \text{.................(274).}$$

Since the kinetic energy must be zero when q is zero, the constant must be $-mc^2$, and we find that the kinetic energy of a charged particle, which has a mass m when it is moving with a velocity very small compared with that of light, is given by

$$E = mc^2 \left(\frac{1}{\sqrt{1 - q^2/c^2}} - 1 \right) \qquad \text{..............(275).}$$

It should be noted that, when q/c is very small, this expression agrees with the ordinary formula $E = \frac{1}{2} m q^2$, and also that we do *not* obtain (275) by substituting for m the value m', corresponding to the velocity q, taken from (273). It is probable in the future that a further test of the theory will be obtained by measurements of the energy of high speed electrons.

A further deduction concerning the energy of moving systems is important. It can be shown from (272) and (266) that if E' is the energy of the electromagnetic field in a system R physically independent of S, then E, the energy of that field as it appears to S, is $\beta E'$. Accordingly, the energy of such a system R appears to S to change according as R is at rest or moving relatively to him; E, the energy of R when R moves relatively to S with a velocity q, is greater than E', that when it is at rest relatively to S, by $(\beta - 1) E'$. In other words, the system R, in virtue of its internal energy, appears to have a kinetic energy $(\beta - 1) E'$, or, comparing the result with (275), it appears to have a mass m given by

$$m = \frac{E'}{c^2} \quad\quad (276).$$

This conclusion that the mass of a system varies with its internal energy may have some bearing upon the atomic weights of the radioactive elements which undergo large changes of internal energy during their disintegration.

The methods of the last two paragraphs also enable us to deduce equations (234), (235) from the corresponding equations (231), (232) which are applicable only to low speeds. But no useful purpose would be served by giving the deduction here.

The new Dynamics.

137. These results represent only a very small portion of the consequences which have been deduced in recent years from the Principle of Relativity. Since (273) does not involve the charge on the moving particle in any way, the suggestion naturally occurs that the mass of all bodies, and not only of charged bodies, will vary with the velocity in the way stated by that equation. Moreover, if we accept the view which is really fundamental in modern electrical theory, that all forces acting on all material bodies are essentially electromagnetic, this proposition would follow at once from those which have been deduced already[1]. And since this proposition is inconsistent with Newtonian dynamics, which states that the mass of a body is a constant property independent of its velocity relative to other bodies, the acceptance of it makes it necessary to revise the entire body of our dynamical principles[2]. The task of this revision has

---

[1] If the view of the Principle of Relativity put forward by Minkowski (see below, § 139) is adopted, the proposition follows without any assumptions as to the nature of the forces acting.

[2] Another respect in which the theory just stated is inconsistent with Newtonian dynamics should be noted. We have concluded that when two systems react and the reaction is observed by an observer whose axes of reference are fixed in one of those systems the mass of the other system observed by him depends only on the properties of that system. According to Newtonian dynamics, the mass of a system will be a constant determined only by that system if, and only if, the axes of reference of the observer are fixed in a system relatively to which the centre of mass of the two systems is at rest; this system cannot be at rest

occupied the attention of a number of investigators and a new set of principles, consistent with the Principle of Relativity, has been elaborated. It is not proposed here to follow out these conclusions; they have no direct bearing upon the main subject of this volume and, in addition, they appear to me a little dangerous. There is not the smallest prospect of our being able to subject any of these conclusions to the test of experimental verification, for in order to do so we should have to be able to make accurate observations on bodies moving relatively to us with velocities approaching that of light. The confusion in which the theory of moving systems was involved a few years ago arose from the attribution of universal validity to principles which had been tested experimentally in only a very limited class of cases; with this warning before us, it appears wiser to confine our application of the Principle of Relativity to those cases which can be investigated experimentally, and all these cases have been considered in the last two paragraphs.

'Time and Space.'

138. Fortunate is the science, as is the country, which has no history. If the theory of the properties of moving systems had no history, there would be no necessity to add anything to our formal statement of the Principle of Relativity and our proof that it leads to results in accordance with experiment. This accordance would be sufficient to prove that it is valuable and true. But when that Principle was first proposed, difficulties were found by some people in accepting it. These difficulties were of two kinds; first, it was said that the Principle appeared so artificial and so complex that it could not be accepted as the basis of a science unless it could be shown to be the consequence of other principles, less artificial, less complex and more self-evident; second, it was said (by the less wise critics) that the Principle could not be universally true, because, though it led to results in accordance with experiment in the cases which have been considered, in other cases it conflicted with propositions

relative to either of the two reacting systems. When the axes of reference are fixed in one of the two reacting systems, the principle of the 'conservation of momentum,' fundamental in Newtonian dynamics, becomes, not merely false, but meaningless, for the expression 'the velocity of an observer relatively to himself' is without significance.

more certain than any others in science, namely those stating the fundamental properties of space and time[1]. It appears to be desirable to devote some little space to the consideration of those objections; but if the reader does not experience these difficulties, he will be wise to omit the next two paragraphs and not to trouble himself with confusions of thought based on the fallacies of a bygone age.

Such objections are best met by comparing the Principle of Relativity with the Principle of Moving Axes, for, according to the objectors, the latter Principle is a notable example of all those virtues which the former lacks. Let us consider why the transformation characteristic of the Principle of Moving Axes appears to them so simple.

If the objectors could have been induced to state their views rather more fully they would probably have said something like this:—'It is perfectly self-evident and necessary that, if at a given moment, the distance of a point A from the point P is a and the distance of P from the point Q is b (A, P, Q lying on the same straight line), then the distance of A from Q is $a+b$. Now this is the only proposition necessary to establish the transformation asserted by the Principle of Moving Axes. For at a moment t the distance of the origin of S' from the origin of S is vt; this is what is meant by saying that the velocity of S' relative to S is v; so that, if x' is the distance of a point from S', x, its distance from S, must be $x'+vt$ and x' must always be identically equal to $x-vt$. Of course if any proposition about x' is true, the same proposition must be true about $x-vt$, for $x-vt$ is identically equal to x'.'

The Principle of Relativity denies that the proposition which is asserted at the beginning of this statement is either self-evident or, as a matter of fact, true, and by this denial it certainly rejects the conceptions of time and space involved in the statement. But

---

[1] The difficulties of the objectors were intensified because they seemed in some doubt whether these very certain propositions were certain because they represented the results of experiment or because they were logical deductions from the definitions of the terms employed in stating them; in other words whether they were physically or mathematically true. I think they were also troubled by a vague remembrance of a perverted form of the Kantian theory that our conceptions of space and time are necessarily implied in all knowledge.

the denial is not a sufficient ground for doubting the Principle of Relativity unless it can be shown that these conceptions are really so necessary that the rejection of them is logically absurd. I shall proceed to try to show that this proposition is no more self-evident than that which the Principle of Relativity proposes to substitute for it; whether it is actually true can, of course, be decided only by experiment, and the agreement of the results of the Principle of Relativity with experiment is sufficient to indicate that, if it is not self-evident, it is probably not true. For this purpose I shall consider not the proposition itself, but three other propositions (A), (B), (C), which are involved in it and would probably appear to the objectors even more self-evident than that which they have been supposed to state; it can easily be shown that if these three propositions are not true, then the original proposition on which the Principle of Moving Axes is founded is not self-evident and is probably not true.

Proposition (A) may be stated thus:—

'If an event P appears simultaneous with an event Q to one observer S, then P must appear simultaneous with Q to any other observer S'.' (The same proposition might be stated with the substitution of 'before' or 'after' for 'simultaneous with.') It is this proposition which gives rise to and renders useful the conception of a 'time at which a thing happens,' this 'time' being the same for all observers.

If we are to discuss this, or any other, proposition, we must first be sure what is meant by the terms used in it. I think any belief in the self-evidence of proposition (A) can be removed by pointing out that there are two distinct senses in which the word 'simultaneous' is used at the present time, and that the most ordinary experience shows that, if this word is used in one of these senses in (A), then (A) is most certainly not true.

If I say that the sound of the explosion in a gun appears to me simultaneous with the accompanying flash of light I may mean, and indeed usually do mean, that there is a certain relation between my sensations of hearing and seeing; this relation is so simple and fundamental that I cannot possibly explain it to any one who says he does not understand what I mean by it. In order to distinguish this sense of the word 'simultaneous' from other senses I shall say that two events happen 'simul' when they are simultaneous in this sense.

It is quite certain that 'simultaneous with' was originally used, and is still used by people untrained in science, in the sense of 'simul,' and it is also certain that an investigator might make a large number of experiments before he suspected that (A) is not universally true when those words are used in that sense. Indeed, if the events which he investigated were such as can be 'seen,' he would need the most delicate instruments to detect a discrepancy. If, however, the events were such as can be 'heard,' he would find out the mistake more easily. He would find that if one observer is situated between the places where two events occur, while another is situated to one side of both of them, then events which were heard simul to one observer would not be heard simul to the other, if the events took place a considerable distance apart. When he found out his mistake two courses would be open to him; he might deny that (A) was true, or he might change the meaning of the words 'simultaneously with' so as to make (A) true. As a matter of fact, men of science have adopted the second alternative and have invented a new and very complicated meaning for 'simultaneously with.' They imagine that they possess a large number of clocks, similar readings of which occur simul, when the clocks are very near the observer comparing them, and that these clocks are distributed so that there is one very near each event which is to be investigated. They suppose that when an event is observed, the reading on the clock near that event is observed also by the same means; then two events P and Q are defined to be simultaneous if the reading on the clock near P which is simul with P is the same as the reading on the clock near Q which is simul with Q. As the result of experiment they now find that two events which are simultaneous to one observer are simultaneous to any other.

The procedure necessary to apply this definition is very cumbrous in practice and hence investigations have been made to discover whether it is not possible by means of some rule to deduce whether or no two events are simultaneous from the relations between the two readings on a single clock near the observer which happen simul with these events. Such a rule has been found; its form depends on the means used in observing the events; we shall assume that the propagation of light through a vacuum is used for this purpose. Then the following rule is found to give conclusions as to simultaneity concordant with those derived from the definition:—If P happens at a distance r_P simul with the reading t_P on the clock near the observer, and Q happens at a distance r_Q simul with the reading t_Q on that clock, then P and Q are simultaneous if $t_P - r_P/c = t_Q - r_Q/c$. If we call the function $t - r/c$ the 'time at which an event happens at r,' then (A) states that the time at which an event happens is independent of the observer; experiment shows that, if the 'time at which an event happens' is defined in this way, then this statement is true. It follows from the definition that the time is the reading on the clock near the event which is simul with the event.

Now still more accurate experiments suggest that, even with this definition of 'simultaneous with,' (A) is not universally true; that it is true only when all the events happen in systems which are relatively at rest. We have therefore again two alternatives, to abandon (A) or to redefine 'simultaneously with' so as to make (A) true. The Principle of Relativity enables us to find a suitable new definition. Let us define the term 'isochronic' by the statement that two events are isochronic if the reading on a clock near P and at rest relative to P which is simul with P is the same as the reading on a clock near Q and at rest relative to Q which is simul with Q, the clocks being such that equal readings are always simul when the clocks are compared by an observer near and at rest relative to both of them. Then the Principle of Relativity states that (A) is true even for systems which are moving relatively to each other if 'isochronic' is substituted for 'simultaneous.' Further it enables us to deduce from the times at which P and Q happen (that is from the readings on the clocks near them but at rest relative to the observer) whether or no P and Q are isochronic; it states that if t_P, t_Q are the times at which P and Q happen, x_P, x_Q the distances from the observer at which they happen, v_P, v_Q the velocities relative to the observer of the systems in which P and Q happen, then P and Q are isochronic if

$$\left(1 - \frac{v_P^2}{c^2}\right)^{-\frac{1}{2}} \left(t_P - \frac{v_P x_P}{c^2}\right) = \left(1 - \frac{v_Q^2}{c^2}\right)^{-\frac{1}{2}} \left(t_Q - \frac{v_Q x_Q}{c^2}\right).$$

If we call the function $\beta\left(t-\frac{vx}{c^2}\right)$ the 'era' at which an event happens at x and t in a system moving relatively to the observer with a velocity v, then (A) may be stated in the form that the era at which an event happens is independent of the observer. The era, of course, is also the reading on the clock near and at rest relative to the event which is simul with the event.

I trust the object of this long disquisition is clear to the reader. It is suggested that, since the change in our ideas which led to the change in the meaning of 'simultaneous with' from 'simul' to its present sense is precisely similar to the change in our ideas introduced by the Principle of Relativity, the second change ought to cause no perturbation of mind to those who accept the first so tranquilly that they do not appear to recognise that it has occurred. The discoveries leading to the Principle of Relativity do not render it necessary to abandon the hope of defining a 'time' valid for all observers, any more than do the discoveries leading to the belief in the finite velocity of propagation of light; the newer discovery, like the older, only makes it necessary to change somewhat the definition of that 'time.'

Let us now consider another proposition (B):—

'The length of a bar is the same by whatever observer it is observed.'

Let us first inquire what is meant by the length of a bar. We suppose the observer provided with a rod on which are made certain marks numbered consecutively; it is found possible to distribute these marks along the rod so that the following proposition is true:—If one end (P) of any bar is made to coincide with the pth mark on the rod, and the other end (Q) of that bar with the qth mark, then $a=q-p$ is determined only by the bar and not by p. a is called the length of the bar.

This definition in its present form is not applicable to the case where the bar is moving relatively to the rod. In this case an indefinite number of p's and an indefinite number of q's may be observed, and an indefinite number of values for the length of the bar could be obtained, unless we fix by further definition which p is to be subtracted from which q in order to obtain the length of the bar. The definition which is commonly adopted is that, if the coincidence of P with the pth mark is simultaneous with the coincidence of Q with the qth mark, then the length of the bar is $q-p$.

From the Principle of Relativity we can easily show that, if this definition is adopted, the length of a bar as determined by any one observer is independent of the particular value of p chosen by him; so far then, this definition is convenient. But it also follows necessarily from the same principle that the length of the bar, as determined by different observers by reference to a rod which is at rest relatively to one of them, may be different. For, if the coincidence of P with the pth mark is simultaneous with that of Q with the qth mark to one observer, these events will not be simultaneous to any other observer moving relatively to the first. It is easy to prove that, if the bar is moving uniformly with velocity v relative to the rod and appears to have to an observer at rest relative to the rod a length a, then to an observer at rest relative to the bar the bar will appear to have a length βa; in this statement it must be remembered that both observers determine the length with reference to the same rod, but that their definitions of simultaneity refer, as stated above, to different clocks.

It may occur to the reader that this difficulty might possibly be avoided and a definition of length obtained which would give the same results when applied to the same bar by all observers, if in the definition given 'isochronic' were substituted for 'simultaneous.' Unfortunately this procedure is impossible, because our definition of isochronic assumes that the event to which it refers happens in a system in a specified state of motion relatively to the observer; the event with which we are here concerned, the coincidence of P with a mark on the rod, does not 'happen in one system,' but is a relation between two systems relatively in motion. The term 'isochronic' has no meaning applied to such an event. But if, in our definition of length, we substitute 'isochronic' for 'simultaneous' and at the same time insert a provision that the measurement shall be made when the rod is at rest relatively to the bar, then, with this revised definition, the length of the bar will turn out to be the same by whatever observer it is observed.

Once more then we see that (B) is true in some cases and not in others merely because the definition of length which we ordinarily apply, selecting it out of an innumerable number of equally possible alternatives, is inconvenient. If we wish (B) to remain true, we have only to change our definition of length; and moreover the change in this definition which is necessary to make (B) true is precisely that which follows naturally from the change in the definition of 'simultaneous with' which was necessary to make (A) true.

Very brief notice may also be taken of a third proposition (C):—

'If v_P is the velocity of a particle P relative to an observer, v_Q the velocity of a particle Q relative to him, then w, the velocity of Q relative to P, is given by $w = v_Q - v_P$.'

According to (259) this proposition is not universally true, for

$$w = \frac{v_Q - v_P}{1 - \frac{v_Q v_P}{c^2}};$$

but here again the truth or falsity of the proposition depends entirely upon a definition. In (259) the quantity which we termed the velocity of Q relative to P, was defined as the velocity of Q relative to P as determined by an observer at rest relative to P. (259) then states a relation between the observation made by this observer and certain observations made by another observer relative to whom P is moving. Now a little reflection will convince anyone that it is not self-evident, apart from all experience, that there must be a certain relation between the observations of two different observers; surely it is remarkable that experience shows that any relation at all can be discovered. (259) can only be tested by comparing the observations made by two different observers who have been travelling relatively to each other and subsequently meet at relative rest. The experiment has never been tried when the velocity v_P is so great that a decision between (C) and (259) could be expected. (C) appears at first sight self-evident, because we are so accustomed to finding that it is true, that we have come to define 'the velocity of Q relative to P' as the quantity obtained by subtracting from the velocity of Q relatively to the observer the velocity of P relative to him; and, of course, if this definition is adopted, (C) is necessarily true. But, if the definition is adopted, the Principle of Relativity does not deny (C); it only denies that this definition leads to the same values as the other.

If the reader is now convinced that there is nothing paradoxical in the denial of propositions (A), (B), (C), he will probably find no difficulty in solving any of the apparent contradictions which will be found among the other consequences of the Principle of Relativity. When any proposition appears paradoxical, he must inquire whether there is not implied in the definition of the terms used in stating it a proposition inconsistent with that which is stated; if such inconsistency is discovered he must re-define his terms so as to remove it and will then find that all appearance of paradox has vanished[1].

Minkowski's treatment.

139. Objections to the Principle of Relativity of the second class are based on the ground that it is artificial and complex. Such objections can only be removed by showing that it is less artificial and less complex than some proposition which is regarded by the objector as a model of simplicity and directness. The brilliant exposition of Minkowski shows that, viewed in the right light, the Principle of Relativity is actually simpler than the Principle of Moving Axes.

As we have seen, the Principle of Moving Axes states that we can deduce from the law $f'(x', y', z', t') = 0$ observed by S' the law $f(x, y, z, t) = 0$ observed by S by substituting the quantities $(x - vt, y, z, t)$ for (x', y', z', t'). But when the Principle is used in dynamics another assertion is added, namely that, if the law is one stating the motion of material particles according to the principles of Newtonian dynamics, then the function f obtained by the substitution will be the same as f'. This proposition is often stated in the form that the laws of any dynamical system

---

[1] These considerations are necessary to resolve the confusions which are raised by those authors who attempt to introduce a distinction between 'what really happens' and 'what appears to happen.' This distinction has never been the source of anything but confusion of thought when applied to scientific propositions. Cf. *Phil. Mag.* 21. 513—515, 1911.

will be unchanged in form if they are transferred from one set of axes to another set, moving relatively to the first with a uniform velocity.

Now it is well known that there is another way in which we can change our axes without altering the form of laws referred to them; we may change from one set of rectangular axes to any other rectangular set and we may displace our origin through any distance. These two kinds of permissible transference are usually regarded as quite independent and unconnected; one kind can be effected without the other. That the first kind is permissible is regarded as due to the fact that the laws of a dynamical system can be expressed as differential equations of the second order with regard to the time; that the second kind is permissible is regarded as due to the fact, quite unrelated to the first fact, that the only function of the coordinates of two particles (x_1, y_1, z_1) and (x_2, y_2, z_2) which occurs in the laws is

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2;$$

so that any such transference of axes which leaves this function unchanged is permissible. For the sake of brevity, changes of origin will be left out of account; the condition for permissible transferences of the second kind is then that $x_1^2 + y_1^2 + z_1^2$ should be unchanged. This condition is equivalent geometrically to the rotation of rectangular axes through any angle; to say that it must be fulfilled is the same thing as saying that if we change our axis of x from OX to OX', we must change our axes of y and z from OY, OZ to OY', OZ', where a certain relation must hold between the angles XOX', YOY', ZOZ'.

Now let us suppose that we represent the complete motion of a particle on a four-dimensional diagram[1], taking three mutually perpendicular lines as our axes of x, y, z and a fourth line, perpendicular to all three, as our axis of t. Then a straight line in this diagram will represent the motion of a particle which is moving relatively to the observer with a uniform velocity; and, in particular, the line which is taken as the axis of t represents the 'motion' of the particle which is always at rest relative to and at

---

[1] Since few people think readily in four dimensions the reader will do well to imagine a simple case in which the motions of all particles are confined to a single line; the diagram then reduces to two dimensions, x, t.

zero distance from the observer. The second kind of change of axes corresponds to the choice of new straight lines for the axes of x, y, z; the first kind of change (a change to axes moving uniformly relatively to the original axes) corresponds to the choice of a new straight line for the axis of t. The Principle of Moving Axes states that, if a new line is chosen for the axis of x, certain other new lines must be chosen for the axes of y, z, so that the condition that $x^2+y^2+z^2$ is unaltered is fulfilled; but it states also that if a new axis of x is chosen, there is no need to choose a new axis of t, and that, if a new axis of t is chosen, there is no need to choose a new axis of x, in order to preserve the form of the laws.

On the other hand, Minkowski points out that what the Principle of Relativity states is simply this, that if we choose a new axis of t we must also choose a new axis of x, just as if we choose a new axis of x we must choose a new axis of y, and that the condition which must be fulfilled in choosing the new axes of x, y, z to suit a new axis of t is that $x^2+y^2+z^2-c^2t^2$ must remain unchanged; the two forms of transference are not unconnected, but are special forms, applicable to special cases, of a single more general transference from one set of axes to another, which can be effected without changing the form of the fundamental laws. For, according to the Principle of Relativity, the laws which must be taken as fundamental[1] are those expressed by (263). The assumption, made in § 135, that these laws are valid for all observers moving relatively to each other with any uniform velocities, is equivalent to the assumption that the laws are unchanged in form by the transformation (258). Since, according to (258), $x'^2+y'^2+z'^2-c^2t'^2=x^2+y^2+z^2-c^2t^2$, the transformation is equivalent to a transference to new axes, such that $x^2+y^2+z^2-c^2t^2$ remains unchanged. By such a transference, then, the form of the fundamental laws is unaltered.

Again the condition that $x^2+y^2+z^2-c^2t^2$ must be unchanged

---

[1] If it is assumed that the laws of the motion of a particle are also unchanged by the transformation (258) and thus the Principle of Relativity made yet more analogous to the Principle of Moving Axes, laws slightly different from those of Newtonian mechanics must be adopted. These are the laws of the 'New Dynamics' to which reference is made in § 137.

is, like the old condition that $x^2+y^2+z^2$ must be unchanged, capable of a geometrical interpretation. Let us suppose that all motions are confined to one line, so that we have only two axes, those of x and t; and let us draw, for all values of $t>0$, the curve $c^2t^2-x^2=1$. This curve is a hyperbola of which the asymptotes are the lines $ct \pm x=0$. Then the condition that $x^2-c^2t^2$ is to be unchanged by our change of axes means that, if we choose as a new axis of t a line drawn from the origin and cutting the hyperbola at P, we must choose for the new axis of x the line through the origin parallel to the tangent to the hyperbola at P.

But what if the line which we propose to take as the new axis of t does not cut the hyperbola? We must remember that the only lines which we shall want to take as new axes of t are those which represent the motion of a particle moving uniformly relatively to the observer. A line which lies outside the asymptotes of the hyperbola, and so never cuts it, corresponds to a particle moving uniformly with a velocity greater than c. The Principle of Relativity asserts that there are no such bodies and therefore that the question does not arise; the velocity c is simply that velocity which is not less than the velocity of any material particle relative to any observer. The Principle of Moving Axes asserts (without any evidence) that this velocity is infinite; if we put $c=\infty$, the hyperbola degenerates into the axis of x and the changes of axes permissible according to that principle become identical with those permissible according to the Principle of Relativity. According to this view the Principle of Relativity differs from the Principle of Moving Axes in the assertion that the greatest possible velocity of one particle relatively to another is not infinity, but the velocity of light. If the reader will bear in mind that we cannot obtain information about distant events by any influence which travels faster than light and will consider the observations which would be possible if there were bodies travelling relatively to him with a greater velocity (the distinctions which we now make between past and future would be absolutely impossible), I think he will accept Minkowski's conclusions; he will agree that the assumption concerning the limiting relative velocity of material particles made by the Principle of Relativity is a great deal simpler than that made by the Principle of Moving Axes and that, if in the past we had discovered Minkowski's method of regarding the permissible transferences of axes and of showing that the two kinds of permissible transference are intimately connected, we should long ago have arrived at the Principle of Relativity and should now regard it as the simplest and most obvious of propositions.

Faraday Lines.

140. One further objection may be considered. Someone may say:—'Your comparison of the change in the meaning of a word from "simul" to "simultaneous with" with the change in the meaning from "simultaneous with" to "isochronic" has left one important feature out of account. I find no difficulty in accepting the rule which you give for determining from observations of events which happen simul those which happen simultaneously, because I can see a physical explanation of the rule; all that it means is that light travels with the velocity c. You have offered me no similar physical explanation of the rule by which I am to determine from events which are simultaneous those which are isochronic; can you give me such an explanation?' 'No,' I reply, 'I cannot; and for a very good reason. If you will think what you mean by a physical explanation, you will find that you mean one which assumes that the Principle of Moving Axes is true. Now I say that that Principle is not true—that is the essence of my contention; and I cannot possibly explain a proposition to you if at the outset you insist on my assuming a proposition which is inconsistent with that I am to explain. Your difficulty arises from the fact that you cannot get out of your head the idea that the Principle of Moving Axes is true; once you realise its falsity, you will no more ask me for a "physical explanation" of why (A) is not true in all cases than you now ask me for a physical explanation of why it is true in some cases.'

But a legitimate question may be asked about the bearing of the Principle of Relativity on the conception of light as propagated along Faraday lines, which has been employed so freely in this volume. It is one of the consequences of that Principle that an event which happens at a distance r from me at a time (*not* at an 'era') t reaches me at a time $t + r/c$ whether or no the source of light is moving relatively to me. Now it would seem that, if the light is propagated along the Faraday lines attached to the source and if, as we have assumed, these lines are carried along by the source in its motion relatively to the observer, the velocity of the light relatively to the observer must vary with the motion of the source relative to him. This difficulty can be avoided if the view is taken (and this view is on general grounds more reasonable) that the disturbances which are observed by the observer are propagated along the lines attached to the observer's system and not along those attached to the source; we may adopt this view because the considerations of § 7 forced us to the conclusion that the lines attached to one charge do not end on another charge but are quite independent of those attached to any other charge. According to this view the disturbances received by the observer from two sources, both at the same point, one of which is at rest relative to the observer and the other in motion, are both propagated along the same line, namely that line attached to the observer's system which passes through the point occupied by the sources; and this line is at rest relatively to the observer. The velocity relative to the observer with which the two disturbances are propagated is the same, but the difference in the relative motion of the sources relative to the line may make the form of the disturbance received from one source different from that received from the other.

The difficulties raised on p. 279 will be found to vanish if Fig. 5 is interpreted afresh according to this view. It will be found that the thick lines in that diagram must not now be regarded as giving the position of the line attached to the electron (for with the position of this line the observers at P and Q' have nothing to do); the direction of the thick line at such a point as P must now be regarded as giving the direction at the time t of that line, attached to the system of the observer at P, which, at the moment when the electron was at O, lay in the direction PO. This interpretation is, of course, rather more complex, but it does not altogether deprive the conception of Faraday lines of its value as a means for visualising electromagnetic phenomena. It must be remembered, of course, that we have used the conception for this purpose only and have made no serious attempt to prove by its use any of the propositions stated; such proofs, like those based on any other conception, would have to be revised in the light of the Principle of Relativity.

Light and Gravitation.

141. The Principle of Relativity enables us to deduce from the laws discovered by an observer S', who is at rest relatively to the system observed, the laws discovered by S who is moving uniformly relatively to that system. The question has been raised recently whether it can provide a similar deduction if the motion of S relatively to the system observed is not uniform but uniformly accelerated. The importance of this question appears at once. Suppose that S is on the earth and the system which he observes, which includes another observer S', is falling towards the earth. Then the system R is uniformly accelerated relatively to S; if we can find (1) the laws observed by S' and (2) a means of transferring those laws into laws observed by S, we can find the laws which will be observed by S in the falling system.

If the acceleration of S' relative to S is so small that the relative velocity of S and S' does not change appreciably during the time necessary for S' to make an experiment (this time is that required for light to travel from any one part of R to any other), then the second part of this problem can be solved. For suppose that S' makes an experiment at era t_1 and finds a law $f_1'(x', y', z', t_1) = 0$, then at the same era, according to the principle of relativity, S finds a law $f_1(x, y, z, t_1, v_1) = 0$, where v_1 is the relative velocity of S and S' at the era t_1. At a subsequent era t_2, S' finds a law $f_2'(x_1', y_1', z_1', t_2) = 0$ and S a law $f_2(x, y, z, t_2, v_2) = 0$. Now v_1 and v_2 are known functions of t_1 and t_2; accordingly the two laws $f_1 = 0$ and $f_2 = 0$, observed by S at the eras t_1 and t_2, can be combined by suitable algebraical treatment into one law, valid for all eras. This law will be the law observed by S in the falling system R.

But what are the laws observed by S', who is falling with R? we have no direct experimental knowledge, for hitherto observers falling freely relatively to the earth have never been in a position to make accurate experiments. When we were dealing with uniform motion, we assumed that S' might always be made physically independent of S in the sense that the observations made on R by S' were the same as those which he or S would make, if R were at rest relatively to S[1]. Einstein extends this assumption to the case of uniformly accelerated motion; he assumes that the observations made by S' on R are the same as those which would be made by S on R, if R were at rest relative to S and there were no gravitational action. For the earlier assumption there is good experimental evidence; for the later there is none, but there is equally no evidence against it.

By this method Einstein has deduced some very remarkable results. He shows that the weight of a body must vary with the amount of energy which it contains in the same way as the mass; and he shows also that c, the velocity of light determined by an observer who is accelerated relatively to the source of light, will not be the same as c_0, the velocity determined by an observer who is moving uniformly relatively to that source, but that

$$c = c_0 (1 + \Phi/c^2) \quad\text{.....................}(277),$$

where Φ is the potential of the gravitational field producing the acceleration. For the earth's field at its surface Φ/c^2 is only 2×10^{-6}, so that no experiments such as we can make could hope to distinguish between c and c_0. But, on the other hand, the conclusion appears that, since the velocity of light varies with the gravitational potential, light travelling through a region in which that potential is not constant should not travel in a straight line; in particular the light from a star, if it passes near the sun on its way to the earth, should not travel in a straight line, and stars which appear to us near the sun but on opposite sides of it should seem displaced from their normal positions. The magnitude of

---

[1] It is this proposition, that the observations made by an observer who is at rest relatively to the system which he is observing can be the same whether or no he is moving uniformly relative to some other observer, which was originally called by Einstein the Principle of Relativity. The name was used because the proposition states that the observations of S' are determined only by his velocity *relative to R*. The name has now generally been extended to include the assertion of the validity of the transformation (258); this proposition implies besides the original Principle of Relativity the assumption that the velocity of light determined by any observer is the same whatever the motion of the source relative to him. This assumption Einstein terms the Principle of the Constancy of the Velocity of Light.

this displacement should be such as to be capable of detection; so far no experiments have confirmed this prediction, but the observations made are probably not sufficiently accurate.

These deductions are then so far entirely unconfirmed. But should subsequent research prove them false our science will be in a great difficulty. Either we shall be forced to deny the universal validity of the transformation of the Principle of Relativity, or we shall have to abandon the assumption that the system R may be physically independent of the observer S, even when it is accelerated relatively to him. Whichever of the alternatives is adopted, the theory of moving systems will again be beset by all the difficulties which surrounded it a few years ago (see Appendix I) but have been removed by Einstein's brilliant theory[1]. It has been urged recently by Abraham that we shall be in as great a predicament if the conclusions are confirmed. He points out that the transformations of the Principle of Relativity necessarily lead to the conclusion that the velocity of light is a constant for all systems and all observers. If (277) is true, this conclusion can only be accurately true if all the systems are in the same uniform gravitational field. Now this condition is not fulfilled by any actual system, for in any system there are gravitational forces between the parts of the system which give rise to a field which is not uniform. It would seem then that the Principle of Relativity is not applicable to any actual conditions!

As a formal objection to the universal validity of the Principle, this argument is indubitably correct. But all physical theories are open to the same objection. There is probably no physical law which is accurately true if applied to any actual system; but the laws are none the less useful. There are a great many systems applied to which they are very nearly true; we apply them to those systems and take account of the fact that they are not accurately applicable by introducing small 'corrections.' Thus

---

[1] It is perhaps well to remark that the conclusions of this paragraph are quite independent of any of the physical theories of gravitation which were discussed in § 125. We are not attempting to explain why bodies near the earth have a uniform acceleration relative to it any more than, earlier in the chapter, we were trying to explain, for example, why the sun seems to have a uniform velocity relative to the stars. We are accepting the fact of this acceleration and trying to discover how our experiments will be thereby affected.

the law that, if a bar A is found to have the same length as the bar B and B is found to have the same length as the bar C, then A will be found to have the same length as the bar C, is not accurately applicable to any actual bars, for—as the matter would be commonly expressed—no two bars are accurately at the same temperature. But we do not abandon the law as useless; we use it every day and, if we are doing extraordinarily accurate work, apply a correction for differences in temperature. In the course of applying this correction, we always assume again that the law is actually true and thereby lay ourselves open to the formal logical objection that we are assuming the law to be both true and false. What we are really doing is to proceed to discover the law applicable to the phenomena within the limits of experimental possibilities by a series of successive approximations.

It is quite probable that in the future there will have to be substituted for the transformation (258) some rather more complicated transformation; it is almost certain that we shall never arrive at a transformation which is applicable without any further correction to all moving systems. But I cannot believe that the method which underlies Einstein's treatment of the problems of this chapter will soon be abandoned; that method is to divide the region of physical phenomena into systems which, as a first approximation, can be regarded as unrelated and then to combine these isolated systems into one physical universe by correcting our laws in the light of the relations which are found to exist between them. It is by this method that physical sciences have progressed; the method of the Principle of Relativity is the method of all successful science.

REFERENCES.

The most complete and the most satisfactory statement of the assumptions and of the simpler consequences of the Principle of Relativity is that made by Einstein himself in *Jahrb. d. Radioakt.* 4. 411, 1908. Minkowski's treatment of the matter, with which the reader should certainly acquaint himself on account of its very great beauty, is contained in *Phys. Zeit.* 10. 104, 1909; that paper is also published separately as a pamphlet under the title *Raum und Zeit.* A recent book by Laue, called *Das Relativitätsprinzip*, gives a very full account of the Principle and its applications; the author pays more attention to the interesting problems in applied mathematics which occur than to the questions to which much the greater part of this chapter has been devoted.

The reader who after reading this chapter is still puzzled by the Principle will probably want to study some other attempt at expounding it, and he will probably prefer that it should be in his own language. During the past four years many such attempts have been published, chiefly in the semi-scientific journals such as *Science*. I refrain from mentioning any of them, except that of Bumstead (*Am. Journ. Sci.* 26. 493, 1908), because on the whole they appear to me better calculated to produce than to remove confusion of thought; others may think differently. But I am compelled to offer one warning; readers should beware of any writer who opens with the specious statement that he is going to treat the subject without any 'metaphysical subtleties.' If by 'metaphysical subtleties' he means any searching inquiry into the significance which is usually attached to the words 'time' and 'length' or into that which it is most convenient to attach to them, then any treatment of the problems of this chapter must introduce metaphysics, though it may be very bad metaphysics. If, however, by 'metaphysics' is meant the study of the conception of reality—and this is the usual sense—then metaphysics are certainly better kept out of scientific treatises. But experience shows, as common sense would suggest, that an author only states explicitly that he is not going to be metaphysical when he wishes to conceal from himself and his readers that he is going to assume tacitly a metaphysical proposition so absurd that it would not stand a moment's examination.

APPENDIX I

THE AETHER

THE developement of the modern theory of moving systems which is sketched in the last chapter is so recent that it can hardly be hoped that the student is free from the confusing ideas connected with the older theory which it displaces. In the first edition of this book a criticism of that theory occupied most of the corresponding chapter; in this edition it is, perhaps, desirable to repeat part of that criticism in order to show that the newer theory is really simpler and more comprehensible than the older; if there is a further edition, there will probably be no need to consider the latter at all.

The older theory arose from an attempt to answer the question, Where is the energy of the radiation localised during the interval between its leaving the source and arriving at the receiver? The answer given by the earliest forms of the undulatory theory of radiation was that the energy was localised in 'the aether.' Radiation was supposed to consist of some processes, probably of a dynamical nature, taking place in a system, the aether, which so far resembled material bodies that it could contain energy and could have a definite position or velocity relative to material bodies; portions of the aether were present at every 'point of space,' and these portions were at relative rest except when the disturbances constituting radiation passed through them. The velocity of a material body A relative to a material body B was considered to be equal to the difference between the velocities of A and B relative to the aether.

On the assumption that the aether through which light is propagated is a dynamical system physically independent of any material system and that the Principle of Moving Axes is applicable to such system, the Doppler effect for the difference in the frequency of the same source of light observed by two observers in relative motion, could be calculated. If certain assumptions were also made concerning the relation between the directions of a 'ray' and of the wave normal, the aberrational change in the direction of the light could also be deduced. But aberration, as well as the Fizeau effect, was treated more rigidly by investigating the refraction of light in a medium moving relatively to the aether. In its latest form (due chiefly to Lorentz), this treatment was based on two assumptions:—(1) that the theory of refraction given in Chapter II is correct, (2) that the fundamental equations of Chapter I are true if the axes to which they are referred are fixed in the aether, so that, if v is the velocity of a charged particle relative to the aether, (5) gives the force acting on it in the electric and magnetic field in the light. From these assumptions, values agreeing with experiment for the aberration and the Fizeau effect could be calculated.

In dealing with astronomical phenomena on these principles, it is necessary to assume that the difference in the velocity of the earth relative to the sun at different points of its orbit is equal to the difference in its velocity relative to the aether, that is to say, the sun must be regarded as moving uniformly relatively to the aether with a velocity which may be finite or zero. From the assumptions adequate to explain the Doppler effect, it follows that, if the velocity of light relative to an observer who is at rest relatively to the aether is c, the velocity of light relative to an observer who is moving relatively to the aether with a velocity v is $c+v$. Since the velocity of the earth relative to the aether in the direction of the earth's orbital motion differs from that in a direction perpendicular to that motion, the velocity of light relative to an observer on the earth should differ according as the light is travelling in the direction of the orbital motion or in a perpendicular direction. The famous experiment of Michelson and Morley showed that this consequence of the aether hypothesis is false and that the velocity of light (from a source at rest relative to the observer) is the same in whatever direction relative to the orbital motion of the earth the light travels.

There were two possible ways (not wholly exclusive of each other) of overcoming this difficulty. First, one or more of the assumptions introduced might be denied; second, some new assumption might be introduced. If we adopt the first alternative, we might deny either (*a*) that the equations of Chapter I are true referred to axes in the aether, or (*b*) that, if two material bodies are relatively at rest, they have the same velocity relative to the aether. The denial of (*b*) seems, doubtless, very startling or almost incomprehensible; nevertheless, it is easy to see that it is a perfectly natural consequence of the aether hypothesis. For we cannot deny (*a*), because (*a*) is not a proposition at all, but a mere definition. Unless (*a*) is true, we have no way whatever of finding out what is the velocity of any body relative to the aether; accordingly, (*a*) is simply equivalent to the following definition:—By a body at rest relatively to the aether we mean a body such that, if our axes of reference are fixed in it, the equations of Chapter I are true. All that the experiments on aberration and those of Michelson and Morley show is that, adopting this definition, the source of light which the observer is observing is at rest relative to the aether. When the stars are observed, the stars are at rest relative to the aether; when a terrestrial source is observed, that source is at rest relative to the aether; an observer has different velocities relative to the aether according as he observes one source or the other.

We might avoid this conclusion by stating some new definition of the aether before we state (*a*), which would then become a proposition and not a mere definition and would be capable of being denied. But there is no object in adopting this course; the conclusion at which we have arrived is quite simple and obvious, if we do not allow ourselves to be confused merely by an unfortunate choice of words. Our original idea of the aether was 'the body in which radiant energy is localised'; if, in that statement, we put 'bodies' for 'body' all our difficulties vanish; if the energy from different sources is localised in different bodies, then it is obvious that the velocity of an observer relative to the body in which the energy of one source is localised may be different from that relative to the body in which the energy from another source is localised. If we speak of 'aethers' and not 'the aether,' all our experiments prove is that the particular aether with which we are concerned in any case is that which is at rest relatively to the source and may be regarded as forming part of it.

This is the simple way out of the difficulties raised by the Michelson-Morley experiment. If from the beginning we had used a plural instead of a singular word to denote the system in which radiant energy is localised (or even a word which, like 'sheep,' might be either singular or plural), those difficulties would never have appeared. There has never been a better example of the danger of being deceived by an arbitrary choice of terminology.

However, physicists, not recognising the gratuitous assumption made in the use of the words 'the aether,' adopted the second alternative; they introduced new assumptions. They first redefined 'velocity relative to the aether' as a velocity differing from velocity relative to the stars only by the addition of a constant, the same for all sources and all observers. In order to explain the Michelson-Morley experiment they adopted another hypothesis, that known by the names of Lorentz and Fitzgerald, to the effect that the length of a body in motion relatively to the aether measured against a standard always at rest relatively to the aether was decreased in the ratio $1/\beta$. It could be shown that it was a result of the second hypothesis that no possible experiment could ever determine what was the velocity of any material system relative to the aether; the conception 'velocity relative to the aether' was one which could not possibly enter into the result of any experiment. It became certain that a theory would soon be proposed which would banish that conception altogether from physics; very shortly Einstein showed that the rules given by this complicated theory for deducing the relation between the laws observed by two observers who are in relative motion are exactly the same as those given by the Principle of Relativity; these rules, of course, involve only the relative velocity of the two observers and not their 'velocity relative to the aether.' Moreover, the form in which those rules are expressed by the Principle of Relativity is much simpler and more convenient than that in which they are given by the aether theory combined with the Lorentz-Fitzgerald hypothesis.

It may be well to point out exactly how the results of the Michelson-Morley experiment are introduced into the arguments based on the Principle of Relativity. They are introduced by the statement made in § 134 that, if R is a source of light, R is physically independent of S. If this assumption is combined with the transformation of the Principle of Moving Axes instead of with (258), the correct formula for the Doppler effect can be deduced, but the conclusion is reached that there should be no aberration. Accordingly, when the Principle of Moving Axes was believed to be true, it was necessary to deny this assumption; it is denied by the aether theory which says that R cannot be physically independent of S, if S is at rest relatively to the aether.

But how does the Principle of Relativity answer the question from which the aether hypothesis sprang? Where, according to that Principle, is the radiant energy localised? The answer has been given in § 140; the energy is localised in the system of the observer who observes it. But then (it may be asked), since an infinite number of observers can observe a finite amount of energy from the same source, the energy given out by the source must be infinite. By no means; the energy of a source is no more infinite because an infinite number of observers can observe it than the energy of a rifle bullet is infinite because an infinite number of observers can observe that it is moving relatively to them. But (the objector may go on) in the case of the bullet all the observers observe the same energy, whereas, according to you, two observers do not observe the same energy, because each observes only the energy which is localised in his own system. At this point I inquire what is meant by 'the same energy' and if the discussion were continued I should discover that the objection was based on the doctrine—which has hampered science ever since it originated in Greek philosophy—that there is 'a real energy' or 'a real body' which 'exists' independently of any observer and of which any observer only observes the 'appearance.' So far as I can understand this doctrine at all, I dissent from it; I deny that this conception of an energy which is 'the same' for all observers has any significance. An observer A observes that the source sends him an amount of energy a and he finds that the energy he receives from the source is the equal to that lost by the source; *B* observes that he gains and the source loses an energy *b*. Now I may be able to find (I can find with the help of the Principle of Relativity) a function of *a* and other properties determined by *A* which has the same value as the same function of *b* and the same properties determined by *B*. If I call this function 'the energy' then it is a function which is the same for all observers; in this sense the conception of 'the same energy' has a meaning and is very valuable. If we ask whether the energy *a* observed by *A* is the 'same energy' as *b* observed by *B*, then, if 'the same energy' means the function just mentioned, of course *a* and *b* are not the same energy. But at the same time it is not legitimate to add *a* and *b*, to call them the 'total energy' and raise difficulties because the total energy is not finite. *a* and *b* are not quantities of the same kind; one is a quantity observed by *A* and one a quantity observed by *B*; they cannot be added to make one total quantity of the same nature as both, any more than one cow and one sheep can be added to make two horses.

The reader will find similar remarks on the same subjects in *Phil. Mag.* 19. 181, 1910 and *Phys. Zeit.* 13. 120, 1912.

APPENDIX II

ABERRATION

In the *Monthly Notices of the R. A. S.* 69. 403, 1909, Prof. Turner raises an interesting question concerning aberration, which is worthy the attention of physicists as well as of astronomers. The question may be stated thus:—

If we want to calculate from the observed position (or rather direction) of a star its true position, the theory of optics indicates that we must introduce a correction depending on u, the velocity of the earth. According to the theory based on the conception of the aether, u is the velocity of the earth relative to the aether; what is the velocity u according to the Principle of Relativity? Should it not be the velocity of the earth relative to the star?

Perhaps we should be inclined at first to answer the last question affirmatively and should be confirmed in this opinion by observing that, according to (260), the Doppler effect and aberration are different aspects of the same phenomenon. Now the correction necessary to determine the true frequency of the light emitted by the star from its observed frequency depends, as observations on double stars show, on the relative velocity of the earth and the star. Ought not the aberrational correction to depend on the same velocity?

But if we accept the view suggested, we at once arrive at a serious difficulty. Consider a close double star, or even a 'spectroscopic double.' The pair of stars appear in the same direction, but they are moving relatively to the earth with different velocities, and therefore different corrections (the difference amounting to several seconds of arc) should be applied to obtain their true directions. The stars which appear close together must really be an enormous distance apart; how are we then to explain their orbits? Conversely, the parts of a rotating star, being all really in the same direction, ought to appear spread out into a band several seconds long!

Prof. Turner gives a solution of this difficulty, but it will be more convenient for us to look at it somewhat differently.

If we accept the Principle of Relativity at all, we must accept it consistently. What according to that Principle do we mean by the true direction of a star? By the true frequency of the light, we mean the frequency observed by an observer at rest relative to the star, for this frequency is determined wholly by the nature of the star and not by the circumstances of the observer. But by the true direction of a star we do not mean the direction observed by an observer at rest relative to the star, for, if we meant that, it would be absurd to speak of the 'motion,' or change of direction, of the star. A clue to what we mean by the 'true position' of a star is given by the fact that we regard it as impossible that the 'true orbit' should not obey Kepler's Laws. A little inquiry will show that what we mean by the true direction of the star is simply the direction relative to an observer to whom the orbit of the pair of stars appears to obey Kepler's Laws. Such an observer is one at rest relative to the centre of mass of the pair; by the true position of the star we mean its position as observed by such an observer, and the aberrational correction must be such as to transform the direction observed by the observer on the earth to that observed by this observer. Now the observations are consistent with the view that the sun is at rest (or in uniform motion) relative to the centre of mass of the pair, and hence the required aberrational correction is that which transforms the direction observed on the earth to that observed on the sun. This correction of course depends on the motion of the earth relative to the sun. The correction which is actually applied in astronomical practice is thus justified by the Principle of Relativity.

Moreover these considerations are precisely the same as those based on the aether hypothesis which led to the view that the aberrational correction ought to be determined by the velocity of the earth relative to the sun. For why did the holders of that hypothesis believe that the aether was at rest (or in uniform motion) relative to the sun, and that the position of a star observed by an observer at rest relative to the aether was the 'true position' of the star? Simply because the aether was believed to be at 'absolute rest,' at rest, that is, relative to the absolute axes which are such that motions referred to them obey the laws of Newtonian dynamics, and because observations showed that relative to these axes the sun was either at rest or in uniform motion. Whether we adopt the aether hypothesis or the relativity hypothesis, by the 'true position' of the star we mean simply its position observed by an observer who is at rest relative to axes such that motions referred to them obey the laws of Newtonian dynamics.

Two further points may be noted. In calculating the aberrational correction a mechanical analogy has sometimes been employed, usually that of the apparent direction of a shot fired from one ship and striking another moving relatively to it. The difficulty which has been raised shows that this analogy is inadequate. Mechanical analogies are not easy to apply to problems of aberration, even if we adopt the analogy of § 140 and consider the 'aether' to be a system at rest relatively to the observer.

Second, the reader may feel that, though the 'true direction' is actually that observed by an observer at 'absolute rest' in the Newtonian sense, this is not what he means by the 'true direction'; by the true direction he means the direction in which the light has travelled from the star to the observer. After what has been said in the last chapter and Appendix I this conception need not be criticised in detail. The reader will find that, if he will consider now exactly what he means by the true direction in this sense and why he believes that the customary aberrational correction will give it, he will either have to conceive himself performing inconceivable experiments or make the assumption that the true position is that referred to an observer at 'absolute rest.'

NAME INDEX

See also Subject Index

SUBJECT INDEX

Numbers in heavy print denote definitions

CAMBRIDGE: PRINTED BY JOHN CLAY, M.A. AT THE UNIVERSITY PRESS.

CAMBRIDGE PHYSICAL SERIES

The Times.—"The Cambridge Physical Series...has the merit of being written by scholars who have taken the trouble to acquaint themselves with modern needs as well as with modern theories."

Volumetric Analysis. By A. J. BERRY, M.A., Fellow of Downing College, Cambridge. Demy 8vo. pp. viii + 138. 6*s.* 6*d.* net.

"Rich in practical teaching and skilfully reflecting the lessons of its writer's practical experience in his college laboratory, the book is also thoroughgoing on its theoretical side."—*Scotsman*

Modern Electrical Theory. By N. R. CAMPBELL, Sc.D. Second edition, largely rewritten. Demy 8vo. pp. xii + 400. 9*s.* net.

"The treatment throughout is admirably clear and readable, the arrangement is logical, and the work of different investigators is carefully considered and given due weight....Possibly the most striking feature is the completeness of the book, for it would be hard to find a section of the subject which has not received due attention."—*Cambridge Review*

Sound. An Elementary Text-Book for Schools and Colleges. By J. W. CAPSTICK, M.A., D.Sc., Fellow of Trinity College, Cambridge. Crown 8vo. pp. viii + 296. With 120 figures. 4*s.* 6*d.*

"An important addition to the *Cambridge Physical Series*....The latter portion is of especial value, and gives an interesting *rationale* of the principal orchestral instruments....Dr Capstick has succeeded in making a remarkably lucid exposition of his theme, and the book will, we have little doubt, promptly take its place as a standard text-book on the subject."—*Guardian*

Mechanics. By JOHN COX, M.A., F.R.S.C. Demy 8vo. pp. xiv + 332. With four plates and 148 figures. 9*s.* net.

"Prof. Cox has his eye specially upon mechanical principles, avoiding merely mathematical difficulties so far as that is fairly possible; he 'starts from real problems, as the subject started, showing how the great investigators attacked these problems, and introducing the leading concepts only as they arise necessarily and naturally in the course of solving them'; he 'brings out incidentally the points of philosophic interest and the methods of science'; he appeals constantly to experiment for verification, 'leading up to an experimental course limited to the most important practical applications,' and eventually embodying a good deal of matter not usually found in the elementary text-books; and he adds limited sets of carefully selected examples for exercise. Students that will not learn mechanics from this work, and be fired with interest in the subject, must be hopeless."—*Educational Times*

5000
11.15

CAMBRIDGE PHYSICAL SERIES

The Study of Chemical Composition. An Account of its Method and Historical Development, with illustrative quotations. By Ida Freund. Demy 8vo. pp. xvi+650. With 99 figures. 18*s*. net.

"The accomplished Staff Lecturer on Chemistry of Newnham College has admirably succeeded in her efforts to produce a book dealing with the historical development of theories regarding chemical composition....No part of the book is dull, and the student who starts on it will be led on to continue to the end, to his own great advantage. Many people have no time or opportunity to refer to a large mass of original papers, and the value of this book of careful selections carefully and cleverly put together will be very great....Students of physical science who wish to form a clear conception of the laws of chemical combination, of the meaning and accuracy of combining weights and combining volumes, of the doctrine of valency and of isomerism, and of the parts played by various investigators and authors in elucidating these and kindred matters cannot do better than refer to this book."—*Athenaeum*

Mechanics and Hydrostatics. An Elementary Text-book, Theoretical and Practical, for Colleges and Schools. By R. T. Glazebrook, C.B., M.A., F.R.S., Director of the National Physical Laboratory and Fellow of Trinity College, Cambridge. Crown 8vo. 6*s*.

Also in separate volumes

Part I. **Dynamics.** pp. xii+256. With 99 figures. 3*s*.
Part II. **Statics.** pp. viii+182. With 139 figures. 2*s*.
Part III. **Hydrostatics.** pp. x+216. With 98 figures. 2*s*.

"A very good book, which combines the theoretical and practical treatment of mechanics very happily....The discussion of force, momentum, and motion is consistent and clear; and the experiments described are rational and inexpensive. This treatment is calculated to give much clearer ideas on dynamical concepts than a purely mathematical course could possibly do."—*Journal of Education on* Dynamics

"A clearly-printed and well-arranged text-book of hydrostatics for colleges and schools....The descriptions are clearly written, and the exercises are numerous. Moreover, the treatment is experimental; so that altogether the book is calculated to give a good grasp of the fundamental principles of hydrostatics."—*Nature on* Hydrostatics

Heat and Light. An Elementary Text-book, Theoretical and Practical, for Colleges and Schools. By R. T. Glazebrook, C.B., M.A., F.R.S. Crown 8vo. 5*s*.

Also in separate volumes:

Heat. pp. xii+230. With 88 figures. 3*s*.
Light. pp. viii+210+vi. With 134 figures. 3*s*.

"The very able author of the treatise now before us is exceptionally well qualified to deal with the subject of theoretical and experimental physics, and we may at once say that he has succeeded in producing a class-book which deserves, and will doubtless receive, a full share of the patronage of our school and college authorities."—*Mechanical World*

CAMBRIDGE PHYSICAL SERIES

Electricity and Magnetism: an Elementary Text-book, Theoretical and Practical. By R. T. GLAZEBROOK, C.B., M.A., F.R.S. Crown 8vo. pp. vi + 440. 6*s.*

"As an elementary treatise on the laws of electricity and magnetism it leaves little to be desired. The explanations are clear, and the choice of experiments, intended to carry home these explanations to the mind of the student, is admirable. We have no doubt that teachers of the subject will find the volume of great use."—*Engineering*

Photo-Electricity. By A. LL. HUGHES, D.Sc., B.A., Assistant Professor of Physics in the Rice Institute, Houston, Texas. Demy 8vo. pp. viii + 144. With 40 text-figures. 6*s.* net.

"The book goes very thoroughly into all modern experimental research related to its subject....The whole question is important and may have far-reaching results, and we consider the author is to be heartily congratulated on the lucid manner in which he has dealt with it."—*Electrical Review*

The Electron Theory of Matter. By O. W. RICHARDSON, D.Sc., F.R.S., Wheatstone Professor of Physics at King's College, London. Demy 8vo. pp. viii + 612. With 58 figures. 18*s.* net.

"This book will serve as an admirably compact and yet very complete account of the subject, leading directly to many points where research, theoretical and experimental, is proceeding."—*Science Progress*

A Treatise on the Theory of Alternating Currents. By ALEXANDER RUSSELL, M.A., M.I.E.E., Principal of Faraday House. In two volumes. Demy 8vo. Vol. I. pp. xiv + 534. Second edition. 15*s.* net. Vol. II. pp. xii + 488. 12*s.* net.

"It reveals the thorough mastery which the author has of the subject, and the capable reader will find much that is of value. Recent contributions are carefully digested and generally elucidated, and, on the whole we may consider the book an up-to-date treatment of the alternating-current theory."—*Electrical Review*

Experimental Elasticity. A Manual for the Laboratory. By G. F. C. SEARLE, Sc.D., F.R.S., University Lecturer in Experimental Physics. Demy 8vo. pp. xvi + 187. 5*s.* net.

"The author...gives us in this book a collection of the experiments which he has found suitable for University students in their second and third years' preparation for the Natural Sciences Tripos....The book can be strongly recommended to all teachers and to University students. It forms an excellent monograph on elementary elasticity."—*Guardian*

Experimental Harmonic Motion. By G. F. C. SEARLE, Sc.D., F.R.S. Demy 8vo. pp. x + 92. With 32 text figures. 4*s.* 6*d.* net.

"The volume is a useful addition to the Cambridge Physical Series, and it can be well recommended to students of mathematical physics."
School World

CAMBRIDGE PHYSICAL SERIES

Air Currents and the Laws of Ventilation. Lectures on the Physics of the Ventilation of Buildings. By W. N. SHAW, Sc.D., F.R.S., Fellow of Emmanuel College, Director of the Meteorological Office. Demy 8vo. pp. xii + 94. 3*s*. net.

"The present volume by Dr Shaw is most welcome, because it draws attention forcibly to the physics of the ventilated space, and deals with the whole subject in a thoroughly scientific manner....The treatment of the subject is quite original....The whole volume is highly suggestive, and constitutes a real advance in the study of this important subject."—*Engineering*

Conduction of Electricity through Gases. By Sir J. J. THOMSON, O.M., D.Sc., LL.D., Ph.D., F.R.S., Fellow of Trinity College, Cambridge, and Cavendish Professor of Experimental Physics. Second edition enlarged and partly re-written. Demy 8vo. pp. viii + 678. 16*s*.

"It is difficult to think of a single branch of the physical sciences in which these advances are not of fundamental importance. The physicist sees the relations between electricity and matter laid bare in a manner hardly hoped for hitherto....The workers in the field of science are to-day reaping an unparalleled harvest, and we may congratulate ourselves that in this field at least we more than hold our own among the nations of the world."—*Times* (on the First Edition)

A Treatise on the Theory of Solution, including the Phenomena of Electrolysis. By WILLIAM CECIL DAMPIER WHETHAM, Sc.D., F.R.S., Fellow of Trinity College, Cambridge. Demy 8vo. pp. x + 488. 10*s*. net. [*New edition in preparation*

The Theory of Experimental Electricity. By W. C. D. WHETHAM, Sc.D., F.R.S. Second edition. Demy 8vo. pp. xii + 340. 8*s*. net.

"This book is certain to be heartily welcomed by all those who are engaged in the teaching of theoretical electricity in our University Colleges. We have no hesitation in recommending the book to all teachers and students of electricity."—*Athenaeum* (on the First Edition)

Experimental Physics. A Text-book of Mechanics, Heat, Sound and Light. By H. A. WILSON, Professor of Physics in the Rice Institute, Houston, Texas, U.S.A. Demy 8vo. pp. viii + 406. With 235 figures. 10*s*. net.

"This book is intended as a text-book for use in connection with a course of experimental lectures on mechanics, properties of matter, heat, sound and light. No previous knowledge of physics is assumed, but nevertheless the book is primarily intended for a first year college course, and the majority of the students attending such a course have studied elementary physics at school."—*From the Preface*

CAMBRIDGE UNIVERSITY PRESS
C. F. CLAY, MANAGER
London: FETTER LANE
Edinburgh: 100, PRINCES STREET
ALSO
London: H. K. LEWIS, 136, GOWER STREET, W.C.