UNIVERSAL
LIBRARY
OU_160262
UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 530.9/G75H Accession No. 31063

Author Wilson, W

Title Hundred years of Physics

This book should be returned on or before the date last marked below.

100 *Years Series*

A HUNDRED YEARS OF PHYSICS

100 Years Series

A HUNDRED YEARS OF PSYCHOLOGY. By J. C. Flugel

A HUNDRED YEARS OF ASTRONOMY. By R. L. Waterfield

A HUNDRED YEARS OF MUSIC. By Gerald Abraham

A HUNDRED YEARS OF THE BRITISH EMPIRE. By A. P. Newton

A HUNDRED YEARS OF ENGLISH GOVERNMENT. By K. B. Smellie

A HUNDRED YEARS OF ARCHAEOLOGY. By Glyn Daniel

A HUNDRED YEARS OF CHEMISTRY. By Alexander Findlay

A HUNDRED YEARS OF ENGLISH LITERATURE. By W. Sherard Vines

A HUNDRED YEARS OF ECONOMIC DEVELOPMENT IN GT. BRITAIN. By G. P. Jones and A. G. Pool

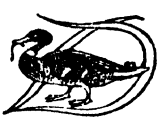
A Hundred Years of Physics

by

WILLIAM WILSON

F.R.S., PH.D., D.SC.

Emeritus Professor of Physics
University of London



GERALD DUCKWORTH & CO. LTD.
3 HENRIETTA STREET, LONDON, W.C.2

First published 1950

All rights reserved

PRINTED IN GREAT BRITAIN
BY WESTERN PRINTING SERVICES LTD., BRISTOL

PREFACE

To understand most of this book the reader needs very little previous knowledge of the science of physics and a very slight mathematical equipment. It is hardly an exaggeration to say that mathematics is absent in most of the chapters. I have been forced to introduce a little in dealing with parts of relativity and quantum mechanics; but even that is of the most elementary kind.

The work is intended, as its title suggests, to present as clearly as possible the essential features in the development of the science during the past hundred years, or—to be more precise—since the beginning of Queen Victoria's reign. The discovery of the cardinal facts of physics and the growth of its unifying theoretical structure are described and discussed together. There is much in it that is addressed to both students and teachers of science and a little of it, here and there, even to the great pundits of physics.

Figures IV–1, VII–5 and XV–1 are borrowed from my *Theoretical Physics*, and I wish to express my thanks to Messrs. Methuen & Co. for permitting me to use them. My thanks are also due to Messrs. Edward Arnold & Co. for Figure VII–1 which appeared in my article on "The Origin and Nature of Wave Mechanics" in *Science Progress*, 32, p. 209 (1937).

W.W.

April 1950

CONTENTS

Chapter		Page
	PROLOGUE	9
I.	THERMAL PHENOMENA	15
II.	ANOTHER ASPECT OF THERMAL PHENOMENA	31
III.	AN EXPERIMENTAL MISCELLANY	46
IV.	THE STATISTICAL THEORY OF HEAT I. KINETIC THEORY OF GASES	62
V.	THE STATISTICAL THEORY OF HEAT II. STATISTICAL MECHANICS	80
VI.	LIGHT	86
VII.	MORE ABOUT LIGHT	106
VIII.	INTEGRATION OF ELECTROMAGNETISM	118
IX.	MAXWELL'S ELECTROMAGNETIC THEORY	137
X.	THE AETHER	148
XI.	NEWTONIAN AND SPECIAL RELATIVITY	155
XII.	EINSTEIN'S GENERAL THEORY	174
XIII.	RADIANT HEAT AND THE EMERGENCE OF THE QUANTUM THEORY	182
XIV.	CONDUCTION OF ELECTRICITY AND ASSOCIATED PHENOMENA	203
XV.	MISCELLANEOUS SUBJECTS	212
XVI.	ATOMIC STRUCTURE	228
XVII.	THE CORRESPONDENCE PRINCIPLE AND ITS APPLICATION	244
XVIII.	QUANTUM MECHANICS	249
XIX.	RADIOACTIVITY AND COSMIC RADIATION	263
XX.	THE ATOMIC NUCLEUS	276
XXI.	ASTROPHYSICS AND COSMOLOGICAL SPECULATION	282
	EPILOGUE	298
	INDEX	309

Dedicated to my former
students at Bedford College
and at King's College, London

PROLOGUE

AT the beginning of Victoria's reign physical and chemical phenomena, that is to say, the phenomena of the inorganic world, were regarded as the manifestation of the motions of material bodies, ranging from John Dalton's atoms to planets and stars, under the influence of forces. In one set of phenomena the forces were gravitational, in another electrical, in still another chemical, and so on. The physicists and chemists of that time had fairly clear notions of what they meant by these terms. They were philosophers—though perhaps not in the special and rather restricted present-day sense of the term. For them the investigator of natural phenomena was a philosopher. Physical science was usually called natural philosophy. Their geometry was Euclidean and the inverse square law of force had for them an almost *a priori* character, and indeed in a Euclidean world it would seem to be as inevitable as it is for the illumination of a surface by a point source of light. The mechanical principles to which the motions of bodies, under the influence of forces, were found to conform were those laid down by Sir Isaac Newton (1642–1727) in the great work *Philosophiae naturalis principia mathematica*, usually referred to as the *Principia*. No physicist of 1837 could contemplate any other attitude to physical phenomena. Lord Kelvin (William Thomson, 1824–1907), one of the greatest scientific geniuses of his century, clung to it tenaciously till the end of his life. The mechanics of Newton impressed the early Victorians as something final which, like Euclid's geometry, was settled for all time. It had acquired an almost perfect form from the progressive contributions of Maclaurin, de Maupertuis, Euler, d'Alembert and Lagrange in the eighteenth century and of Sir William Hamilton and C. G. J. Jacobi in the earlier half of the nineteenth century, and fitted perfectly into its Euclidean frame.

The word "geometry" has now two senses: for the physicist it is a basic part of physical theory, with axioms, or premisses, the

truth of which is guaranteed by observation—though doubtless some of his early Victorian predecessors were convinced, like Immanuel Kant, that these axioms had an apodictic *a priori* certainty which rendered experimental verification unnecessary. For the pure mathematician on the other hand, no question arises as to whether the axioms or premisses of *his* geometry are true or not. That is to say he is quite unconcerned about any such question as: Do they truthfully describe or report features of the physical world? He is only concerned with their logical coherence and that of the structure built up on them.

At the beginning of the epoch in which we are interested this mechanical or matter and motion science was already highly developed and had great achievements to its credit. We cannot regard the fundamental principles on which it was based as crude and immature or as something on which we are entitled to look down with an air of modern superiority. As these principles left the hands of Hamilton and Jacobi they had acquired the character of a finished product, like the geometry of the Euclidean continuum in which they were framed, and indeed, had the essentials of the Hamiltonian expression of the old mechanics been better appreciated when the strange problem of black body radiation appeared, wave mechanics might have emerged directly from it rather than Planck's tentative form of quantum theory. Hamilton's principle, like that of de Maupertuis, which preceded it, is intimately related to Fermat's principle in geometrical optics. Indeed the statement of the latter can be extended so as to take precisely the same form as Hamilton's principle and in this form it provides us, not only with the shape of a ray of light (the path which the light takes), but also the group velocity with which the luminous disturbance is propagated. This indeed is expressed by a formula which is the exact analogue of one of Hamilton's canonical equations. The wave mechanics of de Broglie and Schrödinger emerges when the analogy between geometrical optics and mechanics is widened so as to become an analogy between optics (in the widest sense of the term) and mechanics. Indeed Hamilton's mechanics is a limiting case of quantum, i.e. wave mechanics (from the small-scale side), and it requires extraordinarily little amplification to make it identical with the mechanics of the special theory of relativity. It is valid (to a close approximation) for a vast field

of phenomena and its structural elegance gives it the impress of a supreme work of art.

The outstanding achievements of the physicists, or natural philosophers, before the beginning of our period of approximately one hundred years may be summarized as follows:

The principles of mechanics had been successfully applied to a great variety of phenomena. The constant of gravitation and the mean density of the earth had been determined by Henry Cavendish (1731–1810) with an accuracy which has hardly been surpassed by the latest measurements. Good methods had been devised for the determination of the intensity of gravity (the familiar g of the text-books) by Kater (1818) and especially by Bessel (1826). Elastic phenomena (i.e. phenomena within the scope of Hooke's law) and the propagation of waves in elastic media were well understood. Fresnel's wonderful theory of light was based on the premiss that light was a manifestation of wave motion in an elastic medium (aether) and it has been an object of admiration ever since. He arrived at the familiar sine and tangent formulae for the reflection of light at the interface between two transparent and isotropic media and the associated formulae for refraction and was thus enabled to explain Brewster's law connecting the refractive index with the tangent of the polarizing angle. He constructed, on the basis of this theory, the device now called Fresnel's rhomb for producing circularly polarized light and developed a complete theory of the propagation of light in crystalline media. Nearly all his theoretical results are accepted to-day, though we have been compelled to re-erect them on new foundations.

Quite respectable optical instruments, such as microscopes and telescopes, including astronomical telescopes, existed. Captain James Cook carried quite excellent chronometers with him round the world. The speed with which light travels in interstellar space had been computed, with an impressive approach to accuracy, from astronomical observations (Roemer, 1695, and Bradley, 1727). Wollaston (1802) had noticed that the sun's spectrum was crossed by dark lines. Fraunhofer rediscovered them (1814) and measured their wave-lengths, using wire gratings. Thomas Young had carried out his famous interference experiments and provided a new explanation of the earlier experiment of Newton (Newton's rings). Prévost enun-

ciated his theory of exchanges near the close of the eighteenth century. Fresnel had amplified Huygens' principle and applied it successfully to diffraction phenomena.

Electrostatic phenomena were nearly as well known and understood as they are now. Cavendish had tested the inverse square law by what appears to be the best method. He also arrived (before Faraday) at the notions of electrical capacity and specific inductive capacity (dielectric constant). George Green's famous Essay appeared in 1828. Oersted discovered the basic phenomenon of electromagnetism in 1820 and Ampère's great contribution followed soon after. G. S. Ohm discovered the law which is named after him in 1827. Faraday's law of induced currents was discovered in 1831 and independently by the American Henry, a fact which is commemorated by the adoption of his name for a unit of inductance. Wheatstone's bridge was invented (by someone named Christie, it is believed) in 1833, and in this year appeared the great work of Gauss on magnetism including, among other things, his method of investigating the law of force between magnetic point poles. Thermo-electric phenomena (with the exception of the Thomson effect) were known (Seebeck, 1821, and Peltier, 1834), but not understood. The oscillating currents associated with circuits containing capacity (capacitance) and inductance were still unknown, or at least not yet understood, and the complex exponential, so beloved by our present-day electrical engineers, had not yet acquired much importance in electrical theory. Faraday had enunciated his laws of electrolysis and thus opened the way to the discovery of charged ions. Primary cells for generating electric currents were in existence. Daniell's cell, the prototype of the standard cells of later times, appeared in 1836. The tangent galvanometer was invented by Pouillet in 1837.

The phenomena of heat were however very imperfectly understood and indeed misunderstood. It was still generally believed that heat (caloric) was some kind of imponderable fluid and that it was strictly conserved. Lord Kelvin (William Thomson) held this view till about 1850, though it had long been questioned by, among others, Count Rumford (Benjamin Thompson), Humphry Davy and Thomas Young. We are still reminded of it by the term "*latent* heat".

On the other hand Sadi Carnot (1796–1832) had discovered

the great principle which is still named after him and which is, in effect, the second law of thermodynamics (1824), and Clapeyron (1834) had deduced remarkable consequences from it which were later confirmed experimentally. Carnot's principle and its mathematical formulation by Clapeyron must certainly be reckoned among the greatest achievements in the history of science. The principle cannot be fully applied, nor even rightly appreciated, except in association with the first law of thermodynamics; but this too appears to have been foreseen by Carnot (some short time after the publication of his great work, *Réflexions sur la Puissance Motrice du Feu*) and the mechanical equivalent computed by him—probably from the specific heats of air—as we learn from the manuscript notes deposited with the French Academy in 1878 by his surviving brother, H. Carnot.

The law of Charles (discovered independently by John Dalton and Gay-Lussac) and that of Avogadro were known; but the intimate relationship between them was not yet appreciated. Daniel Bernoulli had shown that Boyle's law could be deduced from a simple form of kinetic theory (1738) and he appreciated that the product pv is proportional to the *vis viva* of the molecules.

Gay-Lussac had carried out (in 1807) the classical experiment in which a gas (air) was allowed to expand into a previously exhausted space, with the same result as Joule obtained about forty years later. Dulong and Petit had discovered (1817) the law, named after them, concerning the atomic heats of solid elements. Some gases had been liquefied, notably by Faraday, and half a century before Dr. Thomas Andrews began his beautiful investigation of the isothermals of carbon dioxide, the existence of a critical state had been demonstrated experimentally by Cagniard de la Tour. Jean Baptiste Fourier published his great work on the theory of heat in 1822 and had developed that wonderful mathematical tool known as Fourier's series (or expansion) as long ago as 1807.

The velocity of dilatational (sound) waves in air had been measured by Newton and the mathematical expression which he found for it, namely, the square root of the quotient of the elasticity of the air by its density, had received its correct interpretation from Laplace.

One of the triumphs for Newtonian mechanics was the pre-

diction by J. C. Adams, and independently by the Frenchman Leverrier, of the existence and position of a hitherto unobserved planet (Neptune). This was actually observed, in the predicted position, by the Berlin astronomer Galle (1846).

Sir William Herschel, in endeavouring to determine stellar distances by parallactic observations, was thereby led to the discovery of double stars, and F. W. Bessel actually carried out the first successful observations of parallax; so that the distances of the nearer stars were beginning to be estimated over a hundred years ago. It is noteworthy that Herschel arrived at the conclusion that the stars occupied a region shaped rather like a biscuit or flat bun. He also observed extra-galactic nebulae and may indeed have suspected them to lie outside our galaxy.

Such, broadly speaking, was the state of physical science at the beginning of the period of one hundred years, or thereabouts, with which this book is concerned.

CHAPTER ONE

THERMAL PHENOMENA

EQUIVALENCE OF HEAT AND WORK

THE development of the science of heat, or thermodynamics, provides some of the most fascinating chapters in the history of physics. In 1837 F. Mohr published views on the nature of heat, very similar to Carnot's later view, in Liebig's *Annalen.** The following passages are from this paper: "Heat is thus not a particular kind of matter, but an oscillatory motion of the smallest parts of bodies." "Heat appears as force [*Kraft*]." It should be explained that the words "force" and "*Kraft*" were frequently used in earlier times as synonymous with "work", although the term "work" (*travail*) with its present-day sense had been introduced into mechanics by Poncelet as early as 1826.

The first serious publication in which the equivalence of heat and work is insisted on and an estimate of the mechanical equivalent of heat given, was by Julius Robert Mayer, a medical practitioner of Heilbronn in Bavaria.†

The question immediately arises: What is meant by saying that heat and work are equivalent? It means first of all that heat is not conserved; but may be used up (or destroyed) in doing work and may be generated when work is done, as for example in overcoming a frictional resistance. It means further that when work is done in generating heat, a quite definite amount of work—the MECHANICAL EQUIVALENT OF HEAT—has to be done to produce one unit of heat, and conversely, when heat is used up solely in doing work, precisely this amount of work is done at the expense of one unit of heat.

Work is measured by the product of force (as used in the sense

* Translated by P. G. Tait for the *Philosophical Magazine*, Ser. 5, Vol. 2, p. 110 (1876).

† "Bemerkungen über die Kräfte der unbelebten Natur," *Ann. der Chemie und Pharmacie*, 42 (1842). Mayer's collected works on Heat were published in 1867 in a book entitled *Die Mechanik der Wärme*, in Stuttgart (J. G. Cotta). A second edition appeared in 1874.

in which a weight is called a force) and distance. Thus we might describe a particular weight (a force directed downwards) as a weight (or force) of 5 lb. and when it descends a distance of, for example, 20 feet, we should say work is done by it and take as its measure $5 \times 20 = 100$ foot-pounds. We have of course introduced other and more suitable units since Mayer's time. Indeed systematic units were already gradually coming into use—units related to the structure of the quantities which they measure.

Mayer computed the mechanical equivalent from the specific heats of air at constant pressure and at constant volume. For the former he used 0·267* given by Delaroche and Bérard, and the specific heat at constant volume he calculated from the ratio of the two, 1·421, given by Dulong. These numbers are appreciably in error and Mayer's result is consequently rather rough. He found 365 (kilogramme weight × metres) per kilogramme calorie—a kilogramme calorie being the amount of heat required to raise the temperature of a kilogramme of water 1°C.—whereas the correct result is near 427. He did the calculation in the right way and evidently understood it; though his knowledge and understanding of the physical science of his day, and of mechanics especially, appears to have been very slender. The calculation in question, which will be more fully discussed later, was based on the result of the experiment of Gay-Lussac.

Mayer's conviction of the equivalence of heat and work, though confirmed by experiments, was due chiefly to quasi-metaphysical fancies. *Kräfte*, i.e. forces, are causes, said Mayer, and therefore the fundamental principle *causa aequat effectum* must apply to them. Hence the equivalence of heat and work, one caused by the other. He contrasts two kinds of causes, namely, force and matter (*Kraft und Stoff*), or, as we should say, in our terminology, energy and matter. This attitude to physical phenomena was rather common a hundred years ago. Even Joule writes: "... but we are not left with this argument alone [i.e. the argument that the powers with which God has endowed matter cannot be destroyed or created by human agency], decisive as it must be to every unprejudiced mind."†

As early as 1839 James Prescott Joule, a brewer of Salford, began a long sequence of experiments in most of which heat was

* The specific heat of water is the unit.

† Joule's *Collected Papers*, 1, p. 268.

generated in all sorts of ways, as a consequence of work done. This sequence of experiments constituted one of the most important contributions ever made to physical science. The best known of them is the famous water-churning experiment. The water was contained in a cylindrical vessel (calorimeter), well protected against loss or gain of heat to or from the surroundings, and churned by a coaxial spindle to which paddles were fixed radially. These passed through openings in vanes which were fixed to the calorimeter and extended radially from its cylindrical wall towards the spindle in the centre and were just wide enough to let the paddles pass through. The purpose of this arrangement was to cause as much friction as possible between the moving paddles and the water. Without some such device the water would have rotated with the paddles, much as does the tea in a teacup with the stirring spoon.* The rotation of the spindle was effected by falling weights suspended by cords which passed over fixed pulleys and round a cylinder fixed to the spindle and coaxial with it. The measure of the work done was the product of each weight and the distance it fell (lb. weight × feet), multiplied of course by the number of descents. Some portion of the work was done, of course, in overcoming the friction of the pulleys, i.e. in producing heat in the bearings. As this failed to get measured in the calorimeter, Joule carried out subsidiary experiments to determine what portion of the work was done in producing it. In fact he experimented with the apparatus when the spindle and paddles were disconnected and subtracted the work done in this way from the total work done in the main experiment. He also guarded against error due to the small loss or gain of heat to or from the surroundings and even estimated the rather small amount of work done to produce the sound emitted by the apparatus. A certain part of the work was done in giving the weights the velocity they acquired in descending. This was duly taken into account. He measured the heat generated in the calorimeter by

$$\left\{\begin{array}{l}\text{Mass of water in lb.+water}\\ \text{equivalent of containing vessel}\end{array}\right\} \times \left\{\begin{array}{l}\text{Rise in temperature}\\ \text{in Fahrenheit degrees}\end{array}\right\}$$

and found for the equivalent the value 772.

* The apparatus was, and probably still is, in the museum at South Kensington.

This type of experiment was repeated by the great American physicist, Henry Augustus Rowland (1848–1901),* with one very great improvement in the mode of carrying it out. Instead of causing the paddles to revolve and keeping the calorimeter fixed, the calorimeter was made to revolve at a steady rate by a small steam engine, the rate being adjusted so that the weights and paddles, which tended to hinder the rotation, remained at rest. With this procedure no heat was generated in the bearings of the pulleys and the whole system of weights, pulleys, spindle and paddles remained at rest. He also caused the heat to be generated at a greater rate, thus reducing the gain or loss of heat from or to the surroundings. Joule himself thought of this in his later experiments. Rowland expressed his temperatures on the ideal gas scale (i.e. Kelvin's work scale) while Joule merely gave the readings of his mercury thermometers. The American's mean result was, in terms of the particular units he used, 427·52 at 14·6°C. He gave the value of the gravitational intensity at Baltimore, where his experiments were conducted, as 9·8005 metres/sec.^{2}, i.e. 980·05 cm./sec.^{2},† which makes Rowland's result almost $4{\cdot}19 \times 10^7$ ergs per 14·6° small calorie.‡

After correcting Joule's thermometer, Rowland found, as the best result he could arrive at from Joule's reports of his experiments, 426·75 as compared with Rowland's own 427·52; impressive evidence of Joule's accuracy.§ Joule carried out an immense variety of experiments which cannot be described here. He generated heat by compressing air, by electric currents in

* *Proc. American Academy*, 15, pp. 72–200 (1880), and 16, pp. 38–45 (1881).

† This is an important datum. Had the experiments been carried out near the equator where the gravitational intensity, g, is only about 978 cm./sec.^{2}, the *weight* of a kilogramme would have been less in proportion and would have meant something different from what it did in Baltimore.

‡ Rowland's units were the kilogramme weight, metre and centigrade degree.

§ I have spoken of Rowland *correcting* Joule's thermometer. This expression may mislead. In Joule's day thermometry was very imperfectly understood and his temperature readings were in terms of the arbitrary scale of his mercury thermometer, though these were taken with meticulous care.

wires, and he used other liquids besides water in the churning type of experiment. In the course of his experiments with electric currents he discovered the law, named after him, that the rate of generation of heat in a wire is proportional to its resistance and to the square of the strength of the current flowing in it.

Two other investigators deserve mention: A. Colding and G. A. Hirn. The former, who was the chief engineer of the city of Copenhagen during the time when Joule was in full activity, convinced himself of the equivalence of heat and work on the ground that, as he expressed it, forces are *spiritual* in their nature and therefore cannot be destroyed. He claimed to have received his inspiration from d'Alembert's *Traité de Dynamique*,* published a hundred years earlier, and he made an estimate, not a very good one, of the mechanical equivalent, on the basis of experiments. G. A. Hirn was also an engineer and belonged to Colmar in Alsace. He carried out a number of interesting experiments, in one of which he estimated the work done, and the corresponding amount of heat consumed, by a steam engine. Neither Colding nor Hirn achieved the extraordinary precision of Joule; but their work confirmed that of the Englishman and was important in helping to bring about the general recognition of the equivalence of heat and work.

PRINCIPLE OF CONSERVATION OF ENERGY

Dr. Thomas Young suggested the term ENERGY for *one half* of the *vis viva*, that is for what we now term KINETIC ENERGY (*vis viva* was defined to be (mass) $\times$ (velocity)^2); hence the term POTENTIAL ENERGY, suggested by Rankine for anything which was not *in fact* energy (i.e. kinetic energy), but convertible into it and therefore *potentially* energy. The notion of energy, in a sense near to the present-day one, began to emerge with the PRINCIPLE OF CONSERVATION OF FORCE (*Prinzip der Erhaltung der Kraft*) of Hermann von Helmholtz (1821–1894);† but the term "energy" in this larger sense was first used by Rankine (1854)

* One of the greatest contributions to mechanics during the eighteenth century and a fine source of inspiration.

† H. Helmholtz: *Wissenschaftliche Abhandlungen*, 1, p. 12.

and adopted by William Thomson (Lord Kelvin).* Helmholtz was influenced partly by the failure of all efforts to produce a *perpetuum mobile* and partly by the belief that the phenomena of heat, like all inorganic phenomena, were mechanical. The *perpetuum mobile* is any device which, if for the moment we assume its possibility, can do work for nothing—a steam engine, for example, which could carry on without the supply of fuel. There is a story of a Viennese clockmaker who claimed to have devised a *perpetuum mobile*. He made a clock which continued to go and to be a satisfactory time-keeper while seemingly receiving no assistance from without. It was never wound up; had no visible connection with anything outside it and no permanent changes occurred within it but only those which constituted the motions of its mechanism. It was of course only a rather good simulation of something which did work for nothing. It did in fact receive assistance from outside, since it was kept going by the upward and downward motion of a barometric column of mercury which was harnessed to its mechanism.

The mechanical view suggested to von Helmholtz that the various forms of "force", for example heat, or the energy (if I may use the present-day term) exemplified by a charged condenser, were just forms, not then very obvious, of mechanical energy. Indeed heat was suspected to be identical with the kinetic energy of the small particles (molecules) of the hot body. The charged condenser, or Leyden jar, was thought to have the strain type of energy (potential energy) such as that of a compressed gas or a stretched spring. When the jar is discharged an equal or equivalent amount of heat, i.e. kinetic energy, is generated in its place. This attitude was reinforced in the mind of von Helmholtz by the study of purely mechanical problems, i.e. problems concerning the motion of material bodies under the influence of forces (planetary motion for example) and unassociated with thermal or other effects not obviously mechanical. In such problems, as von Helmholtz was aware, there is some-

* Macquorn Rankine: "Outlines of the Science of Energetics," *Edinburgh Journal* (2), 2, p. 120 (1855). I feel that Rankine has never received the full credit due to him in connection with the energy principle and its application in thermodynamics. No doubt he was rather overshadowed by the contemporary giants, Helmholtz, Thomson and Clausius.

thing which is conserved and which we now describe as the sum of the potential and kinetic energies of the mechanical system. It was not *quite* so obvious as this to von Helmholtz. This conserved mechanical something must have been presented to him originally in the form of a difference, $T-W$ ($T \equiv$ kinetic energy and $W \equiv$ force function); but he wrote W as $-V$ and thus had a conserved totality $T+V$.

Thus a fundamental principle was recognized as the outcome of the experimental labours of Joule, Hirn and others and the theoretical studies of von Helmholtz and many of his contemporaries: the PRINCIPLE OF CONSERVATION OF ENERGY. A simple illustration is provided by a freely falling body. While descending, work is done on it by the force which is called its weight. This is measured by the product of the mass of the body and its acceleration (in the present illustration it is the familiar g). When the body has descended through the vertical distance h, the amount of work done on it by its weight is therefore

$$mgh.$$

This measures the amount of POTENTIAL ENERGY the body has lost. When work is done *some* form of energy is always used up and this is measured by the work done at its expense. In place of it an equal amount of energy is produced elsewhere and frequently in some other form. In the present illustration a quantity of KINETIC ENERGY equal to

$$\tfrac{1}{2}mv^2$$

is produced and the principle of conservation is expressed by

$$\tfrac{1}{2}mv^2 = mgh,$$

or

$$\left\{\begin{array}{c}\text{Increment of}\\ \text{kinetic energy}\end{array}\right\} = \left\{\begin{array}{c}\text{Decrement of}\\ \text{potential energy}\end{array}\right\}$$

and therefore

Total energy remains constant.

When the body strikes the ground the kinetic energy is lost; but the equivalent quantity of heat appears in its place.

For a long time after the experiments of Joule *work* and *energy* were (rather naturally) identified. Indeed they still are by many (or most) physicists. This seems to be the wrong attitude. It is rather better to speak of energy as *capacity for*

doing work, but even this attitude is not quite the right one, since there are the strongest reasons for associating MASS with energy; so that energy is more substantial, as it were, than work or the capacity for doing work. We may say, however, that work is the measure of the energy used up in doing it, as it is also of the energy generated in consequence. In conclusion, energy is conserved and one of its most important forms is heat.

Heat is still frequently measured in terms of such units as the amount of heat required to raise a gramme of water 1°C. in temperature, or the amount of heat required to melt a gramme of ice, and before its relationship to work was known there was no other way of measuring heat. The significance of the EQUIVALENT (mechanical equivalent or Joule's equivalent) is now clear. It means the number of units of work that generate one (old style) unit of heat, or alternatively are done at the expense of one such unit.

THE FIRST LAW OF THERMODYNAMICS

In the form it takes when applied to *thermal* phenomena the energy principle is called the FIRST LAW OF THERMODYNAMICS, which is just the English equivalent of Clausius' *erster Hauptsatz der mechanischen Wärmetheorie.** Its first application was, naturally, to gases; since their properties are much simpler than those of liquids or solids.

To a first approximation the product of the pressure and the volume of a fixed quantity of a gas remains constant, so long as the temperature of the gas is kept constant (Boyle's law), so that

$$pv = \textit{function of } t, \qquad \text{(I—1)}$$

if t means the temperature. The precise nature of the function depends on the conventions and on the scale adopted for measuring the temperature, and here must follow some explanations in order that this sketch of the development of thermodynamics may be intelligible. It is convenient to use the product pv (pressure × volume) of a fixed quantity of *some chosen gas* (air for example) as a measure of the temperature of the air itself and

* R. Clausius: "Ueber die bewegende Krafte der Wärme und die Gesetze, welche sich daraus für die Wärmelehre selbst ableiten lassen," *Pogg. Ann.*, 79, pp. 368, 500 (1850).

of bodies which are in temperature equilibrium with it. And it seems desirable, *for a little while*, to ignore the fact that Boyle's law is not *strictly* accurate; or alternatively let us imagine we have chosen some rather ideal gas, instead of air, which obeys Boyle's law perfectly. We *define* the measure of temperature by

$$\left\{\begin{matrix}\text{Increment of}\\ \text{temperature}\end{matrix}\right\} = \left\{\begin{matrix}\text{Some conventional}\\ \text{positive constant}\end{matrix}\right\} \times \left\{\begin{matrix}\text{Increment}\\ \text{of } pv\end{matrix}\right\} \qquad \text{(I—2)}$$

It happens, very happily, that when we make this definition the basis of our temperature measurements, they are in quite close agreement with those made with almost any gas. This is merely an unusual way of stating the old law of Charles, Gay-Lussac and John Dalton. It amounts to the same thing as saying that, if fixed quantities of any two gases have the same value of pv at some given temperature, they will have the same value of pv at any other temperature. The law is of course not *exactly* true; but rather near the truth under moderate temperatures and pressures.

The definition (I—2) may be written

$$dt = A d(pv).$$

The letter d is the symbol for increment. This formula is obviously equivalent to

$$a + t = Apv$$

where a is a constant, positive or negative, as you like. It is usual to use the letter R to mean $1/A$, so we get

$$pv = R(a + t). \qquad \text{(I—3)}$$

R is called the gas constant for the quantity of gas involved and the constant a simply fixes the zero from which the temperature is measured. We shall often suppose it to be chosen so that the temperature, t, of melting ice is zero. When we do this a is very nearly equal to 273°C. Such an equation as (I—3) is called an EQUATION OF STATE. It is nearly true (Charles' law) for all gases and it is very convenient to write it in the form

$$pv = RT, \qquad \text{(I—3A)}$$

where T means the temperature as measured from a zero about 273° lower than the temperature of melting ice ($a + t = T$).

It is characteristic of an equation of state (like I—3) that if

we are told the values of any two of the variables of state, e.g. those of v and t, then the equation obviously enables us to find that of the remaining one, p.

Every material has an equation of state involving the temperature, pressure and volume, though it is seldom possible to give it a simple analytical expression; but it can always be expressed, with any desired approach to precision, by graphical methods.

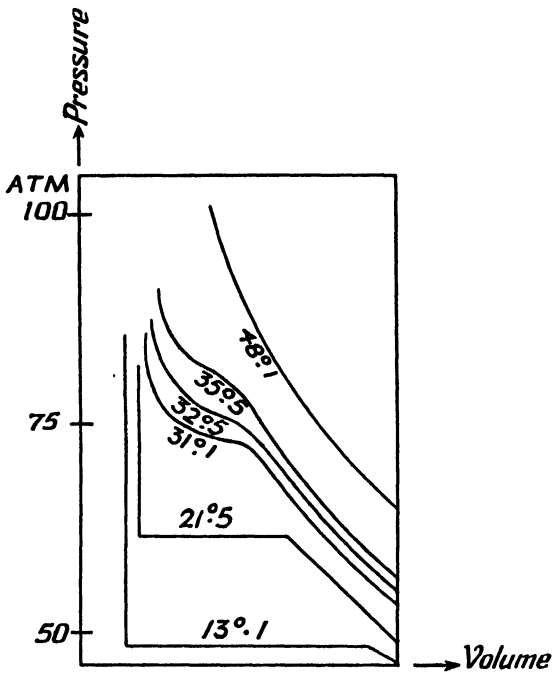


Fig. I—1

A good illustration of this is provided by the work of Thomas Andrews (1813–1885) who, in the years between 1863 and 1869, carried out some remarkable experiments on carbon dioxide to which I shall have occasion to refer again. What he did was to subject the carbon dioxide (liquid, vapour or gas) to some known (i.e. measured) temperature, and while it was maintained at this known constant temperature he measured the volumes it occupied at various measured pressures. He repeated this for a whole range of temperatures.

In this way he obtained a number of graphs (Fig. I—1). Each of them is called an ISOTHERMAL and collectively they constitute, in effect, an equation of state. If for example a particular pressure and a particular volume are specified, we can at once find the corresponding temperature. We simply draw, on the diagram, a horizontal line corresponding to the specified pressure and a vertical one corresponding to the specified volume. Their intersection will be on, or sufficiently near to, one of the pv curves (isothermals) to give the required temperature, provided enough of the isothermals have been constructed.

The energy in a given quantity of some material is called its INTERNAL ENERGY and is fixed or determined by its pressure, volume and temperature. It is in fact fixed by any two of these, since any two fix the value of the third one.

This statement needs amplification. It assumes that the only interference to which the material is subjected is due to some change in these variables. We might however imagine the material to be subjected, for example, to a magnetic field, in which case its internal energy might be altered, even though its pressure and volume remained unchanged. But on admitting such a new physical agency we should of course enlarge the equation of state to include, in addition to pressure, volume and temperature, at least the magnetic field intensity, or possibly the intensity of magnetization of the material, as a further variable of state.

In the special case of a gas it happens that its internal energy is determined *by its temperature alone*—very nearly. This is what the old experiment of Gay-Lussac indicated and what justified Mayer's way of computing the mechanical equivalent. Another characteristic of gases is the (approximate) constancy (at ordinary temperatures) of their specific heats at constant pressure and at constant volume, and it has long been usual to use the term IDEAL GAS or PERFECT GAS for a hypothetical body which conforms exactly to the following three laws:

(*a*) Boyle's law.

(*b*) The internal energy depends on the temperature only (Gay-Lussac and Joule).

(*c*) Constancy of specific heat at constant volume (Regnault and Clausius).

No apology is needed, I think, on account of its great historical and intrinsic importance, for describing Joule's form of Gay-

Lussac's experiment in some detail and for explaining how (*b*) is inferred from it. Two strong vessels, *A* and *B* (Fig. I—2), were enclosed in a calorimeter and connected by a tube provided with a stop-cock, so that *A* could be shut off from *B* when desired.

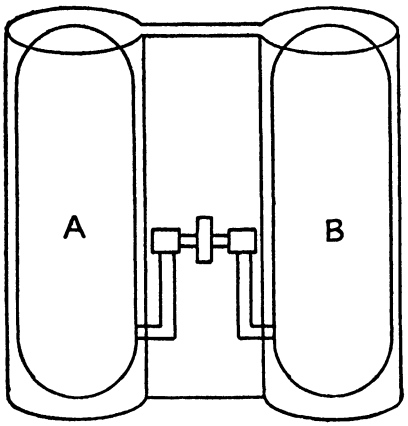


Fig. I—2

The calorimeter was shaped so as to fit as closely as possible round the vessels. This enabled the two vessels and the connecting tube to be completely immersed in a very small quantity of water, so as to make the experiment as sensitive as possible. Initially one of the vessels, *A*, was filled with gas (air) to a moderately high pressure (22 atmospheres) while the other one, *B*, was exhausted. The water in the calorimeter was stirred and the temperature taken with a sensitive mercury thermometer. The stop-cock separating *A* from *B* was then opened so that the air expanded into the previously exhausted *B*. Once again the water was stirred and the temperature taken. Joule could observe no change.* We can draw an inference from this experiment by applying the first law of thermodynamics, namely,

$$\Delta Q - \Delta W = \Delta U. \qquad \text{(I—4)}$$

The meaning of the symbol Δ may be described in the following way: generally I shall use the letter Q for the quantity of heat

* He did not omit to try the experiment with *A* and *B* in *separate* calorimeters. With this modified arrangement he was able to observe an appreciable lowering of the temperature of the water round *A* and a corresponding rise of that round *B*. The explanation is fairly obvious.

communicated to a body (or system of some sort) and it is convenient to regard it as communicated in successive small portions. Each of these is represented by ΔQ and in equation (I—4) we are assuming that it is expressed, or measured, in terms of the same unit as is work. This of course is justified by the equivalence of heat and work. Before the equivalence had been established quite different units were, very naturally, used to measure heat on the one hand and work on the other, and when the equivalence was experimentally demonstrated it was of interest and importance to determine the ratio of, say, the small calorie to the erg (i.e. the number of ergs equivalent to a small calorie—MECHANICAL EQUIVALENT OF HEAT). But of course the equivalence having once been established, the sensible thing is to drop the inconvenient equivalent and to measure both things in terms of one unit—e.g. the erg. Just as in the case of heat (Q or ΔQ), so in the case of the work done by a gas or other body in expanding against an opposing pressure we use a letter, W, for it and think of it (usually) as done in successive small amounts ΔW. Similar observations apply to U and ΔU. The former means the *internal energy* of the substance and the latter represents a small increment of the internal energy. The meaning of (I—4) is then as follows: a quantity of heat, ΔQ, has been given to a body and part of it used up in doing the work ΔW, involved in the expansion of the body against some external pressure. The balance left over represents the energy, ΔU, which stays in the body and constitutes an addition to its internal energy.

Sometimes (I—4) is written as

$$dQ - dW = dU.$$

This happens when it is convenient, or necessary, to deal with infinitesimally small quantities. It should be noted that on occasion work is done *on a body* instead of *by it*. When this happens we continue to use (I—4); but we associate a minus sign with ΔW. Thus if 1/10 of a unit of work were done *on* the body, we should say $\Delta W = -1/10$; just as one might say on receiving £1 that one had spent *minus* £1.

Now in Joule's experiment ΔQ was zero, because the water in the calorimeter surrounding the gas did not change in temperature and did not consequently give any heat to, or receive

any heat from, the expanding gas. Further ΔW was also zero, because the gas expanded into empty space. There was nothing on which it could exert any force or pressure. Thus ΔQ and ΔW being both zero, equation (I—4) tells us that

$$\Delta U = 0.$$

Moreover the gas cannot have changed in temperature, since this would have shown itself by a change in the temperature of the surrounding water. The result of the experiment was that the gas occupied a larger volume at the same temperature and without any change in its internal energy. The conclusion must therefore be that, so long as the temperature of a gas continues unchanged, so does its internal energy, even if the volume of the gas should change. In fact the internal energy can only change when the temperature changes. It is, as we say, a function of the temperature only. This is what Joule found for air. A more sensitive type of experiment does however indicate a *slight* dependence of the internal energies of gases on their volumes, even when their temperatures remain unchanged.

We can now understand Mayer's calculation. What it is desired to find is the number of units of work—ergs for example —which are equivalent, shall we suppose, to the unit of heat represented by the quantity of heat which raises the temperature of one gramme of water from 14·5°C. to 15·5°C. Let us first examine the difference between the specific heat at constant pressure and that at constant volume, while continuing to measure heat in work units—in ergs for example. Imagine a small quantity of heat given to a gramme of air at constant pressure while its temperature rises some small amount, dt. The heat communicated to the air *must be*, by definition of the specific heat, c_p,

$$c_p \times dt.$$

Therefore

$$dQ = c_p \times dt$$

and the work the gas (or air) does in expanding is measured by multiplying the pressure, p, by the small volume increment, dv, so that

$$dW = p \times dv.$$

We have therefore

$$c \times dt - p \times dv = dU.$$

But dU depends *only* on dt, as we have learned from the result of the experiment of Joule and Gay-Lussac. If we were to raise the temperature by the same amount, dt, while keeping the *volume* constant, the increment, dU, of the internal energy would be just the same; it is indifferent to volume changes. So we have

$$c_v \times dt = dU$$

and

$$c_p \times dt - p \times dv = c_v \times dt.$$

But clearly, at constant pressure, by (I—3)

$$p \times dv = Rdt$$

and consequently

$$c_p \times dt - R \times dt = c_v \times dt$$

or

$$c_p - c_v = R. \qquad \text{(I—5)}$$

This is true when we measure heat in terms of the same unit as work. If now we were to measure heat, that is to say $c_p dt$ and $c_v dt$, in terms of the small calorie instead of the erg, then we should have to divide the right-hand side of (I—5) by the number of ergs corresponding to a calorie to keep the equation correct—by the mechanical equivalent J in fact—and so we get the equation

$$c_p - c_v = R/J. \qquad \text{(I—5A)}$$

Experiment gives for air

$$c_p = 0{\cdot}239,$$
$$c_v = 0{\cdot}169,$$

approximately as measured by the ordinary small calorie, i.e. the amount of heat which raises a gramme of water 1°C. The gas constant, R, for a gramme of air is equal, very nearly, to $2{\cdot}9 \times 10^6$ when expressed in terms of ergs per °C. So we find

$$0{\cdot}239 - 0{\cdot}169 = \frac{2{\cdot}9 \times 10^6}{J},$$

which yields for the mechanical equivalent, J, the value

$$4{\cdot}14 \times 10^7 \text{ ergs/cal.}$$

An important point should be noted here, the neglect of which in the past has led to much futile argument and misunderstanding. In the very simple calculation just given (Mayer's

calculation) dW has been taken to be correctly expressed by pdv. This means, in effect, that the expansion, dv, has been assumed to take place very slowly. It is only under this condition that the p in pdv is *the same* p as occurs in the equation of state (I—3). The reason for this is very simple. The pressure, p, as it appears in the equation of state is measured while the gas is *at rest*, and in the pdv the pressure, p, can only have the same meaning as the p in

$$pv = R(a + t)$$

if the expansion dv takes place so slowly that the gas may be regarded as practically at rest. We shall better appreciate the significance of the REVERSIBILITY, which is so prominent in certain thermodynamical arguments, if we bear this in mind.*

* *Vide* Carl Neumann: *Vorlesungen ueber die Mechanische Theorie der Wärme*, § 15, p. 43, and § 37, p. 110 (Leipzig, B. G. Teubner, 1875).

CHAPTER TWO

ANOTHER ASPECT OF THERMAL PHENOMENA

CARNOT'S PRINCIPLE

IN order to appreciate the great contributions which Rudolph Clausius (1822–1888) and William Thomson (Lord Kelvin) made to thermal science, it is necessary to study the work of Sadi Carnot (1796–1832) and E. Clapeyron (1799–1864), to whom the subsequent progress of thermodynamics owes so much. The former of these published, in 1824, a short treatise entitled *Réflexions sur la Puissance Motrice du Feu et sur les Machines propres à Développer cette Puissance.* The investigation described in this little book is purely theoretical and led to one of the most fundamental principles of physical science: CARNOT'S PRINCIPLE, which is effectively the SECOND LAW OF THERMODYNAMICS (Clausius' *zweiter Hauptsatz der mechanischen Wärmetheorie*). The principle emerged before the first law and some quite important and experimentally verifiable consequences can be derived from it without any use, or even knowledge, of the first law. It is an instance of a sound principle inferred from faulty premisses. Carnot's premisses were indeed generally approved by the men of science of his day, and though faulty (one of them at any rate) they do appear to necessitate the result he found. He was interested at first in the practical question of the EFFICIENCY of steam engines; but was inevitably drawn to the more entrancing problem of the efficiency of heat engines in general. He attacked this problem by the device of the REVERSIBLE CYCLE, which has continued to be the most prominent feature of text-books on thermodynamics ever since.

His picture of a heat engine was like that of a water mill. The heat descended, as it were, from a higher level (temperature) to a lower one and was *not* used up (according to the views of that day) any more than is the water which drives the mill wheel. On page 28 of his famous treatise we read: "La puissance motrice d'une chute d'eau dépend de sa hauteur et de la quan-

tité du liquide; la puissance motrice de la chaleur dépend aussi de la quantité de calorique employé, et de ce qu'on pourrait nommer, de ce que nous appellerons en effet la hauteur de sa chute, c'est à dire de la différence de température des corps entre lesquels se fait l'échange du calorique."

His proof of the principle is very simple. It is most easily presented in terms of the mill wheel analogy which was in Carnot's mind. We imagine two mill wheels, each taking water from the same upper level and rejecting all of it at a common lower level. Now imagine one (or both) of them to be exactly REVERSIBLE. That is to say, the wheel, if made to turn once round in the reverse sense, would take the same amount of water from the *lower* level and eject it at the upper level as would be transported from the upper level to the lower one if it turned once round in the direct sense. In fact, let us suppose that all the operations of the mill wheel are literally reversible and that nothing happens when it turns in one direction which is not undone when it turns in the other direction. This means, among other things, *complete absence of friction*, since friction cannot be undone. Let us suppose that, while in the direct operation of the mill wheel the work that it does consists usually in raising a weight, or weights, one of the wheels is arranged to drive the other (reversible one) backwards, in such a way that the same amount of water which descends in driving it is put back by the reversed one. If now the directly functioning mill wheel were able to do more work in one revolution, say, than is needed to drive the reversed one, we might use the combination to raise a weight. Thus we should have in this combination of two mill wheels a *perpetuum mobile*. The impossibility of such a thing was Carnot's chief premiss. The driving mill wheel cannot therefore do more work than is just sufficient to drive the reversed one. In fact its efficiency cannot be greater than that of the reversible wheel and it follows that all reversible mill wheels must have the *same efficiency* when they are working between the same water levels. We might take as a measure of efficiency the work done per gallon of water transported from the upper to the lower level.

All this applies, from Carnot's point of view, to heat engines. If they are reversible they must have the same efficiency when working between the same temperatures. The nature of the

particular working substance, for example, whether air, steam or what, is quite irrelevant. All reversible heat engines working between the same temperatures do the same amount of work when supplied with the same quantity of heat. This is CARNOT'S PRINCIPLE.

In order to translate the water wheel analogy more completely into terms of the reversible heat engine we imagine, as did Carnot, a cylinder provided with a piston which can move smoothly in it, that is without friction. The cylinder contains the working material, which might, for example, be air. A source of heat at the higher temperature, say t_2, is brought into contact with the base of the cylinder. This is imagined to be made of highly conducting material, while the cylindrical wall and the piston are thought of as perfectly non-conducting. The piston moves outwards very slowly. It must be like this or the engine could not be reversible. We are quite unconcerned with the nature of the machinery which is driven by the backward and forward motion of the piston, except that we insist on it being frictionless and arranged, it is convenient to suppose, so that it can raise a weight. During the outward motion of the piston heat is supplied to the working substance, in the cylinder, at the temperature t_2. This part of the cycle of operations is represented by *ab* in Fig. II—1A. From *b* to *c* the expansion continues; but now it is ADIABATIC. That is to say, heat is not being communicated to (nor abstracted from) the working substance, the base of the cylinder being covered, we suppose, by non-conducting material. The next stage is an isothermal *compression* at the lower temperature, t_1, the base of the cylinder being in contact with a body whose temperature is t_1. During this part of the cycle, which is represented by *cd* (Fig. II—1A), heat is given out by, or abstracted from, the working substance. According to Carnot, but not in fact, the amount of heat given out at the temperature t_1 is just the same as that received at the temperature t_2. Finally the compression is continued adiabatically, as represented by *da*, and the cycle completed. The area of the closed figure *abcd* is equal to the work done by the engine during one cycle.*

Carnot's principle is quite sound, notwithstanding the un-

* See, for example, Dr. Barton's *Text Book on Heat* (Longmans, Green & Co.).

tenability of his assumption of the conservation of heat or caloric, and some remarkable and correct results were inferred from it, which are quite independent of the truth, or falsity, of the equivalence of heat and work. This appears in the work of E. Clapeyron* who gave Carnot's theory a good and very simple mathematical form, and later in that of the brothers James and William Thomson.

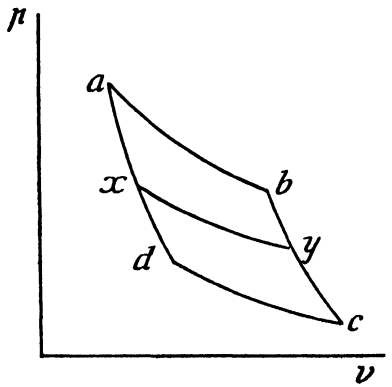


Fig. II—1A

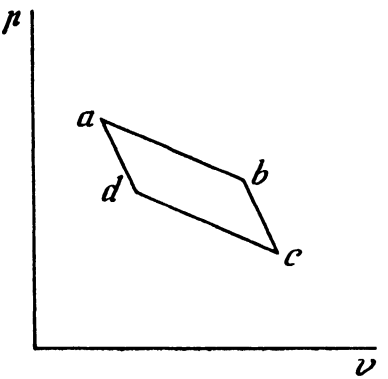


Fig. II—1B

CLAPEYRON'S THEORY

The method of Clapeyron was to apply Carnot's principle to a very small cycle, *abcd* in Fig. II—1B. It was in fact he who first applied the graphical method (no doubt suggested by James Watt's indicator diagram) to Carnot's theory, a method which has been, together with the reversible cycle, a prominent feature in thermodynamical text-books during the past hundred years. The representative closed line *abcd* in Fig. II—1B of such a small cycle is obviously a parallelogram, since the opposite isothermals (and opposite adiabatics) are so near together that their directions cannot deviate. Its area is equal to

(Vertical height) × (Horizontal breadth)

or as we may describe it,

(Pressure increment) × (Volume expansion).

* E. Clapeyron: "Memoire sur la puissance motrice du feu". *Journ. de l'école polytechnique*, 16, p. 170 (1834).

This area, as we have seen, is equal to the work that a Carnot reversible engine does when it completes the cycle *abcd* of Fig. II—1B. Since the efficiency of such an engine is measured by work done per unit quantity of heat supplied we may express it in the form

$$\frac{\text{(Pressure increment)} \times \text{(Volume increment)}}{\text{(Quantity of heat supplied)}} = \text{Efficiency.}$$

It is clear, I think, that when, as we are now supposing, the temperature difference is very small, the efficiency is proportional to it and so, by Carnot's principle, it must be equal to

$$\mu \times \text{(Temperature difference)},$$

where μ depends only on the temperature in some way or another. The factor μ is called CARNOT'S FUNCTION, though this name is sometimes given to its reciprocal. When we combine these statements we have

$$\text{(Heat supplied)} = \frac{1}{\mu} \frac{\text{(Pressure increment)}}{\text{(Temperature difference)}} \times \text{(Increase in volume)}$$

This is Clapeyron's famous formula. It is usually written

$$dQ = \frac{1}{\mu} \frac{\partial p}{\partial t_v} dv,$$

or, as I prefer it,

$$dQ = C \frac{\partial p}{\partial t_v} dv, \qquad \text{(II—1A)}$$

in which C means the same thing as $1/\mu$.

When the heat is supplied during the evaporation of a liquid (or melting of a solid) it is then the so-called LATENT HEAT and the formula becomes

$$\text{(Latent heat)} = \text{(Increase in volume)} \times C \times \frac{\text{(Pressure increment)}}{\text{(Temperature difference)}},$$

more conveniently written

$$L = (v_{\text{vapour}} - v_{\text{liquid}}) C \frac{dp}{dt}. \qquad \text{(II—1B)}$$

The increment dp is here the increase in the saturation vapour

pressure corresponding to the temperature rise *dt*. A similar formula applies, *mutatis mutandis*, to the fusion of a solid. It must once again be emphasized that Carnot's principle requires C to be a *function of the temperature only*.

A REMARKABLE PREDICTION

James Thomson made use of this fact and was thereby led to one of those remarkable predictions which have marked the course of the development of physical science.* He determined the *numerical value* of C (not the *nature of the function*) corresponding to the temperature of melting ice by using the experimental data for water (liquid) and its vapour and substituting in II—1B. The value of C thus obtained is of course equally good for the fusion of ice and he used it to calculate dp/dt for the equilibrium state of ice and water in the same temperature neighbourhood. Now since L (latent heat of fusion of ice) is positive, but $v_2 - v_1$ is negative (volume of water produced is *less* than that of ice melted), Thomson calculated a negative value for dp/dt. For *plus dp* (pressure increased) there will therefore necessarily be a *minus dt* (equilibrium temperature, i.e. melting point, lowered). The amount of this depression (per one atmosphere rise in pressure) Thomson found to be equivalent to 0·0075 of a degree centigrade. It was tested experimentally by his brother William (Lord Kelvin), whose experimental value differed only by about 0·0002°C. from the predicted one.

SCALES OF TEMPERATURE

Obviously there is some advantage in replacing $a + t$, which appears in I—3, by the single letter, T, as we did in I—3A. The latter represents the temperature as measured from the so-called absolute zero. It is even more important and intellectually satisfying to define, if possible, a temperature scale in a way which is independent of the peculiarities of any thermometric substance or property and one of Kelvin's strokes of genius was

* James Thomson: "Theoretical Considerations on the Effect of Pressure in Lowering the Freezing Point of Water," *Trans. Roy. Soc. of Edinburgh*, 16, Part V, p. 575 (1848–1849). This paper was communicated by his brother William Thomson.

his appreciation of the fact that Clapeyron's formula enables this to be done. Why not, one may ask, *define the scale* on which t is measured by laying down that C shall be a positive constant? In other words, why not define the measure of the temperature difference between two temperatures, which are very near together, in such a way that it is proportional to the efficiency of a Carnot engine which uses the two temperatures? This was actually proposed by Kelvin in 1848* when he was still clinging to the principle of conservation of caloric. I like to call this scale the EFFICIENCY SCALE.

No use has been made so far of the equivalence of heat and work; but when we do make use of it and apply Clapeyron's formula (II—1A) to a gas—which means that we identify dQ with $p \times dv$—we find that Carnot's function, C, is identical with the ideal gas temperature, T. Having defined one scale by taking C to be constant, Thomson, after the victory of the equivalence of heat and work, defined another, now often called the WORK SCALE, which in effect makes C a measure of the temperature.† The work scale has a so-called absolute zero (about 273° below the temperature of melting ice) being in fact identical with the ideal gas scale. A gas thermometer, especially one containing hydrogen, which approximates closely to the ideal gas, gives work scale temperatures approximately and needs only small corrections to enable work scale temperatures to be effectively reached. The efficiency scale temperature is proportional to the logarithm of the work scale temperature and has its zero therefore at minus infinity.

ACHIEVEMENTS OF CLAUSIUS AND THOMSON

It is difficult in these days to appreciate what a formidable problem the reconciliation of the equivalence of heat and work and Carnot's principle presented to the physicists of the eighteen-forties. Joule indeed suggested that the principle should be abandoned, while Thomson was equally convinced of its soundness. Clausius appears to have been the first to realize that it was

* W. Thomson: "On an Absolute Thermometric Scale," *Phil. Mag.*, 33, p. 313 (1848).

† "On the Dynamical Theory of Heat, Part V. Thermo-electric Currents," *Trans. Roy. Soc., Edinburgh*, 21, p. 123 (1854).

only Carnot's *premisses* which were incompatible with the results of the experiments of Joule and his contemporaries and he succeeded (as did Thomson independently and possibly just as early) in finding new premisses which led to Carnot's result without being in conflict with that of Mayer and Joule. Clausius adopted the principle of the equivalence of heat and work and combined it with the axiom that heat cannot pass *of itself* from a colder to a hotter body. This axiom is quite true of course; but it is impossible to apply it to the problem in hand without raising, as we shall see, a rather strange difficulty. Clausius had to give it a somewhat sophisticated form: he imagined two reversible cycles of the Carnot type—I prefer to call them "reversible engines"—each doing the same amount of work in one cycle of operations; but one of them taking, if possible, less heat from the source at the higher temperature, t_2, during one cycle than does the other. Imagine now the less efficient engine (which, by hypothesis, is reversible) to be driven backwards by the more efficient one in such a way that each performs its cycle in the same time. The net amount of work done by the two of them during one cycle of the compound engine, which they constitute, is of course zero; but the reversed one now transfers more heat from the sink to the source than is transferred in the opposite sense by the driving engine. So we have an isolated physical process the sole result of which, on completion of one cycle shall we say, is the passage of heat from the sink at the lower temperature to the source at the higher. The axiom adopted by Clausius forbids this, so he reaches the same conclusion as did Carnot.

And now we come to the difficulty which has been mentioned. When the cyclical processes (or engines) are both reversible, transference of heat from a higher to a lower temperature is just as impossible as transference in the opposite sense; since if it were possible the reversal of the whole process makes the transference of heat from the lower to the higher temperature possible. Clearly it cannot be the impossibility of the transference of heat in a *specific direction* which is relevant in discussing the efficiency of reversible engines. I am thus led to the following amendment of the axiom of Clausius: *In any reversible cyclic process, in which the net amount of work done during a cycle is zero, there can be no transference of heat.*

It is noteworthy that Carl Neumann (in the first really

rigorous account of thermodynamics, published more than twenty years before Planck wrote his *Thermodynamik*) suggested* the following axiom: *The results of processes occupying some time interval can never be such that mere contact of two bodies will cause them to be reversed.*

Leaving this for the reader to reflect about, I shall pass on to Thomson's axiom, which is better than that of Clausius. He assumes *the impossibility of a self-acting engine which can do work by cooling a source of heat below the lowest temperature in the surroundings.* It is a very good axiom; but the one adopted by Max Planck (1858–1947) is perhaps the best of them all. It is the axiom of *the impossibility of the perpetuum mobile of the second kind*—a name due to Wilhelm Ostwald (1853–1932). It may be stated as follows: It is impossible to construct a machine, or device, which

(*a*) raises a weight,
(*b*) takes heat from a single source of heat,
(*c*) works in a periodic (or cyclic) way,

and does nothing else, i.e. it brings about no other changes except the raising of the weight and the abstraction of heat from a source. This axiom leads very simply to Carnot's principle. For suppose that, of two reversible engines working between the temperatures t_2 and t_1, one of them, A, is more efficient than the other, B, and suppose further, to simplify the argument, that both take the same quantity of heat, Q_2, from the source, at t_2, during a cycle. A then does more work during a cycle than does B—since by hypothesis it is more efficient and takes the same quantity of heat from the source during a cycle. Clearly A must give less heat to the sink at the temperature t_1. If now A be made to drive B (assumed to be reversible) backwards in such a way that each performs its cycle in the same time—this is possible since A does more work in a cycle than does B—the source of heat will remain unchanged, or it will change periodically. The sink on the other hand will lose heat, since during a cycle B will abstract more heat from it than A supplies to it. The combination of the two engines—A driving B backwards—constitutes a machine which can be applied to

(*a*) raise a weight, since A does more work in a cycle than is needed to drive B backwards.

* C. Neumann: *op. cit.*, § 24, p. 70.

Moreover

(*b*) it takes heat from a single source—what, in the normal working of the engines, was called the sink (at the temperature t_1), and

(*c*) it functions periodically and does nothing else.

The combination of the two engines is in fact a *perpetuum mobile* of the second kind which Planck's axiom declares to be impossible.

It has been taken for granted in the foregoing proof that the efficiency of a reversible engine does not depend on the quantity of heat taken from the source during a cycle, but it is easily proved that this must be the case.*

ENTROPY

The efficiency of an engine is measured by

$$\frac{\text{Work done}}{\text{Heat supplied}},$$

and if Q_2 is the quantity of heat abstracted from the source and Q_1 that rejected to the sink, evidently the work done in a cycle is equal to $Q_2 - Q_1$, assuming as we do that the heat is measured in work units; since no change can have occurred in the internal energy. At the end of the cycle the working substance is in the same state exactly as at the beginning. This is one important reason for the use of a cycle of operations. Thus we arrive at the result that the efficiency of a reversible engine is equal to

$$\frac{Q_2 - Q_1}{Q_2}.$$

Carnot's principle requires this to be the same for *all reversible engines* working between the same two temperatures. Therefore it is the same function of the two temperatures for any working substance, whether air, steam or anything else, which might be used to operate a heat engine.

If we use Kelvin's work scale the function must be

$$\frac{T_2 - T_1}{T_2}.$$

* *Vide* W. Wilson: *Theoretical Physics*, 1, p. 292, or A. W. Barton: *A Text Book on Heat*, p. 286.

The reader will easily assure himself that this statement is in agreement with the definition of the work scale already given. Therefore

$$\frac{Q_2 - Q_1}{Q_2} = \frac{T_2 - T_1}{T_2}. \qquad \text{(II—2)}$$

This statement is in fact closely related to Clapeyron's formula (II—1). It is equivalent to

$$\frac{Q_2}{T_2} = \frac{Q_1}{T_1}. \qquad \text{(II—2A)}$$

This simple formula is the basis of the noble science of thermodynamics. It means this: If we choose any isothermal whatsoever, e.g. the isothermal *xy* shown in Fig. II—1A, joining the two adiabatics, *ad* and *bc*, and if T be the corresponding temperature and Q the associated quantity of heat that must be supplied to the substance to change it isothermally and reversibly from the state x (on the adiabatic *ad*) to the state y (on the adiabatic *bc*); the quantity Q/T *has the same value whatsoever isothermal we may select, provided always we choose the same pair of adiabatics.*

There is another aspect of this quantity Q/T. Think of any two isothermals which are $1°K$ apart (K means work scale temperature). Call the higher temperature T and the lower one, consequently, $T - 1$. Imagine them constructed on the diagram (Fig. II—1A) between the adiabatics *ad* and *bc*. The efficiency of a reversible engine using these two temperatures must be equal to

$$\frac{\varphi}{Q},$$

where φ is the area enclosed between the isothermals T and $T - 1$ and the two adiabatics, and in virtue of II—2

$$\frac{\varphi}{Q} = \frac{1}{T},$$

so that the area φ is equal to Q/T. Temperatures $1°K$ apart enclose equal areas—each equal to Q/T.

We have now reached the great discovery of Rudolph Clausius. We can distinguish the different (reversible) adiabatics from one another in a numerical way, just as we distinguish the isothermals from one another by the numerical quantity temperature. Let

us call the adiabatic, which passes through the point corresponding to 0°C. and the normal pressure, the zero adiabatic, just as we call a particular isothermal the zero (centigrade) isothermal, We may suppose this adiabatic to be represented by *ad* (Fig. II—1A). We can now see that *any* other adiabatic (such as *bc*) is distinguished by the particular value of φ which is associated with it. In fact φ is a definite function of the state of the material or system, just as is its temperature, or its internal energy. The discovery and appreciation of this was the great achievement of Clausius, who gave to φ the name ENTROPY. It is the most important function in thermodynamics; indeed perhaps in the whole of physical and chemical science. The statement

$$dQ = Td\varphi, \qquad \text{(II—3)}$$

in which dQ means a small quantity of heat communicated reversibly to a body or system, is the concise expression of the second law and thermodynamical reasoning is most conveniently and effectively done with its aid and with that of functions built up with it, such as FREE ENERGY. It has two outstanding properties: (*a*) the sum total of the entropy within an adiabatic enclosure *must remain constant when only reversible changes occur within it*, though the entropy may increase in one part of the enclosure at the expense of that in some other part, and (*b*) in any *actual* change or process—an actual process can never be reversible, since it is never a succession of states of equilibrium, though it may approach infinitely near to the limiting case of reversibility; hence the validity of arguments based on reversible processes—the sum total of all the entropy changes must be *positive*. In other words, the sum total of entropy can only get bigger. As Clausius expressed it:

Die Energie der Welt ist constant,
Die Entropie der Welt strebt einem Maximum zu.

Entropy will appear again in a new and fascinating guise (which justifies the remark about its importance) when we study the development of our knowledge of the statistical foundations of thermodynamics.

THE LATENT HEAT EQUATION OF CLAUSIUS AND ITS ELECTRICAL ANALOGUE

In February 1850 Rankine read a paper to the Royal Society of Edinburgh which contained the following formula, the derivation of which can now be found in any elementary book on thermodynamics:

$$(v_2 - v_1)\frac{dp}{dT} = c_1 - c_2 + \frac{dL}{dT}; \qquad \text{(II—4)}$$

v_1 and v_2 are respectively the volumes of the unit mass of a liquid and its vapour, both of these *being in equilibrium with one another*, while c_1 and c_2 are their respective specific heats—not to be confused with specific heats at constant pressure or at constant volume—and L is the latent heat of evaporation.* In the same month of the same year Clausius read a paper to the Prussian Academy which contained the same result; but he took the important further step of eliminating the left-hand side which, as the reader will recognize, is equal to L/T, thus obtaining†

$$c_2 - c_1 = \frac{dL}{dT} - \frac{L}{T}. \qquad \text{(II—4A)}$$

On substituting the values of the latent heat of evaporation of water and its temperature variation, dL/dT (as given by Regnault), he found the paradoxical result that the specific heat of saturated water vapour (steam) is *negative*.‡ This means simply that, in raising the temperature of the vapour or steam *while keeping it saturated*, work is done on it—more than is enough to raise the temperature; so that heat has to be withdrawn.

This explains the behaviour of the CLOUD CHAMBER (C. T. R. Wilson, 1897) which we shall meet with later. When initially saturated water vapour expands adiabatically it becomes *supersaturated*, or, if dust particles or charged ions are present, drops of water condense round them.

As Clausius pointed out, some liquids, ether in particular, behave differently. If we start with saturated ether vapour and

* *Trans. Roy. Soc. Edinburgh*, 20, p. 147.

† *Pogg. Ann.*, 79, pp. 368, 500.

‡ *Vide* the chapter on "Saturated Vapour" in Browne's translation of Clausius' work *The Mechanical Theory of Heat.*

let it expand adiabatically it is certainly cooled; but it becomes *undersaturated*. The specific heat of its saturated vapour is *positive*, i.e. when we lower its temperature adiabatically we have to withdraw heat (at constant temperature) to bring it back into the saturated state.

Examination candidates not infrequently write to the effect that, during the adiabatic expansion of initially saturated aqueous vapour, the vapour is cooled (which is quite correct) and therefore condensation occurs (an illegitimate inference). Clausius knew better than this nearly a hundred years ago.

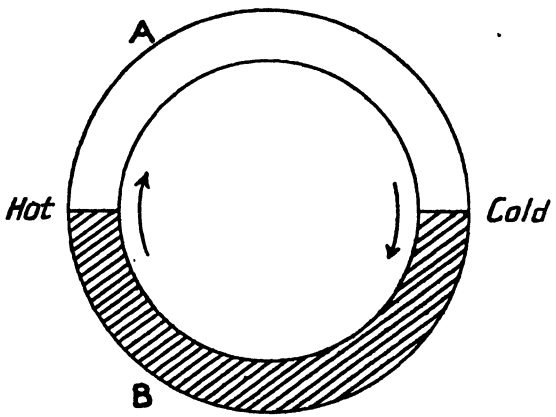


Fig. II—2

There is a remarkable analogy between Clausius' formula (II—4A) and the formula which Thomson (Lord Kelvin) discovered for the passage of electricity across a junction between two metals in a thermoelectric circuit. Imagine a circuit made of two different metals, A and B (Fig. II—2). When one junction is maintained at a higher temperature than the other, then a thermoelectric current flows round the circuit—in the direction of the arrows, shall we suppose. The circumstances are exactly analogous (thermodynamically) to the circulation of vapour and liquid which would occur on the supposition that A and B were the two halves of a closed tube, the lower half, B, being filled with some liquid and the upper half, A, with its vapour. Liquid would evaporate at the hot junction where latent heat would be absorbed, corresponding to the Peltier heat, which I shall call Π.

At the cold junction vapour condenses and latent heat is given out, corresponding to the Peltier heat given out. The two specific heats, c_1 and c_2, have their parallel in the Thomson (Kelvin) coefficients, σ_1 and σ_2, of the two metals—Kelvin called them the specific heats of electricity in the respective metals and he might have called the Peltier coefficients latent heats of electricity. Finally he deduced the formula:

$$\sigma_2 - \sigma_1 = \frac{d\Pi}{dT} - \frac{\Pi}{T}, \qquad \text{(II—4B)}$$

which in fact *predicted* the phenomenon indicated by the coefficients, σ. This is the phenomenon of the reversible generation, or absorption, of heat in a wire for example, when an electric current flows through it, in consequence of a difference in temperature between its ends.*

* *Vide* S. G. Starling: *Electricity and Magnetism.*

. . .

Note added during correction of the proofs.

The apparent incompatibility between the equivalence of heat and work (Joule) and Carnot's principle would have appeared much less formidable to Kelvin had he reflected that conservation of caloric is the limiting case of an infinite value of the mechanical equivalent of heat, for which the two types of *perpetuum mobile* become identical.

CHAPTER THREE

AN EXPERIMENTAL MISCELLANY

ANDREWS' EXPERIMENTS ON CARBON DIOXIDE

BETWEEN 1863 and 1869 Dr. Thomas Andrews (1813–1885) carried out a remarkable series of experiments with carbon dioxide.* They have been described and diagrams similar to Figs. I—1 and III—1 reproduced so often that one might be forgiven for merely referring to them. They are however very important and constitute such a perfect model of this kind of physical inquiry that I feel compelled to describe them in some detail. The apparatus which he used was quite small and has often been reproduced since his day for teaching purposes. It is illustrated in Fig. III—1 and consisted of a narrow tube, *AB*, with a calibrated capillary portion at its upper end for measuring volumes. It contained pure carbon dioxide, which was enclosed in the capillary portion by a short column of mercury, as shown in the figure. The way in which it was introduced was as follows: the top of the tube, at *A*, was originally open and the previously dried carbon dioxide was made to flow through it for some hours to ensure the absence of air. The top, at *A*, was then sealed off and the lower end opened under mercury. The small enclosing column of mercury was introduced by the familiar device of a slight warming which expelled some of the carbon dioxide, so that mercury was drawn in on subsequent cooling. The tube was then placed in the steel chamber, *E*, so that the capillary part projected outside. The whole of the space in the chamber and the part of the glass tube below the mercury was filled with

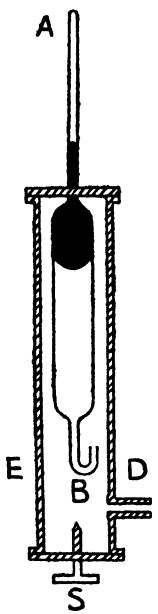


Fig. III—1

* Andrews: *Phil. Trans.*, 159, p. 575 (1869).

water which could be subjected to pressure by turning the screw, *S*.

The side tube, *D*, led to an exactly similar piece of apparatus, not shown in the figure, filled in a similar way; but with air instead of carbon dioxide. Each projecting capillary was enclosed in a water bath with plate-glass sides, so that the tubes could be seen and the volumes of the carbon dioxide and of the air might be read off. One water bath was used of course to give the carbon dioxide various temperatures and the other to keep the air at a constant temperature. The measured volume of the air enabled the pressure, both of the air and of the carbon dioxide, to be determined, since Andrews knew how the pressure of the air depended on its volume.

The results of the experiments are exhibited in Fig. I—1. At the lower temperatures each isothermal curve has a horizontal part associated with the coexistence of liquid and gas. Once liquid has begun to form, the isothermal remains horizontal till all the gas has liquefied. This means that during the transition stage, which begins on the right with the substance wholly in the form of gas and ends on the left with the substance wholly in the liquid form, the pressure remains constant. This is the pressure under which (at the particular temperature marking the isothermal) the gas and liquid are in equilibrium with one another. It is called the EQUILIBRIUM PRESSURE or SATURATION VAPOUR PRESSURE.

The horizontal parts of the isothermals get shorter and shorter, it will be seen, the higher the temperature, until at a certain CRITICAL TEMPERATURE the horizontal part has contracted to nothing. At this critical temperature and higher temperatures there is no longer any distinction between two states (gas or vapour and liquid). The substance can only exist in one state (the gas state in the narrower sense of the term) at temperatures higher than the critical one. The term VAPOUR is applied to the gas state *below* the critical temperature.

It may be remarked parenthetically that, in suitable circumstances, a liquid at constant temperature may have its pressure reduced far below the equilibrium (horizontal line) pressure for that temperature. Indeed it may be made negative, i.e. the liquid may be in a state of tension (Worthington's experimen‘ Drops of water (or small spherules) suspended in a mixture

of appropriate density have been raised to 178°C. at normal pressure before vaporization occurred (Dufour). On the other hand a vapour may be subjected to a pressure well above the equilibrium pressure corresponding to its temperature.

The point on the diagram (Fig. I—1) where the horizontal (liquid-vapour) line has just contracted to nothing is called the CRITICAL POINT and the associated pressure and volume are called the CRITICAL PRESSURE and CRITICAL VOLUME of the quantity of the substance involved respectively. The table below gives the critical data for a few substances.

TABLE III—1

Substance	Critical Pressure (Atmospheres)	Critical Volume (in cubic centimetres per gramme)	Critical Temperature (° C.)
Water ..	218	2·5	374
Carbon dioxide	73	2·17	31
Ethylene ..	51	4·55	10
Oxygen ..	50	2·33	−119
Nitrogen ..	34	3·23	−147
Hydrogen ..	13	33·3	−240
Helium ..	2·3	14·3	−268

Andrews' experiments made perfectly clear what had long been suspected, that a gas can only be liquefied when its temperature has been reduced below a certain critical one characteristic of it. I shall have a little more to say about the significance of Andrews' results when we study the great contribution which J. D. van der Waals (1837–1923) made to the theory of gases a short while after Andrews had finished his work.

DEVIATIONS FROM BOYLE'S LAW

It has long been known that Boyle's law is not exact for any gas. If it were exactly true and the product $p \times v$ (pressure × volume), measured at a particular constant temperature, were plotted against the pressure, p, the resulting graph would of ırse be a horizontal straight line. The first serious investiga- ٢f the extent to which gases deviate from the law was made

by Henri Victor Regnault (1810–1878), who carried out in 1847 a series of constant-temperature measurements of pv (reckoned for a fixed quantity of the gas under experiment) for values of p varying from one to about 30 atmospheres.* His method was to increase the pressure gradually till the volume was exactly halved, while taking observations of the pressure and the volume. He then, while keeping the pressure constant, introduced more gas till the original volume was restored; that is to say he doubled the quantity of the gas under experiment and proceeded further in the same way. In this manner he avoided the difficulty due to the progressive diminution of the volume with increasing pressure. He used for his pressure measurements a mercury manometer which was fixed to the wall of a tower. Many others followed Regnault. In 1850 Natterer discovered that, in the case of air and nitrogen, the product pv at first diminishes (at constant temperature) as p increases; but beyond a certain pressure it begins to increase as the pressure is further raised. He noticed in fact that at a certain pressure pv has a minimum value. Amagat and Cailletet worked at the bottom of a mine in order to have plenty of scope for a manometer of great height. The former began a long sequence of experiments as early as 1870 and continued them till after 1890.† Amagat's manometer, in his mine experiments, was over 300 metres high and his latest methods enabled him to reach 3000 atmospheres. At a temperature in the neighbourhood of 20°C. he found the minimum value of pv to be at a pressure of 50 metres of mercury in the case of nitrogen and 100 metres in the case of oxygen. The most recent investigations have been made by Holborn and his collaborators (1915 and since). He introduced new methods including his PRESSURE BALANCE.‡ The results of these numerous experiments have the same general character for all gases. They are illustrated in Fig. III—2, which is not drawn to any particular scale, nor to represent a particular gas. Each of the full lines, or graphs, such as *abc* or *def*, represents the relation between pv and p of a given mass of the gas at some particular temp
being the isothermal (as it may be called, since

* *Mémoires de l'Institut de France*, 21, p. 329.

† Amagat: *Ann. d. Chim. et d. Phys.* (187
Rendus (1888–1893).

‡ See Dr. Barton's *Text Book*, p. 224.

constant temperature relationship) corresponding to a higher temperature than that to which *def* corresponds, and all the isothermals are plotted for the same mass of the gas. The parts which correspond to the horizontal, liquid-vapour, lines of Andrews' diagram are here, naturally, vertical straight lines,

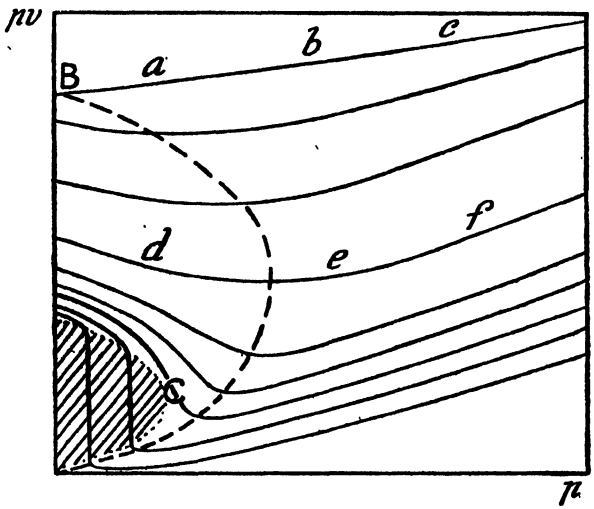


Fig. III—2

shown in Fig. III—2 traversing the shaded area. The critical point is at *C*. The minima lie on a curve, shown in the diagram as a broken line, the shape of which is roughly parabolic. The point, *B*, where this cuts the vertical, *pv*, axis is called the BOYLE POINT, since in its neighbourhood the gas conforms most closely to Boyle's law. The shapes of the curves in the figure are accounted for extraordinarily well, though quantitatively not quite exactly, by van der Waals' theory, as we shall see.

CALORIMETERS AND MEASUREMENTS OF SPECIFIC HEATS

t calorimeters, i.e. heat measurers, functioned by sure of heat, the quantity of ice which it melted. s were used by Joseph Black* and Antoine

st measured quantities of heat and who first and temperature.

Lavoisier. Joule and Mayer used the product (mass of water) × (rise in temperature) as a measure of heat. Joule's calorimeters were metal vessels covered with thermally insulating material. Such calorimeters were always slightly leaky. After Kirchhoff's study of the radiation of heat from material surfaces and their associated reflecting powers, an improved type of calorimeter came into use. It consisted of two well-polished copper or, better still, silver vessels, one inside the other and, so far as practicable, not in thermal contact. That is to say the inner one was suspended by fine threads or perhaps rested on a small piece of badly conducting material, such as felt or ebonite, placed at the bottom of the outer one. A polished metallic surface, especially a silver one, radiates heat very slowly and such poor radiators reflect radiant heat correspondingly well, as Kirchhoff explained. The inner vessel, into which the hot body is introduced, throws out heat only very slowly and the greater part of this is reflected back to it from the highly reflecting outer vessel, so that although heat does escape from such an arrangement (or penetrate into it, if the temperature in its interior is lower than that of the surroundings) it does so remarkably slowly. With such an arrangement, used for example by Regnault* in measuring the specific heats of gases at constant pressure, the corrections for loss or gain of heat are very small and therefore more effective.

The final perfection of this device was achieved by Sir James Dewar (1842–1923). It is well known as the thermos flask and was used by him as a container for liquid air and liquid hydrogen. It is a double-walled glass vessel, silvered inside, and having the interspace exhausted, so that transference of heat by conduction and convection, as well as by radiation, is almost completely eliminated.

Bunsen, the inventor of the Bunsen burner, devised an ingenious form of ice calorimeter in 1870, in which the quantity of ice melted is determined from the resulting contraction in volume. He also devised an equally ingenious method of finding the density of ice, in order to provide himself with a necessary datum for use with his calorimeter. These are described in the elementary text-books.

Joly's steam calorimeter provided, in its differential form

* Regnault: *Mém. de l'Acad.*, 26, p. 1 (1862).

(1894), the first successful direct way of determining the specific heats of gases at constant volume. H. Callendar introduced the continuous flow type of calorimeter in 1902. It is a double-walled tube with the interspace exhausted and with a coil of wire running through it to enable heat to be generated by an electric current. It was used by Callendar's collaborator, Barnes, to investigate how the specific heat of water depended on the temperature.

Special forms of calorimeter have been invented for measuring specific heats at low temperatures—Nernst and Lindemann (1911)* and a very elaborate one by Simon and Lange (1924)† in which heat losses are reduced or eliminated by maintaining the wall of the calorimeter at the same temperature, practically, as its interior.

Much work has been done in recent times on the specific heats of gases at low temperatures (Eucken, Partington, Brinkworth), and the significant result, which the quantum theory predicts, has been found that, as the temperature is lowered, all gases behave more and more like monatomic gases, i.e. the molecules of gases seem to lose their degrees of freedom until these are reduced to three. Below $80°K$ the ratio of the specific heats of hydrogen, for example, approximates almost to $1\frac{2}{3}$, whereas of course it is close to $1\frac{2}{5}$ at ordinary temperatures.

Determinations of the specific heats of solid elements are also of great interest from the point of view of the quantum theory and will be referred to again when we study the law of Dulong and Petit and Debye's great work.

DENSITIES OF GASES AND VAPOURS

Avoga[illegible] and its significance for molecular weight determinations [illegible]he measurement of the densities of gases and vapours a quite [illegible]al interest. The molecular weights of gases and vapours are [illegible]rtional to their densities, under like conditions of pressure [illegible] temperature, in consequence of Avogadro's law. The mole[illegible]r weight of hydrogen may be taken as 2·016 (that of oxygen being fixed, conventionally, as 32). It

* Nernst and Lindemann: *Zeitschr. für Elektrochemie*, 17, p. 817 (1911).

† Lange: *Zeitschr. für Phys. Chemie.*, 110, p. 343 (1924).

follows that that of any other gas or vapour must be equal to

$$2{\cdot}016 \times \frac{\text{Density of gas or vapour}}{\text{Density of hydrogen}}.$$

Purely chemical methods enable the percentage composition of a substance to be determined. That of ether for example indicates that its molecule is $C_4H_{10}O$ or $C_8H_{20}O_2$ or some other integral multiple of $C_4H_{10}O$, so that its molecular weight is shown to be an integral multiple of $(48 + 10{\cdot}08 + 16) = 74{\cdot}08$. A rough estimate of its density indicates that it is equal to or near 37 times that of hydrogen and so definitely decides in favour of 74·08.

Dumas, Gay-Lussac, Hoffmann and Victor Meyer all devised well-known methods for determining vapour densities, and Regnault, who seems to have measured nearly every quantity depending in any way on temperature, determined directly the densities of a large number of gases.

THE POROUS PLUG EXPERIMENT OF JOULE AND KELVIN

The old Gay-Lussac experiment, as repeated by Joule, was not very sensitive for the reason already given. The ideal form of the experiment, had Joule found it practicable, would have been to provide the expansion vessel with an adiabatic covering and thus *compel U*, the internal energy, to remain constant. The temperature change, had one occurred, might have been measured by the change in the electrical resistance of a fine wire spread through both of the vessels *A* and *B* (Fig. I—2).* Kelvin thought of an experiment which, though quite different from the original Gay-Lussac–Joule one, was exactly analogous.† In this experiment it was $U + pv$, that is the sum of internal energy and the product of pressure and volume, that was made to remain constant during adiabatic expansion, and not U itself. This quantity, $U + pv$, is called ENTHALPY (from $\theta\alpha\lambda\pi\omega\varsigma$, warmth, heat) sometimes TOTAL HEAT. It is of interest that it is exactly the same function of φ (entropy) and p (pressure) as U is of φ and v. The Joule–Kelvin experiment is of

* I am not aware that the experiment has ever been attempted in this form.

† Joule's *Scientific Papers*, 2, p. 217, or Kelvin's *Paper's*, 1, p. 333.

interest, not only on its own account, but because it also gives the information about the dependence of U on temperature and volume which was hoped for from the Gay-Lussac–Joule experiment. The reason for this is that the part of the temperature change due to the change of pv in the Joule–Kelvin experiment can be computed separately from the experimental results of Regnault and others on the dependence of pv on temperature and pressure.

Joule and Kelvin experimented on several gases, including hydrogen, air and carbon dioxide, between 1852 and 1862. They found for air, initially at 0°C., a cooling of 0·275°C. per one atmosphere drop in pressure and 1·39°C. for carbon dioxide. Hydrogen on the other hand showed a heating effect of 0·03°C. This is very largely due to the change in pv. Thermodynamical theory indicates (and this is borne out by experiments on hydrogen and helium) that at sufficiently high temperatures all gases *are heated*; but there is a TEMPERATURE OF INVERSION (which depends slightly on the volume), characteristic of each gas, below which it is cooled.

The experiment consisted in forcing the gas, adiabatically, through a porous plug (or throttle) consisting of cotton wool or silk. In the earliest experiments it was a piece of leather. To secure that the expansion was adiabatic, the part of the tube in which the plug was situated was made of non-conducting material—boxwood. On one side of the plug a high pressure was maintained and a low one (atmospheric) on the other; so that the gas was forced slowly through the pores in the plug from the high-pressure side to the low-pressure side. The temperatures on the two sides of the throttle were noted by sensitive mercury thermometers.

It should be noted that the adiabatic expansion, both in this experiment and in the original one of Joule, was an IRREVERSIBLE one. This does not invalidate the application of thermodynamical reasoning to the initial and final states. We imagine the final state to be reached in a reversible way. We cannot of course apply such reasoning to the actual process in the plug itself. The state of affairs there is so far removed from equilibrium that both temperature and pressure are illusory.

The practical details of the experiment are given in the textbooks (see Dr. Barton's book for example). In describing its

nature we may think of the porous plug as a minute aperture in a diaphragm, shown at a in Fig. III—3. The gas to the left of a is maintained at the high pressure, P, and that which has passed through the aperture, a, to the right is at a steady lower pressure, p. It is helpful to think of these pressures as maintained by the

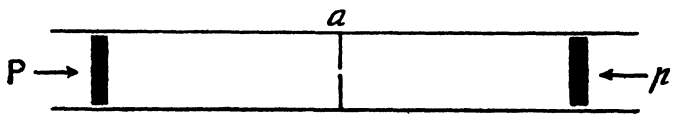


Fig. III—3

pistons shown in the diagram, although they constituted no part of the practical procedure. Let us consider the passage of one gramme of the gas through the aperture, a. The work done *on* the gas by the piston on the left is equal to the product of the pressure, P, and the volume occupied by one gramme of the gas on the left. The work done *by* the gas on the piston on the right is equal to the product of p and the volume a gramme of the gas occupies on the right of the aperture, a. Let us call the former product $(pv)_L$ and the latter $(pv)_R$. So the net work done *by* the gas is

$$(pv)_R - (pv)_L.$$

It is in fact the increment of pv (for one gramme) and we shall represent it by

$$\Delta\,(pv).$$

If we further represent the increment (positive or negative) of the internal energy by

$$\Delta\,U,$$

the first law (I—4) will take the form

$$0 - \Delta\,(pv) = \Delta U$$

or

$$\Delta\,(U + pv) = 0, \qquad \text{(III—1)}$$

since $\Delta\,Q$ is zero (adiabatic process).

Therefore $U + pv$ remains unchanged. Compare this with the

$$\Delta\,U = 0$$

of the form of the Gay-Lussac–Joule experiment described above.

It should be observed that the cooling (or heating) effect to be expected in the Joule–Kelvin experiment is quite small by comparison with that in the more usual type of adiabatic expansion (or compression). The reason is of course that the work done on the gas and that done by it nearly compensate one another.

The application of thermodynamics to the passage of a fluid (gas or liquid) through a minute aperture under adiabatic conditions with constant, but of course different, pressures on the two sides of the aperture, leads to the formula

$$\Delta T/\Delta p = V\{\alpha T - 1\}/c_p,$$

in which ΔT is the drop in temperature corresponding to the pressure drop Δp, α is the coefficient of expansion (i.e. $\partial V/\partial T_p V$) of the fluid, V is its specific volume (volume of one gramme), and c_p is its specific heat at constant pressure. The temperature is expressed on Kelvin's work scale. The corresponding formula for the earlier Gay-Lussac–Joule experiment, when carried out in the way outlined above, is

$$\Delta T/\Delta v = p(1 - \beta T)/c_v,$$

in which β means the temperature coefficient of pressure increase (i.e. $\partial p/p\partial T_v$) and c_v is the specific heat of the fluid at constant volume. It will be seen that ΔT vanishes in both experiments when the fluid has the properties of an ideal gas, since for an ideal gas $\alpha = \beta = 1/T$.

LIQUEFACTION OF GASES

The liquefaction of any gas necessarily involves, as we have learned, lowering its temperature below the critical temperature characteristic of it. The earliest really effective method of doing this was by the use of the so-called CASCADE PROCESS of the Genevese chemist, Pictet, who employed it to liquefy oxygen (1877). It consists in liquefying, by the application of pressure, some easily liquefiable gas with a relatively high critical temperature. The liquid produced is made to boil under reduced pressure, thus producing a considerably lower temperature, which is used to bring some other gas below its critical temperature. After liquefaction its ebullition under reduced pressure produces a lower temperature still and so on. Pictet employed in succession methyl chloride and ethylene and was thus enabled to cool

oxygen below its critical temperature (now known to be about − 119°C.), so that the application of pressure liquefied it.

The next important step in the improvement of the methods for liquefying air and other gases is usually ascribed to Linde (1895). It is sometimes called the regenerative method. It is probable that it was thought of independently and used by Dewar, Hampson and Kamerlingh-Onnes about the same time. The functioning part of Linde's apparatus is illustrated in Fig. III—4.

The air is compressed to about 180 or 200 atmospheres and the heat thereby generated drawn off by passing it through a copper pipe which is coiled in a tank through which cold water is kept running. The water vapour and carbon dioxide in the air (or other gas) are removed by passing it over suitable absorbing materials. It then enters the tube, *A*, of the liquefier (Fig. III—4). This is in fact a long coiled tube, shown in the diagram for convenience as straight. At its end is a very narrow aperture, *B*, the size of which is adjustable. Beyond the aperture, *B*, the pressure is approximately atmospheric and in passing through it the air (or other gas) suffers the Joule–Kelvin drop in temperature. This cooled gas returns to the compressing pump by way of the wider tube, *C*, which encloses *A*. The oncoming air is thus cooled before it reaches the aperture, where, of course, on expansion it is cooled still further, and so it continues progressively. The whole of this part of the apparatus is thoroughly well lagged with thermally insulating material.

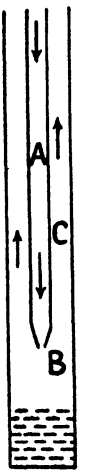


Fig. III—4

In Claude's method the compressed air passes into the cylinder of an engine and is cooled in the "ordinary" adiabatic way in pushing out the piston. The air, or gas, so cooled is made to cool the oncoming air which has not yet suffered the adiabatic cooling; so that the lowering of the temperature is progressive, as in Linde's process.

Naturally the methods, such as Linde's, which depend on the Joule–Kelvin cooling effect cannot be used *directly* to liquefy hydrogen or helium, since the inversion temperatures of these gases are far below ordinary temperatures; but Dewar succeeded in liquefying hydrogen (1898) in this way *after first cooling it below its temperature of inversion* (− 80°C.) by means of liquid

air or oxygen. The boiling point of hydrogen, it may be noted, is − 258°C.

The Dutch physicist Kamerlingh-Onnes (1853–1926) was the first to liquefy helium (1908).* He used Linde's method, first cooling the helium below its temperature of inversion by passing it through hydrogen, boiling under reduced pressure. The experiments of Onnes and his collaborators indicated the existence of two modifications of liquid helium and this has been definitely established by Keesom and others. Helium was solidified in 1926 by Keesom.

LOW TEMPERATURES—DEBYE'S SUGGESTION

By boiling liquid helium under reduced pressure Keesom and his collaborators got down to the temperature 0·844°*K*. Temperatures down to this point have been determined by the helium gas thermometer. In the actual practical procedure of measuring low temperatures, or for that matter of measuring temperatures generally, a gas thermometer is an awkward kind of thing. In practice other devices are much more convenient. The resistance of a lead wire and thermocouples have been used down to 14°*K*. and below this the vapour pressure of helium; but all these, in the last resort, have to be calibrated by comparison with a gas thermometer and this must be a helium one at these temperatures.

P. Debye (1884–) made a brilliant suggestion for reaching (and measuring) temperatures still lower than those attainable by boiling helium (1926). When a paramagnetic material is magnetized adiabatically it is slightly heated, very naturally, since work is done on it. And conversely, when it is demagnetized adiabatically, by the removal of the magnetizing field, it is slightly cooled. This is a very trivial thing at ordinary temperatures, but not when it begins at a temperature so low as 1°*K*, and since 1933 extraordinarily low temperatures have been reached by this magnetic method. Moreover, provided the initial temperature, say 1°*K*, from which the magnetic cooling starts, is known (by the use of a helium gas thermometer) the magnetic method provides the means of determining the temperatures on the work scale.

* Onnes, *Leid. Comm.* 108, Supp. 21A (1908).

There is a remarkable parallelism between this process and the reversible adiabatic cooling of a gas. The relationship between T, H and I (temperature, magnetizing field intensity and intensity of magnetization respectively) of a *paramagnetic* substance has a formal similarity to the relationship between T, p and v (temperature, pressure and volume respectively) of a gas. Thus the internal energy of a gas depends only on its temperature to a close approximation; so too the internal energy of magnetization of a paramagnetic material depends only on its temperature, to a close approximation of course. Thermodynamical reasoning leads in the two cases to

$$v = F\left(\frac{p}{T}\right), \text{ for the gas}$$

and

$$I = f\left(\frac{H}{T}\right), \text{ paramagnetic substance,}$$

F and f being functions which the thermodynamical reasoning tells us nothing about. But experiment indicates that

$$v = (\text{a constant, } R) \div \left(\frac{p}{T}\right)$$

and

$$I = (\text{a constant, } C) \times \left(\frac{H}{T}\right)$$

The constant, C, is called CURIE'S CONSTANT, after Pierre Curie (1859–1906), who discovered this law.* As is well known, I/H is called the MAGNETIC SUSCEPTIBILITY of the material and usually represented by κ. Therefore

$$\kappa = C/T.$$

It may be remarked that the significance of I (intensity of magnetization), or κH, is MAGNETIC MOMENT PER UNIT VOLUME. Sometimes it is convenient to use the magnetic moment per unit mass instead and the quantity, χ, defined by

$$\chi \times H = \text{Magnetic moment per unit mass.}$$

is called the MASS SUSCEPTIBILITY of the material. Obviously

$$\chi = A/T$$

* Curie: "Propriétés magnetiques des corps à diverses températures," *Ann. de Chim. et Phys.*, 7, pp. 289–405 (1895).

where A is another constant. Thus we have in Debye's method, not only a means of producing unheard of low temperatures, but a simple means of estimating them, at least approximately, since it is a comparatively simple process to measure susceptibilities. Moreover the analogue of Clapeyron's equation (II—1) is also available for estimating the actual Kelvin work scale temperature and for testing Curie's law.

THERMAL CONDUCTIVITY

The scientific interest of a phenomenon and its measurement sometimes extends far beyond it. This is the case with the conduction of heat. For Jean Baptiste Fourier (1768–1830) it provided interesting mathematical problems and led to the famous theorem, or expansion,* which is quite as important from the point of view of pure mathematics as from that of physical science. Fourier's definition of the measure of THERMAL CONDUCTIVITY is concisely expressed in the formula

$$K = Q/A\frac{\Delta\tau}{\Delta x},$$

in which K is the conductivity, Q is the quantity of heat passing in the unit time through a small area, A, which is at right angles to the direction of flow and $\Delta\tau$ means the drop in temperature which occurs in a short distance Δx measured in this direction.

Ångström devised an interesting method of finding the conductivity of a long rod or bar which involved the use of Fourier's expansion. It was suggested by the still earlier work of Forbes and of F. E. Neumann. One end of the bar was heated and cooled in a strictly periodic fashion, thus causing a temperature wave, of diminishing amplitude, to travel along the bar. Both the conductivity and the emissivity from the surface of the bar could be found from suitable observations at two points on it. (See Tait's *Heat*, Chap. xiv.)

Lees devised methods for determining the conductivities of materials available in the form of thin plates, e.g. thin crystalline plates, and for liquids and poor conductors.†

* *Théorie Analytique de la Chaleur* (1822), translated by A. Freeman (Cambridge).

† C. H. Lees: *Phil. Trans. A.*, p. 481 (1892) and p. 399 (1898).

In the case of gases difficulties are presented by radiation and convection. The former is easily dealt with and the latter has been eliminated in various ways, e.g. by giving the gas the form of a Lees disc and heating it from above. (For further information see Roberts' *Heat and Thermodynamics.*)

It is of great interest that Maxwell's prediction, that the thermal conductivity of a gas (like its viscosity) is independent of the pressure, has been confirmed by experiment (see Kinetic Theory). The relationship between the thermal and electrical conductivities is of great interest and will be found in a later chapter.

CHAPTER FOUR

THE STATISTICAL THEORY OF HEAT

I. KINETIC THEORY OF GASES

PRELIMINARY SURVEY

THE thermodynamics of Clausius and Thomson has the remarkable feature that it makes no assumptions of a constitutive kind. It reaches valid conclusions without assumptions about the fine structure of material systems; without assuming the existence of molecules at all and without the adoption of any view even as to the precise nature of the form of energy called heat; that is to say, without such an assumption, for example, as that heat is the kinetic (or mechanical) energy of molecules.

At the time when Clausius and Thomson were giving definite form to the first and second laws, the kinetic theory of gases was already rapidly developing and was eventually to expand into the more comprehensive STATISTICAL MECHANICS—L. Boltzmann (1844–1906) and J. Willard Gibbs (1839–1903)—which takes account of the molecular constitution of bodies and reveals a deeper foundation on which the laws of thermodynamics rest. Daniel Bernoulli's explanation of Boyle's law (*Hydrodynamica*, 1738) seems to mark the initiation of the kinetic theory of gases. It is of interest to note that he appreciated that the pressure is proportional to the mean square of the particle velocity.

The kinetic theory was revived by Waterston in 1845. Krönig and Clausius followed a little later (1857). In its earliest and simplest form the particles (molecules) were treated as so small that collisions might be ignored, as were also intermolecular forces, and the important assumptions were made (*a*) that the heat energy in the gas might be identified with the kinetic energy of the molecules and (*b*) that, at constant temperature, the total energy of the molecules remained constant. Waterston anticipated the PRINCIPLE OF EQUIPARTITION OF ENERGY, or at least that particular case of it which lays down that, in all

gases at the same temperature, the average kinetic energy of translation of a molecule is the same.*

Clausius and Krönig established the formula

$$pv = \tfrac{1}{3}Nm\overline{c^2}, \qquad \text{(IV—1)}$$

in which p means the pressure of the gas, v its volume, m the mass of a molecule, N the number of molecules in the gas and $\overline{c^2}$ is the average value of the squares of the velocities of the molecules in the gas.

It may be noted that (IV—1) is equivalent to

$$p = \tfrac{1}{3}\rho\overline{c^2},$$

where ρ is the density of the gas (i.e. the number of grammes per c.c.) or

$$\sqrt{\overline{c^2}} = \sqrt{\frac{\overline{3p}}{\rho}},$$

so that it is very easy to find the mean velocity (or, strictly speaking, the root of the mean of the squares of the velocities) of the gas molecules. For hydrogen, oxygen and nitrogen at 0°C. these mean velocities are 184,400, 46,100, and 49,200 cm./sec. respectively.† Such estimates appear to have been first made by Joule.

TEMPERATURE

The equipartition theorem enables us to define a scale of temperature. Since the average kinetic energy of translation of the molecules of all gases, at the same temperature, has the same value, and as this average is all the greater the hotter is the gas, we may use it to *define* a scale of temperature. The simplest way of doing this is to lay down that

$$T = \mathrm{A} \times \tfrac{1}{2}m\overline{c^2}, \qquad \text{(IV—2)}$$

* This anticipation is contained in a paper communicated to the Royal Society in 1845. The paper was rejected by reason of some errors it contained; but published in 1892 at the instance of Lord Rayleigh, who was then secretary. He wrote concerning it that "it marks an immense advance in the direction of the now generally received theory".

† We get the mean velocities in the strict sense if we multiply each of these by $\sqrt{8/3\pi} = \sqrt{28/33} = 0{\cdot}921$.

where T is the temperature and A is a conventional positive constant. Eliminating $\overline{mc^2}$ from (IV—1) and (IV—2) we get

$$pv = \left(\frac{2}{3A}\right) NT.$$

It is customary to write k for $2/3A$ and therefore

$$pv = kNT. \qquad \text{(IV—3)}$$

We see that kN is identical with the gas constant, R, and therefore k may be described as the gas constant reckoned per molecule. It is usually called BOLTZMANN'S CONSTANT, but also the ENTROPY CONSTANT, for reasons which will appear later.

The significance of the statement that A in IV—2 is a *conventional* positive constant is that it is perfectly legitimate to give it any positive value we like. It might be chosen to be $\frac{1}{2}$ or to be 1000. Whatever choice we may make, it fixes the value of k and hence of R. In the practice of thermometry R is more prominent than k or A. It has been chosen so as to make the temperature difference between melting ice and water boiling under normal pressure equal to 100. But conceivably the physicists of the future may think it desirable to make $A = 1$, or alternatively (and more probably) $k = 1$. In either case the constant R would have to conform to the choice made. The special choice, $k = 1$, has the merit of making the temperature, T, equal to Willard Gibbs' modulus Θ, which we shall meet later. The temperature as defined by (IV—2) is clearly the ideal gas temperature, or Kelvin's work scale temperature.

PHENOMENA OF DIFFUSION

The simple gas theory accounts not only for Boyle's law, but also for the laws of diffusion of gases through minute apertures, discovered by Thomas Graham (1805–1869) about the year 1846. Two different gases at the same temperature have molecules with the same average kinetic energy; therefore

$$\tfrac{1}{2} m_1 \overline{c^2}_1 = \tfrac{1}{2} m_2 \overline{c_2}^2,$$

if we distinguish them by the subscripts 1 and 2. So that

$$\frac{\overline{c_1}}{\overline{c_2}} = \sqrt{\frac{\overline{m_2}}{\overline{m_1}}}$$

The rates at which the gases get through a very small aperture are proportional to their molecular velocities c_1 and c_2, other things being equal, and therefore inversely proportional to the square roots of the masses of the molecules of the respective gases. This is the same thing as inverse proportionality to the square roots of their densities under like conditions of temperature and pressure (by Avogadro's law, which follows from the equipartition theorem as does that of Charles). It should be noted that rate of diffusion, as we have just used the expression, means numbers of molecules passing through the small apertures per second and clearly this is proportional to volumes (measured of course under like temperature and pressure conditions) diffusing per second.

It is often overlooked that, if we measure the rate of diffusion by the total *mass* diffusing per unit time, it is *directly* proportional to the square root of the density of the gas. In fact, if n is the number of molecules of some gas which pass through a minute aperture per second, then

$$n = \frac{\text{Constant}}{\sqrt{m}},$$

m being the mass of the molecule. Clearly then

$$nm = \text{Constant} \times \sqrt{m}$$

and nm is the total mass which has diffused per second.

This simple theory of diffusion also indicates that the rate of diffusion, other things being equal, is proportional to the square root of the temperature, since it is proportional to the mean velocity, $\bar{c}$, of agitation of the molecules.

SPECIFIC HEATS OF A GAS

By the number of degrees of freedom of a body (rigid body) or of a mechanical system is understood the number of independent numerical data needed to specify completely its situation and configuration. In the special case of the rigid body six such data are needed—three to fix the position of its centre of mass; two more to fix the direction of some straight line fixed in the body and passing through its centre of mass, and last of all, after we have fixed the values of these five, we can still contemplate rotation

about this line. So we need six in all. The rigid body is said to have six degrees of freedom. The equipartition theorem, in its full generality, requires that the average kinetic energy, *reckoned per degree of freedom*, of the molecules or systems in an assemblage of molecules or systems, such as a gas, shall be the same for all gases at the same temperature. Now since the average kinetic energy of translation is equal to $3kT/2$, evidently the average kinetic energy of a molecule, reckoned per degree of freedom, must be equal to $kT/2$. If now the molecules in a gas were really small rigid bodies we should expect their average kinetic energy to be

$$6 \times \tfrac{1}{2}kT$$

or

$$3kT$$

and the *total* kinetic energy in the gas to be

$$3NkT \text{ or } 3RT.$$

We are supposing a rather ideal gas in which the molecules exert no forces on one another. There is therefore no potential energy in the gas, or its potential energy is a conventional constant. The total energy of the gas must therefore be

$$3RT + (\text{a conventional constant}).$$

This will be the energy of a gramme of the gas if R is the gas constant for one gramme. Consequently its specific heat at constant volume must be

$$c_v = 3R.$$

To get the specific heat at constant pressure we must add R. Therefore

$$c_p = 3R + R^*$$

In such a case the ratio of the specific heats is

$$\frac{c_p}{c_v} = \frac{4R}{3R} = \frac{4}{3}$$

Let us consider a more general case. A molecule has, shall we

* It is perhaps needless to say that we are measuring heat in work units.

say, n degrees of freedom. Then the sum total of the kinetic energy of the gas must be

$$N \times n \times \frac{kT}{2}$$
$$= \frac{n}{2} RT$$

and its specific heat at constant volume must be

$$c_v = \frac{n}{2} R$$

and that at constant pressure,

$$c_p = \frac{n}{2} R + R.$$

Their ratio is therefore

$$\frac{c_p}{c_v} = 1 + \frac{2}{n}. \qquad \text{(IV—4)}$$

This formula appears to have been given first by Clerk Maxwell (1866).

One can easily imagine some of the degrees of freedom of a rigid body becoming, as it were, ineffective. If a molecule were perfectly spherical in shape, perfectly uniform in density and perfectly smooth, it could not be set in rotation by collisions with its neighbours and if actually in rotation it would be impossible for its angular velocity to be changed by such collisions. In fact the number of its degrees of freedom would be effectively reduced to three. The very simplest type of molecule is like this; *but it must have at least three degrees of freedom*—corresponding to the three co-ordinates of its centre of mass—and the ratio of its specific heats would be

$$\frac{c_p}{c_v} = 1 + \frac{2}{n} = 1\tfrac{2}{3}.$$

This is the greatest value which the ratio of the specific heats of a gas can reach. It is just what is found experimentally for gases and vapours whose molecules are believed to be monatomic. Those gases which have two atoms in the molecule behave as if they had five degrees of freedom and the ratio is $1\frac{2}{5}$, e.g. hydrogen, oxygen and nitrogen.

FREE PATH PHENOMENA

In order to avoid the rather formidable mathematical problems which quickly presented themselves in the early days of the kinetic theory, Clausius and his contemporaries made many simplifying, though quite reasonable, assumptions. The relationship (IV—1), for example, was reached by completely ignoring the fact that a molecule is something more than a mere material point. He then advanced to the point of regarding a gas as constituted of spherical molecules, all having the same diameter, σ. While imagining the molecules to be moving about in all directions he made the simplifying assumption that they all moved with the same speed, which he took to be $\bar{c}$, the average of their actual speeds. Let us follow Clausius rather naïvely at first. Think of one particular molecule, A, travelling among the others, B, which at first we shall rather crudely imagine to be at rest. The molecule A has a diameter, σ, and travels a distance, $\bar{c}$, in a second. It must therefore collide with all the molecules B, which are in a volume equal to

$$\pi\sigma^2\bar{c},$$

that is to say it must make

$$\pi\sigma^2\bar{c}n$$

collisions in one second, if n is the number of the molecules, B, in one cubic centimetre. Clausius did the calculation more perfectly than this. Instead of crudely imagining all the molecules, B, as at rest, he worked out the average speed with which A *overtakes* the others and found it to be $4\bar{c}/3$. So the amended calculation of the number of collisions per second must give

$$\pi\sigma^2 n \times \frac{4\bar{c}}{3}.$$

If now we divide $\bar{c}$, i.e. the average distance a molecule like A moves in a second, by the (very big) number of collisions it makes, we get the average distance it travels between two consecutive collisions. This is called its MEAN FREE PATH and is usually represented by λ. Consequently λ is equal to

$$\bar{c} \div \frac{4\pi\sigma^2 n\bar{c}}{3}$$

or

$$\lambda = \frac{3}{4\pi\sigma^2 n} \qquad \text{(IV—5)}$$

The conception of the mean free path is one of Clausius' chief contributions to the kinetic theory. Clerk Maxwell, making fewer approximations than did Clausius, found the more accurate but not very different formula

$$\lambda = \frac{1}{\sqrt{2}\pi\sigma^2 n} \qquad \text{(IV—5A)}$$

by the application of his famous distribution law, which gives the relative frequencies of all possible velocities among the molecules.

Clausius' expression for the mean free path of a molecule made it possible to give quantitative expression to the COEFFICIENT OF VISCOSITY, or briefly the VISCOSITY, of the gas. Viscosity is just another name for internal friction. The velocity which is represented by c is the velocity of agitation of the molecules. Its average value is quite big, even when the gas as a whole is not in motion; but when the gas is flowing along a pipe there is superposed on the velocity of agitation, c, of every molecule the STREAM VELOCITY of the gas. Now the molecules close to the wall of the pipe have a very small, almost zero, stream velocity. They are hindered by frequent contact with the wall. These molecules, with small stream velocity, are often brought by their velocity of agitation (which is quite big) into the interior of the pipe and they naturally tend to reduce the stream velocity of the molecules there. The molecules in the interior (with their big stream velocity) are frequently brought near to the wall of the pipe and tend to add to the stream velocity of the molecules there. In fact these occurrences *constitute friction* in the gas, tending to retard the faster moving portion of the gas in the interior and to accelerate the more slowly moving portion near the wall. The measure of the viscosity, which we shall call η, is the frictional force over the unit area (in the gas) when the gradient of the stream velocity is unity, and a very simple calculation, which cannot be given here, shows that

$$\eta = \tfrac{1}{3} nmc\bar{\lambda},$$

or, substituting Clausius' expression for λ,

$$\eta = \frac{m\bar{c}}{4\pi\sigma^2} \qquad \text{(IV—6)}$$

Maxwell noticed something extraordinary about this expression.

It is quite independent of the pressure of the gas. This had never been suspected and Maxwell himself confirmed its correctness by experiments which he carried out in the garret at the top of his house (8 Palace Gardens, Kensington).* But (IV—6) contained another prediction: $\bar{c}$ is obviously proportional to the square root of the temperature, so that (IV—6) suggests the proportionality of viscosity and square root of temperature. This is not confirmed by experiment—unless indeed the temperature is sufficiently high—but turns out to be more nearly proportional to $T^{\frac{3}{2}}/(T + C)$, where C is a constant characteristic of the particular gas and known as SUTHERLAND'S CONSTANT. This deviation from what (IV—6) predicts is due to the fact that the molecules actually exert attractive forces on one another. This was not taken into account in the derivation of the formula. It is easy to see that two *slowly moving* molecules may deviate from rectilinear paths when passing one another on account of their mutual attraction, so that the effect is the same *as if* a collision actually occurred. The effect is in fact the same *as if* the diameters of molecules *become bigger at lower temperatures.* At high temperatures we should expect (IV—6) to approximate more closely to the truth and indeed $T^{\frac{3}{2}}/(T + C)$ does approach $T^{\frac{1}{2}}$ when T gets very great.

The conduction of heat in gases and the diffusion of one gas into another are free path phenomena; but they can only be mentioned here on account of the vast number of other things of even greater importance. Perhaps one should mention that the simple theory leads to

$$K = \eta c_v \qquad \text{(IV—7)}$$

where K is the thermal conductivity of the gas, η is its viscosity and c_v its specific heat at constant volume. It falls short of complete success in much the same way as does (IV—6).

THE THEORY OF VAN DER WAALS

An illuminating contribution to the study of the nature of gases was made in 1873 by the great Dutch physicist Johannes Diderik van der Waals (1837–1923). The main features of his theory are quite simple: In arriving at the gas equation (I—3) or

* Bakerian lecture, *Phil. Trans.* (1865).

(I—3A) the actual proper volumes of the individual molecules have been completely ignored. Van der Waals saw that when the molecular volumes are taken into account the inference must be drawn that the number of collisions between the molecules and the wall of the containing vessel is increased. The effect is the same as if the volume of the vessel had been reduced in size and a better formula than (I—3) or (I—3A) is obviously

$$p(v - b) = RT,$$

where b is a small constant. But van der Waals made a further improvement. He assumed that the molecules exert attractive forces on one another and that these forces are only appreciable over very short distances; so that in the interior of the gas the forces exerted on a molecule are on the average equally distributed in direction and magnitude. Their influence on the pressure is therefore nil. It is different with the molecules near the boundary. They must, obviously, all of them be dragged inwards and thus add to the pressure that is applied through the wall of the vessel. The number of molecules being dragged inwards, as well as the number of those dragging them inwards, must both be proportional to the density of the gas; that is to $1/v$; so that the addition to the pressure, p, can be expressed as a/v^2, where a is a small positive constant. Thus we obtain van der Waals' equation

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT. \qquad \text{(IV—8)}$$

When this equation is used to plot p against v, for various given values of T, we get curves like those shown in Fig. IV—1, assuming we have given suitable small positive values to the constants a and b. It will be seen that they strongly resemble those of Andrews (Fig. I—1); but instead of the horizontal portions, which are characteristic of Andrews' isothermals and represent the transition between liquid and vapour states, there are portions like $BCDEF$ with a double bend. This does not in the least discredit van der Waals' theory. The theory does no doubt represent approximately the states of the substance to which the constants a, b and R belong, *when all of it is in the same state at the same time*; but states represented by points such as H, where the isothermal slopes upwards from *left to*

right are *unstable*—as a little reflection will reveal—and if the whole of the substance were imagined to be in such a state at some instant, parts of it would immediately change to a stable state represented by some point on *EFG* (vapour) and the rest to a stable state represented by a point on *ABC* (liquid). The

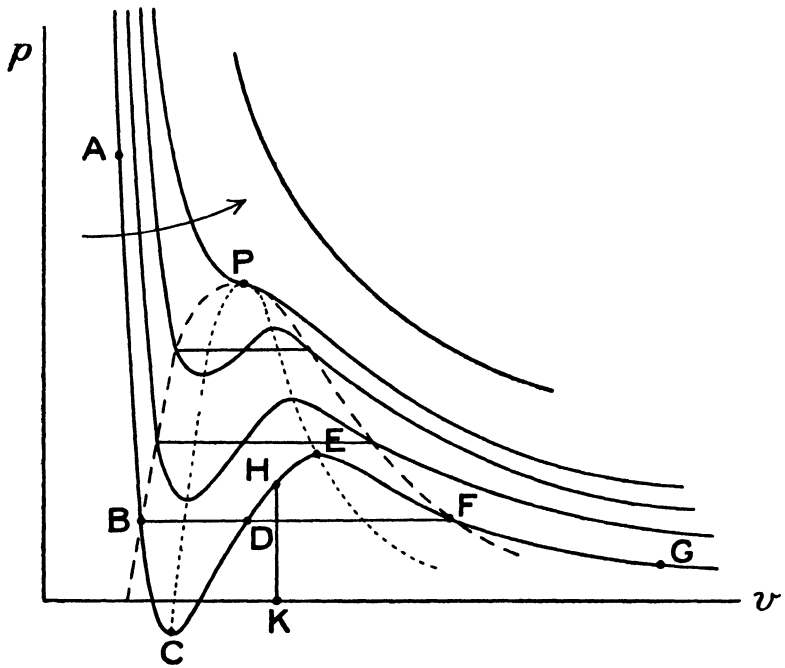


Fig. IV—1

necessary consequence of this is indeed a horizontal line, *BDE*, the situation of which very simple thermodynamical reasoning shows to be such as to make the areas *BCD* and *DEF* equal to one another.*

We have already seen that states corresponding to *EF* (supersaturated vapour) and *BC* (superheated liquid) can be realized experimentally and James Thomson was led by this in 1871 to suggest isothermals like those of van der Waals. There is a critical point represented by *P* in the diagram and the shapes of

* See W. Wilson: *Theoretical Physics*, 1, p. 317.

the isothermals are very close to those determined by experiment. The critical values can easily be shown to be

$$v_c = 3b,$$

$$p_c = \frac{a}{27b^2},$$

and

$$T_c = \frac{8a}{27Rb}.$$

When these critical values are introduced as *units* for volume, pressure and temperature, respectively, van der Waals' equation becomes

$$\left(\pi + \frac{3}{\omega^2}\right)\left(\omega - \frac{1}{3}\right) = \frac{8}{3}\tau$$

(π = pressure, ω = volume, and τ = temperature in terms of these special units). Instead of the constant, a, we now have 3; instead of b we have $\frac{1}{3}$, and R is now replaced by 8/3. They are the same for all gases (THEOREM OF CORRESPONDING STATES).

When we construct the isothermal graphs in which pv is plotted against p—the equation (IV—8) is easily adapted for this purpose by simply replacing v, wherever it occurs, by pv/p—we get curves like those in Fig. III—2 with minima lying on a parabola corresponding to the broken line in the figure.

Many other gas equations have been proposed since that of van der Waals, by Clausius, Dieterici, Callendar and others. They have all their peculiar merits and make up for particular deficiencies in that of van der Waals; but it is true, I believe, that none of them has an approximate validity for so many substances and over such a wide range of temperatures and pressures as that of the great Dutchman.

LOSCHMIDT'S NUMBER

The kinetic theory enables us to make an estimate of the number of molecules in a given mass of a gas. Though such an estimate is rather rough, it is not to be despised. It appears to have been made first by a Viennese, Joseph Loschmidt (1821–1895), about the year 1865. It is convenient to express Loschmidt's number (or Avogadro's number, as it is sometimes called) as the number of molecules in a gramme molecule (or mole) of

the gas; since when thus expressed it is the same number for all elements and compounds. As already explained the atomic weight of oxygen is 16 and its molecular weight 32; the molecular weight of hydrogen is 2·016. Consequently a mole of oxygen is 32 grammes and a mole of hydrogen is 2·016 grammes. The volume of a mole of any gas is near 22,400 cubic centimetres under normal temperature and pressure. There are two principal ways in which we may proceed to utilize experimental data to estimate Loschmidt's number. One is to estimate the mass of a single molecule, e.g. that of hydrogen, and then divide the mass of the mole (gramme molecule)—in this case 2·016 grammes—by it. The other is to estimate k, the gas constant reckoned per molecule, and divide R (for a gramme molecule, namely $8{\cdot}315 \times 10^7$ in our usual units) by it. We have already seen how $\bar{c}$, the mean velocity of agitation of a molecule—e.g. a hydrogen molecule—can be found. We can then use (IV—6) to find m/σ^2, since the viscosity of the gas is given experimentally. The molecular diameter, σ, can be estimated from the volume occupied by the gas when in liquid form, its molecules being then closely packed together. Thus knowing m/σ^2 and σ we can calculate m, the mass of a molecule, and so Loschmidt's number.

The reader may compute it for himself from the following data for hydrogen:

Mass of gramme molecule $= 2{\cdot}016$ gm.
Gas constant for one gramme molecule $= 8{\cdot}3 \times 10^7$ ergs/°K.
Viscosity at zero centigrade $= 86 \times 10^{-6}$ gm. cm.$^{-1}$ sec.$^{-1}$
Total proper volume of the molecules $= 4{\cdot}93$ c.c.

Zero centigrade may be taken to be 273°K.

From these data, with the help of a table of logarithms, I find:

Diameter of a molecule, σ $= 2{\cdot}74 \times 10^{-8}$ cm.
Mass of a hydrogen molecule $= 4{\cdot}38 \times 10^{-24}$ gm.
Loschmidt's number $= 4{\cdot}6 \times 10^{23}$.

The first fairly accurate estimate of this number was made by Max Planck (1858–1947) near the end of the year 1900 (see the chapter on Radiant Heat). We are convinced that it was fairly accurate, because it is almost the same as many later estimates made by very different methods. He calculated the value of

Boltzmann's constant, k, from the data provided by experiments on black body radiation. The best estimate of Loschmidt's number that can be given at the present moment is perhaps

$$6{\cdot}02 \times 10^{23},$$

probably in error by appreciably less than one per cent.

BROWNIAN MOVEMENT

In 1827 the botanist, Robert Brown (1773–1858), observed that the pollen grains of *Clarkia pulchella*, when suspended in water, were in constant irregular motion.* Regnault (1858) suggested that it might be due to temperature irregularities, but Christian Wiener (1826–1896) concluded in 1863 that it could not be explained in this way and he appears to have been the first to suspect that it might be due to collisions between the pollen grains and the molecules of the liquid in which the grains were suspended. By degrees it was appreciated that the movement persisted when every kind of external disturbance to which it might conceivably be due was completely eliminated. It was also found to be quite independent of the nature or chemical composition of the particles, pollen grains or what not; though its violence depended on their size or mass and Jean Perrin (1870–1942) characterized it by the sentence: *Il est éternel et spontané.* Delsaux (1877) and Sir William Ramsay (1879) thought, like Christian Wiener, that it might be due to the molecules of the suspending liquid. The complete elucidation of the nature of the phenomenon was due mainly to the theoretical work of M. Smoluchowski (1872–1917) and A. Einstein (1879–) in the earliest years of this century, followed by a remarkable experimental study of it by Jean Perrin who established beyond all reasonable doubt that the movement of small particles, suspended in a liquid—or indeed in a gas—was in fact due to collisions with the molecules of the suspending medium, which have of course the irregular motion corresponding to the temperature.

We have seen that the principle of equipartition of energy requires that at a given temperature the average kinetic energy of translation of the ultimate particles or molecules of gases,

* *Phil. Mag.* 4, p. 161 (1828).

liquids and solids must have the same definite value and we should expect it to be true of the particles participating in the Brownian movement. The average kinetic energy of translation of a particle we expect to be equal to $3kT/2$. Perrin verified the equipartition principle by actual measurements on the small spherules which he used and thus estimated Boltzmann's constant, k, or its equivalent, and so Loschmidt's number. He carried out a great variety of experiments with spherules of gamboge or mastic suspended in water by which he tested Einstein's formulae and made estimates of Loschmidt's number, which he found to range in the moderately near neighbourhood of 6×10^{23}. His work furnishes convincing evidence of the soundness of the kinetic theory and of the reality of molecules.*

OSMOSIS

Thomas Graham noticed that water passes quickly through porous membranes into alkaline solutions and solutions of crystalline substances.† In fact it was he who first suggested the classification of substances into CRYSTALLOIDS and COLLOIDS, the former associated with appreciable OSMOTIC PRESSURE; the latter with hardly any.

A great advance was made towards understanding the nature of osmosis by the German botanist, W. Pfeffer‡ (1845–1920), who experimented with cane sugar solutions contained in a vessel with a SEMI-PERMEABLE wall; that is a wall through which water, but not the dissolved sugar, can pass. The vessel was made of porcelain (like the inner vessel in a Daniell cell), which is porous, and had the semi-permeable material, copper ferrocyanide, built up inside it. The semi-permeability of this substance was discovered in 1867 by the plant physiologist, Moritz Traube (1826–1894). Pfeffer's method was to fill the semi-permeable vessel with the sugar solution and attach to it a

* Einstein's theoretical work is contained in *Investigations on the Theory of the Brownian Movement* (Methuen). Perrin's work is described in *Les Atomes* (Paris, Alcan). Earlier and very interesting work on the Brownian movement was carried out by M. Gouy: *Journ. d. Phys.* 7, p. 561 (1888).

† T. Graham: "On Osmotic Force," Bakerian Lecture, *Phil. Trans.* (1854).

‡ *Osmotische Untersuchungen* (1877).

suitable mercury manometer. The vessel was immersed in pure water contained in a larger vessel. The water passed through into the solution until a certain excess of pressure (the OSMOTIC PRESSURE) was developed inside it. The significant result of his experiments may be expressed in the form:

$$(\text{Osmotic pressure}) \times (\text{Volume of solution})$$

remains constant at a given temperature—at least approximately. The explanation of the phenomenon is really very simple. By way of illustration imagine two rooms connected by one or more very small and narrow open doors; but otherwise cut off from the rest of the world. Suppose now that a large number of Lilliputian creatures are distributed in any sort of way (to begin with) in one or both rooms. Imagine them to be blindfolded and continually moving about in all directions at random and of course now and again colliding with the walls or passing through the open doorways. Lastly, imagine a very small number of Gullivers in *one only* of the two rooms. They are also blindfolded and moving about in all directions in a perfectly random way. What will be the final state of affairs? Clearly the Lilliputians will, sooner or later, become uniformly distributed through both rooms, since they can pass freely through the open doors. The few Gullivers, on the other hand, will remain in the one room, the doorways being too narrow to let them pass through, and will become more or less uniformly spread through it. One other thing is obvious. The Lilliputians will occasion a nearly uniform pressure—by their collisions—on all the walls; but there will be an extra pressure on the walls *in the interior* of the room where the Gullivers are—the OSMOTIC PRESSURE of the Gullivers.

An experiment similar to Pfeffer's can be carried out with two gases. A palladium tube, sealed to a longer glass tube, contains nitrogen gas which is enclosed by mercury, constituting a simple manometer, such as students use in verifying Boyle's law. An atmosphere of hydrogen is maintained round the palladium tube, the latter being kept sufficiently hot by an electric current. The hydrogen diffuses through the hot palladium, but not the nitrogen, and the end result is an excess of pressure in the inside exactly equal to that of the nitrogen if it filled the volume by itself.

Both Pfeffer and van't Hoff (1852–1911) showed* that

$$\frac{(\text{Osmotic pressure}) \times (\text{Volume of solution})}{T} = \text{Constant},$$

provided the solution is sufficiently dilute. Moreover van't Hoff noticed that the constant is, in general, equal to the *gas constant*, R, proper to the quantity of the dissolved substance, if it occupied the volume of the solution in the form of gas. It did not escape observation that in the case of dilute solutions of sodium chloride and similar electrolytes the constant was *twice* the appropriate value of the gas constant. This was to be expected because electrolytic phenomena indicate that in dilute solution sodium chloride, NaCl, for example, dissociates almost completely into Na^+ and Cl^- ions and thus has the number of its "molecules" virtually doubled.

The distinction between crystalloidal and colloidal solutions appears to be due to the fact that in the case of the former the ultimate particles of the dissolved substance are molecules or ions, whereas in the latter case they are much larger. Since the equation (IV—3), i.e.

$$pv = kNT$$

holds generally (at least approximately) for a given quantity then if the substance be a colloid its particles are large and N is consequently small; therefore p (osmotic pressure) will be small for a given volume. Examples of colloidal solutions are: solutions of silicic acid, ferric hydrate and solutions of metals, e.g. gold.

RADIOMETER PRESSURE
KNUDSEN'S ABSOLUTE MANOMETER

When two surfaces in a gas at low pressure are placed near together and are at different temperatures, forces, called RADIOMETRIC FORCES, are exerted on them. They may be accounted for in the following simple way: the molecules leaving the colder surface have the velocities of agitation, and therefore the energy and momentum, proper to the lower temperature. When the gas pressure is low enough the mean free path of the molecules is greater than the distance between the surfaces, so that the surface at the lower temperature is bombarded by molecules from

* van't Hoff: *Zeits. f. Physikal. Chem.* i, p. 481 (1887).

the hotter surface. These have of course the momentum appropriate to the higher temperature and consequently the colder surface, facing the hotter one, is receiving more momentum than it loses. In fact, it is subject to a pressure in excess of that which it would experience under conditions of uniform temperature. Similar reasoning applies to the hotter surface. Sir William Crookes devised a simple piece of apparatus to exhibit this pressure—Crookes' radiometer—and it was formerly a familiar object in the windows of opticians' shops. Martin Knudsen developed it into an absolute manometer for measuring very low pressures (of the order of 10^{-7} mm. of mercury).* A simple description of it is given in *The General Properties of Matter* by Newman and Searle (Ernest Benn). Knudsen carried out a great deal of important theoretical and experimental work on gases which cannot be described here.†

* M. Knudsen: *Ann. d. Physik.* 2, p. 809 (1910).

† See, for example, Knudsen's *Radiometer Pressure and Coefficient of Accommodation.*

CHAPTER FIVE

THE STATISTICAL THEORY OF HEAT

II. STATISTICAL MECHANICS

MAXWELL'S LAW OF DISTRIBUTION

THE kinetic theory of gases was greatly advanced and the way prepared for something much more general by Maxwell's discovery of the law of distribution of velocities among gas molecules (1866). This is the law which lays down what fraction of the total number of molecules in a gas have velocities between specified limits, e.g. between 1000 cm. sec.$^{-1}$ and 1001 cm. sec.$^{-1}$, or also what fraction of the total number of molecules have one of the components of their velocities, e.g. the X component, between specified limits. The law is very like the Gaussian law for the distribution of errors in measurements, or the very similar law according to which the shots fired at a target are distributed. Very likely these laws suggested to Maxwell his law of distribution of velocities. It will be described and its *raison d'ètre* discussed when we study the more general approach to the statistical problems which gases, i.e. assemblages of particles or molecules, and other assemblages have suggested.

BOLTZMANN'S H THEOREM

The great Viennese mathematical physicist, Ludwig Boltzmann (1844–1906), encountered certain difficulties in studying Maxwell's law, some of which, I think, were due to wrongly framed questions; but one of his questions was certainly very important: Is Maxwell's law the only possible one, or is STATISTICAL EQUILIBRIUM possible for more than one law of distribution? In order to appreciate Boltzmann's method of finding the answers to these questions let us imagine, for the sake of argument, *any* kind of distribution of velocities among the molecules at the outset. For example a certain fraction of the total number

of molecules have their X components of velocity between

1000 cm. sec.$^{-1}$ and 1000·1 cm. sec.$^{-1}$,

their Y components between

1500 cm. sec.$^{-1}$ and 1500·15 cm. sec.$^{-1}$

and their Z components between

1200 cm. sec.$^{-1}$ and 1200·05 cm. sec.$^{-1}$.

We might represent this fraction by

$$f \times 0{\cdot}1 \times 0{\cdot}15 \times 0{\cdot}05,$$

where f is some function of the numbers 1000, 1500 and 1200. This is the appropriate way of representing such a state of affairs, because if we were to imagine any one of the small ranges, say the Y one, to be doubled so as to become 0·3 instead of 0·15, the fraction of the molecules involved would obviously be doubled, so long of course as the range is small. To take an illustration: Think of a large population, approaching 50 millions; the fraction of the total number of individuals with ages between 40 years and 40 years plus three weeks might be expected to be very nearly three times the number of those whose ages lie between 40 years and 40 years plus one week; so that we might express such a fraction as

$$f \times \text{range in weeks},$$

where f depends, according to some law, on 40. This was the way in which both Maxwell and Boltzmann dealt with gases and the latter was led to construct a certain mathematical function with these quantities, f, which he represented by the symbol, H*, and I shall try to give at least a picture of it nearer the end of this chapter. Meanwhile it may be said that however the distribution function, f, may change in consequence of the motions and collisions of the molecules the function H is so constructed that it *cannot increase*: it is only possible for it to become smaller or to remain unchanged. Obviously the steady state of the gas, i.e. the state of STATISTICAL EQUILIBRIUM—which of course must be identified with the state of temperature equilibrium—will have set in when H has its smallest value and Boltzmann found that this required that Maxwell's law should hold. Maxwell's law and other laws appear to be *sufficient* for statistical equi-

* Vide L. Boltzmann: *Gastheorie* (Johan Ambrosius Barth, Leipzig, 1896).

librium; but, as the H theorem indicates, Maxwell's is *necessary*. The most important thing to be said about the H function is that, provided we change its sign—i.e. put a minus sign in front of it—it can be identified with the natural logarithm of the probability of the state of the gas. The function H *without* a minus sign in front of it is already negative, consequently *minus* H is positive. We feel fairly safe in assuming that a gas, or any assemblage of an enormous number of things, will change from a less probable to a more probable state and settle down in the most probable one. For a gas this means that *minus* H has its biggest value. In fact H (with the minus sign) can be identified with the ENTROPY of the gas. The appreciation of this by Boltzmann and his successors marked the beginning of that wonderful generalization of the gas theory called STATISTICAL MECHANICS, which is the real basis of thermodynamics and the creation of which was largely the work of the New England genius, Josiah Willard Gibbs.* It may be added parenthetically that Willard Gibbs made many important contributions to thermodynamics outside those relating to statistical thermodynamics. He discovered the famous PHASE RULE as early as 1876. Much of his greatest work was published in the *Transactions of the Connecticut Academy* and did not become known till much later.

The investigations of Maxwell, Boltzmann and their predecessors were concerned with a very special kind of assemblage, namely, a number of particles (molecules). In theoretical investigations in physics it sometimes happens that a very general kind of approach to some problem is easier than the study of a special case, and the more general subject of statistical mechanics is in some ways easier than the direct approach to the particular case of it which the kinetic theory of gases exemplifies. In this more general study Boltzmann's H theorem, or its equivalent, emerges naturally and dominates the subject.

ASSEMBLAGE WITH NO INTERACTION

Let us give our attention at first to an assemblage of mechanical systems which are all exactly alike and CONSERVATIVE. This means that each system in the assemblage has kinetic and

* J. Willard Gibbs: *Elementary Principles in Statistical Mechanics* (1901).

potential energy and that the sum of the two is constant. It is therefore implied that the systems *do not interact* with one another. As an illustration let us think of a vast number of ideal simple pendulums, each vibrating in a plane and sufficiently separated from one another to make interaction impossible. We shall not worry about their positions. Their points of suspension are fixed and sufficiently far apart. The state of such an assemblage at any moment may be represented on a very simple diagram (Fig. V—1). Each system in the assemblage, i.e. each pendulum, is represented by a point, like P, on the diagram.

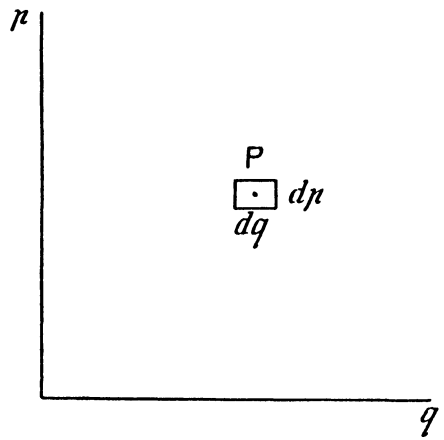


Fig. V—1

The co-ordinates of this point are q, the distance of the bob of the pendulum (+ or − as the case may be) from its position of equilibrium, and p, its momentum, mv (likewise + or −). The position of each point, P, will be continually changing (unless the pendulum is not in motion at all). In fact in our illustration of the pendulums each point will travel along a curve approximating to an ellipse. There is, we may suppose, an immense number of these points P, each representing a different pendulum. Those points which, at some instant, happen to be within the small element $dp \times dq$—and we are supposing an enormous number of them even in such a small element—will be in stream motion in some direction and it can be demonstrated that in the course of their motion *they will continue to occupy the same area* ($= dp \times dq$). This is called LIOUVILLE'S

THEOREM. Another remarkable thing can be demonstrated: there will be STATISTICAL EQUILIBRIUM, i.e. the number of the points P in a particular element—any one that may be chosen in some fixed position on the diagram—will remain unchanged, provided the number of points P per unit area at every place on the diagram is *any* function of their energy that you may like to select. For example, there will be statistical equilibrium if the density of the points P is proportional to corresponding energy. There is in fact *no unique* law of distribution when statistical equilibrium exists. This justifies Boltzmann's question: Is Maxwell's law the only possible one? *It is only when there is some interaction between one system and another that there is a unique law of distribution.*

ASSEMBLAGE OF INTERACTING SYSTEMS

The problem of an assemblage of interacting systems is not so terribly difficult if the interaction is slight. Keeping to the illustration of the pendulums, let us suppose that now and again a pendulum collides with one of its neighbours. While the two pendulums are interacting they constitute a compound system. We are going to deal with the case where the interaction is so rare that the number of such compound systems is extremely small and we may equate the total energy in the assemblage to the sum of the energies of individual free pendulums. In these circumstances the following law of distribution is found to be necessary and sufficient for statistical equilibrium. The fraction of the total number of systems (pendulums) whose representative points are in $dp \times dq$ is expressed by

$$Be^{-E/\Theta}\, dp\, dq. \qquad \text{(V—1)}$$

Note that $Be^{-E/\Theta}$ is the function f already mentioned. E is the mean energy of a system in $dp \times dq$; while B and Θ are constants which depend on the total energy of the assemblage. Each of our pendulums has (we have assumed) only one degree of freedom—one q and its corresponding p. For several degrees of freedom the formula (V—1) is correspondingly generalized.

The distribution represented by (V—1), or its gene ralization

was called by Willard Gibbs a CANONICAL DISTRIBUTION. It includes Maxwell's law as a special case—the case where E is the kinetic energy of a molecule in an assemblage of molecules (gas). Willard Gibbs called Θ the modulus of the distribution. It can be identified with temperature; in fact

$$\Theta = kT,$$

k being Boltzmann's constant.

The method of arriving at (V—1) is, roughly, as follows: We find an expression for the *probability* of *any* state of the assemblage. We take its logarithm, ψ, because its additive property introduces simplicity and we arrive at (V—1) by inquiring under what circumstances ψ has its greatest value. We can identify ψ with $-\mathrm{H}$. It is easy to show that

$$\text{Entropy}\,(\varphi) = k \times \psi,$$

which explains why k is sometimes called the ENTROPY CONSTANT.

Of course statistical mechanics changes in some ways when we give our attention, for example, to an electron gas, or to a photon gas (black body radiation); but happily the general statistical method remains unchanged when we replace Newtonian mechanical principles by those of quantum mechanics.

CHAPTER SIX

LIGHT

PRELIMINARY EXPLANATIONS

SUCH words as "light", "heat" and "sound" have a subjective significance which does not belong to the parts of physical science which they label. The objective phenomena which we study under these titles did indeed first force themselves on our attention through our faculties of sight, touch and hearing; but physics is mainly, or perhaps one should say only, concerned with the purely objective phenomena which convey to us our knowledge of the physical world. These phenomena constitute the mechanism, as it were, by means of which all the information that goes to build up natural science is conveyed to us. It would not indeed be inappropriate to define the scope of physics as the study of the phenomena which enable us to observe the world.

A blind man might very well carry out "optical" experiments. He might use a spectrometer the telescope of which is provided with a thermo-junction in place of the usual fine hair in the eye-piece. He would be able to recognize the coincidence with the image of the slit, which constitutes a spectral line, by means of a suitable amplifying device, which would enlarge the consequent thermo-current so that it might, for example, ring a bell. He would be able to determine angular values accurately by rotating drums. In fact we are not concerned in physics with any of those subjective concomitants, such as sight, which, though immensely convenient and helpful, are not essential in the observation of the phenomena with which we associate them. The essential part played by one or other of the senses is confined, one might say, to reading the record which the apparatus delivers.

If I may anticipate the description of the work of Clerk Maxwell and his successors, which appears in a later chapter, "light", in its widest objective sense, may be said to embrace electromagnetic waves ranging from the wave-length infinity

through that of broadcasting waves, infra-red, visible light, ultra violet, X-rays and gamma rays to the wave-length zero at the other extreme. It is characterized by a unique and very great velocity of propagation in free space (i.e. *in vacuo*). Before Clerk Maxwell produced his electromagnetic theory, light was believed to be an undulatory phenomenon in a medium called the AETHER, with properties like those of an elastic solid. Sound has always, at least since Newton's time, been regarded as an undulatory disturbance in ordinary elastic material media, more especially in air and gases, propagated in the form of a *dilational* wave; but light waves are of the transverse or distortional type. The aether was destined, as we shall see, to raise formidable difficulties, first of all for the elastic solid theory and later for the much more satisfactory Maxwellian theory.

The most prominent of the interesting features of light are:

(i) The immense speed with which it travels, and

(ii) the fact that it travels freely through empty regions, meaning regions which are devoid of everything which the exhausting devices of the physicist can remove and which, I believe, are in fact empty while no light is passing through them; though this view could hardly have been entertained even so recently as thirty years ago.

VELOCITY OF LIGHT—FIZEAU

The earliest measurement of the velocity of light, in which astronomical distances or phenomena were not used, was made by H. L. Fizeau (1849),* who succeeded in a comparatively simple way in measuring the very short time required by light to travel the double distance between Suresnes and Montmartre, 17·266 kilometres in all. The time-measuring device which he used was a toothed wheel with 720 teeth round its circumference. This was made to rotate by a suitable clockwork in such a way that the number of complete revolutions per second could be determined. A rough description of Fizeau's procedure is as follows: A beam of light, travelling in a direction parallel to the

* A splendid account and critical discussion of the whole series of measurements (exclusive of astronomical measurements) of the velocity of light in free space is contained in a publication by N. Ernest Dorsey: *Trans. American Phil. Soc.*, New Series, Vol. 34, Part I (1944).

axis of the wheel, was projected so that it was alternately blocked by a tooth of the rotating wheel or permitted to pass when a gap between two teeth replaced the tooth. The light which emerged between two teeth (at Montmartre, say) travelled on across Paris and was reflected by a mirror at Suresnes (8·633 kilometres away). This reflected beam was received by the observer (at Montmartre) if the wheel was either at rest or, alternatively, rotating so fast that, in the time taken by the light for the double journey, the next gap between two teeth, or indeed any succeeding one, had just managed to occupy the place of that through which the light had originally passed. In fact the returning light was seen by the observer when the wheel was rotating at certain definite speeds. There were of course intermediate speeds when the returning light was blocked by a tooth which had just managed to reach the place of the gap, through which the light had started, during the time it took for its double journey.

The calculation of the speed of the light from the observations was very simple. For illustration let us take the following case: The wheel is gradually speeded up until at last the observer sees the light which has returned, i.e. he sees it for the second time. The first time was when the wheel was moving so slowly that the light gets back before the gap has gone. Clearly the time the light has taken to do the double journey is just that in which a gap gets into what was the position of its predecessor when the light started off on its journey. Suppose now Fizeau found that his wheel was making 25 turns per second. This would in fact be very nearly what he did find. He did not of course have to follow directly a speed like 25 turns per second. He could, for example, observe another wheel of his clock work, rotating with a known fraction of the required speed. For *one* turn per second, the time required by a gap to get into the place of its predecessor would be 1/720 second, and this must be divided by 25, thus giving 1/18,000 second. The total distance is 17·266 kilometres and therefore in one whole second the light travels

$$17{\cdot}266 \times 18{,}000,$$

or approximately 310,800 kilometres per second. Fizeau actually found 315,000 kilometres per second, or

$$3{\cdot}15 \times 10^{10} \text{ cm./sec.},$$

which was creditable enough for the first direct method; but we know now that Fizeau's result is a bit too large.

The latest and best measurements lead to the velocity (*in vacuo*)

$$2{\cdot}99773 \times 10^{10} \text{ cm./sec.},$$

and it is unlikely to be less than

$$2{\cdot}99763 \times 10^{10} \text{ cm./sec.}$$

or greater than

$$2{\cdot}99783 \times 10^{10} \text{ cm./sec.}^*$$

The description just given of Fizeau's measurement is very rough. Had the experiments been attempted *exactly* in the way described, the light would have spread so much that the returning portion of it would have been too faint to be seen at all.

The light was in fact concentrated practically at a point between two teeth of the wheel by a suitable converging lens system and, after passing on, was rendered *parallel* by another such lens system, so that it went on its journey as a parallel beam and was thus prevented from spreading out in all directions.

Similarly the optical system at the reflecting mirror end made the light return as a similar parallel beam. Finally it converged to a point (nearly)—between two teeth when the speed of the wheel was suitable—and was reflected by a plane piece of glass to the observer's eye.

FOUCAULT'S METHOD

As early as 1834 Charles Wheatstone used a rotating mirror to measure what he interpreted to be the velocity of electricity along a cable, and he is said to have had the idea of applying this device to measure the velocity of light; but the performance of this was left to J. L. Foucault (1819–1868) who carried out a successful series of determinations between 1850 and 1862.† The principle of the mirror method is the measurement of the short time involved by observing the small angle through which a rapidly rotating mirror turns while the light travels outwards from it over a measured distance and back again. A beam of light was directed to the rotating mirror, so that when the latter

* N. Ernest Dorsey, *loc. cit.*

† Foucault: *Comptes Rendus*, 30, p. 551 (1850).

was in a suitable position, it reflected the light across the measured distance to a fixed mirror, which in turn reflected it back again. Meanwhile the rotating mirror has turned through a small angle, so that the observer receives the light along a slightly different direction. The measurement of the angle between the two directions gives the angle through which the mirror has turned; in fact this latter is one-half the former, in virtue of the law of reflection: angle of incidence equals angle of reflection. Foucault overcame his quite formidable difficulties with great ingenuity and obtained the final result:

$$2{\cdot}98 \times 10^{10}\,\text{cm./sec.}$$

The next to measure the velocity of light was again a Frenchman, Cornu, whose measurements were made between 1872 and 1876.* His earliest result was

$$2{\cdot}984 \times 10^{10}\,\text{cm./sec.}$$

which is equivalent to

$$2{\cdot}985 \times 10^{10}\,\text{cm./sec.}$$

in vacuo. His repetitions of these measurements in 1874 and 1876 led him to the final value

$$3{\cdot}004 \times 10^{10}\,\text{cm./sec.}$$

(reckoned for *vacuo*).

Since Cornu's time the velocity of light has been determined many times with steadily increasing precision. About half of these were made by the great American physicist, Albert Michelson (1852–1931) who began as long ago as 1878. He used the rotating mirror method. In his later work he had, instead of the old rotating mirror, a rotating prism, originally eight-sided, but in his last determination 32-sided, each side being a mirror. It made of course many hundreds of revolutions per second; the number per second being determined by stroboscopic methods ($\sigma\tau\rho\acute{o}\beta o\varsigma$ = a whirling round).

Sometimes at the cinema we observe the wheels of an automobile to be quite still or perhaps even moving backwards. In the former case evidently the interval between two consecutive photographs is equal to the time taken by a spoke to get into the position its predecessor had occupied and we might determine this interval from the speed of the automobile and the dimen-

* Cornu: *Annales de l'Observatoire de Paris, Memoires*, Tom. 13 (1876).

sions of the wheel. This would be a stroboscopic method. The light travelled a distance of the order of 5 miles along a pipe which was exhausted as far as practicable, thus reducing the correction needed to get the velocity *in vacuo*. Michelson's method of observation was to arrange that the returning light could only be seen by the observer when a face of the prism occupied the place of its predecessor during the interval of time required by the light to traverse the measured distance in both directions.

The methods used in the very latest measurements may be described as improved forms of the original method of Fizeau. The first to use such methods were Karolus and Mittelstaedt (1929). Then came the American, Anderson, and lastly, A. Hüttel. Instead of the mechanical device of the toothed wheel, a Kerr cell was used. Starting with plane polarized light this was made to pass through the Kerr cell (in which an electric field is oscillating with so-called radio frequency). The effect of this on the light was to cause its state of polarization to vary periodically —from plane polarization to elliptical, circular, elliptical, plane and so on. After travelling some distance—which may be a few yards instead of miles, on account of the high frequency of the electrical alternations in the Kerr cell—the light passed through another, suitably oriented Kerr cell, then through a Nicol's prism to the observer's eye, or to a photo-electric cell. According to the phase of the field in the cell, the light may be blocked altogether (as if a tooth of Fizeau's wheel were in the way) or it may pass through. As the length of the path was changed the response of the photo-electric cell was periodic, the corresponding distances being traversed by the light during the very minute (and, of course, known) period of the Kerr cell.

An alternative procedure was to keep the distance traversed unchanged and vary the period of the Kerr cell.

The final estimate made from all these later methods is that already given above.

GROUP AND PHASE VELOCITY

Any one can convince himself, in a very simple way, that *two* velocities are, in general, associated with wave propagation. If a tiny pebble be dropped into (initially) still water a group of

ripples spreads out with a definite velocity. Its boundaries are not sharply defined; the waves (crests and troughs) fade out gradually in front and behind and it will be seen that the crests and troughs *travel through the group*, new crests being formed behind and fading out in front after traversing it. They travel *faster* than does the group, though it may be remarked parenthetically that, with sufficiently short waves-length, they travel more slowly than the group. There is a particular wave-length for which PHASE VELOCITY (i.e. the velocity of the crests and troughs) and GROUP VELOCITY are equal. So in the case of light we have to distinguish between these two velocities. It seems to have been appreciated first by Sir G. G. Stokes. The relationship between them was first given by Lord Rayleigh and is explained later. Naturally we ask which of these the experimental measurements of the velocity of light yield. The answer is the GROUP VELOCITY. One reason is that crests and troughs are unable to stay the whole course, unless group and phase velocities happen to be equal to one another. Group and phase velocities are always equal in a medium in which there is no dispersion—as we shall see—e.g. in free space. Moreover none of the methods used is capable of picking out a crest or trough and finding how long it takes to travel over a measured distance. Even stellar aberration gives the group velocity—though it happens that in free space phase velocity is equal to group velocity. It is the *phase velocity* which is referred to in such a statement as

$$\frac{\text{Velocity of light } \textit{in vacuo}}{\text{Velocity of light in glass}} = \text{Refractive index of glass.}$$

CONVECTION OF LIGHT

Already in the time of Fresnel (1818) experiments had been carried out by the French mathematician and physicist, Francois Arago (1786–1853), to ascertain whether the refraction of the light from a star (by a prism) was influenced by the motion of the earth through the luminiferous medium. Finding no such influence he applied to his protégé, Fresnel, for an explanation and the answer he received is one of the historic documents of physical science.* The obvious explanation would appear to be

* Larmor: *Aether and Matter*, p. 320 (Cambridge University Press).

that the aether travels along with the earth, but, as Fresnel wrote: "*Il parait impossible d'expliquer l'aberration des étoiles dans cette hypothèse.*" The whole subject is discussed in the chapter on the aether. Meanwhile it may be said that Fresnel came to the conclusion that the motion of the earth, or of any body, transparent or other, does not influence the aether outside it—the aether remains "at rest". In the case of a transparent body which is travelling with some velocity, v, relative to the quiescent aether outside it, and through which a beam of monochromatic light is travelling, in the same direction shall we say, Fresnel inferred that the velocity of the light relative to the quiescent aether was given by

$$\frac{c}{n} + \left(1 - \frac{1}{n^2}\right)v, \qquad \text{(VI—1)}$$

in which expression n means the refractive index of the transparent material. We should *now* say that v is the velocity of the transparent body relative to the *observer* and it may be remarked that we now appreciate that it is the *phase velocity* of the light which is expressed by (VI—1). The crests and troughs are, as it were, convected by the moving medium and the factor $(1 - 1/n^2)$ is called FRESNEL'S CONVECTION COEFFICIENT or DRAGGING COEFFICIENT. This prediction of Fresnel's (VI—1)—which will be expressed with rather more precision later—was tested by Fizeau in 1859* and again in 1886 by Michelson and quite recently by Zeeman. All three of them confirmed Fresnel's prediction. Fizeau's method is illustrated by Fig. VI—1. A beam of monochromatic light starts from a slit, S, is then reflected from a plate of glass, P, to the converging lens, A, which renders it parallel. The upper part of the parallel beam travels along ab, through which water is flowing in the same direction as the light is travelling. The parallel beam is made to converge to a line (image of the slit) on the mirror, M, and on reflection the part of it which has already traversed ab now passes along cd, again passing through the water which is flowing in the same direction. The lower portion of the beam from S takes in succession the paths dc and ba, travelling in the opposite

* Fizeau: *Ann. d. Chimie et de Physique*, Ser. 3, Tom. 57, p. 385 (1859); Michelson and Morley: *Amer. Journal of Science*, 31, p. 377 (1886).

sense to that in which the water is flowing. The two portions are reunited at S', where interference fringes are produced and can be seen by the observer. The amount by which the fringes are displaced as the velocity of the water is raised from zero to some final value, v, enables the change produced in the *phase* velocity of the light, in either direction, to be computed, since c, the velocity of light in the free aether, and n, the refractive index of the water, are known in advance.

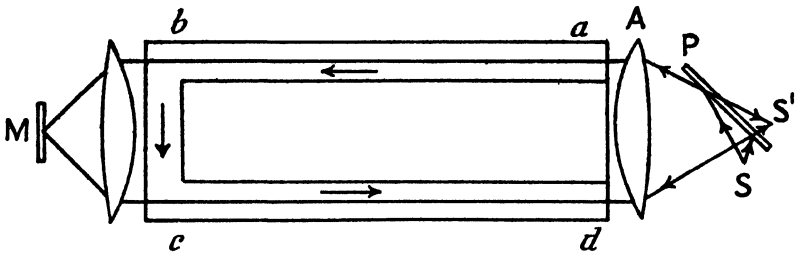


Fig. VI—1

Fresnel seems to have regarded the problem of the convection of light in the following way: He had to assume that the aether outside material media is at rest in order to account for stellar aberration and his theory of light—as indeed any theory which regarded light as a wave propagation in a material, or quasi-material, medium—required the formula

$$\text{Velocity of light} = \sqrt{\frac{\text{Elasticity of aether}}{\text{Density of aether}}}$$

the elasticity being, of course, that which is associated with a shearing stress, often called RIGIDITY. He supposed this elasticity to have the same value for the aether in free space as for the aether in any material medium and consequently he had to account for the different velocity of light—we should now say phase velocity—in different media by supposing the density of the aether in material media to vary from one medium to another. The present-day use of the term "optical density" has its origin in this hypothesis. For free space, therefore, according to Fresnel,

$$c = \sqrt{\frac{e}{\rho_0}},$$

e being the elasticity and ρ_0 the density of the aether in free space. Further

$$u = \sqrt{\frac{e}{\rho}},$$

where u is the velocity of light in any material medium, for example in water, at rest, and ρ is the density of the aether in it. Therefore

$$\frac{c}{u} = \sqrt{\frac{\rho}{\rho_0}},$$

or

$$n^2 = \frac{\rho}{\rho_0}, \qquad \text{(VI—2)}$$

n being the refractive index of the transparent medium. Now think of this transparent medium as a rectangular block of glass, *abcd*, its end faces *ab* and *cd* having each the unit area (Fig. VI—2). It is travelling, we suppose, with the velocity, v, in the

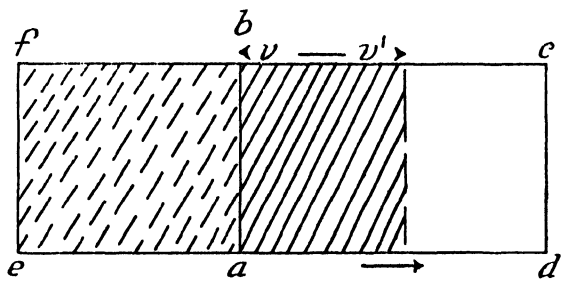


Fig. VI—2

direction of the arrow, and we represent the velocity of the convected aether in the glass by v' (still unknown of course) in the same sense as v. Since the aether outside the block is at rest, a quantity of the aether occupying the volume *efba*, the length of which is v, must be left behind in one second; that is to say the quantity

$$v \times \rho_0$$

But this emerges from the block, and as the aether in the block has the velocity v' it is being left behind at the rate $v - v'$, so that the quantity of the emerging aether must be equal to

$$(v - v') \times \rho,$$

ρ being the density of the aether in the glass. We have therefore

$$v \times \rho_0 = (v - v')\rho,$$

and consequently

$$\frac{\rho_0}{\rho} = 1 - \frac{v'}{v},$$

and by (VI—2),

$$\frac{1}{n^2} = 1 - \frac{v'}{v},$$

so that

$$v' = v\left(1 - \frac{1}{n^2}\right)$$

The velocity, v', of the aether in the glass must be added on to c/n, which is the velocity of light in the resting material; just as in the case of sound travelling in the direction of the wind, the velocity is that in still air plus the velocity of the air (wind). The result, therefore, is that expressed by (VI—1). Thomas Young described this view of Fresnel's about the relationship between the aether and moving materials by comparing it to the wind blowing through a grove of trees.

Fresnel's formula has turned out to be correct; but his method of arriving at it is open to criticisms which need not be discussed here,* but that based on Fermat's principle appears to be quite sound, if we accept the premiss of the validity of Fermat's principle in the special form which Arago's experiments appear to justify.

One of Fresnel's predictions, based on (VI—1), was that the angle of aberration should be independent of the nature of the transparent material—whether air, water, or what not—filling the observing telescope. This prediction was verified about 1870 by Sir George Airy, the Astronomer Royal, who had a telescope filled with water (and its optical system suitably adjusted of course). He found the same aberration as that observed with normal telescopes.† Sir Joseph Larmor records‡ that Fresnel

* One which occurs to me is that the density of the aether in any material medium would seem to depend on the wave-length of the light.

† G. B. Airy: *Proc. Roy. Soc.*, 20, p. 35 (1871); *Phil. Mag.*, 43, p. 310 (1872).

‡ *Aether and Matter*, p. 10.

pointed out that some one named Wilson (about whom I have been unable to find any information) had shown that no effect was to be expected on the corpuscular theory as well as on the undulatory theory.

DISPERSION

This is the name of the phenomenon first investigated by Newton when, as he wrote in his *Opticks*, "In a very dark Chamber, at a round Hole, about one-third Part of an Inch broad, made in the Shut of a Window, I placed a Glass Prism. . . . ".* The refractive index of a transparent substance, water or glass for example, depends on the colour, or, speaking more precisely, on the wave-length of the light; so when ordinary white light is passed through a prism of glass its constituent parts of different colours are separated.† A so-called SPECTRUM is produced. In the case of a glass prism and light belonging to the visible part of the spectrum, the deviation produced is the greater the shorter the wave-length. This was long regarded as the normal type of DISPERSION, as it is called, until Fox Talbot (1840) noticed an exception to it. About 1860 le Roux observed‡ that iodine vapour transmitted only the extreme ends of the visible spectrum, red and violet, and that the refraction of the red (longer wave-length) was *greater* than that of the violet. Kundt§ observed that when a substance absorbed some part of the spectrum, the deviation or refraction of the light is greatly increased as the absorption band is approached from the long-wave side and greatly diminished as it is approached from the short-wave side. (KUNDT'S LAW). The term ANOMALOUS DISPERSION was intro-

* Newton's *Opticks*: p. 26 in the reprinted 4th Edition (G. Bell and Sons, 1931).

† I do not feel happy about the use of the word "constituent" in this context; but perhaps it is well to approach precision of statement gradually and I am hoping that the precise meaning of the sentence will emerge later.

‡ Le Roux: *Ann. de Chimie et de Physique*, Ser. 3, Tom. 61, p. 285 (1861).

§ Kundt: *Pogg. Ann.*, 1871–1872.

duced for this departure from what was regarded as normal.*

The theoretical explanation of dispersion appears to have been suggested first by Maxwell in the form of a question in the Cambridge Mathematical Tripos Examination in 1869. The same explanation was given independently by a German called Sellmeyer in 1872. Briefly it is based on the supposition that there are in the dispersing material small particles capable of simple harmonic vibrations about fixed points. These particles are to be thought of as molecules or ultimate particles of the material embedded in the aether and their oscillations are relative to the aether. When monochromatic light is traversing the material they are set in forced vibration with the same period as the light waves, which is of course different from their natural periods in general. The consequence of this is that the light traverses the material with a definite phase velocity which is determined by the two periods, that of the light wave and that of the oscillating particles. This fixes the refractive index of the material for light of a particular wave-length, and since it varies from one wave-length to another, dispersion is the inevitable consequence. The theory is very simple, but cannot be given here.† If we neglect frictional (dissipative) forces and assume only one group of these vibrating particles with a natural period corresponding to the wave-length, λ_0, *in vacuo*, then we arrive at the following formula:

$$n^2 - 1 = \frac{Nm}{\rho} \times \frac{\lambda^2}{\lambda^2 - \lambda_0^2}, \qquad \text{(VI—3)}$$

for the refractive index n; N and m mean respectively the number per unit volume and the mass of one of the vibrating particles; ρ is the density of the free aether and λ is the wave-length of the light travelling through it. Of course a substance might have two or more groups of such particles, each associated with

* Kundt made very thin prisms of silver by sputtering silver from a silver cathode in a discharge tube. These transmit light of the blue end of the visible spectrum and the refraction is *towards the edge of the prism*, indicating a refractive index less than unity and therefore a phase velocity *greater than that of light in vacuo*. This is not in conflict with the pronouncements of the theory of relativity which sets no limits to the *phase* velocity of light.

† It is presented very simply in E. Edser's text-book on *Light* (Macmillan & Co.).

its own special set of values of N, m and λ_0; in which case the expression on the right of (VI—3) would be replaced by a sum; for example, if there were two such sets the formula would become:

$$n^2 - 1 = \frac{Nm}{\rho} \frac{\lambda^2}{\lambda^2 - \lambda_0^2} + \frac{N'm'}{\rho} \frac{\lambda^2}{\lambda^2 - \lambda_0^{-2}} \qquad \text{(VI—4)}$$

The general consequences of this theory can be seen by studying the simplest case (VI—3). If we begin with λ greater than λ_0 we see that as λ gets smaller, n^2 (and consequently n) gets progressively bigger, tending to infinity as λ approaches λ_0. For illustration imagine $\lambda = 500$ and $\lambda_0 = 300$, then

$$\lambda^2/(\lambda^2 - \lambda_0^2) = 250{,}000/160{,}000 = 25/16 = 1\tfrac{9}{16}.$$

Now suppose λ to decrease to 301; then $\lambda^2/(\lambda^2 - \lambda_0^2)$ becomes 90,601/601, which is more than 150. The infinite value corresponding to $\lambda = \lambda_0$ is really due to the fact that in deducing (VI—3) frictional forces have been ignored. This defect in the original, very simple, theory was repaired by von Helmholtz.

When λ is in the *neighbourhood* of λ_0 (but $\lambda > \lambda_0$), the frequency of the light wave is close to the natural frequency of vibration of the particles (corresponding to λ_0) and we have RESONANCE, nearly, which means very strong absorption. Imagine, for example, the point of support of a pendulum to be agitated with a frequency near to the natural one of the pendulum; it will rapidly get into violent oscillation. So when the frequency in the light waves is near to the natural frequency characteristic of one of the sets of particles they will quickly develop oscillations of large amplitude; that is to say they will absorb much energy from the light waves. When λ is just a trifle less than λ_0 we see that n^2 is minus and so n is purely imaginary (in the mathematical sense of the word). The interpretation of this is that, in the narrow range of wave-lengths where n is imaginary, light cannot penetrate the material at all; it is completely reflected.

Moreover when λ is extremely short n is very nearly equal to unity. All these features of dispersion are observed. The general character of the dispersion curve (n plotted against λ) for a material with two sets of particles, with characteristic natural periods, is illustrated by Fig. VI—3. The shaded parts

show the regions of total reflection. The part *ab* illustrates dispersion, as it is exhibited by glass in the visible part of the spectrum—refractive index getting steadily less as we pass from shorter to longer wave-lengths. The "anomalous" behaviour of n is shown in the neighbourhoods of the wave-lengths λ_0 and λ_0' which correspond to the natural periods of the two groups of vibrating particles.

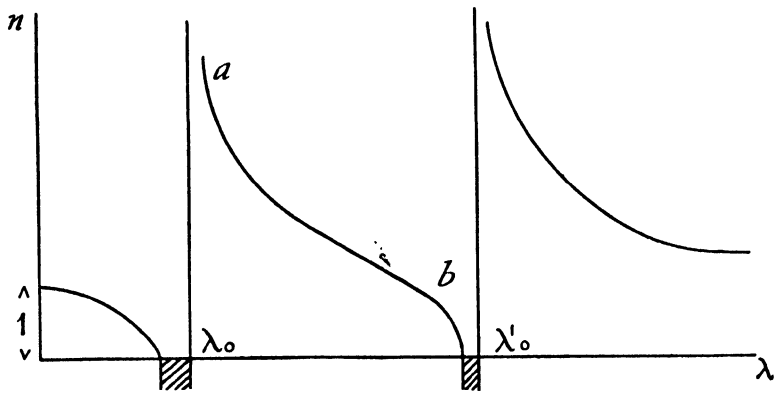


Fig. VI—3

R. W. Wood has described* a simple way of observing anomalous dispersion. A prism of cyanine is made by pressing a small fragment of it between two glass plates hot enough to fuse it. The observations are carried out by placing together the cyanine prism and a glass one; the former having its refracting edge horizontal and the latter its edge vertical. An appearance rather like that of Fig. VI—3 will be observed when white light passes through the combined prisms.

It follows from this simple theory of dispersion that

$$\frac{n^2-1}{\text{(density of material)}}$$

remains constant for the same material and the same wave-length. Consequently when n is close to unity, as with gases, $n - 1$ is proportional to the density. This is the law of Gladstone and Dale.†

* R. W. Wood: *Physical Optics*, p. 96 (1905).
† *Phil. Trans.*, p. 887 (1858) and p. 317 (1863).

The essentials of this theory of dispersion still remained when the old "elastic solid" type of theory was replaced by Maxwell's electromagnetic one. In the newer theory there are sets of vibrating IONS (charged particles) each with its characteristic natural period, instead of the vibrating particles of Sellmeyer's theory. It turned out to be even more successful than the older one and accounted for facts which the elastic solid theory could not explain. In particular it leads to the law discovered by Lorenz of Copenhagen (1829–1891) and H. A. Lorentz of Leyden (1853–1928), namely that

$$\frac{n^2 - 1}{n^2 + 2},$$

if divided by the density of the material, has the same value in its different phases, e.g. the same value for the liquid as for the vapour.*

RESIDUAL RAYS

The shaded portions of the diagram (Fig. VI—3) mark spectral regions, as we have seen, where there is the total reflection characteristic of so-called anomalous dispersion. This leads me to a remarkable investigation of the reflection of radiation belonging to the infra-red part of the spectrum (roughly between 10μ and 70μ where $\mu = 1/1000$ mm.) by Heinrich Rubens (1865–1922) and his pupils† which is of immense interest. Their method of experiment consisted in throwing a beam of radiation from a hot body such as a heated Nernst rod—this of course included visible radiation as well as infra-red; indeed the Nernst rod was a feature in the lamps once used in the London A.B.C. shops for lighting—on to the flat surface of a piece of quartz, or of rock salt, or other crystalline material. The beam of radiation was first rendered parallel by reflection from a concave mirror, at the principal focus of which the incandescent rod was placed, and after successive reflections from several plates of the particular material, was found to be

* H. A. Lorentz: *Ann. d. Physik*, 9, p. 641 (1880); L. V. Lorenz: *Ann. d. Physik*, 11, p. 70 (1880).

† H. Rubens and E. F. Nichols: *Wied. Ann.*, 60, p. 418 (1897); H. Rubens and E. Aschkinass: *Wied. Ann.*, 65, p. 241 (1898).

either monochromatic, i.e. to consist of a very narrow range of wave-lengths, or a small number, two or three, of such narrow ranges; all the rest of the radiation having been absorbed or transmitted. Rubens called this radiation which was left over after many reflections: RESIDUAL RADIATION (*Reststrahlung*). The phenomenon is just what the theory of *dispersion* predicts (strong reflexion of certain wave-lengths). Among the residual wave-lengths measured by Rubens and his pupils may be mentioned:

Quartz	.	$8{\cdot}25\mu$ and $20{\cdot}75\mu$,
Fluorspar	.	$23{\cdot}7\mu$,
Rock salt	.	$51{\cdot}2\mu$,
Sylvin	.	$61{\cdot}1\mu$.

The reader may be reminded that

$$1\mu = 1/1000 \text{ mm.},$$

sometimes called the MICRON.

These wave-lengths correspond, at any rate according to the views of that time, to the natural frequencies of vibration of the ions in the materials. We know that the relationship between wave-length in free space, velocity of light in free space, and frequency is expressed by

$$c = \nu\lambda,$$

therefore the frequency corresponding to the wave-length $51{\cdot}2\mu$ of rock salt, for example, must be given by

$$3 \times 10^{10} = \nu \times 51{\cdot}2 \times 10^{-4},$$

$$\text{therefore} \qquad \nu = \frac{3}{51{\cdot}2} \times 10^{14},$$

$$\text{or} \qquad \nu = 5{\cdot}86 \times 10^{12} \text{ vibrations per second.}$$

Rubens made use of his residual radiation, or I should say, rather, his method of producing it, in the investigation of the distribution of energy in the spectrum of black body radiation (*vide* the chapter on Radiant Heat and Quantum Theory).

DISPERSION AND SPECTROSCOPY

Spectroscopy began seriously about 1859 with R. W. Bunsen (1811–1899) and G. R. Kirchhoff (1824–1887), and their form of prism spectroscope was the prototype of all the prism spectroscopes, or spectrometers, that have been constructed since. The

spectroscope analyses the light which passes through it into its constituents of various wave-lengths. In Fig. VI—4 S represents a very narrow slit (perpendicular to the plane of the paper).

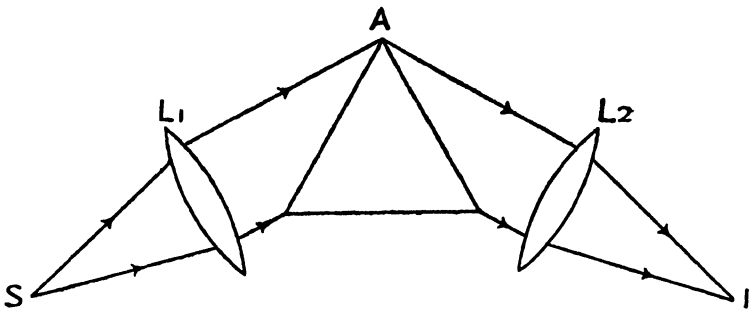


Fig. VI—4

It is illuminated with light from some source—it might be, for example, from hydrogen made to emit light by the passage of an electric current through a tube containing the gas at low pressure. The light passing through the slit, S, is made into a parallel beam by a converging lens, L_1, called a collimating lens. Now we have learned that the refractive index of the prism varies with the wave-length of the light. The light from the source (e.g. the glowing hydrogen) will emerge from the prism, A, which has its refracting edge parallel to the slit, S, as a number of parallel beams, each corresponding to a different wave-length. A converging lens, L_2, suitably placed perpendicularly to any one of these emerging parallel beams, will form a real image of the slit at I with the light of the associated wave-length. The totality of such images is the SPECTRUM of the source of light. In the special case of the glowing hydrogen, and in many other cases, these images which constitute the spectrum are discrete (for the most part), i.e. they do not overlap and the spectrum has the appearance of bright lines of various colours (emission line spectrum). Bunsen and Kirchhoff began the serious study of line spectra, the interest of the former being mainly bound up with chemical investigations. A spectrum consisting of an unbroken band—no gaps between the individual images of the slit—is called a CONTINUOUS SPECTRUM. The outstanding example is the spectrum of BLACK BODY RADIATION.

Of course, in the complete spectroscope, the slit, S, and the collimating lens, L_1, are built into a tube called the collimator; while L_2 forms the object glass of a telescope, so mounted that it can turn about a vertical axis parallel to the edge of the prism, A, and thus receive any one of the emerging parallel beams.

Bunsen and Kirchhoff noticed (what I think had been observed still earlier) that, when common salt was volatilized in the flame of the burner devised by the former, the characteristic pair of yellow lines occupied exactly the position of the dark solar line which Fraunhofer had distinguished by the letter D. Kirchhoff explained the production of the dark lines in the sun's spectrum by the hypothesis that the light from the incandescent solar interior, which would normally give rise to a bright D line (strictly speaking, D lines), passes through a colder atmosphere containing sodium vapour or the vapour of sodium compounds, in which it is absorbed. He and Bunsen confirmed this view by an experiment which has often been repeated since for the benefit of students. They successfully produced the dark D lines by passing the light from incandescent lime through an alcohol flame in which sodium was volatilized.*

EARLY APPLICATIONS OF THE SPECTROSCOPE

By 1862 Sir William Huggins (1824–1910), using a small spectroscope attached to an 8-inch telescope, had noted the positions of the more prominent lines in the spectra of about forty stars. It was he who first observed (in the case of light) the phenomenon which Doppler (1803–1853) foresaw for waves of any kind (1842). When a star is receding, its spectral lines are displaced towards the red end. It is the same phenomenon (essentially) as the drop in the pitch of a locomotive whistle when it is receding from the listener.

In 1878 Sir Norman Lockyer (1836–1920) observed certain lines in the solar spectrum that had not been seen in any terrestrial source. He ascribed them to a new element (unknown on the earth) which he named helium (from ἥλιος, the sun). It was discovered in the mineral cleveite about seventeen years

* The light from incandescent lime (limelight) was used in projection lanterns long before the carbon arc light.

later (1895) by Sir William Ramsay (1852–1916) and Sir William Crookes (1832–1919).

Victor Schumann (1841–1913), a Leipzig engineer and amateur scientist, extended spectroscopy far into the ultra-violet part of the spectrum. His house was filled with a vast variety of green plants which he thought to be good for his invalid wife. He became very interested in them and started to photograph them. He was thus led to study photography, which he very much improved, and in consequence light and spectra. By 1892 he reached the wave-length 127 $\mu\mu$.* He was an expert mechanic and made his own spectroscopes, which were provided with finely made slits, and in his latest investigations were *in vacuo*, since the light of the spectral region in which his measurements were made is highly absorbable.

* 1 $\mu\mu = 1/1000$ micron and is called the MILLIMICRON.

CHAPTER SEVEN

MORE ABOUT LIGHT

LIGHT AS AN UNDULATORY PHENOMENON

IF, in the investigation of the phenomena of light, observations of the effects it produces were confined to such as might be used to determine its intensity—such for example as the heating of a thermo-junction or the blackening of a photographic plate or the visual comparison of intensities—we should probably remain convinced that it is purely a wave phenomenon; but light produces certain other effects which we shall study later—the so-called photo-electric emission of electrons for example—which have rather forced us to regard it as a stream of quasi-particles (photons) and I believe this is the correct view to take of a beam of light. More than a quarter of a century ago the Duc de Broglie, elder brother of Louis de Broglie, the author of wave mechanics, wrote: " . . . *on serait bien prés d'une théorie corpusculaire de la lumière.*"*

I will try to explain later how this corpuscular aspect of light can be reconciled with its equally prominent undulatory characteristics. Meanwhile it is the latter with which we are concerned.

CHARACTERISTICS OF WAVES

Certain features of waves, which are important from the point of view of understanding optical phenomena and also from the point of view of appreciating the fundamentals of wave mechanics, can be illustrated by a stretched cord. We imagine the cord to be horizontal and, to avoid irrelevant complications, we confine our attention to up and down motions of the parts of the cord. The simplest kind of wave that can travel along it has the shape of a sine (or cosine) curve (Fig. VII—1 *a*). As the wave outline

* Maurice de Broglie: *Les Rayons X*, p. 157 (1922).

progresses along the cord the up and down motion of any point on it is simple harmonic—the simplest kind of vibration of all. The period of a complete up and down vibration may be called τ and the wave-length, i.e. the distance from crest to crest, may

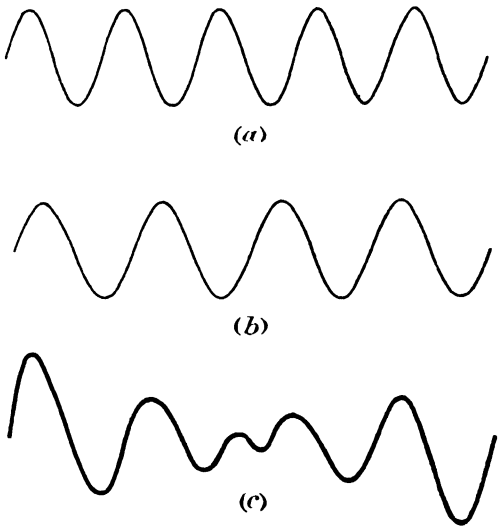


Fig. VII—1

be called λ. The phase velocity (velocity with which crests or troughs travel along) is expressed by

$$u = \lambda/\tau, \qquad \text{(VII—1)}$$

or by

$$u = \nu/\nu', \qquad \text{(VII—1A)}$$

if ν $(= 1/\tau)$ is the frequency of vibration and ν' $(= 1/\lambda)$ is the wave-number (number of waves in one centimetre). Expressed in words:

(Phase velocity) = (Frequency)/(Wave-number)

Any kind of curve into which the cord may be bent—at least any kind that concerns us—can be represented as a superposition of sine (or cosine) curves. This is a purely mathematical pronouncement (Fourier's theorem, 1807); but it may be explained here that the analysis which the prism makes of the light from a source such as a Bunsen flame in which some salt is volatilized, or of ordinary white light, is not correctly described as a decom-

position of the light, from the source, into its constituents. The prism really manufactures, from the original light, a number of beams of light, each of which has a sine (cosine) form and whose superposition would build up the original light. For example, the curve in Fig. VII—1 (*c*) is a superposition of the two sine curves Fig. VII—1 (*a*) and (*b*).

Of great interest, both from the point of view of light and from that of wave mechanics, is something which I call a SIMPLE GROUP of waves. It has approximately a sine or cosine form; but occupies a limited region—a limited length on the cord, or a limited area in the case of ripples on the surface of water, or a limited volume in the case of light waves. Its crests and troughs (or wave-fronts) are parallel to one another and the wave-length is very short compared with the dimensions of the group. It may be regarded as a superposition of sine (or cosine) waves all of which, whose amplitudes are appreciable, have very nearly the same wave-length. In fact the wave-lengths will range from

$$\lambda \text{ to } \lambda + \Delta\lambda,$$

$\Delta\lambda$ being very small compared with λ; or, what amounts to the same thing, we may think of the wave-numbers as included in the narrow range between

$$\nu' \text{ and } \nu' + \Delta\nu',$$

and associated with this will be a corresponding narrow range of frequencies between

$$\nu \text{ and } \nu + \Delta\nu.$$

So-called monochromatic light consists of such simple groups, or a superposition of such simple groups. The group velocity can be shown to be expressed by

$$v = \Delta\nu/\Delta\nu', \qquad \text{(VII—2)}$$

or, in words,

(Group velocity) = (Frequency range)/(Corresponding wave-number range)

An equivalent, though formally different, expression was first given by Lord Rayleigh, namely:*

$$v = u - \lambda \frac{du}{d\lambda} \qquad \text{(VII—2A)}$$

* *Vide* W. Wilson: *Theoretical Physics*, 1, p. 136 and 3, p. 189.

A simple group travels in a direction perpendicular to its wave fronts (in isotropic media), as if it were a rigid configuration, though its boundary is not sharply defined and it can be shown that

$$\frac{\Delta\lambda}{\lambda}\left\{\begin{array}{c}\text{is of the same order}\\ \text{of magnitude as}\end{array}\right\}\frac{\lambda}{L} \qquad \text{(VII—3)}$$

where L is the distance across the group perpendicular to the wave-fronts. Another way of saying the same thing as (VII—3) is:

$$\Delta\nu' \times L\left\{\begin{array}{c}\text{is of the same order}\\ \text{of magnitude as}\end{array}\right\}1 \qquad \text{(VII—3A)}$$

DIFFRACTION AND INTERFERENCE

When monochromatic light (i.e. light consisting of simple groups) falls perpendicularly on a screen with a slit in it, then, if the width of the slit be very small by comparison with the wave-length, λ, the waves, after passing through the slit, will proceed as shown in Fig. VII—2. If on the other hand the wave-length is a first order small quantity compared with the width of the slit, they will proceed, after passing through it, in the way indicated in Fig. VII—3. They bend round slightly, as it were, into the geometrical shadow. In the former case (Fig. VII—2) this is very pronounced. In the latter case most of the light travels straight on and there is a pronounced, but not quite sharply defined, shadow. This phenomenon is called DIFFRACTION and the way in which we have here regarded it is based on a principle due to the great Dutch mathematician and philosopher Christiaan Huygens (1629–1695)—without all the details that are contained in it. This famous principle was greatly improved by Fresnel, who took into account the phenomenon of INTER-

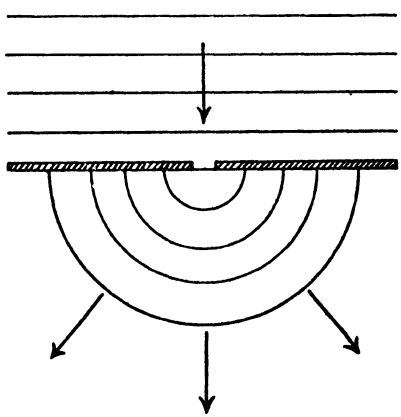


Fig. VII—2

FERENCE. For instance, when two wave crests, or two wave troughs, come together at the same time, they reinforce one another and produce an amplified crest, or trough, the resulting amplitude being the sum of the amplitudes of the individual waves. If on the other hand a crest and a trough come together at the same time their interference is destructive and the resulting amplitude is the difference of the amplitudes of the individual waves. All cases are comprehended in the single expression, or rule, that the resulting amplitude or displacement is the algebraic sum of the individual amplitudes or displacements. This is Fourier's theorem in operation—validated by the fact that light waves are described by a *linear* partial differential equation.

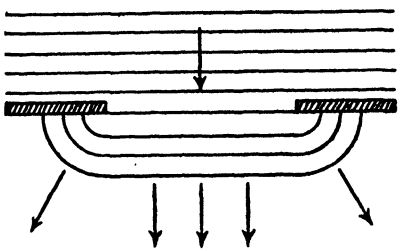


Fig. VII—3

There is the *extreme* case which may be mentioned here, namely that in which the wave-length of the light is a *second order* small quantity compared with the breadth of the slit in Fig. VII—3. In this case there is no diffraction. The light travels on in a strictly rectilinear fashion and the associated phenomena belong *par excellence* to GEOMETRICAL OPTICS (*vide* Fermat's principle).

An important application of these reflections must be explained here in order that what follows may be intelligible. We can deal with it, in its essentials, after the manner of Fresnel. Again we have a parallel beam of monochromatic light falling perpendicularly on a screen with a slit in it, the wave-length being rather small compared with the breadth of the slit (say 1/100,000 of it or less). It is illustrated in Fig VII—4. We appreciate already that *most* of the light, after passing through the slit or aperture, will travel straight on; but not all of it. Let us give our attention to the light which travels in the direction represented by the arrows on the side Y of the aperture. Suppose bc to be equal to the wave-length of the light. Then this direction makes an angle with that of the incident beam equal (approximately) to λ/B radians, if B is the breadth of the

aperture ($B=ab$). The length de (d is the mid-point of ab) shown in the figure must be equal to $\lambda/2$, and therefore if the light travelling in this direction is brought to a focus by a converging lens, the part travelling along bc, being out of step with that travelling along de by the amount $\lambda/2$, will destroy it by interference and it is not difficult to see that every part of the beam travelling in this direction will be destroyed in this pair-wise fashion. If now two beams of light, emerging, shall we suppose, from a spectroscope, give images of the slit of the spectroscope, i.e. spectral lines, whose wavelengths are λ and $\lambda + d\lambda$, and are separated by the angle

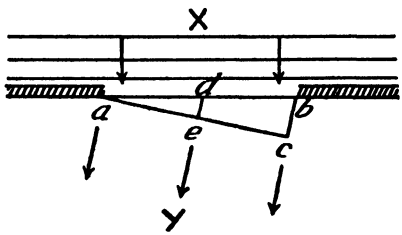


Fig. VII—4

$$d\theta = \lambda/B,$$

B being the breadth of either beam, then the two lines will be distinguishable from one another. They will not overlap sufficiently to look like one line. They will be RESOLVED. This was shown by Lord Rayleigh.*

Two lines will appear to be resolved even when they are separated by a smaller angle than that which we have called $d\theta$, but this separation is the conventional criterion of resolution and

$$\frac{\lambda}{d\lambda} = B \times \frac{d\theta}{d\lambda} \qquad \text{(VII—4)}$$

or (Breadth of beam) × (Dispersion)

is the conventional measure of the RESOLVING POWER of the instrument.

Fresnel's form of Huygens' principle, which is only indicated in the foregoing description, was a great improvement on the original. It was, however, imperfect in certain respects. The perfect form of the principle was discovered by Kirchhoff (1882)†

* Lord Rayleigh: *Phil. Mag.*, Ser. 5, Vol. 9, p. 271 (1879), and "Wave Theory", *Encyclopaedia Britannica*, Vol. 24 (9th Edition).

† *Vide* Drude's work on *Optics* or W. Wilson: *Theoretical Physics*, Vol. 2.

THE GRATING

The grating is a device for measuring wave-lengths by means of diffraction. In its earliest form, as used by Fraunhofer to measure the wave-lengths corresponding to the dark lines in the solar spectrum, it consisted of parallel, equally spaced fine wires, made by winding a long wire over the threads of two parallel screws. A. J. Ångström (1814–1874), Professor of Astronomy at the University of Uppsala, had measured many lines of the solar spectrum by 1868, using plane gratings constructed by a contemporary, Norbert. For many years his measurements constituted the standards for wave-length determinations, e.g. when spectra were observed by prism spectroscopes, which are not adapted for the absolute determination of wave-lengths.

A plane grating is usually a piece of glass, or, in the case of a reflecting grating, a piece of speculum metal, on which equally spaced parallel lines are ruled. It is immaterial how arbitrary may be the shape of the groove made by the diamond which scratches it out. The important thing is that each line or groove shall be exactly like every other one and that they are equally spaced. In the case of the plane transmission grating on which monochromatic light falls perpendicularly there are definite directions in which the transmitted light travels strongly. If these directions make an angle, θ, with that of the incident light, then, as shown in the text-books,

$$n\lambda = e \sin\theta, \qquad \text{(VII—5)}$$

where n is an integer, λ is the wave-length of the monochromatic light and e is the distance separating two consecutive lines, or grooves, of the grating. In the general case, i.e. whether the light is monochromatic or not, a whole spectrum is produced for each value of n (other than zero) and the particular (integral) value of n is called the ORDER of the spectrum, e.g. first order, second order, and so on.

We obtain an expression for the dispersion produced by the spectrum, i.e. for $d\theta/d\lambda$, by differentiating (VII—5) with respect to λ, thus

$$n = e\cos\theta \times \frac{d\theta}{d\lambda},$$

or

$$\frac{d\theta}{d\lambda} = \frac{n}{e\cos\theta}. \qquad \text{(VII—5A)}$$

Now the breadth of the emerging beam (of a particular wave-length, λ) is easily seen to be expressed by

$$B = Ne \cos \theta, \qquad \text{(VII—5B)}$$

if N is the number of grating lines which the incident light spreads over. The combination of the three formulae (VII—4), (VII—5A) and (VII—5B) gives us very simply

$$\text{Resolving power} = Nn. \qquad \text{(VII—6)}$$

It is the product of the number of grating lines over which the beam spreads and the order of the spectrum.

THE CONCAVE REFLECTING GRATING

The American Rowland, the same who determined Joule's equivalent, made great improvements in wave-length measurements.* First of all he made a much more perfect screw than had ever been used before for the ruling machine—the screw which pushes the ruling point on from one line to the next. Consequently the lines on his gratings were more equally spaced than those of any earlier gratings. Secondly, he ruled the lines on a concave reflecting surface of comparatively soft speculum metal. The concave surface of his grating, which was mounted in an ingenious way, enabled him to dispense with lenses for focusing purposes, and the softness of the metal made it possible for him to rule many more lines without wearing down the ruling diamond point to an appreciable extent and so changing the shape of the grooves. It will be remembered that the resolving power of a grating is proportional to the number of grating lines over which the light spreads.

THE ECHELON GRATING

A. Michelson invented a piece of apparatus with the characteristics of a grating and possessing enormous resolving power. It is known as the echelon grating, and its design is amazingly simple; though the practical construction of one is rather difficult. It consists of a number of glass (or quartz) plates, all of

* Rowland: *Phil. Mag.* Ser. 4, Vol. 13, pp. 469–474 (1882), *Phil. Mag.*, Ser. 5, Vol. 16, pp. 197–210 (1883), also *Physical Papers*, beginning at p. 487.

the same thickness and built up like the steps of a staircase (hence its name). The breadths of the steps are also made as near equal as possible.* The reason for its extremely great resolving power lies mainly in the fact that the dispersion it produces, i.e. $d\theta/d\lambda$, see (VII—4)) is so great. It can in fact only be used for light which is already nearly homogeneous, so that it is usually placed in the parallel beam of a spectrometer, between the prism and the telescope of the constant deviation type of instrument.

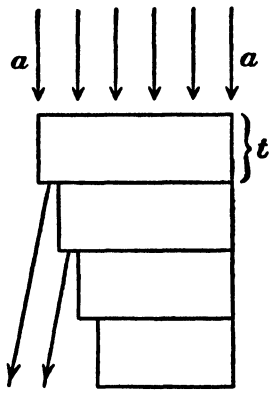


Fig. VII—5

In recent times W. E. Williams has succeeded in constructing a reflecting echelon grating and applying it to the investigation of the hyperfine structure of spectral lines (1927–1933).

INTERFEROMETERS

Diffraction gratings might well be classed under this heading, since the phenomenon of interference plays an essential part in their functioning. They are adapted almost exclusively to the purpose of measuring wave-lengths or differences of neighbouring wave-lengths. The class of instruments usually called INTERFEROMETERS make little use of diffraction and their application is not confined to wave-length measurements. They have been aptly divided by W. E. Williams into two categories:†

(*a*) Those in which the *wave-front* is divided and the resulting separate beams of light, after travelling along separate roads, reunited to produce interference fringes; and

(*b*) those in which the *amplitude* of a beam of light is divided and the resulting beams ultimately reunited to produce fringes. To the class (*a*) belong Fizeau's apparatus for investigating the convection of light and the famous biprism apparatus of Fresnel. Jamin (1818–1886) devised an interferometer of the type (*b*) and used it between 1858 and 1861 to determine refractive indices of

* A. Michelson: *Astrophysical Journal*, 8, p. 37 (1898).

† W. E. Williams: *Applications of Interferometry* (Methuen's Monographs on Physical Subjects).

gases, for which purpose it is rather well adapted.* It consists essentially of two thick parallel glass plates. The incident beam is partly reflected from the front of one of these; the rest of the light is refracted into the glass and reflected from the other side of the plate. The two beams emerge parallel to one another and travel an appreciable distance to the second plate. The first-mentioned beam is reflected from the back face of this plate and the other one from its front face. They reunite to form interference fringes which are conveniently observed through a small telescope. The details of the use of the apparatus are described in the text-books on physical optics.

By far the most important types of interferometer are those of Michelson (1881) and Fabry and Perot (1900). The former will be described and an important application of it discussed in the chapter on the aether. The latter consists of two parallel, half silvered glass plates. A very fine screw is provided for varying the distance between them; though in some forms of the apparatus the plates and the distance separating them are fixed (Fabry-Perot *étalon*). Unlike the interferometers of Jamin and Michelson, this interferometer has an extremely high resolving power, the fringes appearing as bright *thin lines* separated by *broad dark spaces*. This is due to the fact that each incident ray such as *a* in Fig. VII—6 is transformed by multiple reflection into a wide emerging beam *b* . . . *b*. The consequence of this is that the factor *B* in (VII—4) is very great. With their apparatus Fabry and Perot were able to investigate the fine structure of

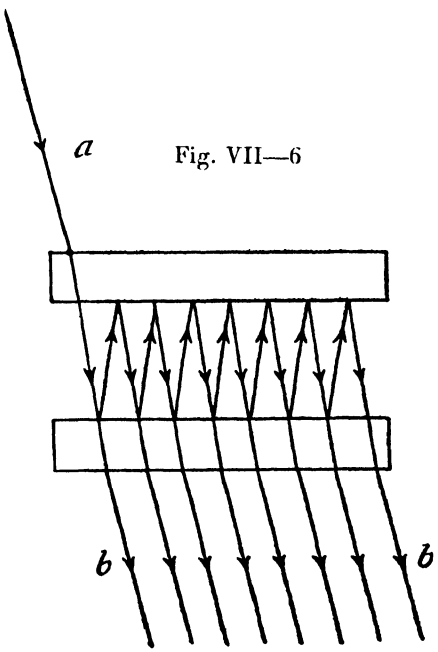


Fig. VII—6

* Jamin: *Cours de Physique.*

spectral lines directly—the sharpness of the fringes enabling the separate components to be distinguished from one another. Michelson also used his interferometer to investigate fine structure (1892); but in consequence of its low resolving power the fringe systems of the components of a line usually overlapped and made it impossible to distinguish one from another. He overcame this difficulty most ingeniously by observing carefully how the VISIBILITY of the fringes varied as the difference in the distances travelled by the two interfering beams of light was changed and was able to infer the fine structure from this.

Mention should be made of the interferometer of O. Lummer and E. Gehrcke (1902). It is a sort of blend of a Fabry-Perot *étalon* and an echelon grating; multiple reflections occurring inside a glass plate.*

Michelson counted the number of wave-lengths of the highly homogeneous cadmium red line in the Paris metre (1894) by an ingenious adaptation of his interferometer. Fabry and Perot repeated this very soon afterwards by an equally ingenious (and rather similar) use of their *étalons*.† Michelson's results were:

For the cadmium red line,		$\lambda = 6438{\cdot}4722$ A.U.
,,	,, green line,	$\lambda = 5085{\cdot}8240$ A.U.
,,	,, blue line,	$\lambda = 4799{\cdot}9107$ A.U.

in air at 15°C. and under a pressure of 760 mm. of mercury. A.U. of course means Ångström unit (10^{-10} metre). It will give some idea of the precision of these measurements to observe that Fabry, Perot and Benoit found for the cadmium red line

$$\lambda = 6438{\cdot}4696 \text{ A.U.}‡$$

W. E. Williams has proposed as a standard, in place of the cadmium red line, the green line of a certain isotope of mercury (Hg. 198) which is produced by exposing gold to bombardment by slow neutrons§—the meanings of the terms "isotope" and "neutron" are explained later.

One further achievement of Michelson's may be mentioned

* O. Lummer and E. Gehrcke: *Ann. d. Physik*, 10, p. 457 (1903).

† *Vide* Schuster and Nicholson, *The Theory of Optics* (Arnold), from which the wave-lenths given below are taken.

‡ Compare this with Michelson's value.

§ The reason being of course the greater homogeneity of this line.

here, namely, his STELLAR INTERFEROMETER, by means of which he measured the angular diameters of (not too distant) stars (*vide* Chapter XXI).

MISCELLANEOUS WORK

It is well known that when two exactly similar sets of waves—waves of the same wave-length and amplitude—meet one another, as for instance waves travelling in opposite directions on a stretched cord, STATIONARY waves are formed with nodes separated from one another by half the wave-length. Otto Wiener (1890) succeeded in demonstrating stationary light waves.* He threw a beam of monochromatic light perpendicularly on a strongly reflecting plane surface. The resulting system of nodal and internodal planes was exhibited by means of a sensitized film of collodion, the plane of which formed a minute angle with the reflecting plane. The internodes appeared in the developed film as lines of deposited silver. The experimental realization of colour photography by Lippmann was suggested by Wiener's experiment.

In 1875 Ernst Abbé developed a theory of microscopic vision. According to this the image of an illuminated structure, produced by a lens, will not correspond to the object unless light from the whole of the associated diffraction pattern goes to form the image. If the image is formed by light from some portion only of the diffraction pattern which the minute structure of the object produces, then it will look like an object which produces a diffraction pattern like this portion.

* O. Wiener: *Wiedemann's Annalen*, 40, p. 203 (1890).

CHAPTER EIGHT

INTEGRATION OF ELECTROMAGNETISM

ABSOLUTE MEASUREMENTS

IN studying the history of electromagnetism during the nineteenth century we learn that in 1849 Gustav Kirchhoff made, or is believed to have made, the earliest determination of electrical resistance in ABSOLUTE MEASURE. What is meant by an absolute measurement and what made such measurements seem important? Why not, one might ask, adopt as a unit the resistance of a piece of platinum wire of say one metre in length and a square millimetre in cross-sectional area and leave it at that? We can easily use a bridge to determine other resistances by comparison. The unit just suggested may be called an ARBITRARY UNIT, since no consideration, except that of convenience, determines its choice. Other such units are, for example, the pint measure for volumes, the weight of a pound avoirdupois as a unit of force, or the small calorie as a unit of heat. But in measuring volumes, for example, our knowledge of the relationship between the linear dimensions of a body of well-defined geometrical shape—a rectangular block for instance—and its volume, makes it convenient to proceed in another way. We may measure the length, breadth and height of the block in terms of some arbitrary unit, e.g. the inch, and express the volume of the block as their product. The unit of volume is now the cubic inch. It is not arbitrary: it is a DERIVED UNIT—derived from the inch. In terms of this unit the volume of a sphere is $4\pi r^3/3$, where r is its radius in inches. Derived units are often an immense convenience. Among such units are units of velocity, or speed, e.g. 1 cm./sec. or 1 mile/hour; the DYNE, a unit of force, namely, the force producing an acceleration of 1 cm./sec.^{2} in a mass of one gramme; or the poundal, a unit of force producing an acceleration of 1 foot/sec.^{2} in a mass of 1 lb., and so on. There is still a certain kind of arbitrariness (seldom mentioned in the textbooks) associated with derived units. If in measuring volumes,

for example, we were mostly concerned with spherical bodies, we might very well adopt the volume of a sphere of unit radius as the unit of volume. The volume of any sphere would then be expressed by r^3, and if L, B and H are the length, breadth and height of a rectangular block, its volume, in terms of such a unit, would be

$$\frac{3}{4\pi} \times L \times B \times H$$

The cubic inch, or cubic centimetre, is so much more convenient than the last-mentioned type of unit of volume (which is $4\pi/3$ times the cubic centimetre or cubic inch, as the case may be) that we never even contemplate it; but we see that in the choice or specification of a derived unit for some physical quantity an arbitrary pure number is always involved. The volume of a rectangular block, for example, may be written

$$k \times L \times B \times H,$$

where k is a numerical constant which has very naturally been chosen to be unity; but which might have been $3/4\pi$, had we felt it sufficiently important that the unit volume should be that of a sphere of unit radius, or $1/\pi$, were the unit of volume chosen to be that of a circular cylinder of unit radius and unit length, or any other number.

In purely mechanical problems and measurements only three units need be arbitrary—all other units can be derived from them. They have been chosen, for the purposes of physical science, to be those of length, mass and time, and are respectively the centimetre, gramme and mean solar second. They are called the FUNDAMENTAL UNITS and the whole ensemble of units derived from them is known as the C.G.S. system of units. There is of course a FOOT, POUND, SECOND system of units, hardly ever used now, except occasionally as exercises for students. Engineers use a system of units based on the foot (unit of length), WEIGHT of one pound (unit of force) and the mean solar second. Their unit of mass is a derived unit* based on

$$\text{Force} = \text{Mass} \times \text{Acceleration}$$

or

1 lb.	=	Mass	×	32·2.
↓		↓		↓
(weight)		(in engineers' units)		(feet per sec. per sec., the intensity of gravity)

* I am informed while reading the proofs that it is called the SLUG.

Hence the thing we call a lb., whose mass = 1 in the foot, pound, second system, has a mass of 1/32·2 in engineers' units. A numerical constant, k, like that already mentioned in connection with volume, can be introduced in the definition of force. We might lay down that

$$\text{Force} = k \times \text{Mass} \times \text{Acceleration},$$

and if we felt it desirable that the mass of a body should be expressed by the same number as its weight, at a particular place, we should require that

$$1 = k \times 1 \times g,$$

where g is the intensity of gravity at the place in question, for instance 980 cm./sec.^{2}, and in such a case k would have to be chosen to be 1/980. These reflections are important in studying the units of electromagnetism.

One reason for suspecting that electromagnetic phenomena cannot be purely mechanical is that NONE of the units for electromagnetic quantities can be derived from one or more of those of length, mass and time only. We are compelled to adopt one of our electrical or magnetic units arbitrarily as a fourth fundamental unit. In the so-called electrostatic system of units—I should say systems, because there is more than one electrostatic system—the fundamental unit chosen is that of DIELECTRIC CONSTANT for which the dielectric constant of free, or empty, space has been adopted. Thus the equation

$$\text{Force (in dynes)} = \frac{Q_1 \times Q_2}{Kr^2}, \qquad \text{(VIII—1)}$$

where Q_1 and Q_2 are the charges on two very small bodies (particles), r is the distance (in centimetres) separating them and K is the dielectric constant of the medium in which they are situated, becomes

$$\text{Force} = \frac{Q_1 \times Q_2}{r^2},$$

when the particles are situated in empty space and the charges are measured in (ordinary) electrostatic units.

It is open to us, of course, to introduce in the formula (VIII—1) a numerical constant like the k in the volume expression, $k \times L \times B \times H$. I represent this number by A, so that formula (VIII—1) becomes

$$\text{Force} = \frac{A \times Q_1 \times Q_2}{Kr^2} \qquad \text{(VIII—1A)}$$

In the old style of electrical units, which I shall consistently use, A is chosen to be unity; but H. A. Lorentz and Oliver Heaviside preferred to give A the value $1/4\pi$, a choice which has certain advantages that cannot be discussed here. It will be seen that the old electrostatic unit of quantity is equal to the Lorentz-Heaviside one multiplied by $\sqrt{4\pi}$.

The equation (VIII—1) or (VIII)—1A enables quantity of electricity to be expressed in terms of the units of length, mass, time and dielectric constant and there is no formula or relationship which enables us, for example, to eliminate K—without introducing another electromagnetic quantity in its place—and so express a charge in terms of length, mass and time only. When we eliminate K in any of the possible ways we *necessarily* introduce another electromagnetic quantity. Every equation containing an electromagnetic quantity necessarily contains two of them at least.

In the electromagnetic system of units the MAGNETIC PERMEABILITY of empty space is chosen as a fundamental unit. This quantity occupies an analogous position in magnetostatic phenomena to that occupied by dielectric constant in electrostatic phenomena; but it is not in general a constant. Thus we have for the force between two point poles,

$$\text{Force} = \frac{m_1 \times m_2}{\mu r^2}, \qquad \text{(VIII—2)}$$

where m_1 and m_2 are the STRENGTHS of the poles, r is the distance separating them and μ is the permeability of the medium. The permeabilities of empty space and of PARAMAGNETIC media are constants.

Now an ABSOLUTE MEASUREMENT of some physical quantity is one which gives its value in terms of a unit derived from one or more of the adopted fundamental units. It would give a volume as so many cubic centimetres or a velocity as so many centimetres per second. The former is said to have the DIMENSIONS of the cube of a length, the latter those of a length divided by a time. Equation (VIII—1) indicates that the dimensions of an electrical charge can be expressed by

$$\sqrt{\text{Force}} \times \sqrt{\text{Dielectric constant}} \times \text{Length}\,(r),$$

or, since force, being the product of a mass and an acceleration, has the dimensions

$$\frac{\text{Mass} \times \text{Length}}{(\text{Time})^2},$$

therefore those of a charge are:

$$(\text{Mass})^{\frac{1}{2}} \times (\text{Length})^{\frac{3}{2}} \times K^{\frac{1}{2}} \times (\text{Time})^{-1}.*$$

With a little more trouble it may be shown that an electrical resistance has the dimensions

$$(\text{Permeability}) \times (\text{Velocity}).$$

Now in experiments *in vacuo*, or even in air, in which $\mu = 1$ (in air, nearly 1) in the electromagnetic system of units, a resistance presents itself as a velocity and a certain unit of resistance, later called the OHM in honour of G. S. Ohm, the discoverer of OHM'S LAW, was described about a hundred years ago as

ONE EARTH QUADRANT PER SECOND.

The unit represented by

ONE CENTIMETRE PER SECOND

is the ordinary electromagnetic unit of resistance and since the earth quadrant (distance from equator to pole) was taken to be 10^7 metres (for the purpose of the definition of the ohm), i.e. 10^9 centimetres, therefore

ONE OHM = 10^9 × ONE ELECTROMAGNETIC UNIT.

Absolute measurements of electromagnetic quantities are very troublesome, involve a lot of operations and take up much time; whereas a direct comparison of an unknown quantity with a known one of the same kind is usually a fairly simple and short operation. So a few important absolute measurements have been made with great care and standards set up. From this rather practical point of view electrical resistance is perhaps the most important electrical quantity; partly because a standard resistance is something which can easily be preserved for an indefinite period without change and partly because the precise comparison of two electrical resistances is one of the simplest of all laboratory operations.

Once again may it be emphasized that the description of a

* The notion and theory of dimensions begin with Fourier; *vide* Freeman's translation of his *Théorie de la Chaleur*, p. 128.

resistance as a velocity is incomplete. There is in it the permeability factor which, in those experimental measurements of resistance in which it occurs, is equal to unity or nearly so in electromagnetic units.

ABSOLUTE MEASUREMENT OF ELECTRICAL RESISTANCE

The *simplest* way of finding the value of a resistance in absolute measure is one due to Joule who carried it out in 1867, long after the original measurement made by Kirchhoff. Because of its simplicity I prefer to describe it first. It consists in maintaining a steady current in the wire, the resistance of which is to be determined, while the wire is immersed in the water (or other liquid) in a calorimeter. Thus the amount of heat generated in some chosen period of time can be estimated and the equivalent amount of work computed by multiplying by Joule's equivalent. And here a few words of explanation seem to be called for. Resistance, as the name indicates, is a measure of some kind of hindrance which the conductor (wire) presents to the flow of electricity through it. It is intimately related to friction. We may profitably compare the flow of electricity along a wire with that of water along a tube. Let us think of a horizontal tube (horizontal in order to exclude the irrelevant complication of gravity) in which water is flowing, preferably in a closed circuit as indicated in the diagram (Fig. VIII—1) and kept in motion by some sort of pump, shown at *A*. On account of friction, or viscosity, work has to be done to keep the water moving and the amount of work done, reckoned for each unit volume of water that has completed the circuit, simulates almost perfectly the ELECTROMOTIVE FORCE in the analogous electrical circuit, with a cell (or battery) in place of *A* and resistance (corresponding to the tube) in series. The work done on the unit volume when it travels from *B* to *C*, for example, is equal to the

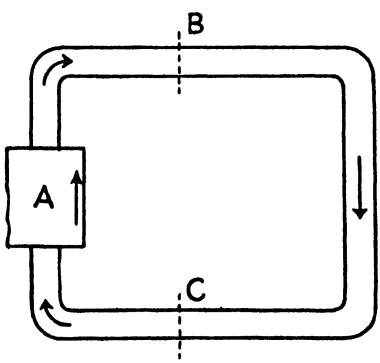


Fig. VIII—1

pressure difference between B and C and simulates the POTENTIAL DIFFERENCE between the corresponding points in the electrical case. WHEN THIS WORK IS DONE IN GENERATING HEAT (I am speaking of the electrical case) the measure of the resistance is defined by

$$\text{Resistance} = \frac{\text{Work done per unit quantity}}{\text{Strength of current}} \qquad \text{(VIII—3)}$$

This relationship is often mistakenly called OHM'S LAW, but it is obviously only a definition. The law discovered by Ohm was the observational fact that, in the case of many conducting materials (but not all), the resistance, i.e. the ratio which expresses it, is independent of the current strength, subject to other conditions, e.g. temperature, remaining unchanged. In general, it should be noted, potential difference divided by current strength is *not* equal to the resistance of the corresponding part of the circuit, as for instance when an electric motor is included in it. In such a case the work is nearly all done in driving the motor and little of it in generating heat.

Returning to Joule's experiment, we start from

$$V = Ri,$$

where V is the work done per unit quantity of electricity in producing heat in the wire and calorimeter, R is the resistance to be measured, and i is the current strength, i.e. the quantity passing per second. During t seconds the quantity $i \times t$ passes through the wire and the work done on it must be equal to

$$Vit,$$

which is equivalent to

$$Ri^2t.* \qquad \text{(VIII—4)}$$

The strength of the current was measured by a tangent galvanometer. This is also an absolute measurement and several pages might be written about it. The electromagnetic unit of current strength is so chosen that the magnetic field intensity generated at the centre of a circle of wire in which a current, i (in electromagnetic units) is flowing, is expressed by

$$f = 2\pi i/r,$$

f being the magnetic field intensity (in electromagnetic units)

* This expression was first given by Kelvin, *Phil. Mag.* (1851).

and r the radius of the circle in centimetres. The experimental fact is that

$$f \text{ is proportional to } i/r,$$

whatever units may be adopted, and we might (after having defined the units for f and r) decide to fix the unit for i by assigning any positive value to the number k in

$$f = k\,i/r.$$

In fact k has been chosen to be 2π in defining the ordinary electromagnetic unit. With the tangent galvanometer

$$f = H \tan \theta,$$

where H is the horizontal component of the earth's magnetic field and θ is the angle of deflection of its needle, and so we get

$$i = \frac{rH}{2\pi n} tan\, \theta,$$

the number n being the number of circles of wire in the instrument. Joule equated the total work done, as determined by the calorimeter, call it W, to (VIII—4), or to an equivalent expression. Thus

$$W = Ri^2t$$

and R is easily computed, since W, i and t are known.

THE METHODS OF WEBER AND KIRCHHOFF

Wilhelm Weber (1804–1891) was one of the famous *Göttinger Sieben*—the Seven of Göttingen—who rebelled against the unwise interference of King Ernst August of Hanover (1837) in the affairs of their university. On leaving Göttingen he received hospitality in the ancient University of Leipzig, where a small laboratory (entirely free from iron) was built for him in a garden off the Thalstrasse.

Weber used a large coil, whose dimensions were carefully measured and which was connected in series with a tangent galvanometer. The coil was mounted so that it could be turned about a vertical axis. When it was turned through 180°, starting and ending with the planes of its windings perpendicular to the horizontal component of the earth's field, there was a calculable change in the magnetic flux linked with it. equal in fact to

$$2gH,$$

if g means the sum of the areas of the windings of the coil and H the intensity of the earth's horizontal component. In consequence a quantity of electricity, equal to

$$2gH/R,$$

is discharged through the galvanometer, R being the resistance of the whole circuit. This quantity must, of course, be equal to

$$\frac{TH}{\pi G} \sin \tfrac{1}{2}\theta,$$

θ being the consequent angular throw of the galvanometer needle, corrected for damping, T the time of a complete oscillation of the needle, and G the galvanometer constant.* So R is given by the equation†

$$\frac{2g}{R} = \frac{T}{\pi G} \sin \tfrac{1}{2}\theta.$$

Maxwell points out in his *Treatise* that the two H's (namely H in the neighbourhood of the rotating coil and H in the neighbourhood of the galvanometer) have been assumed by Weber to have the same value and it is of course well known that H may have appreciably different values at two places even within the same room. Perhaps Maxwell was unaware of the generous provision of Leipzig University. The absence of magnetizable materials in Weber's laboratory undoubtedly ensured that the two H's did not differ to an extent that mattered.‡

A complete description of the earliest method, that of Kirchhoff (1849), would take up too much space.§ Suffice it to say that he used two coils, whose mutual inductance he calculated; one, the secondary, connected in series with a ballistic galvanometer, was used to measure the quantity of electricity discharged round the secondary circuit when a current was started in the primary one. A further experiment enabled the unknown things, e.g. the constant of the galvanometer and the unknown primary current, to be eliminated. This consisted in linking the terminals of the resistance to be measured, which is in the primary circuit, with

* The galvanometer constant is G as it appears in the formula for a steady current i, namely $i = \frac{H}{G} \tan \theta$.

† *Poggen. Annalen*, 82, p. 337 (1851).

‡ Weber's unit of resistance was 1 mm. per second, i.e., 10^{-10} ohm.

§ *Poggen. Annalen*, 76, p. 412 (1849).

the secondary circuit and observing the steady deflection of the galvanometer when the current in the primary circuit is steady.

The experiment constitutes a valuable exercise for present-day students, who however are given the value of the resistance which it was Kirchhoff's purpose to determine and are assigned the task of finding experimentally the mutual inductance which Kirchhoff calculated from the dimensions and positions of the coils.

LATER METHODS

Two further methods, developed later, of measuring resistance absolutely may be referred to: Kelvin's rotating coil method and the method of Lorenz (1873).* The latter method has been improved by F. E. Smith (Sir Frank Smith).† Kelvin's method was originally carried out by Balfour Stewart, Fleeming Jenkin and Clerk Maxwell in 1863 for the Standards Committee of the British Association. Unluckily a small error slipped into the computation of the result of their experiments and consequently the British Association ohm was a little below the correct value. Lord Rayleigh‡ used this method (1882) and found the B.A. ohm to be equal to

0·98651 earth quadrant per second;

while R. T. Glazebrook§, who used the original method of Kirchhoff, found it to be

0·98665 earth quadrant per second.

The error had at least one good consequence: the comparison of the B.A. ohm with the legal ohm by means of Carey Foster's bridge has provided a fine exercise for students since these standards were set up.

In Kelvin's method the wire whose resistance was to be measured was in the form of a circular coil which was caused to rotate about a vertical axis in the earth's field with a constant measured angular velocity. A small magnet (with a big moment of inertia) was suspended at the centre of the coil—the suspending fibre being as nearly torsionless as possible—and its angular

* *Poggen. Annalen*, 149, p. 251 (1873).
† *Dictionary of Applied Physics.*
‡ *Phil. Trans.*, 173, p. 661 (1882). § *B.A. Report*, p. 97 (1890).

deflection observed.* The equation for calculating the result contains another unknown quantity besides the resistance of the wire, namely, the self-inductance of the coil. This can of course be calculated from its dimensions; but Lord Rayleigh devised a simple approximate method of determining this quantity in terms of the resistance that is being measured. The approximate character of his method does not lower its value, since the term containing the self-inductance is a very small one.

In the method of Lorenz an induced electromotive force is balanced against the potential difference between the terminals of the resistance through which a steady current (on which the induced electromotive force depends) is passing.†

ABSOLUTE MEASUREMENT OF ELECTRIC CURRENTS

The earliest device for this purpose was of course the familiar tangent galvanometer—the sine galvanometer is a simple modification of it. One of the weaknesses of this instrument lies in the fact that the magnet, however short it may be, extends some way from the centre of the coil. This causes the well-known formula to be slightly in error. The reason is very simple. The familiar formula assumes that every part of the small magnet is subjected to a magnetic field of the same intensity as that at the centre of the coil, and of course this is only approximately true. Von Helmholtz greatly improved the instrument by having two coaxial coils, so constructed that each circular turn of one coil was separated from the corresponding one in the other coil by a distance equal to the radius of either. The effect of this is that the magnetic field is sensibly uniform over a region of quite appreciable dimensions within which the magnet of the instrument is pivoted and the size of the magnet is therefore unimportant provided it is well within this region.

Weber's electrodynamometer came soon after 1840. It is like Helmholtz's galvanometer with the magnet replaced by a small suspended pair of coils (bi-filar suspension) which are like the large ones in miniature. The current is made to pass in series through both the fixed coils and the suspended coils. The great

* The theory of the experiment is presented very clearly in S. G. Starling's *Electricity and Magnetism* (Longmans, Green & Co.).

† Lord Rayleigh and Mrs. Sidgwick, *Phil. Trans.*, 174, p. 193 (1883).

advantage of Weber's instrument lies in the fact that it is easy to use it in such a way that no knowledge of H, the horizontal component of the earth's magnetic field, is needed. Such knowledge is essential with the old tangent galvanometer and also with the improved one of Helmholtz.

One more type of instrument for the absolute measurement of currents should be mentioned. It appears to have been developed by Kelvin and Joule in 1883. It is usually described as Kelvin's current balance or current weigher or, when adapted to give the current in amperes, the ampere balance. In its simplest form it is a sensitive balance with a coil, whose windings are horizontal, hanging from one arm and a coaxial fixed coil below it. The current to be measured is passed through these coils, in such a sense that they attract one another. The force of attraction is counterbalanced by weights in the scale pan depending from the other arm of the balance. This force, or weight, is obviously proportional to the square of the current and the constant of proportionality can easily be computed.* For practical purposes a unit of current strength called the AMPERE (after the great French savant of that name) is usually used. It is defined to be one-tenth of the electromagnetic unit of current strength and the OHM (one earth quadrant per second) is the associated practical unit of resistance. The practical unit of electromotive force, or of potential difference, is called the VOLT (after Alessandro Volta, a great Italian man of science). It may be defined in the following way: When a steady current is flowing in a wire, the product of the current strength, in amperes, and the resistance of the wire, in ohms, gives the potential difference between the ends of the wire, in volts.† Remembering that the ohm is equal to 10^9 electromagnetic units of resistance and that the ampere is equal to 1/10 (i.e. 10^{-1}) of an e.m. unit of current strength, it is not difficult to see that the volt must be equal to 10^8 e.m. units of potential difference, or electromotive force.‡

It may be added here that the practical unit of quantity of

* Maxwell: *Treatise*, Vol. 2, Art. 701.

† This definition is equivalent to the following one: there is a potential difference of one volt between two points when one joule of work (10^7 ergs) is done on one coulomb in passing from one point to the other.

‡ In order to enable students to remember this number easily, it has sometimes, in my experience, been described to them as equal to the number of the saints in Heaven—ten thousand times ten thousand.

electricity, called the COULOMB, is the quantity passing any point on a conductor during one second when the current strength is one ampere. The practical unit of work or energy is the joule, which is defined to be 10^7 ergs. The systematization of electromagnetic units is due largely to the efforts of the British Association, which appointed a Standards Committee for this purpose in 1861; but it should not be overlooked that the principles on which the systematization is based are those laid down by Wilhelm Weber (1851) who received his inspiration from Gauss with whom he had worked in his youth.

OTHER ABSOLUTE MEASUREMENTS

Absolute measurements of potential difference have been carried out, notably by Lord Rayleigh and Sir Frank Smith, and the Weston (cadmium) cell, the E.M.F. of which is accepted as 1·0183 volts at 20°C., is a recognized standard of potential difference, for use with a potentiometer for example. Another important standard, which has been carefully determined, is one of quantity of electricity. One coulomb passing through a silver voltameter causes 0·0011183 gramme of silver to be deposited on the cathode.

While we are dealing with absolute methods and standards, it may be mentioned that Kelvin invented an absolute electrometer (1851) suitable for the absolute measurement of potential difference in electrostatic units. It consists essentially of a circular conducting plate near a larger parallel one—both in air, though in the ideal instrument they would be *in vacuo*. When a difference of potential is established between them they attract one another. The force of attraction can be measured and from this and the dimensions and separation of the plates the potential difference can be calculated.

THE RELATIONSHIP BETWEEN ELECTROSTATIC AND ELECTROMAGNETIC UNITS

W. Weber and F. Kohlrausch carried out a most interesting experiment just about the middle of the century.* They measured the ratio of the electromagnetic unit of quantity of electricity to the electrostatic one. It was in any case important

* *Poggen. Annalen*, 1856.

that the relationship between these units should be investigated, but the ratio in question turned out later to have a significance which perhaps they never dreamt of; as we shall appreciate when we come to study Clerk Maxwell's views about electromagnetism. The method they used was a very direct one: they determined the charge on a condenser in electrostatic units by measuring the potential difference, in such units, between its plates with an absolute electrometer, the type of instrument invented by Kelvin. They calculated the capacity of the condenser from its dimensions. In the case of a condenser consisting of two parallel plates,

$$\text{Capacity} = \frac{KA}{4\pi d},$$

where K is the dielectric constant of the intervening insulator, A is the appropriate area and d is the distance separating the plates. When the insulating medium is air, $K = 1$ (very nearly) in electrostatic units and the capacity of the condenser becomes

$$A/4\pi d,$$

apart from a small correction. Thus

$$Q = AV/4\pi d,$$

if Q means the charge and V the measured difference of potential between the plates. In this connection another of the numerous electrical devices invented by Kelvin may be mentioned. The simple formula just given is deduced from the assumption (among others) that the field is perfectly uniform everywhere between the plates and ends sharply at the edges; but in fact this is not strictly correct. The correct formula is very complicated and in order that the simple formula may be available Kelvin introduced the device of the GUARD RING which prevents this departure from uniformity, at least sufficiently to reduce any error to second order dimensions. He also used the guard ring with his electrometer.

To get the measure of Q in electromagnetic units, Weber and Kohlrausch discharged the condenser through a ballistic galvanometer and applied the familiar formula

$$Q = \frac{TH}{\pi G} \sin \tfrac{1}{2}\theta.$$

They found for the ratio

$$3{\cdot}1074 \times 10^{10}$$

and one of the tests of Maxwell's theory, as we shall see later, is the closeness with which this number approaches the velocity of light (in cm. sec.$^{-1}$) in free space. The measurement of the ratio has been repeated many times since; by Clerk Maxwell, Kelvin, Rowland and by Sir J. J. Thomson and Dr. G. F. C. Searle.* It has turned out to be identical with the velocity of light in free space to within the narrow limits of unavoidable experimental error.

We have noticed that when magnetic permeability is unity a resistance assumes the guise of a velocity, and we are not surprised to read in Maxwell's treatise that the value of the ratio, as found by Thomson (Kelvin), is 28·2 ohms and by Maxwell himself 28·8 ohms (1869). At that time Rowland's result would doubtless have been described as 29·8 ohms. This means nearly 30 earth quadrants per second, i.e. nearly $7\frac{1}{2}$ times round the earth in a second.

THE SIMPLEST OSCILLATING CIRCUIT

Kelvin appears to have been the first to study the character of the current in what we should now call an INDUCTIVE circuit.† He worked out theoretically what was to be expected when a condenser (Leyden jar) is discharged through a conductor with resistance and INDUCTANCE in series. When the magnetic flux linked with a loop or coil of wire is changing, an electromotive force is induced in it, which is measured, in E.M. units, by the rate at which the flux changes. In the case we are studying the magnetic flux is equal to

$$L\,i,$$

where i is the current strength and L is a constant, called the SELF-INDUCTANCE of the wire.‡ The rate of change of the flux, and consequently the induced electromotive force, is therefore

* J. J. Thomson and G. F. C. Searle: *Phil. Trans.* 181, p. 583 (1890). They found for the ratio $2{\cdot}9955 \times 10^{10}$.

† W. Thomson: *Phil. Mag.*, 5, p. 393 (1853).

‡ The practical unit of inductance is called the HENRY in honour of the American Joseph Henry, a contemporary of Faraday and independent discoverer of electromagnetic induction.

equal to the rate of change of Li. It is convenient to write this in the form

$$L\dot{i},$$

where $\dot{i}$ means the rate of change of i. Therefore

$$\text{Induced E.M.F.} = L\dot{i}$$

Turning now to the discharge of a Leyden jar when its coatings are joined by a wire, or coil of wire, we find the effective electromotive force to be

$$V - L\dot{i},$$

that is to say, the potential difference between the coatings, less the induced E.M.F., since the latter tends to hinder the growth of the current (law of Lenz). We may therefore write

$$V - L\dot{i} = Ri,$$

if R is the resistance of the connecting wire. When R happens to be very small this simplifies to

$$V = L\dot{i}.$$

If Q is the charge of the jar at any instant, the current is numerically equal to its rate of change and therefore

$$\dot{Q} = i.$$

Consequently

$$\ddot{Q} = \dot{i}.$$

That is to say that $\dot{i}$ means (apart from sign) the rate of change of the rate of change of Q, and since

$$V = Q/C$$

(C is the capacity of the jar) we have

$$\frac{Q}{C} = L\ddot{Q}$$

or

$$\frac{Q}{\ddot{Q}} = LC.$$

Now imagine a distance, equal to Q, measured off along a straight line from some chosen point on it—to the right if Q is positive and to the left if Q is negative. The point reached will be in motion, since Q is constantly changing, and its acceleration will be $\ddot{Q}$. Moreover the acceleration is directed to the zero point (law

of Lenz). The equation we have found tells us that the displacement divided by the acceleration (i.e. $Q/\ddot{Q}$) is constant, equal in fact to LC. This kind of motion is called simple harmonic. It is periodic and the period is equal to

$$2\pi\sqrt{\frac{(\text{Displacement})}{(\text{Acceleration})}}$$

which means that the current we have been studying is an alternating one with a period equal to

$$2\pi\sqrt{LC}$$

Of course the resistance is not always negligible and Kelvin's treatment of the problem was more complete than what has just been given.

The discharge of the Leyden jar was studied experimentally by a young German student, Wilhelm Feddersen (1832–1918). He observed the spark discharge of an initially charged jar with a rotating mirror and noticed of course a multiplicity of images, each spark image representing a discharge across the gap in one direction or the other. The speed of his rotating mirror enabled him to compute the period of the oscillations and to confirm Kelvin's theory.*

MAGNETISM

Ampère imagined magnetization to be due to molecular currents or currents in the ultimate particles of magnetizable materials, and this view, in an elaborated form, is still held. Indeed an electric current simulates a magnet. Weber (1854) adopted Ampère's view. He supposed that in the unmagnetized state of a piece of material, e.g. a rod of iron, the molecular magnets were linked together in small closed rings; every molecular north pole being in immediate contact with a south pole of equal strength. In the presence of an external magnetic field these rings were supposed to be broken up and the molecular magnets aligned to form long chains with free north poles at one end of the rod and free south poles at the other. This is, broadly

* W. Feddersen: *Beiträge zur Kenntnis des elektrischen Funkens*, Dissertation, Kiel, 1857.

speaking, still our view about paramagnetic and ferromagnetic materials.

A diamagnetic material is one which is repelled when a magnetic pole is brought near to it. Weber also arrived at the correct view of the nature of this phenomenon. He imagined it to be due to the induction of currents in the molecules (or atoms) of the diamagnetic material. Such currents will, in accordance with the law of Lenz, produce magnetic fields opposing the inducing one.

The phenomenon of HYSTERESIS does not appear to have been noticed till Warburg discovered it in 1881. It was studied in great detail by Ewing between 1885 and 1890 and the name hysteresis (ὑστερέω, to come late) is due to him. It refers to the characteristic of a certain class of materials, among which iron is the most prominent, of retaining much of their magnetization after the magnetizing field has been removed (FERROMAGNETISM). Ewing showed that the behaviour of ferromagnetic materials could be simulated by a large number of magnets (representing the molecular magnets) pivoted close together. Most materials are only slightly magnetizable, in what may be called the normal sense of the term, and show no hysteresis (PARAMAGNETISM). Their permeabilities (and susceptibilities) are constants (independent of the magnetizing field).

The theory of magnetization owes most perhaps to Paul Langevin (1905) and Pierre Weiss.* The former of these has given a very beautiful theory of a paramagnetic gas which may be concisely described as the kinetic theory of the gas, amplified by taking into account the fact that the molecules are magnets and including in the Maxwell-Boltzmann exponential distribution the directions of their magnetic axes and the associated energy terms under the influence of an applied magnetic field.

The theory of Weiss is based on that of Langevin, but he amplified the latter by supposing an *internal* field to be superposed on the applied external one. He was thus able to give a very beautiful account of the main features of ferromagnetism.

* *Journal de Physique*, 6, p. 661 (1907). A very complete account of their work will be found in O. W. Richardson's *Electron Theory of Matter* (Cambridge) and in E. C. Stoner's *Magnetism and Atomic Structure* (Methuen) and the latter's smaller work (Methuen's Monographs) is also very informative.

Weiss came to the conclusion that there is a natural unit of magnetic moment, the MAGNETON, just as there is a natural unit electric charge (e.g. that of an electron, or a hydrogen ion). This appears to be the case; the quantum theory leads to such a unit. The latter, however, is just five times as great as the magneton of Weiss. The explanation of the discrepancy undoubtedly lies in the fact that Langevin's theory needs to be supplemented (and corrected) by the appropriate Wilson-Sommerfeld quantum condition, without which the computation of the magnetic moment of the magneton of Weiss is in error in the same sense and to the same extent as has been observed.

CHAPTER NINE

MAXWELL'S ELECTROMAGNETIC THEORY

DIELECTRIC MEDIA

IN common with the physicists of his day, James Clerk Maxwell (1831–1879), a Scotsman and the first occupant of the Cavendish Chair of Experimental Physics in the University of Cambridge, believed that electromagnetic phenomena, like all other phenomena of the physical world, were matter and motion phenomena of some sort or other and that they came entirely within the scope of Newtonian mechanics. Yet it was his theory which was to give the old mechanics its first shock. He was greatly influenced by Faraday's view of the nature of electromagnetic phenomena and of electrostatic phenomena in particular. He wrote in his great *Treatise*: "We may conceive the physical relation between electrified bodies, either *as the result of the state of the intervening medium* [my italics] or as the result of a direct action between electrified bodies at a distance." Faraday ascribed the force of attraction or repulsion between two electrified bodies to a state of stress in the insulating medium (dielectric) separating them and Maxwell proceeded, as he expressed it, "to investigate the mechanical state of the medium" and used such illustrations as the "tension of a rope or the pressure of a rod". For him, as for Faraday, the medium was in "a state of mechanical stress". This attitude was supported by the observational fact that when the materials of the conductors in an electrostatic field are replaced by different conducting materials, e.g. copper replaced by silver, leaving unchanged the charges, the shapes and configurations of the conductors and the dielectrics the observed electrostatic phenomena are unchanged. On the other hand a replacement of the dielectrics makes all the difference in the world.

There is a well-known mathematical expression for the force exerted on the unit volume of a piece of material under stress—or, shall we preferably say, for the X component of the force per

unit volume. It is much simpler than it looks and it will be helpful to examine it. A rectangular block of some elastic material is shown in Fig. IX—1 with three of its edges coincident with the X, Y, Z axes of a rectangular system of co-ordinates. The

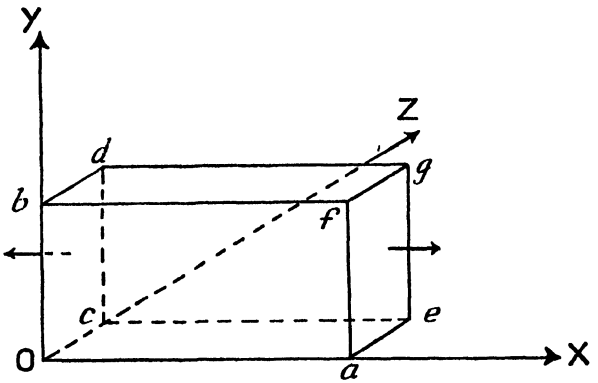


Fig. IX—1

eight corners of the block are marked o, a, b, c, d, e, f and g. Now imagine the material to be in a state of stress—rather a simple kind of stress—namely, a tension in the sense indicated by the arrows. The face *aegf* is pulled to the right by a force at right angles to its surface. Similarly there is a force exerted over *obdc* in the opposite sense. If they are equal the resultant force on the rectangular block is just zero. It is convenient and usual to describe the state of stress by the tension or force per unit area. The force over *aegf*, for example, may be written

$$t \times (\text{area } aegf).$$

Now suppose the tension (force per unit area) over *aegf* to be greater than that over *obdc* and distinguish them as t_2 (the greater) and t_1 (the lesser). Then obviously the resultant force (in the X direction of course) on the block is

$$(t_2 - t_1) \times (\text{area } aegf),$$

since both areas are the same. To get the force per unit volume we divide by the volume of the block, which is clearly

$$(\text{length } oa) \times (\text{area } aegf)$$

and thus obtain

$$\frac{t_2 - t_1}{(oa)}.$$

If we wish to have the force per unit volume in the immediate neighbourhood of some point, we imagine the block containing the point to be very small, approaching zero in the limit, so that $t_2 - t_1$ becomes a very small increment of the tension, t, over a very small distance *oa*. The X component of the force per unit volume is therefore usually written

$$\frac{\partial t}{\partial x}.$$

But we may have other contributions to this X component. For instance, we may have stresses (called shearing stresses) exerted over the areas *oaec* and *bfgd* and directed oppositely, like the tensions already studied, the one over *bfgd* in the X direction and the other, over *oaec*, in the opposite sense. Clearly we must do something to avoid confusing these rather different kinds of stresses with one another. So we shall distinguish the t of the former kind of stress by writing it as

$$t_{xx}$$

and so also the associated force per unit volume as

$$\frac{\partial t_{xx}}{\partial x}.$$

The shearing stress we shall represent by

$$t_{xy}$$

and its contribution to the X component will be

$$\frac{\partial t_{xy}}{\partial y}.$$

The former of the subscripts, i.e. the x (e.g. in t_{xy}) marks the fact that we are dealing with an X component, the latter (e.g. the y in t_{xy}) indicates that t_{xy} is a stress over a face perpendicular to the Y axis. Lastly it should now be obvious, since we have to deal with faces perpendicular to *three* axes, that the total X component of the force per unit volume, i.e. f_x, is expressed by

$$f_x = \frac{\partial t_{xx}}{\partial x} + \frac{\partial t_{xy}}{\partial y} + \frac{\partial t_{xz}}{\partial z}. \qquad \text{(IX—1)}$$

Maxwell succeeded in expressing the forces in an electrostatic field (and also in a magnetic one) in this form, and it seemed, at the time, that he had made an important step in the direction of

the desired mechanical explanation, or interpretation, of electromagnetic phenomena.*

The actual expressions for the stress components, t_{xx}, t_{xy}, t_{xz}, etc., will not be given here. When the system of stresses in an elastic medium is not in equilibrium, the medium acquires momentum and kinetic energy. Expressions can easily be found for the electromagnetic momentum and energy per unit volume of a dielectric medium in an electromagnetic field.

Something may be learned from the following illustration about the nature of the state of stress and strain, whether it be in an elastic medium or in an electrical or magnetic field. Imagine two parallel plates being drawn towards one another. If F be the force dragging one towards the other, then the work done as the distance between them is reduced by the small distance, d, is

$$F \times d.$$

If this work is done at the expense of energy seated in the space swept out, then

$$\text{Energy per unit volume} = \frac{F \times d}{(\text{Plate area}) \times d}$$

or

$$\text{Energy per unit volume} = \frac{F}{\text{Area}} = \text{Tension}.$$

This is the case, for instance, with the plates of a parallel plate condenser: the tension on either plate is equal to the energy per unit volume in the neighbourhood.

It is very easy to show that the energy per unit volume in an electrostatic field is

$$\frac{KE^2}{8\pi},$$

E being the field intensity and K the dielectric constant.† For

* No doubt Ernst Mach (if he were still living) and his disciples would criticize the use of the word "explanation" in this context; but the word "description" cannot be substituted for it without some enrichment beyond its normal endowment.

† The expression for the energy density depends on the kind of units we use. It might be $\frac{1}{2}KE^2$ (Lorentz-Heaviside units). It is part of the definition of the units I am using that the following formula always holds:

$$\text{Force in dynes} = Q_1 \times Q_2 / Kr^2,$$

where Q_1 and Q_2 are the charges on two particles separated by the distance r cm., in a medium of dielectric constant K.

instance, the energy associated with a charged conductor is

$$\tfrac{1}{2}CV^2,$$

where C is its capacity and V its potential. This energy is seated, according to the views of Maxwell and Faraday, in the dielectric over which the associated electric field extends. Think of the conductor as the inner sphere of a spherical condenser. The energy in the interspace depends on V, the potential difference between the spheres and on the capacity of the inner one. This latter is equal to

$$\frac{Ka^2}{d},$$

where a is the sphere's radius and d (assumed to be very small compared with a) is the distance between the charged sphere and the inner surface of the outer sphere. The energy in the interspace is therefore

$$\tfrac{1}{2} \times \frac{Ka^2}{d} \times V^2$$

and the volume of this interspace is $4\pi a^2 d$. Therefore

$$\text{Energy per unit volume} = \tfrac{1}{2} \times \frac{Ka^2}{d} \times \frac{V^2}{4\pi a^2 d}$$

But

$$V = E \times d,$$

where E is the electric field intensity. Consequently

$$\text{Energy per unit volume} = \tfrac{1}{2} \times \frac{Ka^2}{d} \times \frac{E^2 d^2}{4\pi a^2 d}$$

$$= \frac{K E^2}{8\pi} \qquad \text{(IX—2)}$$

We conclude then that the tension along the lines of force must be equal to $KE^2/8\pi$, as Maxwell found. This is what is symbolized above by t_{xx}. A more complete statement would be the following: The forces on the conductors in an electrostatic field are such as would result if the state of stress in the dielectric medium were a tension, in the direction of the electric field, equal to the energy per unit volume, and an equal pressure perpendicular to the direction of the field. It occurred to Maxwell to put the energy per unit volume in the form

$$\tfrac{1}{2}E \times D,$$

where E is the field intensity and D is some sort of displacement.* Evidently

$$D = \frac{KE}{4\pi}. \qquad \text{(IX—3)}$$

It is like a STRAIN corresponding to the STRESS E and when we write

$$\frac{\text{Stress}}{\text{Strain}} = \frac{E}{D} = \frac{4\pi}{K},$$

we are tempted to regard $4\pi/K$ as a kind of modulus, like a modulus of elasticity, and indeed Maxwell called it the COEFFICIENT OF ELECTRICAL ELASTICITY of the medium (a term which has now fallen into disuse). He was led to appreciate the significance of D by the following kind of reflection. Imagine a spherical surface of radius r in a dielectric medium and a particle with a charge Q, placed at its centre. The value of the electric intensity at any point on its surface must be expressed by

$$E = \frac{Q}{Kr^2}$$

and since $E = 4\pi D/K$, we find

$$D = \frac{Q}{4\pi r^2}.$$

THE DISPLACEMENT HYPOTHESIS

Thus was Maxwell led to the hypothesis that, when an electric field is established in an insulating medium, a quantity of electricity, equal to $K/4\pi$ times the field intensity, is displaced, or pushed as it were, through every unit area perpendicular to the direction of the field. This simple DISPLACEMENT HYPOTHESIS† together with the analogous thing in a magnetic field, had momentous consequences.

When the medium is a conducting one there is a continuous flow, or movement, of electricity in the direction of the field to which the medium may be subjected and the current density, i

* It is analogous to the statement that the energy in a stretched wire (or spring) is equal to one-half stretching force times extension.

† "A Dynamical Theory of the Electromagnetic Field," *Phil. Trans.* (1864).

(i.e. current per unit area perpendicular to the direction of flow), is equal to

$$\sigma E,$$

by definition of the electrical conductivity, σ. When the medium is an insulating one, there may also be a current (a DISPLACEMENT CURRENT), but only while E is changing, and the current density is equal to the rate of increase of the displacement, D, or, as we may write it,

$$\frac{\partial D}{\partial t} \text{ or } \frac{K}{4\pi}\frac{\partial E}{\partial t}.$$

This is of course quite compatible with perfect insulation, since when E is constant, so is D.

There is also an analogous magnetic displacement current, namely

$$\frac{\mu}{4\pi}\frac{\partial H}{\partial t} \text{ or } \frac{1}{4\pi}\frac{\partial B}{\partial t}$$

where H is the magnetic field intensity, μ is the permeability of the medium, and $B\ (= \mu H)$ is the magnetic induction.

When Maxwell's hypothesis is combined with the familiar relations (*a*) between an electric current and its associated magnetic field (Ampère) and (*b*) between a changing magnetic flux and the associated electric field (Faraday and Neumann) the familiar partial differential equation representing a wave emerges and the phase velocity of this wave is

$$\frac{1}{\sqrt{\mu K}} \qquad \text{(IX—4)}$$

This expression has the same value in either electrostatic or electromagnetic units and Maxwell noticed that when it is reckoned for free space it is equal to

$$\frac{\text{Electromagnetic unit of Quantity}}{\text{Electrostatic unit of Quantity}}$$

It is easy to see that this is the case. In the electromagnetic system μ is equal to unity in free space and therefore the velocity, u, of E.M. waves *in vacuo* is

$$u = \frac{1}{\sqrt{K_{\text{e.m.u.}} \times 1}}$$

$K_{\text{e.m.u.}}$ meaning the dielectric constant of free space in E.M.

units. Therefore, since $K_{\text{e.s.u.}}$ (dielectric constant of free space in E.S. units) is equal to unity, we may write

$$u^2 = \frac{K_{\text{e.s.u.}}}{K_{\text{e.m.u.}}}.$$

But now the force in dynes between two equal point charges is

$$F = \frac{Q^2_{\text{e.s.u.}}}{K_{\text{e.s.u.}}\, r^2} = \frac{Q^2_{\text{e.m.u.}}}{K_{\text{e.m.u.}}\, r^2}$$

and so

$$\frac{Q^2_{\text{e.s.u.}}}{Q^2_{\text{e.m.u.}}} = \frac{K_{\text{e.s.u.}}}{K_{\text{e.m.u.}}}.$$

Consequently

$$u = \frac{Q_{\text{e.s.u.}}}{Q_{\text{e.m.u.}}},$$

which is equivalent to the statement

$$u = \frac{\text{E.M. unit for } Q}{\text{E.S. unit for } Q}.$$

It is usual to employ the letter c for this particular velocity. It appears to us, i.e. to the reader and myself, at this stage, as the velocity of E.M. waves in the free aether and relative to the aether; but we shall learn more about it in subsequent chapters.

As we have learned, the ratio of the E.M. and E.S. units had already been determined by Weber and has been determined since his time by many others. It is not possible to distinguish it from the velocity of light in free space (expressed of course in the same fundamental units of length and time, i.e. cm./sec.). Maxwell naturally and correctly inferred that light waves are electromagnetic waves and his theory is usually called the ELECTROMAGNETIC THEORY OF LIGHT. It united the two great groups of physical phenomena, electromagnetic and optical. That they were related in some way was foreseen by Faraday when he observed that a magnetic field caused a rotation of the plane of polarization of plane polarized light.

Maxwell's theory has been abundantly confirmed since his day. The comparatively long waves of broadcasting are Maxwell's electromagnetic waves and he may be said to have predicted them about sixty years before it became possible to bring them into efficient use.

FLOW OF ENERGY IN THE ELECTROMAGNETIC FIELD

A very important consequence of Maxwell's theory was deduced by J. H. Poynting (1852–1914).*

Poynting's theorem may be described in the following way. The rate at which the electromagnetic energy within a closed surface diminishes is equal to the sum of two things: (*a*) the rate at which heat, or any form of energy other than electromagnetic energy, is generated within the surface at the expense of the electromagnetic energy, and (*b*) the rate at which electromagnetic energy travels outwards through the surface in the form of an electromagnetic wave, this latter being the same as if the quantity of energy, equal to

$$\frac{1}{4\pi} E \times H,$$

were flowing through every unit area of the surface. This expression requires some amplification, which would take up too much space to be fully dealt with here. It is correct as it stands when E and H are perpendicular to one another and both in the surface. It is known as POYNTING'S VECTOR. If we imagine E to be in a northerly direction and H in an easterly one, then the flow of energy is perpendicularly downwards.

VERIFICATION OF MAXWELL'S THEORY

Perhaps the earliest experimental confirmation of Maxwell's theory was carried out by O. J. Lodge (later Sir Oliver Lodge).† He joined parallel wires to the inner and outer coats of a Leyden jar and arranged the length of the system so that when the jar discharged (and consequently produced the oscillations of Kelvin and Feddersen) there was a maximum spark at the free ends of the wires, in addition of course to the discharge in the immediate neighbourhood of the jars. He assumed the distance between the two sparks to be half a wave-length of the wave which travelled between the wires and was reflected at the end of the system. Multiplying the wave-length thus found by the known frequency of oscillation of the jar he estimated a velocity which was near to that of light.

* *Phil. Trans.*, 2, p. 343 (1884).
† *Phil. Mag.*, 26, p. 217 (1888).

E. Lecher greatly improved this device of two parallel wires and confirmed that the velocity of E.M. waves in air is equal to or near that of light.*

Undoubtedly the strongest support was given to Maxwell's theory by Heinrich Hertz (1857–1894), in his youth a student at Berlin under von Helmholtz, and later professor of physics in Bonn. He used a Ruhmkorff induction coil to produce sparks between two metal knobs connected by metal rods to two large plates (40 cm. square). The total length of the rods joining the plates to the knobs was about 60 cm. The system constituted effectively a condenser with a definite frequency of oscillation and sent out waves of a corresponding wave-length. He carried out a great variety of very convincing experiments. He studied the reflection of the waves and their refraction through prisms, measured wave-lengths and so on. An interesting feature in his experimental procedure was the receiver he used. It consisted of a piece of wire (or rod) bent into a circle except for a small gap. He was able, when his waves were reflected from a conducting wall, to locate nodes and internodes by observing the spark across the gap of his receiver, which of course was caused to oscillate strongly at an internode.

EARLY WIRELESS TELEGRAPHY

Sir Oliver Lodge appears to have noticed† that conductors in loose contact were caused to adhere to one another more firmly when electromagnetic waves fell on them. Branley utilized this in his coherer, which consisted of a mass of finely divided nickel filings with a few per cent of silver filings between silver plugs in an exhausted glass tube. This formed part of a circuit containing a battery and an electric bell or a Morse recorder. In the ordinary way the current in the circuit was too feeble to ring the bell or operate the receiving device; but immediately the E.M. wave arrives the filings cling together, the resistance of the circuit drops and the bell rings. The filings continue to adhere after the waves cease to fall on them and consequently an automatic tapping device was provided to shake them up and detach them from one another.

* E. Lecher: *Wied. Ann.*, 41, p. 850 (1890).
† *Journal I.E.E.*, 19, p. 346 (1890).

The old wireless telegraphy, which was developed mainly by the Italian G. Marconi, has been obsolete since the invention of the electron valve. It was very useful in its day, especially in aiding ships in distress at sea and in conveying information to them.

RADIATION PRESSURE

Maxwell predicted that, as a consequence of his electromagnetic theory, electromagnetic waves, including of course light waves, falling perpendicularly on a plane surface, must exert on it a pressure equal to the density of the electromagnetic energy per unit volume. This has been confirmed experimentally by the Russian P. Lebedev* and the Americans E. F. Nichols and G. F. Hull.†

If the electromagnetic radiation, instead of travelling in one specific direction, is travelling in all directions and uniformly distributed as regards direction (isotropic radiation), then the radiation pressure—as can easily be shown—is equal to *one-third* of the energy density. We shall meet with a most important application of this fact in the chapter on black body radiation.

THE IONOSPHERE

Oliver Heaviside suggested the existence of a conducting layer in the atmosphere to account for the fact that electromagnetic waves keep so closely to the earth's surface. E. V. Appleton (now Sir Edward Appleton) definitely proved the existence of this HEAVISIDE LAYER and found its height to vary from 80 to 100 kilometres. He also found another layer, the APPLETON LAYER, estimated to be about twice as high. These conducting layers constitute, collectively, the IONOSPHERE. Its conductivity is due to ionization by solar and cosmic radiation.

* P. Lebedev: *Rapp. Congrès Internat. d. Phys.*, 2, p. 133 (1900).
† E. F. Nichols and G. F. Hull: *Proc. Amer. Acad.*, 38, p. 559 (1903).

CHAPTER TEN

THE AETHER

MICHELSON'S EXPERIMENTS

UNTIL the end of the eighteenth century the natural philosophers adhered generally to Newton's theory of light rather than to that of Christiaan Huygens, who was the original author of the wave theory of light. Under the influence of Dr. Thomas Young and of Augustin Fresnel, both of whom produced very convincing experimental evidence for the correctness of the wave theory, Newton's views about light were confidently given up and until this century it was believed to be finally settled that a beam of light was a wave propagation in a strange medium called the AETHER, which, before Maxwell produced his electromagnetic theory, was thought to have properties like those of an elastic solid and later to be a medium in which electric and magnetic displacements could be produced and propagated.

In order to account for the phenomenon of stellar aberration, it will be recalled, Fresnel felt compelled to assume that the aether outside materials was permanently at rest—relative to some rather vague system of reference such as Newton's *spatium absolutum*. This view was still strongly held in the opening years of this century, notably by H. A. Lorentz, of Leyden. On the other hand, Fizeau's experiment seemed to provide evidence for Fresnel's prediction that the aether *inside* material media partook in some degree of their motion. About 1881 Albert Michelson, one of the most distinguished experimenters in the history of physical science, devised an experiment which seemed calculated at that time to pronounce decisively about Fresnel's view. It was designed to reveal and to determine the relative velocity of the earth and the aether outside it. He used a form of interferometer to which reference has been made in Chapter VII. It is illustrated in Fig. X—1, and consists of two plane mirrors, M_1 and M_2, whose reflecting surfaces are at right angles to one another. At O is a plane piece of glass, semi-

silvered on its lower face, which is at 45° to the planes of both mirrors. The light from a source of light on the left, which may be regarded as a nearly plane illuminated surface, is divided at *O* (Williams' division of amplitude), part of it travelling straight

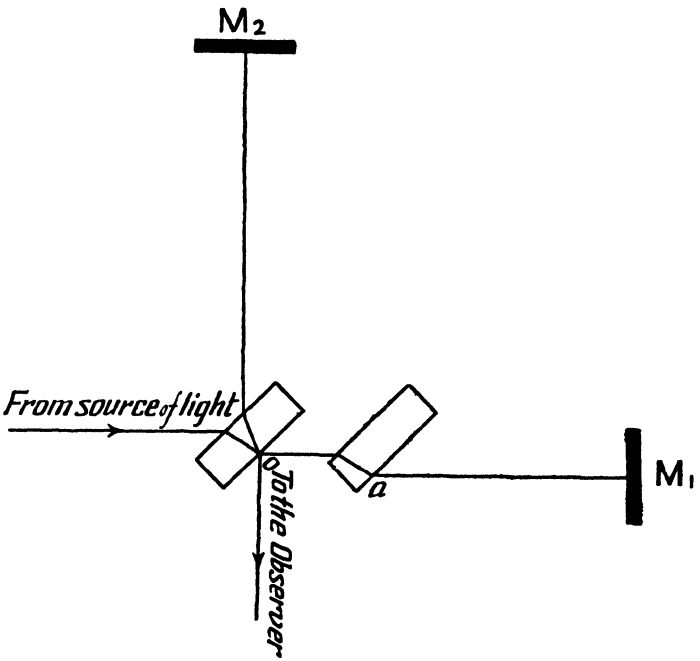


Fig. X—1

on to M_1 and the remaining part reflected to M_2. At *a* is an exactly similar "compensating" plate of glass whose function is to make the two paths, *O* to M_1 and *O* to M_2 (the two directions are perpendicular to one another), which have the same measured length, also have the same "optical" length for light of any wave-length.* The light which travels from *O* to M_1 and back to *O* will pass twice through the plate at *a*, while that which travels from *O* to M_2 and back to *O* will travel twice through the plate at *O*, both beams travelling the same distance through glass. The beams are reunited at *O* and interfere, the interference fringes being seen by the observer as indicated in the figure.

* Two paths have the same "optical" lengths when they contain the same number of waves.

For some of Michelson's purposes the possibility of having two equal optical paths was essential. When the two are equal the fringes (with monochromatic light) are straight, or nearly so, in the centre of the field. This is because the images of the nearly plane source of light are never strictly coincident, but intersect in an approximately straight line. One of the mirrors, e.g. M_2, can be moved by a screw nearer to or farther away from O. When the optical distance OM_2 is a little greater than, or a little less than, OM_1, the fringes appear slightly curved in opposite senses in the two cases. When the two paths are exactly equal coloured fringes can be seen with white light—but not when they are unequal—and Michelson's way of making this adjustment was to superpose faint white light on the monochromatic beam while turning the screw slowly in the direction required to change the sense of the curvature of the monochromatic fringes which were visible all the time. At a certain point the coloured fringes suddenly appear and then it is known that the two paths must be equal.

The whole apparatus was mounted on a stone slab which floated on mercury while the slab and the interferometer on it rotated continuously and slowly in a horizontal plane. *No change whatever could be observed in the position of the monochromatic fringes.* This simply meant that the light travelled each of the double journeys in the same time *whatever the directions of* OM_1 *or* OM_2 *relative to the earth's presumed motion through the aether.*

Imagine the earth to be moving through the aether with the velocity v, in the direction O to M_1. If L be the length of the path OM_1, the time (call it T_1) taken by the light for the double journey, O to M_1 and M_1 to O, is clearly

$$\frac{L}{c-v}+\frac{L}{c+v},$$

so that

$$T_1=\frac{2Lc}{c^2-v^2}$$

(c of course means the velocity of the light relative to the aether). For our purposes it is best to express this in the form

$$T_1=\frac{2L}{c}\,\frac{1}{\left(1-\dfrac{v^2}{c^2}\right)},$$

conveniently written

$$T_1 = \frac{2L}{c}\gamma^2, \qquad \text{(X—1)}$$

where γ means $(1 - v^2/c^2)^{-\frac{1}{2}}$.

The journey from O to M_2 and back is more complicated. The light must travel through the aether in the way shown in Fig. X—2. While it is going from O to M_2 and back, the point O gets

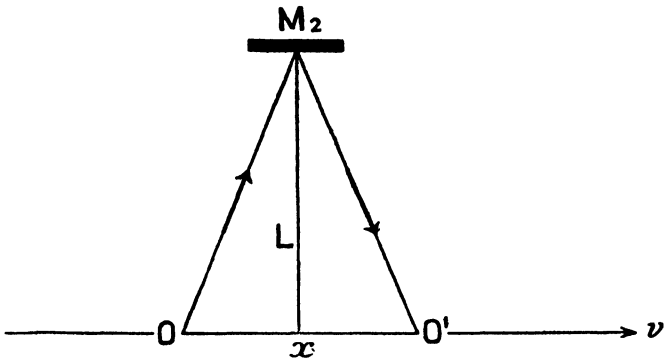


Fig. X—2

to O'. Obviously Ox (i.e. $\frac{1}{2}OO'$) is equal to $vT_2/2$, if T_2 is the time the light requires for its journey. The distance xM_2 is equal to L and therefore

$$(OM_2)^2 = L^2 + \frac{v^2 T_2^2}{4}$$

and consequently

$$\frac{c^2 T_2^2}{4} = L^2 + \frac{v^2 T_2^2}{4}.$$

Thus

$$T_2 = \frac{2L}{c}\gamma, \qquad \text{(X—2)}$$

γ having the same meaning as before.

Michelson's negative result makes the times T_1 and T_2 equal to one another, which means that $\gamma = 1$ and therefore $v = 0$; otherwise expressed, the aether is travelling along with the earth and this was the inference drawn by Michelson himself. The experiment was repeated by Michelson and his pupil Morley* with the same result; but with such refinements that, if the aether were as Fresnel and Lorentz contemplated, the fringe shift corresponding to the difference between T_1 and T_2 could easily have been observed.

* *Phil. Mag.*, 24, p. 449 (1887).

THE CONTRACTION HYPOTHESIS

Since the adoption of what seemed to be the most obvious interpretation of the negative result of this experiment of Michelson and Morley, namely, that the aether moves along with the earth, raised insuperable difficulties in the way of accounting for stellar aberration—and, incidentally, for the result of Fizeau's convection experiment—it seemed necessary to find an interpretation of their negative result which would permit the aether to remain "at rest". If we may not suppose

$$\gamma^2 = \gamma = 1,$$

there is still the alternative suggested by G. F. FitzGerald (1851-1901) and independently by H. A. Lorentz, namely, that bodies contract in length in the direction in which they are moving through the aether. Let us suppose that OM_1, when it is in this direction, contracts to a shorter length, L^1. Then the formula (X—1) becomes

$$T_1 = \frac{2L^1}{c}\gamma^2$$

and we can explain the result

$$T_1 = T_2$$

by supposing

$$\gamma L^1 = L.$$

That is to say, the modified length is equal to the original length, L, divided by the factor γ which is slightly bigger than unity. This is the CONTRACTION HYPOTHESIS of G. F. FitzGerald* and H. A. Lorentz. The latter had noticed that the familiar equation of Poisson in electrostatics changes its form when one passes from one system of rectangular co-ordinates to another moving relatively to the former with some velocity v. On the other hand there is no experimental indication that electrostatic phenomena are different when referred to the new co-ordinates. Strangely enough the contraction hypothesis is just what is needed to make the equation have the same form in both systems.†

* The contraction hypothesis of FitzGerald is mentioned by O. Lodge in a paper, "On Aberration Problems," *Phil. Trans. A.*, 184, p. 749 (1893).

† H. A. Lorentz: *Versuch einer Theorie der elektrischen und optischen Erscheinungen in bewegten Körpern* (Leiden, 1895).

THE EXPERIMENTS OF OLIVER LODGE

After Michelson and Morley had carried out their earlier experiments, Sir Oliver Lodge attacked the same problem in a different way. He used two heavy steel discs, mounted close together on a common axle. A beam of monochromatic light was divided, as in Michelson's interferometer for example, one part being sent round a closed loop between the discs (by suitably arranged mirrors) and the other made to travel in exactly the opposite sense round the loop. The two parts of the original beam were reunited to produce interference fringes. Now it was to be expected that the fringes would shift on setting the discs in motion, provided of course that they dragged the aether along with them; since the velocity of light would become greater in one direction than in the other. Lodge could observe no change whatsoever, indicating that the aether was stagnant. He repeated the experiment with a spheroid of iron weighing half a ton and with a sort of canal or groove cut out along its equator. Round this, by suitably disposed mirrors, he sent the beams of light in opposite senses. The result was just the same and the only way of reconciling it with that of Michelson and Morley appeared to be the adoption of the contraction hypothesis. The situation was not yet desperate, but further experiments brought the physicists of the opening years of this century into a state of complete bewilderment.

THE EXPERIMENT OF TROUTON AND NOBLE*

A charged parallel plate condenser was suspended, by a delicate bifilar suspension, so that its plates were vertical. It was expected that the motion of the charged plates through the aether, if there were such relative motion, would give rise to a couple which would tend to turn the condenser towards the position in which the plates were parallel to the direction of motion, with the positive plate on the left when viewed by some one looking in the direction of the motion. Trouton and Noble found no sign of such a couple, thus supporting Michelson's view that the aether moves along with the earth. This experimental

* *Proc. Roy. Soc.*, 72 (1903).

result *cannot be accounted for by the contraction hypothesis* of FitzGerald, who suggested the experiment to Trouton.

OTHER EXPERIMENTS

If such a contraction of materials as contemplated by the FitzGerald–Lorentz hypothesis really did occur, it should have consequences. A transparent isotropic material, glass for example, might be expected to show double refraction. Lord Rayleigh failed to observe this* and Brace, who repeated his experiments with more refinements, also failed.†

Another consequence that might be anticipated would be a slight variation of the electrical resistance of a piece of wire with its direction. Trouton and Rankine failed to observe this, though their experimental method was amply sensitive enough to detect a variation of the expected magnitude.

A SUMMARY OF PERPLEXITIES

Young, Fresnel and all their successors through the nineteenth century held light to be an undulatory phenomenon requiring a very special LUMINIFEROUS MEDIUM (the aether) which extended, not only through otherwise empty space, but which permeated material media and occupied the space between the atoms. Stellar aberration required this aether to be "at rest" or stagnant, outside materials, and Fizeau's experiment did not appear to be in conflict with this view. Difficulties began with Michelson's experiment. These, it seemed, could be removed by the rather artificial contraction hypothesis. Lodge's experiment favoured the stagnant aether view. On the other hand all the obvious consequences of contraction failed to manifest themselves and the experiment of Trouton and Noble seemed to admit of no other interpretation except the naïve one of Michelson. These perplexities were the outcome mainly of conflicting *negative* results. We shall see how they were removed, along with other still older difficulties, by the theory of relativity.

* Lord Rayleigh: *Phil. Mag.*, 4, p. 678 (1902).
† Brace: *Phil. Mag.*, 7, p. 317 (1904).

CHAPTER ELEVEN

NEWTONIAN AND SPECIAL RELATIVITY

OLD DIFFICULTIES

IT was Eddington, I think, who made the remark that the Inquisition and Galileo were agreed on one point, namely that there was indeed a certain question to be answered, and on this point, about which they were agreed, they were both wrong. The Inquisition contended that there were reasons so authoritative and compelling for belief in the immobility of the earth that it was indeed impious not to accept them. Galileo on the other hand was entranced by the beautiful order and simplification introduced into the apparently erratic behaviour of planetary bodies, by abandoning the ancient belief. We cannot wonder of course that he still retained some prejudices and misconceptions, inherited from an earlier time, as did indeed the great Newton who followed him. The chief of these was that it had a meaning to say that a body was at rest or in motion quite independently of anything relative to which it was at rest or in motion. In the earliest times, naturally, the earth was at rest; then the belief came (to Aristarchus, now almost forgotten, in the third century B.C.—revived by Copernicus nearly 2000 years later) that the earth was in motion and immobility was transferred elsewhere—to the stellar system or (Newton) to absolute space. For Newton space was something which was there in its own right. It was an entity (not unlike the aether which played so great a part in later times) which existed and persisted independently of the material things "contained in it." "*Spatium Absolutum, natura sua sine relatione ad externum quodvis, semper manet similare & immobile. . . .*" His definition of time begins in the same way, but after *quodvis* there follows in its description the words *aequabiliter fluit*. Newton's mechanical and gravitational theory *seemed* to rest on this as the deepest part of its foundations, but when we scrutinize the applications which Newton and his successors made of his mechanical prin-

ciples we fail to find his absolute space and time. They are in fact *never involved.* Indeed neither his absolute space nor his absolute time can be observed and the only way of defining a particular straight line, for example—and I am not here speaking of that quite different thing, the pure mathematician's straight line—is in terms of something that can be observed; perhaps I should say, in terms of an asymptotic sequence of observations and operations involving material things.*

The great German mathematician, Carl Neumann, dealt with these things in his inaugural lecture on appointment as Professor of Mathematics in the University of Leipzig (1869).† This remarkable address contains, I think, the earliest serious examination of the principles of mechanics which dominated physical science till the close of last century. He examined, in particular, the axiom with which we are all familiar under the name of Newton's FIRST LAW OF MOTION. As he said, it is in fact without content and incomprehensible until we can provide acceptable definitions of "straight line", "equal distances" and "equal times". He pointed out that motion in a straight line, defined with reference to some framework fixed in the earth, would appear as motion along a curved line if referred to a framework fixed in some other planet. We must therefore, since we are convinced of the importance of the law, ask ourselves what is meant by *straight line* and by *motion in a straight line,* and so on.

Neumann came to the conclusion that there must exist somewhere in the universe an absolutely rigid body (*absolut starrer Körper*)—a body whose figure and dimensions remain unchanged for all time. He called it the body ALPHA. It is this body which determines whether anything is at rest or in motion and

* Professor E. T. Bell writes in his fascinating book on *The Development of Mathematics* that "nothing so exasperates a mathematician who knows his trade as to hear geometry called a physical science". Of course we grant him the profound difference between his geometry and that of the physicist; but the pure mathematician may be reminded that he borrowed the term "geometry" from the physicist, who has used it quite legitimately for more than two millennia as the name of a fundamental part of his science. One is tempted to ask whether the pure mathematician is becoming exasperated by the physicist's use of the term "measure" which is so prominent in present-day pure mathematics.

† Carl Neumann: *Ueber die Principien der Galilei—Newtonschen Theorie* (Leipzig, B. G. Teubner, 1870).

which settles authoritatively what straight lines are and what are motions along straight lines. One gathers from his lecture that this body Alpha is not necessarily to be regarded as an actual substantial rigid lump of material, as it were. He had in mind the existence of a system of reference (e.g. a system of rectangular axes of co-ordinates) definable by observational procedure and he suggested the principal axes of inertia of the stellar system, as he thought of it. He also appreciated that each of the infinity of systems of, say, rectangular axes of co-ordinates, each moving with a constant velocity of *translation* relatively to the body Alpha, and consequently with constant velocities of translation relatively to one another, is just as fitted to be a system of reference as the body Alpha itself. Some of them would of course be more convenient to use for particular problems than others, but each one of them equally valid as a system of reference for the purposes of Newtonian mechanics, if the body Alpha has this validity. This amounts to saying that the laws of mechanics have the same form (i.e. the equations which express them have the same form) in all systems which have a constant velocity of translation relative to Neumann's body Alpha. Such systems of reference are now called INERTIAL SYSTEMS.

Newton's mechanical principles fit a large class of phenomena extraordinarily well—those which may be described as matter and motion phenomena. They fail with very small things and to some extent with gravitational phenomena and the extra-galactic world. Their application to electromagnetic phenomena is also rather imperfect. Newton's absolute space is not only not needed for the purposes of his mechanics; it is in fact meaningless. Space (that is to say the space of the physicist) is no more than a complex of relationships of a certain type—relationships between observable material things, expressed in terms of measured lengths, angles and constructions built up with them. If we were to imagine all the material bodies and observable things, which are said to "occupy space", to be annihilated, space too would vanish. Indeed we might make a similar remark about time as well. The two, as will be better understood when we study the further development of the theory of relativity, are inseparable. H. G. Wells appears to have appreciated this. His *Time Traveller* anticipated Minkowski some fifty or sixty years ago.

Neumann's body Alpha is an advance beyond Newton's absolute space, since it can be effectively defined by an asymptotic sequence of observations. What this means may be illustrated in the following way. Even quite rough approaches to simple mechanical problems are attended by appreciable success. If we examine the behaviour of a simple pendulum, it agrees quite well with what Newtonian theory predicts even when we refer it to a reference system fixed relatively to the earth, provided our observations are not exceedingly good. But when we proceed very carefully (like Foucault) we notice some deviation from what our naïvely applied Newtonian mechanics predicts and to obtain agreement we must refer our pendulum to a system relatively to which the earth is in rotation. In this way we find an inertial system which Newton would have regarded as moving with a constant velocity of translation in his *spatium absolutum*, or which Neumann would have regarded as moving in such a way relatively to his body Alpha.

We come then to the conclusion that the equations of Newtonian mechanics have the same form in a multiply infinite number of systems (inertial systems) the definition of which presents no formidable problems. This is of course THE PRINCIPLE OF RELATIVITY, as Einstein calls it, in a very restricted form. In its full generality it would assert that the equations of physics, properly expressed, have the same form in ALL reference systems. I am inclined to adopt the view, with which Henri Poincaré would probably have agreed, had he lived longer, that the principle is a convention.

NEWTONIAN RELATIVITY

It is illuminating and, in view of what is to follow, helpful to consider how a simple problem in Newtonian mechanics appears when referred to *different* inertial systems. Almost the simplest we can contemplate is that of a particle moving under the influence of a constant force, for example the problem of a pebble thrown into the air. For our present purpose we may neglect the resistance of the air and we may adopt, as a sufficiently good first approximation to an inertial system, a system of rectangular axes of co-ordinates fixed relatively to the earth, since the acceleration of such a system, relatively to an inertial

one, is very small compared with that of the pebble. The simplest choice of co-ordinate axes is one in which one of the axes (e.g. the X axis) has the same direction as the force acting on the pebble (its weight). That is the direction we usually describe as vertically downward. The X component of the force would then be mg (mass $\times$ acceleration) and the remaining components zero. But in general the X component of the force, F_x, is equal to $mg \cos \theta_1$, where θ_1 means the angle between the direction of the X axis and that of the force (weight). There are similar expressions for the Y and Z components, so that

$$F_x = mg \cos \theta_1,$$
$$F_y = mg \cos \theta_2,$$
$$F_z = mg \cos \theta_3.$$

The equations of motion of the pebble are then

$$m \frac{d^2x}{dt^2} = F_x,$$
$$m \frac{d^2y}{dt^2} = F_y, \qquad \text{(XI—1)}$$
$$m \frac{d^2z}{dt^2} = F_z,$$

the components F_x, F_y and F_z being constants in this simple problem.

The important thing which this illustration is going to elucidate does not require the actual solution of these equations. That may be left as an exercise for the reader. It is that the equations *have the same form*, in any one of the infinity of inertial reference systems, as that of (XI—1). We may write them down for another system as follows:

$$m \frac{d^2x'}{dt^2} = F_x',$$
$$m \frac{d^2y'}{dt^2} = F_y', \qquad \text{(XI—1A)}$$
$$m \frac{d^2z'}{dt^2} = F_z'.$$

We now inquire what this identity in form (covariance)

implies. It simply means that, when we pass from the system in which

$$m\frac{d^2x}{dt^2} = F_x$$

holds to that in which

$$m\frac{d^2x'}{dt^2} = F_x'$$

holds, we have to apply the *same rule* in calculating F_x' from F_x, F_y and F_z as we use in calculating x' from x, y and z. Sets of quantities, F_x, F_y and F_z, which transform like the co-ordinates x, y and z of a point, are called VECTORS. It is precisely this which such propositions as the triangle and polygon of vectors (e.g. of forces) express. It is implicit in Newtonian mechanics that the mass, m, and the time, dt, are not changed in passing from one system to another. They are INVARIANT or SCALAR quantities. It is important to notice—what the definition of a vector, just given, indicates—that it is quite insufficient to define a vector simply as a quantity associated with a direction. The definition of a vector may also be put in the following way: A displacement is a vector and anything with the same transformation properties as a displacement is also a vector.

VECTORS AND TENSORS

There are other sets of quantities, such for example as those describing the state of stress at a point in a material medium, which have NINE COMPONENTS and follow the same rules of transformation as the products

$$\begin{array}{lll} xx, & xy, & xz, \\ yx, & yy, & yz, \\ zx, & zy, & zz. \end{array}$$

It will be remembered that the stress components of (IX—1) were in fact distinguished as

$$t_{xx}, \quad t_{xy}, \quad t_{xz}$$

and so on, with subscripts marking pairs of co-ordinate directions. Such things are called TENSORS and the term TENSOR has gradually come to be used as a general one for all things of this kind. Scalars are tensors of RANK ZERO. Such a tensor has only ONE COMPONENT, i.e. the number of its components is 3^0. A vector is a tensor of RANK ONE; it has 3^1 components. A state of stress (as

also one of strain) is represented by a tensor of RANK TWO; it has 3^2 components, and so on. The exponents, 0, 1, 2, etc., represent the rank, while 3 is the number of dimensions of the Euclidean continuum. It should now be intelligible that the equations expressing physical laws have to be relations between tensors in order that they may be covariant, i.e. retain the same form in different reference systems.

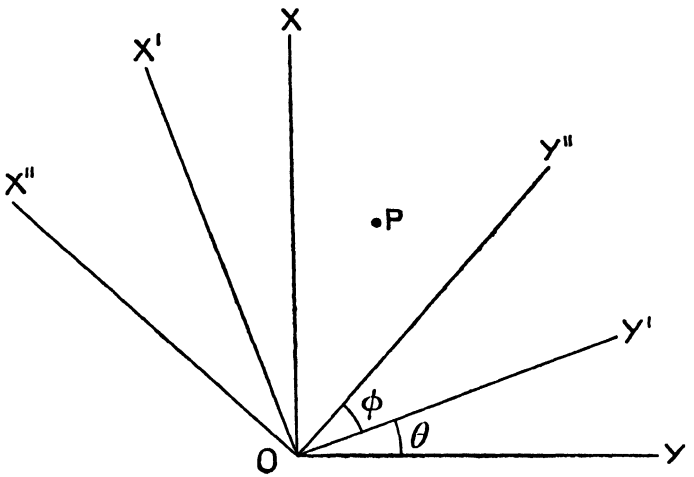


Fig. XI—1

THE SCOPE OF NEWTONIAN RELATIVITY

It will be helpful to examine the simple transformation from a rectangular co-ordinate system, X, Y, Z, to another, X', Y', Z', both having a common origin and, for simplicity, a common Z axis. The equations of transformation are

$$\begin{aligned} x' &= x\cos\theta - y\sin\theta, \\ y' &= x\sin\theta + y\cos\theta, \\ z' &= z, \end{aligned} \qquad \text{(XI—2)}$$

in which x, y and z are the co-ordinates of some point, P (see Fig. XI—1), in one system and x', y' and z' its co-ordinates in the other.

Now it follows that

$$\begin{aligned} F'_x &= F_x\cos\theta - F_y\sin\theta, \\ F'_y &= F_x\sin\theta + F_y\cos\theta, \\ F'_z &= F_z, \end{aligned} \qquad \text{(XI—2A)}$$

since the equations of transformation of a vector must have *exactly the same form* as those for the co-ordinates of a point. Furthermore, since the product $x'y'$ is related to the products xx, xy, xz, yx, etc., according to the formula

$$x'y' = (x\cos\theta - y\sin\theta)(x\sin\theta + y\cos\theta),$$

or

$$x'y' = xx\cos\theta\sin\theta + xy\cos^2\theta - yx\sin^2\theta - yy\sin\theta\cos\theta,$$

so it follows that the component, t'_{xy} (for example) of the stress tensor of Chapter IX is given, in this reference system, by

$$t'_{xy} = t_{xx}\cos\theta\sin\theta + t_{xy}\cos^2\theta - t_{yx}\sin^2\theta - t_{yy}\sin\theta\cos\theta \qquad \text{(XI—3)}$$

This is the kind of calculation we do when we change from the one co-ordinate system to the other.

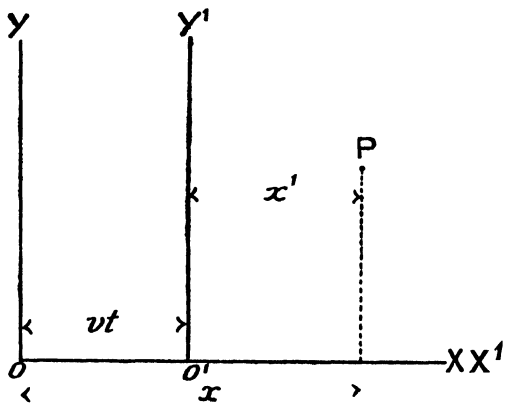


Fig. XI—2

We turn now to something still more important, namely, the change from one (inertial) co-ordinate system to another, the latter of which is travelling with a constant velocity of translation relatively to the former. The Newtonian equations of mechanics have the same form in both systems, as we have seen.

The special case illustrated in Fig. XI—2 is not only very simple, but of great interest. The two systems X, Y, Z and X', Y', Z' have their X axes coincident. The Y and Y' are parallel to one another, and shown in the plane of the paper. The Z and Z' axes are likewise parallel to one another and perpendicular to the plane of the paper, so that they do not appear in the

diagram. Finally the two systems (and therefore also their origins O and O') are exactly coincident at the instant $t = 0$; so that

$$OO' = vt$$

if, as we are going to suppose, the system X', Y', Z' is moving relatively to X, Y, Z, in the direction of the common X axis with the constant velocity, v. If now the co-ordinates of a point, P, are x, y, z and x', y', z' in the two systems, it is obvious that

$$\begin{aligned} x' &= x - vt, \\ y' &= y, \\ z' &= z, \\ t' &= t. \end{aligned} \qquad \text{(XI—4)}$$

It should be carefully noted that

$$\begin{aligned} y' &= y \\ \text{and } z' &= z \end{aligned}$$

only because of our special choice of co-ordinate systems, but

$$t' = t$$

whatever choice we may make. It has been added only for the sake of the comparison we shall make later with the famous LORENTZ TRANSFORMATION.

If the point, P, happens to be moving with a velocity whose components are u_x, u_y and u_z and u'_x, u'_y and u'_z in the respective systems, then obviously

$$\begin{aligned} u'_x &= u_x - v, \\ u'_y &= u_y, \\ u'_z &= u_z. \end{aligned} \qquad \text{(XI—5)}$$

GROUPS OF TRANSFORMATION

Turning back to the transformation (XI—2), let us make a further transformation to a third system X'', Y'', Z'', as indicated in Fig. XI—1. We get, of course,

$$\begin{aligned} x'' &= x' \cos\varphi - y' \sin\varphi, \\ y'' &= x' \sin\varphi + y' \cos\varphi, \\ z'' &= z'. \end{aligned}$$

We should be greatly embarrassed if the result turned out to be different from that due to the direct transformation from

X, Y, Z, to X'', Y'', Z''; but on substituting for x' and y' in terms of x and y (using (XI—2)), we do in fact get

$$x'' = x\cos(\theta + \varphi) - y\sin(\theta + \varphi).$$

This property of the transformation is called the GROUP PROPERTY and such transformations are said to constitute a GROUP OF TRANSFORMATIONS.

One last remark: The equations (XI—2) may be expressed in another way which has a special appeal of its own:

$$x' + iy' = (x + iy)e^{i\theta},$$
$$x'' + iy'' = (x' + iy')e^{i\varphi}.$$

On multiplying these two together we find the group property exhibited in a compact and elegant form,

$$x'' + iy'' = (x + iy)e^{i(\theta+\varphi)}.$$

Needless to add, i means $\sqrt{-1}$ and e is the familiar number 2·718....

THE LORENTZ TRANSFORMATION

The Newtonian transformation (XI—4) and its consequence (XI—5) are satisfactory—within fairly wide limits—when applied to mechanical phenomena and Newtonian theory and observation are in quite excellent agreement. There are only minute discrepancies between theory and observation.

Grave difficulties arise with electromagnetic phenomena. While on the one hand *experiment* indicates that such phenomena are represented by equations of the same form in different inertial systems; on the other hand Maxwell's equations do *change their form* under the transformation (XI—4). There is also the difficulty raised by Michelson's experiment. Imagine the X, Y, Z, system of Fig. XI—2 to be fixed in the (presumed) stagnant aether and the X', Y', Z', system to be fixed in the earth. We should expect from (XI—5), for light travelling in the common X direction,

$$u = c - v,$$

if u is the velocity of light in the system X', Y', Z' and c that in the X, Y, Z, system; but Michelson found a universal

$$u = c$$

and concluded that $v = 0$; but as we have seen, this raises formidable difficulties.

The FitzGerald–Lorentz contraction hypothesis seemed to Lorentz not only to explain Michelson's result, but to promise a way of meeting the difficulties raised by electromagnetic phenomena generally. Why not, one might ask—putting oneself for a moment in the situation of Lorentz—why not replace the first equation of (XI—4) by

$$x' = \gamma(x - vt)\,?$$

where γ is the contraction factor which we have already encountered (see (X—1) and (X—2)). It will be remembered that Poisson's equation, which governs electrostatics, retains its form under the influence of the factor γ. To cut a long story short, Lorentz replaced the Newtonian (XI—4) by

$$\begin{aligned} x' &= \gamma(x - vt), \\ y' &= y, \\ z' &= z, \\ t' &= \gamma\left(t - \frac{vx}{c^2}\right), \end{aligned} \qquad \text{(XI—6)}$$

or, to be strictly accurate, he adopted a set of equations *almost* equivalent to these.* It will be noticed that he was forced to change the time equation as well as the X co-ordinate one. He called the new time LOCAL TIME. These equations (XI—6) are now known as the LORENTZ TRANSFORMATION and they replace the Newtonian one (XI—4). It is easy to see that they lead to

$$\begin{aligned} u'_x &= (u_x - v)/(1 - vu_x/c^2), \\ u'_y &= u_y/\nu\,(1 - vu_x/c^2), \\ u'_z &= u_z/\nu\,(1 - vu_x/c^2), \end{aligned} \qquad \text{(XI—7)}$$

which now replace (XI—5). These last so-called kinematical equations were first given by A. Einstein (1905). It will be noticed—and this is of the very greatest interest—that, when

* H. A. Lorentz: *Proc. Acad. Sci.*, Amsterdam, 6, p. 809 (1904). Lorentz remarks, in one of his later papers on this subject, that he had been anticipated by W. Voigt in a paper published in the *Göttinger Nachrichtungen*, p. 41 (1887).

u_x is equal to c (velocity of light in free space, or, if you like, in the aether), we get from (XI—7)

$$u'_x = \frac{c - v}{1 - \frac{v}{c}}$$

or

$$u'_x = c,$$

in agreement with Michelson's result. There is another, very striking, consequence of (XI—7). Suppose some transparent material, glass or water, at rest in the X', Y', Z' system of Fig. XI—2 and a beam of monochromatic light travelling through it in the X' direction. Its velocity relative to X', Y', Z', in which the transparent material is at rest, is of course equal to c/n, where c is the velocity of light *in vacuo* and n is the refractive index of the material. We may then substitute c/n for u'_x in the first equation (XI—7), so that

$$\frac{c}{n} = \frac{u_x - v}{1 - \frac{u_x v}{c^2}}$$

This gives

$$u_x = \frac{\frac{c}{n} + v}{1 + \frac{v}{nc}}$$

and consequently, on neglecting second order small quantities,

$$u_x = \frac{c}{n} + v\left(1 - \frac{1}{n^2}\right)^*$$

This of course is Fresnel's formula. The formal coherence of these things is very impressive and compels us to take the Lorentz transformation very seriously. The vision of Lorentz was undoubtedly obscured by his belief in the reality of the aether, which played for him much the same rôle as did absolute space for Newton. He seems to have regarded the transformation as expressing an actual distortion in the dimensions of bodies, which was brought about somehow by their motion through the

* In this derivation of the formula, as well as in the earlier one, the small Doppler change in n as between X, Y, Z and X', Y', Z' has been ignored.

aether—yet a distortion which was hidden from experimental observation—and a more mysterious analogous distortion in the temporal intervals between successive events.

The whole subject appeared in a new light in 1905, when Albert Einstein (1879–) published his great paper on the "Electrodynamics of Moving Bodies."*

THE SPECIAL THEORY OF RELATIVITY

Einstein's investigation was based on two premisses:

(*a*) The equations of electrodynamics have the same form in all systems in which the equations of mechanics (i.e. Newton's mechanics) hold.

(*b*) The velocity of light in empty space has under all circumstances the same value (now usually represented by c and a very little below 3×10^{10} cm. sec.$^{-1}$).

The premiss (*a*) he called the PRINCIPLE OF RELATIVITY (*Prinzip der Relativität*) and later the principle of special relativity or the SPECIAL PRINCIPLE OF RELATIVITY, to distinguish it from a more general form which he successfully developed some ten years later. These premisses led to the Lorentz transformation (XI—6) and among other things to the conclusion that the mass of a body, as defined by the formula

FORCE = RATE OF CHANGE OF MOMENTUM,

or

$$F = d(mv)/dt,$$

ries with its velocity according to the law

MASS = $\gamma \times$ (MASS OF BODY WHEN AT REST)†

Einstein also made the great discovery, which in fact is a consequence of the last statement, that the energy of a body is proportional to its mass:

ENERGY = MASS $\times c^2$.

A remarkable feature of his theory is its complete ignoration of the aether. It will be remembered that all attempts to observe it have failed and Einstein was at that time, I think, a disciple of the great Austrian philosopher and physicist, Ernst Mach, who sternly condemned the assumption of anything that defied

* *Ann. d. Physik*, 17, p. 891 (1905).

† This, I believe, was known to Lorentz.

observation and *a fortiori* anything the definition or description of which actually implied its unobservability. The aether of Lorentz belonged to this latter category. It still eludes observation, and the fine coherence of physical theory, in which it has (now at any rate) no place, may be regarded as evidence that there is no such thing. The aether is the Mrs. Harris of physical science.

The reader will very properly ask what becomes of the electromagnetic stresses, momentum and energy which are associated with electromagnetic waves, even in empty space, if there is no medium, like the aether, to support them? The answer to this question was forthcoming much later in the wave mechanics which we owe to the genius of Prince Louis de Broglie and Erwin Schrödinger.

MINKOWSKI'S INTERPRETATION OF THE LORENTZ TRANSFORMATION

In the transformations of Newtonian mechanics time is, as we have seen, an invariant. If the time interval between two events is 20 seconds when we refer it to one system (X, Y, Z) it is also 20 seconds when referred to any other system (X', Y', Z'). This is obviously not so when we adopt the Lorentz transformation. We are reminded by this change in the time of the change in the X co-ordinate of a point, which can happen when we introduce a new set of axes of co-ordinates, and there is a strong suggestion that we are now involved in a wider continuum—a contin embracing the time as well as the three spatial dimensi* This occurred to Hermann Minkowski (1864–1909), who, it be remarked, was one of Einstein's mathematical teachers in Polytechnicum at Zürich. The fourth sentence of his famous address to the former German equivalent of our British Association at the Cologne meeting on the 21st September, 1908, runs:

"Von Stund an sollen Raum für sich und Zeit für sich völlig zu Schatten herabsinken und nur noch eine Art Union der beiden soll Selbstständigkeit bewahren."†

* The reader should turn to the initial pages of H. G. Wells' *Time Machine.*

†"From now on space by itself and time by itself fade away to shadows and only a sort of union of the two remains."—*Die 80te Versammlung Deutscher Naturforscher und Aertze zu Cöln.*

It is fairly easy to describe Minkowski's contribution—one of the greatest—to relativity. Let us boldly introduce a fourth axis—these are almost his words—distances measured along which shall be proportional to the time. I shall call this axis W, and distances measured along it, w. Since w is always proportional to the time,

$$w = st,$$

where s is some constant (evidently with the dimensions of a velocity) which we already suspect to have some intimate relationship to c, the velocity of light in free space. This fourth axis is related to our Euclidean space just as the Z axis in a

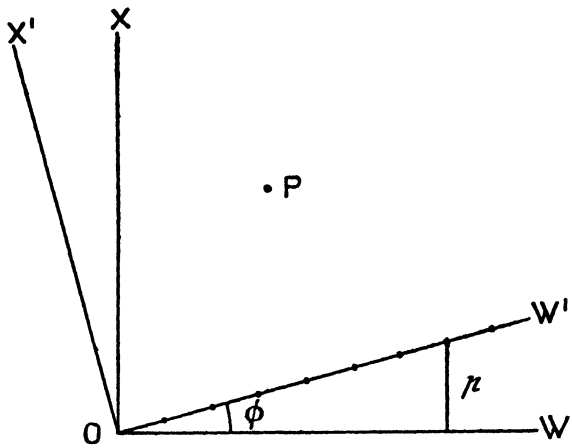


Fig. XI—3

Euclidean space co-ordinate system is related to the plane containing the X and Y axes. Consider now a point event, P (which simply means a point in our enlarged continuum), occurring at a place, i.e. at a spatial point, whose X co-ordinate is x and whose Y and Z co-ordinates are both zero, but whose W co-ordinate is w, and let us study a transformation to another system, X', Y', Z', W', such that $y' = y$ and $z' = z$, and try to find out what emerges when we assume our four-dimensional continuum to have a Euclidean character. We adopt the picture indicated in Fig. XI—3 and are able to write down at once the following equations of transformation:

$$\begin{aligned} x' &= x\cos\varphi - w\sin\varphi, \\ w' &= x\sin\varphi + w\cos\varphi, \\ y' &= y, \\ z' &= z, \end{aligned} \qquad \text{(XI—8)}$$

which should be compared with (XI—2).

Consider now the points on the W' axis. They represent perhaps a particle whose X', Y' and Z' co-ordinates remain constant and equal to zero, but in the other system its X co-ordinate increases as w increases (and therefore as t increases). The perpendicular, p, represents the X co-ordinate of such a particle for a particular value of w, i.e. at a particular time, t. Obviously p grows bigger and bigger as w increases and therefore as t increases. In fact the points on the W' axis represent a particle at rest in X', Y', Z' and in motion in X, Y, Z with a velocity x/t. Now

$$\tan\varphi = x/w,$$

that is

$$\tan\varphi = x/st = v/s. \qquad \text{(XI—9)}$$

It easily follows that

$$\cos\varphi = (1 + v^2/s^2)^{-\frac{1}{2}},$$

and

$$\sin\varphi = (v/s)(1 + v^2/s^2)^{-\frac{1}{2}}.$$

Let us write, for brevity,

$$\gamma = (1 + v^2/s^2)^{-\frac{1}{2}},$$

then

$$\begin{aligned} \cos\varphi &= \gamma, \\ \sin\varphi &= \gamma v/s. \end{aligned} \qquad \text{(XI—10)}$$

If now we substitute these expressions for cos φ and sin φ in (XI—8), we find for the first equation

$$x' = \gamma x - \gamma\frac{vw}{s},$$

or, since $w = st$,

$$x' = \gamma(x - vt),$$

which has exactly the form of the first equation of the Lorentz transformation (XI—6). For the next equation we easily get

$$w' = x\frac{v}{s}\gamma + w\gamma$$

or

$$t' = \gamma\left(t + \frac{vx}{s^2}\right),$$

which becomes the corresponding Lorentz equation when we identify

$$s^2 \text{ with } -c^2;$$

so that

$$s = \sqrt{-1} \times c$$

and

$$w = \sqrt{-1}\, ct.$$

We need not worry about the seemingly strange apparition of $\sqrt{-1}$. It is a very useful symbol and plays a great part in very matter-of-fact and practical things, e.g. in the theory of alternating electric currents.

THE THEOREM OF PYTHAGORAS

The equations (XI—8) imply

$$x^2 + w^2 = x'^2 + w'^2$$

and in the special case where $x/t = \pm c$, i.e. the case where the particle or photon is travelling with velocity c, we have

$$x^2 + w^2 = x^2 - c^2t^2 = 0.$$

Consequently

$$x'^2 + w'^2 = x'^2 - c^2t'^2 = 0$$

and therefore

$$x'/t' = \pm c.$$

This means that when anything is travelling with the particular velocity c (the velocity of light) as referred to the spatial co-ordinates X, Y, Z it must be travelling with the same velocity when referred to X', Y', Z'. Thus in Minkowski's four-dimensional space–time continuum, Michelson's experimental result may be interpreted as a consequence of the ancient theorem of Pythagoras.

MISCELLANEOUS CONCLUSIONS

This four-dimensional interpretation of special relativity presents both mechanics and electromagnetism in a very beautiful and indeed rather simple form; but its complete presentation would take us too far afield. Let it suffice to point out how some of our difficulties have now vanished. Two events occurring at different

places, but at the same time when referred to one system of co-ordinates, are in general no longer simultaneous when referred to another system, as a simple application of the Lorentz transformation indicates. When we examine this in the light of Minkowski's theory we find nothing puzzling about it. We are familiar with the fact that the X co-ordinates of two points may be equal in one system and very different in another. This is equally true of the W co-ordinates (that is the time co-ordinates) of a point event. The FitzGerald–Lorentz contraction too is no longer mysterious for a similar reason. The rod which is said to "contract" in fact *remains unchanged*, but the difference between the co-ordinates of its ends has not the same value in different co-ordinate systems. There is in fact *no physical change* in a material when the system of reference is changed; hence the negative result of such experiments as those of Rayleigh and Brace.

When we refer a plane light wave (in free space) first to the X, Y, Z system and then to the X', Y', Z' system, we find its directions to be slightly different in the two systems. This is the phenomenon of aberration. Suppose the light is travelling towards a terrestrial observer in a direction at right angles to his motion *when referred to a reference system fixed relatively to the star*, it is found by a simple application of the Lorentz transformation to be travelling in a slightly different direction *when referred to a reference system in which the observer is at rest.* In fact the sine of the angle between the two directions is equal to v/c, where v means the relative velocity of the two systems. This angle is not dependent in the slightest degree on the nature of the transparent medium filling the observer's telescope (Airy's experiment).

We may conclude this chapter by reference to a certain type of conundrum which bothers some people. A wheel is set in motion of rotation about its axle, which is at rest (i.e. in the particular reference system we are using). Now it is argued that the circumference of the wheel contracts in consequence, but the spokes do not change in length since the motion of a spoke is at right angles to its length. Therefore, we are told, the ratio of the circumference to the diameter of the wheel becomes, not only different from π, but depends on the speed of the circumference of the wheel. The answer seems to me to be as follows:

It is implicit in the special theory of relativity that the spatial geometry, as referred to a particular inertial system, is Euclidean. A change in dimensions of the kind suggested does not occur at all. The FitzGerald–Lorentz change is due to referring the material system—rotating wheel or what not—to *another system of reference.* It represents no physical change in the body in question. A supplement to this answer is that there is no transformation *within the scope of the special theory of relativity* which enables us to pass from a system in which the wheel is at rest to one in which it is rotating. This is one of the limitations of the theory. It is imperfect in much the same sense as is Newtonian theory; but not nearly to the same extent. We shall get some idea in the next chapter how Einstein generalized and greatly improved it, with consequences even more striking than those we have been discussing.

CHAPTER TWELVE

EINSTEIN'S GENERAL THEORY

THE PRINCIPLE OF EQUIVALENCE

THE last chapter has revealed, I hope, that the characteristic features of relativity, and in fact the principle of relativity itself, are already implicit in Newton's mechanics. The advance to the special theory of relativity, with which is associated a wider (four-dimensional) continuum, solved many of the problems which baffled the nineteenth-century physicists, but, quite apart from the strange microphysical quantum phenomena, there were still unsolved problems of the macrophysical world. The most important of these was that of correlating gravitational phenomena with other physical phenomena and there was the puzzling advance of the perihelion of the planet Mercury. Even without these problems to worry about it seemed desirable to inquire into the consequences that might ensue from a widening of the scope of the principle of relativity.

In order to get some notion of the nature of the gravitational problem and some insight into Einstein's way of approaching it—as I have understood it—it is desirable to examine those characteristics which distinguish universal gravitation from other phenomena. It is not the inverse square law which distinguishes gravitation. Electrostatic phenomena also conform to an inverse square law. Let us examine the gravitational force between two bodies whose masses are M and M_1 and which are small in their linear dimensions compared with the distance, r, which separates them—a supposition needed to give precision to r. It is convenient to imagine their centre of mass fixed in the origin of an inertial system. According to Newton the force on either body is expressed by

$$\text{Force} = G\frac{MM_1}{r^2},$$

G being the constant of gravitation. This is, for example, the force exerted on M_1 by the other mass, M. Thus

$$M_1 \times \text{acceleration of } M_1 = G\frac{MM_1}{r^2}.$$

Therefore

$$\text{Acceleration of } M_1 = G\frac{M}{r^2}.$$

It is clear that, if M_1 were replaced by some other mass, M_2, the acceleration of this latter would also be equal to GM/r^2.

A remarkable characteristic of gravitation is thus revealed. If we place any body whatsoever at a particular point—always supposing its linear dimensions to be so small that it has a meaning to speak of it was being at a point—whatever its mass may be, or, for the matter of that, whatever its chemical composition may be; in every case the gravitational acceleration of the body will be the same. Then there is the further fact that Newton's gravitational constant is a *universal* constant, quite independent of the masses and the nature of the materials that may be in the neighbourhood of the two gravitating bodies.

Let us for a moment regard a gravitating particle rather naïvely from the old Newtonian point of view. It is referred to an inertial system in which it has a certain acceleration which, we appreciate, is *identical* with that of *any other* particle which we might imagine to replace it. If now we refer the particle to another reference system, not an inertial one, but a system accelerated with reference to our inertial system and having the *same* acceleration as the gravitating particle, then in this new system the particle has no acceleration. It is therefore either at rest or moving (at least momentarily) with a constant velocity in a straight line. Its motion is like that we associate with Newton's first law. And note that, if our particle were replaced by *any other* particle, this would also move in such a way, or remain at rest. In fact the gravitational field has *vanished*—at least locally and temporarily. Similarly, if we contemplate a particle which is at rest, or moving in a straight line with a constant velocity, relatively to an inertial system, such a particle, if referred to another system which is accelerated relatively to the inertial one, will appear to be accelerated, and any other particle—not accelerated in the inertial system—would have the same accelera-

tion. In fact this change from the inertial to the accelerated system *has given rise to a field of the same type as a gravitational field.* It has, as it were, *created* a gravitational field. Reflections of this kind, I imagine, led Einstein to another principle which he called the PRINCIPLE OF EQUIVALENCE, which lays down that gravitational forces are equivalent to, or of the same nature as, the forces which may be created artificially, as it were, by changing the system of reference. This kind of change, as we have just seen, when it annuls the gravitational acceleration of a particle, causes it to move, for a short time and while it is confined to a limited spatial region, as if it were following Newton's first law of motion. The suggestion is that perhaps some generalization of our relativity theory will represent all gravitating bodies as obeying a law which is a generalization of Newton's first law of motion.

THE GENERALIZATION OF NEWTON'S FIRST LAW

A straight line drawn from a point A to a point B—in a Euclidean plane, shall we say, for simplicity—is shorter than any other line that may be drawn from A to B. If we take two points, A and B, on the surface of a sphere there is, in general, provided we keep on the spherical surface, a line from A to B which is shorter than any other from A to B. Such lines are called GEODESICS. A geodesic in a non-Euclidean space is the generalization of, or what corresponds to, a straight line in Euclidean or Minkowskian space. The definition given does not include all geodesics. It is rather difficult to define a geodesic satisfactorily in a simple way and I prefer not to attempt it. When a geodesic joins two points, A and B, its length, from A to B, is said to have a STATIONARY value. It may be a minimum or a maximum, but there are other cases.*

Now Einstein imagined a four-dimensional continuum of space and time whose geometry was related to that of Minkowski in a way which resembles the relation between the geometry of a spherical (or curved) surface and that of a plane surface. This more general continuum is called a RIEMANNIAN CONTINUUM, after Bernard Riemann (1826–1866), a great German mathematician and indeed one of the greatest of all mathematicians.

* *Vide* W. Wilson: *Science Progress*, 32, p. 209 (1937).

Einstein assumed that the motion of a particle in a gravitational field is represented by a geodesic in a Riemannian space-time. Another way of expressing this is to say that the "force" acting on the particle—this means the appropriate *mathematical generalization* of Newton's force—vanishes.[1] A warning must be given here: it has become customary to say that force has been eliminated from dynamics, especially of course from relativistic dynamics. This is misleading. In the case before us the vanishing thing is something which, although a generalization of Newton's force, is very different from it. Newton's acceleration is *time* rate of change of velocity, but in the generalization *the time has been replaced by* PROPER TIME, which is very different. Newtonian force of course is still a very real fact.

If we consider a *very small* portion of a curved surface—a spherical surface for example—it is practically plane and we can imagine a tangential plane with which it coincides. So too a sufficiently small portion of Riemannian space-time coincides with a (tangential) Minkowskian space-time and in this region of coincidence Newtonian (gravitational) force vanishes—but in general not elsewhere. We have such a case when in Newtonian physics we change from an inertial system to co-ordinates accelerated in such a way that gravitational acceleration vanishes locally and temporarily. This is represented by the tangential region just described. In it geodesics become straight lines. In Einstein's theory the motion of a gravitating particle is represented by a geodesic and the associated Newtonian force vanishes when this geodesic is a straight line.

EINSTEIN'S LAW OF GRAVITATION

Since, according to the theory now being described, the motion of a particle in a gravitational field is represented by a space-time geodesic, then of course the law of gravitation must be sought in those mathematical features of the continuum which determine the shapes of geodesics. Our experience, on the other hand, indicates that gravitational fields are determined by the distribution of masses and so we conclude *that this determines the geometry of the continuum.* It is not possible to describe Einstein's law of gravitation shortly and without rather elaborate

[1] *Vide* W. Wilson: *Theoretical Physics*, Vol. iii (Methuen).

mathematical statements,* but something of its nature may be indicated in the following way. The Riemannian vector corresponding to the Newtonian md^2x/dt^2 may be written

$$m_0 \left\{ \frac{d^2x_i}{ds^2} + \varphi_i \right\} \qquad \text{(XII—1)}$$

The subscript i in this equation indicates the component, e.g. if $i = 1$, then x_i means what we have usually called x; if $i = 2$, then x_2 means y. Similarly x_4 means w.†

More precisely, (XII—1) corresponds to

$$m_0 \frac{d^2x_i}{ds^2}$$

in Minkowski's space-time. The symbol ds, which takes the place of dt, is in fact a short interval in the space-time continuum. It is called the PROPER TIME. Now the above expression vanishes in Minkowski's space-time *along a straight line* (not otherwise of course) and (XII—1) is the corresponding expression, generalized for the Riemannian continuum: it vanishes for a geodesic. It should be remarked that m_0 means the mass of our particle reckoned for zero velocity. Now $m_0 d^2x_i/ds^2$ measures (very nearly) the i component of the ordinary Newtonian force which is therefore equal to (approximately)

$$- m_0 \varphi_i.$$

Since the expression (XII—1) vanishes we have

$$\frac{d^2x_i}{ds^2} + \varphi_i = 0. \qquad \text{(XII—1A)}$$

This is the equation of a geodesic in Riemannian space-time. φ_i has its origin in the curvature‡ of the continuum. Its very

* *Vide* A. Einstein: *Ann. d. Physik*, 49, p. 769 (1916); A. Eddington: *The Mathematical Theory of Relativity* (Cambridge); W. Wilson: *Theoretical Physics*, Vol. iii (Methuen).

† The φ in (XII—1) represents what is sometimes expressed in the form

$$\Gamma^{i}_{mn} \frac{dx_m}{ds} \frac{dx_n}{ds} \quad \left(\begin{array}{c} \text{Summation with respect} \\ \text{to } m \text{ and } n \end{array} \right)$$

Since it is not possible to deal adequately with this here, I have felt it best to replace it by as simple a symbol as possible.

‡ The word "curvature", it is perhaps unnecessary to state, is used here in a very Pickwickian sense. It means something whose mathematical expression has a similar *form* to that for the curvature of a curved Euclidean surface.

complicated mathematical expression will not be given here; it will be found in the works already referred to (Einstein, Eddington and Wilson). It was already known to Riemann and Christoffel in the middle of last century.

To get some idea of the approach to the law of gravitation, i.e. the law which determines the φ_i of the geodesics, let us first examine the square of a short interval as it appears in Euclidean and Minkowskian continua:

$$ds^2 = dx^2 + dy^2 + dz^2 \qquad \text{(Euclid)}$$
$$ds^2 = dx^2 + dy^2 + dz^2 + dw^2 \qquad \text{(Minkowski)}$$

In each case ds^2 is an invariant, that is to say, it has the same value in any co-ordinate system (theorem of Pythagoras). In Riemannian space-time it is expressed as follows:

$$ds^2 = g_{11}dxdx + g_{12}dxdy + g_{13}dxdz + g_{14}dxdw + g_{21}dydx + g_{22}dydy + g_{23}dydz + g_{24}dydw + g_{31}dzdx + \text{ and so on to } g_{44}dwdw.$$

The g's are functions of x, y, z and w. Our law of gravitation is obviously a law which determines the g's. Once we know them we can find out, if we are clever enough, all about the φ_i and the geodesics. Now Einstein was led to his law of gravitation in the following way. There is, as he was aware, a certain TENSOR—the meaning of this term has already been elucidated—whose vanishing is the necessary and sufficient condition that space-time shall be like that of Minkowski (or like that of Euclid). Its geodesics are straight lines and there is no gravitation. This Riemannian tensor, as it is called, obviously cannot be permitted to vanish in Riemannian space—not in general—and Einstein's stroke of genius was the deriving of another tensor from the Riemannian one, the vanishing of which gave such values to the g_{11}, g_{12}, etc., that when the φ_i are calculated from them the equations (XII—1A) of the geodesics represent the gravitational trajectories of particles even better than does the old Newtonian law.

The application of all this to a body (planet) in the field of a massive central body (sun) leads to equations of motion which differ only very very slightly from those we derive from Newton, but the minute differences are highly important. They account beautifully for the behaviour of the planet Mercury. The reader should refer to the works already cited for the details. Perhaps I

may add that the perihelion of a planet's orbit is the point on it which is nearest to the sun. After allowance is made for the perturbing influence of the other planets, the perihelion should not shift, according to Newton's theory; but Mercury's perihelion does definitely advance about 42″ or 43″ every hundred years. The motion of the perihelion depends partly on the ellipticity of the orbit and Mercury's orbit is more elliptical than the others. That is one of the reasons why we never hear of the advance of, say, the perihelion of Mars.

Another consequence of Einstein's theory was in the nature of a prediction. If instead of an ordinary particle we think of a photon, which is a "particle" travelling with the unique velocity, c, starting out from some distant star and passing near the sun, then we expect it to be deviated by the gravitational field of the sun whether it be subject to Newton's law or to Einstein's. In the latter case the calculated deviation (refraction) amounts to 1·75 seconds of an arc and in the Newtonian case to half as much.

Observations carried out on May 29th, 1919, by two British solar eclipse expeditions were decisively in favour of Einstein's theory. The observers at Sobral in Brazil found a deviation of 1·98 seconds and those at Principe off the west coast of Africa found 1·61 seconds. Another of Einstein's predictions has also stood the test of observation: the so-called red shift of the spectral lines (spectral lines displaced towards the red end of the spectrum) of the light from a star, or more precisely of the light from a place where the gravitational potential is lower than that where the observer is situated. The expected shift was very small and the earlier attempts to observe it were inconclusive. The actual formula is

$$\lambda_s - \lambda_e = GM\lambda/c^2r,$$

where λ_s is the wave-length of some definite spectral line in the light emitted from the star and λ_e is the wave-length of the same line in light emitted near to observer. M is the mass of the star, G the constant of gravitation, r the radius of the star, and λ is the mean of the two wave-lengths. Happily about this time the dark companions of Sirius and Procyon were discovered. They are distinguished by their exceedingly *small radii*, r, while they have masses, M, comparable with those of other stars, which

means that their material is packed into very small space—about a ton or so in a cubic inch. The observations on the light from these stars leave no doubt about the reality of the shift and the correctness of Einstein's formula.

GENERAL RELATIVITY AND ELECTROMAGNETIC PHENOMENA

Gravitational phenomena are an expression of the geometrical structure of the Riemannian space-time continuum. Not so electromagnetic phenomena. These indeed find a place quite comfortably in the continuum; but it merely provides them, as it were, with a home. Unlike gravitational phenomena, they are not in the least bit a necessary consequence of the general theory of relativity. Naturally attempts have been made to widen relativity theory in such a way that it will necessitate electromagnetic phenomena as a consequence of the metrical character of the continuum and thus link them up with gravitational phenomena. Such unifying theories have been developed by Einstein himself, by the mathematician Hermann Weyl and by Kaluza. It is the last of these which, in my view, proceeds along the right lines. It has the great merit of uniting all the conservation principles, so that the conservation of momentum, of mass and energy and of electric charge are just different aspects of the conservation of an extended momentum in a five-dimensional continuum (see page 247). It is not possible to go into it in detail here, but it may be said that the charge on a charged particle is proportional to its component of momentum along the new fifth axis. This gives, as it were, a reason for conservation of charge.

We shall meet with relativity again in studying cosmological speculations and perhaps in some aspects of quantum mechanics.*

* H. Weyl: *Sitzungsberichte*, p. 465 (Berlin, 1918). Kaluza: "Zum Unitätsproblem der Physik," *Preus. Akad. der Wissenschaften*, 54, p. 966, (1921). See also A. Eddington's *Mathematical Theory of Relativity*; W. Wilson: *Proc. Roy. Soc. A.*, 102, p. 478 (1922), and *Proc. Roy. Soc. A.*, 118, p. 441 (1928); O. Klein: *Zeits. für Physik*, 46, p. 188 (1928).

CHAPTER THIRTEEN

RADIANT HEAT AND THE EMERGENCE OF THE QUANTUM THEORY

BLACK BODY RADIATION

HEAT travels from one place to another in three quite different ways, as the text-books tell us. Maxwell's definitions of them are as follows:

(i) "CONVECTION is the motion of the hot body itself carrying its heat with it."

(ii) "CONDUCTION is the flow of heat through an unequally heated body from places of higher to places of lower temperature."

(iii) "In RADIATION, the hotter body loses heat, and the colder body receives heat by means of a process occurring in some intervening medium which does not itself thereby become hot."*

Dr. Barton quotes these definitions in the very admirable text-book to which I have already occasionally referred and remarks, in speaking of the last one, that it is "a splendid example of the caution and power to suspend judgment in face of insufficient evidence which are essential to the scientific temper". This kind of caution was held in very high esteem throughout the nineteenth century. It was the most prominent element in the scientific ethic of the period. How often have examples been given us, such as Darwin's long-continued patient accumulation of evidence before the cautious spirit permitted him to draw his great conclusions? Possibly this kind of scientific temper was inherited from the exasperation of Galileo and Newton and others of their time, when the wildest hypotheses and fancies were freely adopted about all sorts of phenomena. Perhaps this was the origin of Newton's *hypotheses non fingo* and the almost religious caution of a later time. But important as it is to avoid error, it is even more important to run some risk for the prospect of a great truth. It is difficult to see how Young, Fresnel or

* Clerk Maxwell: *Theory of Heat* (Longmans, Green & Co.).

Maxwell himself could possibly have made their really great discoveries about light and electromagnetism without assuming an aether, the existence of which is not only doubtful, but now known to be incompatible with a host of observational facts. It is now, for example, quite certain that heat, or caloric, is not conserved, but the assumption of its conservation nevertheless led Carnot to one of the greatest of all physical discoveries. The scientific ethic of to-day insists rather on the rejection of things which observation makes untenable than on the too severe avoidance of speculation.

Radiation of heat, as Maxwell knew, is just the same process as the transference of energy in the form of light waves. Experiment established long ago that radiant heat is reflected and refracted according to the same laws as light. In fact it may take the form of visible light and of radiation outside both limits of the visible spectrum; though it is much more obvious, at those temperatures reached in a terrestrial laboratory, beyond the red end of the spectrum. Moreover, observations made during total solar eclipses have shown that the radiant heat reaching us from the sun travels with the speed of light. It is simply light and electromagnetic radiation, which, in so far as it may be called "heat", is bound up with temperature equilibrium.

In the earlier investigations of heat radiation the centre of interest was the heat-radiating and absorbing properties of surfaces and all students of physics will remember that old experiment of Ritchie* which demonstrated in a simple way the proportionality of the emitting and absorbing powers of a surface. Kirchhoff (1859) was able to show that this was a consequence of thermodynamical principles and that it must be true, in the case of any surface, not only for the totality of the radiant heat emitted and absorbed by it, but also in detail for any range of wave-lengths, however narrow. Kirchhoff's law includes even more than this. Tourmaline, for example, has the peculiarity that it strongly absorbs light, or radiation, which is polarized in a certain way and it emits correspondingly copiously light which is polarized in just this way.

We get some insight into the emission, absorption and reflection of radiation in the following way. Let us give our attention to a closed surface, *abc* in Fig. XIII—1, which is maintained at a

* Ritchie: *Pogg. Ann.*, 28 (1833).

uniform temperature, T. In the case A the space within is filled with any sort of material. Eventually temperature equilibrium gets established throughout the interior and thereafter no further temperature change can occur (second law of thermodynamics)

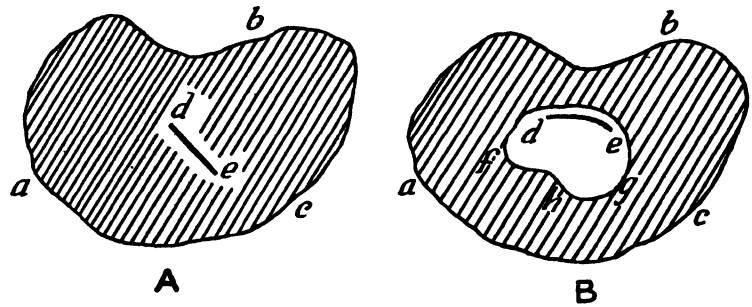


Fig. XIII—1

so long as the temperature of the enclosing surface is maintained constant. Under these circumstances no heat passes through any surface such as *de* in the interior—or, if it does, it passes in both directions at the same rate. Now consider the case B of an evacuated enclosure, *fgh*, within *abc*, in which nothing of a material kind is left. Imagine a surface, *de*, which, while *within* the vacuous enclosure, is quite close to its wall. The heat energy passing through it is now radiant heat (like light waves) and it is bound to pass through *de* at the same rate in both directions. The amount of radiant heat passing through any unit area per second has the same value everywhere in the vacuous enclosure. Let us call it I, the INTENSITY of the radiation. Then I units of radiant energy *reach* the unit area of the wall per second and also *leave* the unit area per second. The part of I which on reaching the wall is absorbed or passes into it, never to return, we may represent by αI. The remainder, rI, is reflected. Therefore

$$\alpha I + rI = I,$$

or

$$\alpha + r = 1.$$

The fractions α and r are respectively the ABSORBING and REFLECTING POWER of the surface.

Of the radiation leaving the unit area per second some

portion, ε (EMITTING POWER), is actually emitted, the rest, rI, is reflected or scattered. Therefore

$$\varepsilon + rI = I$$

and obviously

$$\varepsilon = \alpha I$$

or

$$\frac{\varepsilon}{\alpha} = I.$$

A BLACK SURFACE is, by definition, a hypothetical surface which absorbs ALL the radiation falling on it, so that, for such a surface, $\alpha = 1$. Clearly, the emitting power of a black surface is equal to I. If we represent the emitting power of a black surface by E, then

$$E = I$$

and

$$\frac{\varepsilon}{\alpha} = E;$$

or, in words, the ratio of the emitting power of any surface to its absorbing power is the same for all surfaces (Kirchhoff) at a given temperature and is equal to the emitting power of a black surface. Kirchhoff also established a corresponding law for any narrow range of wave-lengths between λ and $\lambda + d\lambda$ If we represent the intensity of the radiation in such a narrow range by

$$I_\lambda d\lambda$$

and the associated absorbing, reflecting and emitting powers by α_λ, r_λ, and ε_λ (and E_λ for a black surface) we have

$$\frac{\varepsilon_\lambda}{\alpha_\lambda} = E_\lambda.$$

This is included in Kirchhoff's law.*

Every part of the surface bounding an evacuated enclosure simulates a black surface, since the same amount of radiation leaves it in the unit time as would leave it if it were black—the only difference is that it is usually only partly actually emitted, the rest being reflected. There would in fact be no difference in the state of affairs within the enclosure if the containing surface were replaced by a black one. In fact the character of the radia-

* G. Kirchhoff: *Pogg. Ann.* 109, p. 275 (1860).

tion inside a vacuous enclosure is independent of the nature of the enclosing wall.

The discovery and appreciation of these things led to a gradual change in the attitude of physicists. Their interest naturally shifted from the study of the properties of *surfaces*, e.g. their emitting, absorbing or reflecting powers, to the study of the *character of the radiation itself*, which they called BLACK BODY RADIATION and later FULL RADIATION. Undoubtedly the best name for it is the old one, RADIANT HEAT, since it is energy associated with statistical equilibrium, i.e. temperature equilibrium, and heat is just that sort of energy.

THE STEFAN–BOLTZMANN LAW

The Viennese physicist, Stefan (1831–1897), noticed in 1879 that some measurements of the rate of loss of heat from a hot platinum wire, made by John Tyndall (1820–1893), suggested that loss by radiation was proportional to the fourth power of the K temperature of the wire. Guided by this and by the Maxwellian pressure of radiation (equal inside the enclosure to one-third of the energy density) Boltzmann was able to prove thermodynamically that the energy density within a vacuous enclosure is proportional to the fourth power of the K temperature of the enclosing wall. His proof is very simple. It is simply an application of Clapeyron's formula (II—1B). Instead of L we have

$$u(v_2 - v_1) + \frac{u}{3}(v_2 - v_1),$$

for, since u is the energy per unit volume in the enclosure, the quantity of heat $u(v_2 - v_1)$ has to be supplied to fill up the volume increment $(v_2 - v_1)$ and since the pressure is equal to $u/3$ the amount of work done by the radiation pressure is

$$\frac{u}{3}(v_2 - v_1),$$

and the corresponding quantity of heat has to be supplied at constant temperature. Therefore

$$\frac{4u}{3}(v_2 - v_1) = T\frac{dp}{dT}(v_2 - v_1),$$

since, as we have seen, the factor C of Clapeyron's formula is

equal to the Kelvin work scale temperature. When we cancel out $(v_2 - v_1)$ and replace p by $u/3$, we get

$$4u = T\frac{du}{dT}.$$

Thus it follows that

$$u = aT^4, \qquad \text{(XIII—1)}$$

where a is a constant, called STEFAN'S CONSTANT. This law has been confirmed experimentally and the constant a turns out to be near to

$$7{\cdot}67 \times 10^{-15}$$

in our usual units (ergs $\times$ T^{-4}/c.c.).

The method of experiment consists in making an opening in the enclosure, *very small* in comparison with its dimensions, so that the equilibrium in the enclosure may not be appreciably upset and so that the emerging radiation shall retain the character of black body radiation. Such a minute aperture obviously simulates a black surface, both as regards emission and absorption. The rate at which energy is emitted per unit area of the aperture (unit area of a black surface) can easily be shown to be

$$\sigma T^4,$$

where σ is equal to $ac/4$, and therefore if $a = 7{\cdot}67 \times 10^{-15}$,

$$\sigma = \frac{7{\cdot}67 \times 3}{4} \times 10^{-15+10} = 5{\cdot}75 \times 10^{-5}$$

in our units. In the experimental measurement the radiation enters another vacuous enclosure containing a bolometer or other suitable device at room temperature for measuring the radiant heat received. Naturally the net amount of heat measured by the bolometer or other device is proportional to $(T^4 - T_0^4)$ where T_0 is the temperature of the bolometer enclosure and T that of the received radiation (emitting enclosure). One of the most interesting of the measuring devices is Callendar's RADIO-BALANCE.* The name STEFAN'S CONSTANT is usually given to σ in England.

* *Vide* Starling's *Electricity and Magnetism* for a description of the RADIO BALANCE. Coblentz: *Dict. of Applied Physics*, 4, p. 541 (Macmillan, 1923). Hoare: *Phil. Mag.*, 6, p. 828 (1928); 13, p. 386 (1932).

DISTRIBUTION OF ENERGY IN THE SPECTRUM OF BLACK BODY RADIATION

The way in which the radiant energy is distributed among different wave-lengths began to be investigated in the eighties. Perhaps one should mention the American Langley (1834–1905) first, although, so far as I know, he did not actually take observations of black body radiation; but he did devise the most frequently used measuring instrument, the BOLOMETER.* This is essentially a thin metallic strip (iron in Langley's own experiments). It is raised in temperature when radiant heat falls on it and the consequent rise in its electrical resistance can be used to compute the heat received. Langley employed it to investigate the distribution of energy in the sun's spectrum, an experimental problem rather similar to the corresponding black body one. Langley's curves are indeed rather like those for black body radiation. In fact the sun's radiation does approximate, if rather roughly, to black body radiation. In consequence of this it is easy to make an approximate estimate of the temperature of the sun's surface. Experimental measurements give us an estimate of the amount of energy leaving the unit area of the sun's surface per second and the known value of σ enables T to be computed. It is in the neighbourhood of $6000°K$. Wien's displacement law, which we shall study later, also gives us the solar temperature in agreement, roughly, with what is given by Stefan's law.

The experiments of Lummer and Pringsheim, Kurlbaum and Rubens at the end of the nineties yielded fairly accurate information about the distribution of the energy among the wave-lengths for various temperatures. When this is represented graphically it has the appearance shown in Fig. XIII—2. The full line shows the distribution corresponding to the particular temperature, T. The ordinates, E_λ, are so defined that $E_\lambda d\lambda$ means the energy per unit volume, in the vacuous enclosure, which is associated with wave-lengths between λ and $\lambda + d\lambda$, so that the shaded area in the figure is equal to the energy per unit volume associated with the wave-lengths between λ_1 and λ_2. For a particular temperature there is a maximum at a wave-length which may be represented by λ_m. The higher the temperature the more the maximum shifts to the left (to shorter

* Langley: *Proc. Amer. Acad. Arts and Sciences*, 16, p. 342 (1881).

wave-lengths) and it is characteristic of these maxima that, for any temperature,

(Wave-length) × (Temperature) = a constant, (XIII—2)

the value of the constant being near to 0·288 (in terms of centimetres and Kelvin work scale temperature). This is known as WIEN'S DISPLACEMENT LAW and is part of more general theoretical deductions of W. Wien which we shall presently study.

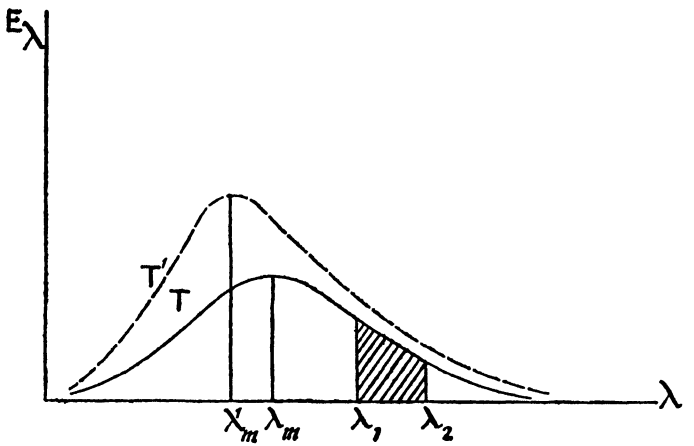


Fig. XIII—2

It may be pointed out that the total area under the curve of Fig. XIII—2 must be equal to aT^4 since it represents the total energy (over all wave-lengths) per unit volume.

The experimental investigation presented a straightforward and soluble, if not awfully easy, problem.

Really great interest began to be associated with black body radiation when the *reason* for its peculiar constitution began to be sought.

THE PROBLEM

When the earliest efforts were made to solve the distribution problem, very likely nobody foresaw, not even Planck himself, how their consequences would reverberate through the world of physical science. They were destined to reveal an extraordinary and quite unsuspected feature of the physical world, with which

Newtonian mechanical principles were quite unable to cope.

W. Wien (1864–1928) arrived theoretically at the formula

$$E_\lambda = \frac{1}{\lambda^5} f(\xi), \qquad \text{(XIII—3)}$$

in which ξ means the product λT, and f was a then unknown function. This leads to the displacement law already mentioned which has been amply confirmed experimentally.* Somewhat later Wien deduced an expression for the function, f, namely,

$$f = Ae^{-B/\xi}. \dagger \qquad \text{(XIII—4)}$$

Wien's deduction of this formula was faulty, but nevertheless it represented a most valuable contribution to the solution of the problem, for two reasons:

(i) It does in fact, as Lummer and Pringsheim showed,‡ fit the observations exceedingly well when ξ, i.e. λT, is sufficiently small.

(ii) It helped in some degree to guide Planck to what, we can scarcely doubt, is the correct form of the distribution law.

Near the end of the century, Lord Rayleigh (1842–1919) deduced the following law from Newtonian dynamical principles:§

$$E_\lambda = 8\pi kT/\lambda^4. \qquad \text{(XIII—5)}$$

According to this, Wien's function, which I have represented by f in (XIII—3) and (XIII—4), should be expressed (in our present-day symbolism) by

$$f = 8\pi k\, \lambda T,$$

or

$$f = 8\pi k\, \xi.$$

* W. Wien: *Sitzungsberichte d. Akad. d. Wissenschaften*, Berlin, 9th Feb., p. 55 (1893), *Wied. Annalen*, 52, p. 132 (1894). F. Paschen: *Sitzungsberichte d. Akad. d. Wissenschaften*, Berlin, pp. 405, 959 (1899), *Ann. d. Physik*, 6, p. 657 (1901). O. Lummer and E. Pringsheim, *Verh. der Deutschen Phys. Gesell.*, 1, pp. 23, 215 (1899), *Ann. d. Phys.*, 6, p. 192 (1901).

† W. Wien: *Wied. Ann.*, 58, p. 662 (1896). In expressing his formula he used the letters c_1 and c_2 instead of my A and B and did not employ my device of representing λT by the single symbol ξ.

‡ O. Lummer and E. Pringsheim: *Verhand. d. Deutschen Phys. Gesell*, 1, pp. 23, 215 (1899), *Ann. d. Phys.*, 6, p. 192 (1901).

§ Lord Rayleigh: *Phil. Mag.*, 49, p. 539 (1900).

Max Planck (1858–1947), who solved the radiation problem, apparently did not become aware of Lord Rayleigh's contribution till somewhat later, but he became acquainted with the experimental results of Rubens and Kurlbaum while their work, I believe, was still in progress, and these indicated that E_λ was proportional to T/λ^4 in the region of sufficiently large values of λT. Rayleigh did not express his law in quite the same way as it is given in (XIII—5). Indeed it was Planck who first introduced the constant k (BOLTZMANN'S CONSTANT) into radiation and statistical theory, though of course it is implicit in Boltzmann's work. Planck inquired after a single formula which would, as it were, unite the extremes (XIII—4) and (XIII—5) and eventually (before he actually succeeded in deducing it from fundamental principles) found it in the form

$$E_\lambda = \frac{A}{\lambda^5} \frac{1}{e^{B/\lambda T} - 1}. \qquad \text{(XIII—6)}$$

This is obviously indistinguishable from Wien's formula (XIII—4) when ξ or λT is very small, while it approximates to Rayleigh's formula (XIII—5) when λT is large.*

In this latter case Planck's formula becomes in fact

$$E_\lambda = \frac{A}{B} \frac{T}{\lambda^4},$$

which, of course, is Rayleigh's formula.

Planck's problem was to deduce the formula (XIII—6) from fundamental principles, if possible, or at least a formula which differed from it sufficiently little, since the experimental results were in excellent accord with it.

Before describing Planck's attack on the problem, it will be illuminating to give a little attention to the two extremes:

(i) λT very small, i.e. ν/T *very large*, and

(ii) λT *very large*,

and we bear in mind that the old principles do very well with large things. Now the exponential in Wien's formula (XIII—4) —this formula fits the former case (i) quite well—suggests the exponential in Maxwell's law of the distribution of velocities among the molecules of a gas, or that in the more general

* When λT (in XIII—6) is large $B/\lambda T$ is small, and when x is small $e^x - 1$ approximates to x.

distribution law of statistical mechanics (V—1), namely

$$e^{\frac{-\text{Energy}}{\theta}} \text{ or } e^{\frac{-\text{Energy}}{kT}},$$

and we are tempted to follow, rather experimentally and tentatively, the suggestion that we are dealing with a strange sort of gas. Other phenomena, notably those of photo-electricity, tend to confirm this suggestion and such a view of black body radiation was taken seriously by Louis de Broglie. The molecules of this "gas" are indeed now well recognized and are called PHOTONS. We may be sure that *some* rather drastic interference with old-established principles is necessary, since it is universally agreed that Rayleigh's deduction of his formula is *sound*—only the Newtonian premisses which he used being inadequate. So we proceed to identify

$$\frac{\text{Energy of photon}}{kT} \text{ with } \frac{B}{\lambda T}.$$

Therefore

$$\text{Energy of photon} = \frac{Bk}{\lambda};$$

or, since it is more convenient to use frequencies, ν, rather than wave-lengths, λ, we write

$$\lambda = c/\nu$$

and so obtain

$$\text{Energy of photon} = \frac{Bk}{c} \times \nu. \qquad \text{(XIII—7)}$$

Planck came on the constant Bk/c from a different direction and called it h, so we may write

$$\text{Energy of photon} = h\nu. \qquad \text{(XIII—7A)}$$

We might indeed build up a "gas" theory and find a "Maxwellian" law giving an energy distribution agreeing with Wien's law.

Turning to the other extreme in which we assume λT very large, or perhaps one ought to say, in which we assume that the large-scale methods will apply. It can be shown that the number of vibrations between ν and $\nu + d\nu$ in the unit volume is $8\pi\nu^2 d\nu/c^3$ and the statistical methods of Chapter V lead to kT as the average value of the energy of one vibration, so we get

$$u_\nu d\nu = \frac{8\pi\nu^2 kT d\nu}{c^3}$$

for the energy per unit volume between the frequencies ν and

$\nu + d\nu$. This is Rayleigh's formula in another form. Thus macroscopic methods applied in different ways, (i) to a photon gas and (ii) to waves or vibrations, lead to formulae (Wien and Rayleigh) which represent the extreme limits of the distribution curve satisfactorily.

The complete solution of the problem can be achieved in two ways: (i) by applying an improved form of statistics to a photon gas (Bose and Einstein), or (ii) by applying somewhat modified statistics to the wave view of radiation. The problem was first solved by Planck* who used a method equivalent to (ii), since in the year 1900 a corpuscular or quasi corpuscular constitution of black body radiation would have appeared to be just midsummer madness. Indeed Planck's theory was regarded by many, or most, physicists as something that could not be taken seriously—in its earlier years at any rate. Planck's innovation was to assume that the energy of any simple harmonic vibration is expressed by

$$(\textit{Whole number}) \times h\nu,$$

where h means the same constant as it has already represented in (XIII—7A).† In consequence of this assumption the smallest quantity of energy that could be emitted, or absorbed, by the systems he contemplated is $h\nu$. This quantity is sometimes called a QUANTUM OF ENERGY and h itself Planck called the QUANTUM OF ACTION. He discovered, or appreciated, later that his radiation formula, which he wrote in the equivalent forms

$$u_\nu d\nu = \frac{8\pi h\nu^3}{c^3} \frac{d\nu}{e^{\frac{h\nu}{kT}} - 1}, \qquad \text{(XIII—8)}$$

and

$$E_\lambda d\lambda = \frac{8\pi ch}{\lambda^5} \frac{d\lambda}{e^{\frac{ch}{k\lambda T}} = 1},$$

* Max Planck: "Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum," *Verh. d. Deutschen Phys. Gesellschaft*, 2, p. 237 (1900). "Ueber das Gesetz der Energieverteilung im Normalspektrum," *Ann. der Phys.*, 4, p. 553 (1901).

† Planck argued, correctly, that, since the character of the radiation in the vacuous enclosure is quite independent of the nature of the material constituting the enclosing wall, he was at liberty to imagine it constituted of simple harmonic OSCILLATORS which were the simplest absorbing and emitting systems he could think of. This device has been widely misunderstood to mean that he assumed that all materials do in fact contain such oscillators.

could be derived by laying down that every small element $dp \times dq$ of Fig. V—1, instead of having the limit zero, must have the definite fixed value h. In fact, according to classical Newtonian theory the energy of a simple harmonic vibration is expressed by

$$Energy = Area \times \nu,$$

in which the area referred to is that of the closed curve (ellipse) which we construct when we plot the momentum, p, against the displacement, q. Now according to Planck's new theory the energy of such a vibration is expressed by

$$Energy = (Whole\ number) \times h\nu$$

and on comparing the two expressions we see that the area in question is equal to a

$$(Whole\ number) \times h.$$

The classical statistical method developed by Maxwell, Boltzmann and Willard Gibbs takes into account every possible distribution of the systems (or molecules) over the p, q diagram of Fig. V—1 in the process of finding the maximum value of the logarithm of the probability of the distributions, and such cases as the following are distinguished from one another:

(a) the system a is in the $dp \times dq$ element number 1 and the system b is in the element number 2;

(b) the system a is in the $dp \times dq$ element number 2 and the system b is in the element number 1;

but in the statistics of Bose and Einstein,* especially when it is applied to photons, which cannot possibly be distinguished from one another (when associated with the same value of ν) as macroscopic individuals, or even the molecules of an ordinary gas can be, such distinctions are not made. This new Bose–Einstein statistics leads to Planck's law (XIII—8).

Obviously the number of "molecules" or photons in a photon gas filling a given volume at a given temperature is quite indeterminate and practically infinite. Einstein appears to have

* Bose: *Zeits. für Physik*, 26, p. 178 (1924); Einstein: *Preuss. Akad. d. Wiss.*, Berlin, 22, p. 261 (1924). See also de Broglie: *J. de Physique et le Radium*, 3, p. 422 (1922), for a derivation of Wien's law by the application of the methods of statistical mechanics to a photon gas. The conception of black body radiation as a photon gas is one of de Broglie's great contributions to radiation and quantum theory.

been the first to apply the theory to an "ordinary" gas, i.e. a gas with a precise finite number of molecules in a given mass of the gas.

THE REALIZATION OF KELVIN'S WORK SCALE OF TEMPERATURE

We have already learned that Kelvin's scale is identical with that determined by the product

$$(Pressure) \times (Volume)$$

of a fixed quantity of an *ideal* gas. This product therefore gives the K temperature *very nearly* when *actual* gases, such as hydrogen, are used and it is possible to make the small corrections needed to approach indefinitely close to the ideal K temperature. It is impossible, however, to use gas thermometers at very high temperatures, but we have in full radiation a means of measuring, on Kelvin's scale, temperatures to which no upper limit need be assigned. We have at our disposal the law of Stefan and Boltzmann, the displacement law of Wien, or indeed Planck's law, which includes them. Various forms of pyrometer, such as the disappearing filament pyrometer and Féry's pyrometer, etc., are in use for measuring the temperatures of furnaces by means of the radiation they emit. They are mostly calibrated by using a number of known temperatures—known by the use of the properties of black body radiation.

LOSCHMIDT'S NUMBER AND THE ELEMENTARY IONIC CHARGE

At the end of the century, when Planck was completing the solution of the great problem of radiant heat, the best estimates of Loschmidt's number, though of the correct order of magnitude, still varied between rather wide limits, roughly between 4×10^{23} and 8×10^{23}, and similarly the experimental methods of determining the elementary ionic charge directly were rather rough and gave values between 3×10^{-10} and 6×10^{-10} electrostatic units. One of Planck's greatest triumphs was to produce estimates of these numbers which differ extraordinarily little from what we now know to be very near the correct values.

He was able to compute the values of h and k from the data furnished by Lummer and Pringsheim and others. The former turned out to be about

$$6{\cdot}54 \times 10^{-27} \text{ ergs} \times \text{sec.}$$

and the latter

$$1{\cdot}34 \times 10^{-16} \text{ ergs per } 1°\text{C.}$$

The constant k is now believed to be near

$$1{\cdot}372 \times 10^{-16} \text{ ergs per } 1°\text{C.}$$

while

$$h = 6{\cdot}6 \times 10^{-27} \text{ ergs} \times \text{sec.}$$

We have seen how a knowledge of the value of k enables Loschmidt's number to be found. We simply divide the gas constant, R, for a gramme-molecule (mole) by k. So Planck divided

$$8{\cdot}3 \times 10^{7}$$

by

$$1{\cdot}34 \times 10^{-16}$$

and thus found for it

$$6{\cdot}2 \times 10^{23}.$$

He used this number to find the value of the ionic charge. When a gramme of hydrogen is liberated in electrolysis (of dilute sulphuric acid, for example), experiment shows that a quantity of electricity equal to

$$2{\cdot}896 \times 10^{14}$$

in ordinary electrostatic units is carried to the cathode by the hydrogen ions. Now the number of hydrogen atoms in a gramme equivalent of hydrogen is identical with Loschmidt's number, so Planck obtained the charge on a single hydrogen atom (ion) by dividing

$$2{\cdot}896 \times 10^{14} \text{ by } 6{\cdot}2 \times 10^{23},$$

thus obtaining

$$4{\cdot}67 \times 10^{-10} \text{ E.S. units.}$$

It is now believed to be very near to

$$4{\cdot}803 \times 10^{-10} \text{ E.S. units.}$$

The closeness of Planck's approach to it is even more impressive than appears from the actual figures. His error is due solely to the slight inaccuracy of the experimental data at his disposal.

THE LAW OF DULONG AND PETIT AND DEBYE'S THEORY

The product of the specific heat (at constant volume) of an element and its atomic weight is called its ATOMIC HEAT. Dulong and Petit found the atomic heats of most *solid* elements to have the same value—roughly about 6 in terms of our ordinary small calorie. Thus copper has an atomic weight of 63 and its specific heat is near to 0·095. The law is approximately true for elements whose atomic weights are not too small. It fails for elements of low atomic weight, e.g. for carbon; but in these cases it approximates to the truth at high temperatures, as experiments on graphite at *circa* 1000°C. indicate. Classical methods account quite simply for this law. We may regard the heat in a solid element as the sum of the energies of its individual atoms and each of these may be regarded as in simple harmonic vibration about a position of equilibrium. In temperature equilibrium the average kinetic energy of an atom, reckoned per degree of freedom, is $kT/2$ (principle of equipartition of energy). In a simple harmonic vibration the average potential energy is equal to the average kinetic energy and therefore the average energy of an atom must be equal to kT per degree of freedom.

Each atom has three degrees of freedom and so the total average energy of an atom must be equal to

$$3kT.$$

The number of atoms in a gramme-atom is identical with Loschmidt's number, which we may represent by L. We see then that the heat energy in a gramme-atom is equal to

$$3LkT.$$

Now the product Lk is equal to the gas constant for a gramme-molecule. Therefore

$$Lk = 8{\cdot}3 \times 10^7 \text{ ergs per } 1°\text{C.}$$

Thus the heat in a gramme-atom of any solid element is equal to

$$3 \times 8{\cdot}3 \times 10^7 \times T \text{ ergs.}$$

This gives an atomic heat—for *any* solid element—equal to

$$24{\cdot}9 \times 10^7 \text{ ergs per degree.}$$

To express it in calories per degree we divide by Joule's equivalent, $4{\cdot}19 \times 10^7$, and thus get very nearly 6.

After a tentative application of the quantum theory to the problem by Einstein,* P. Debye (1884–) succeeded in a rather wonderful way in accounting for the atomic heats of solid elements and their dependence on the temperature.† His theory is like Planck's theory of cavity radiation, but in addition to the electromagnetic waves and vibrations which carry the energy of the radiant heat in the exhausted cavity, Debye considers the energy associated with elastic waves in the material. Indeed he was able to neglect the energy associated with electromagnetic waves altogether because of the small velocity of elastic waves as compared with that of electromagnetic waves. Moreover there are two kinds of elastic waves: longitudinal waves (like sound waves) and transverse waves. It will be remembered that in the cavity radiation there were

$$8\pi\nu^2 d\nu/c^3$$

vibrations per unit volume with frequencies between ν and $\nu + d\nu$, and when we look at Planck's law in the form (XIII—8) we see that the average energy of a vibration is

$$h\nu/(e^{\frac{h\nu}{kT}} - 1).$$

Debye used these formulae for his transverse elastic vibrations, with of course the difference that he replaced the velocity c by that of transverse elastic vibrations, namely

$$\nu_\tau = \sqrt{n/\rho},$$

n being the rigidity of the material and ρ its density. In addition to this he included longitudinal vibrations, replacing the velocity c by that of the longitudinal waves, namely

$$\nu_\lambda = \sqrt{(k + 4n/3)/\rho},$$

k here meaning the bulk modulus. In Debye's theory therefore the expression corresponding to Planck's $u_\nu d\nu$ (XIII—8) is

$$\left(\frac{8\pi h\nu^3}{\nu_\tau^3} + \frac{4\pi h\nu^3}{\nu_\lambda^3}\right)\frac{d\nu}{e^{\frac{h\nu}{kT}} - 1}. \qquad \text{(XIII—9)}$$

The 4π in the second term corresponds to the fact that there are only half as many longitudinal vibrations as transverse ones. To get the total energy in the evacuated cavity Planck summed up (integrated) the energy associated with *all* frequencies from

* Einstein: *Ann. der Phys.* 34, pp. 170, 590 (1911).
† Debye: *Ann. der Phys.* 39, p. 789 (1912).

0 to ∞. In Debye's case however there must be an upper limit of frequency and this he chose so as to make his theory conform to the classical one at high temperatures (strictly at $T = \infty$).

The resulting expression is rather complicated.* It enables the atomic heat of a solid element, e.g. copper, to be calculated from the values of its elastic moduli, since ν_τ and ν_λ are determined by them.

Debye's theory has been amplified by Max Born and Kármán, who took into account the natural vibrations of the atoms in the crystal lattices.†

At low temperatures Debye's expression for the heat energy in a solid element simplifies to the product of a constant and the fourth power of the temperature. In this respect it resembles black body radiation and the atomic heat (so long as the temperature is low enough) is proportional to the cube of the absolute temperature. This is DEBYE'S LAW.

NERNST'S HEAT THEOREM

In the case of a substance which obeys Debye's law the increment of its entropy (at constant volume), when its temperature is raised from T_1 to T_2, is proportional to $T_2^3 - T_1^3$, provided T_1 and T_2 are low temperatures. The increment tends to zero as T_2 approaches zero. The condition of constant volume has been imposed because Debye's theory is a theory of atomic heats at constant volume. The distinguished physical chemist Walther Nernst (1864–1941) assumed this result to be true for all substances‡ and the form he gave it may be expressed by saying that the increment of the FREE ENERGY of a system is equal to the increment of its internal energy, in the neighbourhood of the absolute zero. He restricts its application to condensed systems (this is virtually equivalent to the condition already mentioned—constant volume) and to chemically homogeneous systems.§

* *Vide* W. Wilson: *Theoretical Physics*, Vol. iii, p. 154 (Methuen).

† Born and Kármán: *Phys. Zeits.*, 13, p. 297 (1912); 14, pp. 15 and 65 (1913).

‡ *Nachr. Gesell. d. Wiss.*, Göttingen; *Math.-phys. Klasse* (1906). Nernst: *The New Heat Theorem* (Methuen, 1926).

§ For the meaning of free energy see J. R. Partington: *A Text Book of Thermodynamics* (Constable), or W. Wilson: *Theoretical Physics*, Vol. i, p. 310 (Methuen).

It will be remembered that before the advent of Einstein's relativity theory only *differences* in the energy of a system were defined. It had no absolute value. The theory of relativity enables us, as we have seen, to assign an absolute value to energy. For instance, the energy associated with a mass, m, is equal to mc^2. Now Nernst's theorem assigns an absolute value to entropy and its universal validity can be justified by the statistical methods of the quantum theory.

THE WILSON–SOMMERFELD QUANTUM CONDITIONS

We have seen that the point, P (Fig. V—1), which represents the state of a simple harmonic system, moves along an ellipse

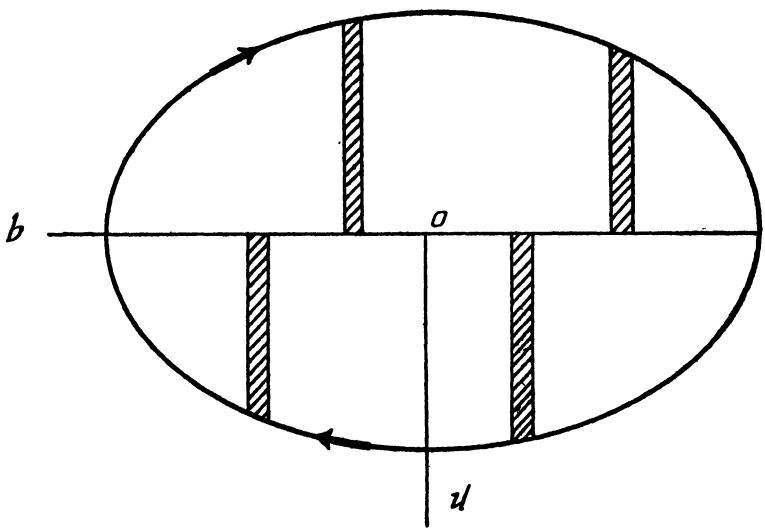


Fig. XIII—3

(Fig. XIII—3) and that the energy of the system is equal to the product of the area of the ellipse and the frequency of the simple harmonic vibration, or, as we may express it,

$$(\textit{Area of ellipse}) \times \nu = (\textit{Energy});$$

so that Planck's quantum hypothesis may be expressed in the form

$$(\textit{Area of ellipse}) = n \times h,$$

n being a whole number—for simplicity we ignore the possibility of some constant being added to n. Now this area may be regarded as the sum of an enormous number of very narrow strips, four of which are indicated in the diagram—only indicated, since their actual breadths are infinitely narrow. The area is therefore a sum of such products. The conventional way of representing any one of these is

$$p \times dq,$$

p being the vertical height of the strip and dq its infinitely narrow width. It is usual to represent such a sum by the symbol $\int$ (this symbol is a form of the letter s, the initial letter of sum) and to indicate the sum which is involved when we go exactly once round the ellipse some conventional modification of the symbol, is made. Planck's quantum condition may therefore be expressed in the following form:

$$\int pdq = nh. \qquad \text{(XIII—10)}$$

It is often convenient to represent the integral by a single letter, usually the letter J. Therefore

$$J = nh. \qquad \text{(XIII—10A)}$$

This simple-looking formula is the expression of one of the very greatest of all scientific discoveries. It has an even more profound significance—which I hope to discuss later—than the whole of the Newton–Einstein relativistic theory and its discovery places Planck among the half-dozen or so most notable human individuals of all time.

Unknown to each other, Arnold Sommerfeld (1868–) and William Wilson (1875–) generalized the form of statement (XIII—10) so that it may be applied to a great variety of systems of several degrees of freedom, very nearly about the same time.*

The nature of this generalization will be indicated in a simple application of it to the specific heats of gases and more fully, when we come to Bohr's theory of the hydrogen atom. Suffice it to say now that, in the case of many mechanical systems it is

* A. Sommerfeld: *Ann. der Phys.*, 51, p. 1 (1916). See also Sommerfeld's book: *Atomic Structure and Spectral Lines* (Methuen, 1923). W. Wilson: *Phil. Mag.*, 29, p. 795 (1915), and 31, p. 156 (1916). The p and q notation for the momentum and positional co-ordinates was introduced by the distinguished Königsberger mathematician, C. G. J. Jacobi, *Vorlesungen uber Dynamik* (1842).

possible to choose co-ordinates—the positional co-ordinates are represented by $q_1, q_2, q_3, \ldots q_n$, where n means the number of degrees of freedom—in such a way that every co-ordinate, q_i, passes periodically through the same sequence of values and is associated with a constant value of the sum $\int p_i dq_i$. The generalization of Wilson and Sommerfeld consisted in laying down that each such sum or integral is equal to the product of an integer, n_i, and h.

$$J_i = n_i h. \qquad \text{(XIII—10B)}$$

SPECIFIC HEATS OF GASES

The ratio of the specific heats of a gas increases as the temperature is lowered, tending to the limiting value $1\frac{2}{3}$. This behaviour is easily explained by the quantum theory. At rather low temperatures that part of the energy of a molecule which is associated with rotation may well be so small that the corresponding values of J (XIII—10B) never reach so high as h and the only possibility for them is $J = 0 \times h = 0$. The molecule has in fact, at low enough temperatures, no energy of rotation at all. It is quite different with *translational* kinetic energy. While the molecule is traversing a free path its velocity is practically constant. It is in fact moving as if it were in simple harmonic motion of infinitely long period, or *zero frequency*. And since its energy must be equal to $nh\nu$ for each of the three translational degrees of freedom, even a minute quantity of energy makes n enormous. The quantum discontinuity is, as it were, wiped out, and we may in fact deal with the energy of the molecule as we did in Chapter IV, where we learned that the ratio of the specific heats is $1\frac{2}{3}$ when we had only translational energy to consider.

CHAPTER FOURTEEN

CONDUCTION OF ELECTRICITY AND ASSOCIATED PHENOMENA

ELECTROLYSIS

THIS is the name given to the conduction of electricity in acids, salts, etc., more especially aqueous solutions of them. It is accompanied by chemical changes at the electrodes of a characteristic kind. These result in decomposition of the electrolyte, though secondary chemical changes may reform it, e.g. in the case of sulphuric acid in aqueous solution and with electrodes (e.g. platinum electrodes) which do not interact chemically with the *immediate* products of the chemical changes. The old hypothesis of Theodor von Grotthus (1785–1822) accounted for some of the obvious features of electrolytic conduction and in some degree anticipated Faraday's laws, discovered later, but as Clausius pointed out, electrolytes obey Ohm's law and this is incompatible with Grotthus' assumption that the molecules, H_2SO_4 for example, are disrupted by the electric field between the electrodes, as we should now say, into two positively charged H ions and a negatively charged SO_4 ion.

Johann Wilhelm Hittorf (1824–1914), professor at the former Academy of Münster in Westphalia, noticed that different quantities of the electrolyte were lost round the two electrodes and ascribed it, correctly, to the different rates at which the two sorts of ions, the ANIONS and the CATIONS travel: the former in the direction of the anode and the latter towards the cathode.* The ratio of the velocity of the positive ions (cations) to that of the negative ions (anions) is equal, as Hittorf supposed, to the ratio of the loss round the anode to that round the cathode.

Friedrich Wilhelm Kohlrausch (1840–1910), professor in the University of Würzburg, showed how the actual ionic velocities could be determined absolutely from the ratio of the velocities as determined by Hittorf's method and their sum which can

* *Ann. der Phys.*, 89, p. 177 (1853); 98, p. 1 (1858); 106, p. 337 (1859).

be found by measuring the electrical conductivity of the electrolyte.* The understanding of electrolytic conduction and allied phenomena was much further advanced by van't Hoff (1852–1911) and Svante August Arrhenius (1859–1927). The latter introduced the hypothesis that an electrolyte, when dissolved in water, dissociates partially into its separate ions, complete dissociation being approached at extreme dilution.† This seems to have been suggested much earlier by Clausius. It removed, as I have already explained in Chapter IV (I) the difficulty encountered by van't Hoff's view that the osmotic pressure of a substance in solution should be the same as the pressure it would exert if it filled, in the form of gas, the same volume as the solution at the same temperature. Indeed approximately

$$PV = iRT,$$

where i is van't Hoff's factor and P and R respectively the osmotic pressure and the gas constant appropriate to the substance in its undissociated condition. If we have, for example, N molecules in the solution and if the fraction α (degree of dissociation) dissociates into ions, each molecule breaking up into n ions, then the number of particles (complete molecules and ions together) in the solution must be equal to

$$\alpha Nn + (1 - \alpha) N.$$

Evidently this is equal to

$$iN$$

where i is van't Hoff's factor. Therefore, on cancelling out N,

$$1 + \alpha(n - 1) = i.$$

As we already appreciate, i can be found by measurements of osmotic pressure. It can also be determined by measuring the electrical conductivity of the solution and in other ways, e.g. by finding the elevation of the boiling point of the solution above that of the pure solvent. All these different methods are in satisfactory agreement.

* *Göttinger Nachr.*, p. 213 (1876); *Ann. der Phys.*, 6, p. 167 (1879).

† Arrhenius: *Zeits. für Phys. Chem.*, 1, p. 631 (1887). No doubt the high value of the dielectric constant of water facilitates dissociation and the consequent low values of the inter-ionic forces explain the comparative success of electrolytic theory without taking them into account.

The forces exerted on one another by the charged ions in an electrolytic solution have some influence on the behaviour of electrolytes, and in recent times P. Debye and E. Hückel have improved the theory of electrolysis by taking this into account.*

CONDUCTION OF ELECTRICITY IN GASES

The fact that gases do conduct electricity—though nearly complete insulators when subjected to rather small potential gradients—has long been known, and the success of the ionic hypothesis with electrolytes naturally suggested that gaseous conductivity is due to charged ions. When air or any gas is subjected to a potential gradient which is not too great, a minute current passes for a short time and then the air becomes nearly a perfect insulator. This is due to the fact that all the ions have been removed from it. Under ordinary conditions a gas never insulates quite perfectly; ions are always being formed—e.g. by cosmic radiation—unless rather special precautions are taken.

Gases may be ionized in many ways—e.g. by hot bodies, by a flame, and by X-rays. When ions are produced in this way, a potential difference between two electrodes in the gas gives rise to a current which, for small potential differences, obeys Ohm's law, but when the potential difference is big enough the current is almost independent of it. As the potential difference is increased the current remains almost constant. The explanation is that the ions which the ionizing agent produces are now being removed as fast as they are formed. The current is now said to be SATURATED. Further increase in the potential difference between the electrodes causes the current to increase again and results finally in an electric spark between the electrodes. This is explained by the supposition that the ions already present in the gas now acquire such high velocities that their collisions with the molecules of the gas cause the latter to be broken up into charged parts or ions, so that new ions are now being formed at a rate which is all the greater the greater the potential difference between the elctrodes.

The phenomena of conduction in gases have been studied in great detail by Sir Joseph Thomson (1857–1940), Cavendish

* *Physikal. Zeitsch.*, 24, p. 305 (1923); 25, p. 145 (1924). *Trans. Faraday Soc.*, 23, p. 334 (1927).

Professor at Cambridge, and his pupils, and the aspects just outlined have been minutely investigated by Sir John Townsend (1868–), lately professor at Oxford.*

CONDUCTION IN GASES AT LOW PRESSURES

Imagine the gas to be contained in a glass tube provided with two metal electrodes and with a side tube connecting it to a pump. If now an electric discharge is maintained through the gas, it is found that its resistance diminishes as the pressure gets lower, and as low pressures are reached the glow in the tube takes a characteristic form. At pressures of the order of 1 mm. of mercury and lower a long column of luminosity extends from the anode towards the cathode. Beyond this column is a fairly large dark space, the FARADAY DARK SPACE. Near the cathode is a glowing stratum of gas (cathode glow) and between this and the cathode is a rather narrow dark space called the CROOKES DARK SPACE. As the pressure gets still lower the positive column becomes striated, giving the appearance of thin glowing layers separated by dark spaces. At very low pressures the positive column contracts until the Faraday dark space and indeed the cathode glow reach the anode. The resistance of the tube begins to increase as this stage is approached and a remarkable fluorescent glow is now produced on the walls of the glass tube opposite the cathode, the colour of which depends on the composition of the glass. Usually it is greenish. This phenomenon was studied by Plücker (1801–1868), Geissler (1814–1879)—the glass tubes in which the discharge occurs were often called Geissler tubes in Germany—Hittorf and by Sir William Crookes. Plücker observed the phenomenon about 1859 and noticed that the fluorescent glow shifted when a magnet was moved into the neighbourhood of the tube. Hittorf observed in 1869 that an object placed in the tube throws a shadow as if it screened the part of the glass wall from something projected from the cathode. This radiation from the cathode was later called CATHODE RAYS by Goldstein (1880), whose experiments indicated that they might be negatively charged particles. It may be remarked parenthetically that Goldstein made small holes in the cathode of the tube he used and thus discovered the CANAL

* *Phil. Mag.*, 6, p. 598 (1903); 8, p. 738 (1904).

RAYS which proceeded through these holes in the direction away from the anode. Both sorts of rays were later discovered to be corpuscular, the canal rays to be positively charged atoms or molecules of the gas in the tube and the cathode rays to be corpuscles of quite a new kind, later called ELECTRONS. The corpuscular character of cathode rays was insisted on by Crookes who investigated them in the seventies. He demonstrated that they proceed in straight lines perpendicularly to the surface of the cathode. Goldstein noticed that they were deflected by an electrical field as well as by a magnetic field. There was every indication in fact that they were negatively charged particles of some kind thrown out from the cathode.

X-RAYS

In 1895 Wilhelm Conrad Röntgen (1845–1923) was occupied in investigating the emission of ultra-violet light from a discharge tube and used, as a detector, crystals of barium platinocyanide which fluoresce under the influence of light. While this work was in progress he made the startling observation that his platinocyanide crystals fluoresced if there were no light at all and while the discharge tube was covered with opaque paper, provided they were in its neighbourhood. This was the discovery of X-rays, as Röntgen called them.*

CONDUCTION IN METALS

Metals are by far the best conductors of electricity, but the nature of metallic conduction was not so easily open to experimental investigation, or rather, it was not easy to interpret the results of such experiments. G. Wiedemann (1826–1899) and R. Franz discovered the remarkable law that the ratio of the thermal conductivity of a pure metal to its electrical conductivity is proportional to the absolute temperature (K temperature). Moreover the constant of proportionality is very nearly the same for all (pure) metals.† If we write it in the form

$$K/\sigma = AT,$$

* W. C. Röntgen: *Sitzungsber. der Würzburger Physikalisch-Medizinische Gesell.* (1895); *Ann. der Phys.*, 64, p. 1 (1898).

† *Pogg. Ann*, 89, p. 498 (1853).

then A is very nearly the same constant for different pure metals, K and σ being the thermal and electrical conductivities respectively. Paul Drude (1863–1906) assumed the conductivity of metals to be due to free electrons (the corpuscles already mentioned, which constitute the cathode rays). It is very easy to obtain expressions for the electrical and thermal conductivities, assuming the electrons to constitute a gas in the metal and applying the simple kinetic theory. Very simple considerations lead to

$$\sigma = \frac{e^2 nc\,\lambda}{6kT}$$

as an approximate expression for the electrical conductivity. The meanings of the symbols are:

$e \equiv$ charge on one electron,

$n \equiv$ number of electrons per unit volume,

$\lambda \equiv$ mean free path of an electron,

$\bar{c} \equiv$ average velocity of agitation of an electron,

and k is Boltzmann's constant. The thermal conductivity is expressed by

$$K = \tfrac{1}{3} mn\bar{c}\lambda \times C_v,$$

where C_v is the specific heat at constant volume of the electron gas. From this we find

$$K = \tfrac{1}{2} kn\bar{c}\lambda,$$

and on dividing it by the expression for σ we get

$$\frac{K}{\sigma} = 3\frac{k^2}{e^2} T.*$$

Thus the constant, A, is equal to $3k^2/e^2$. Its value is known, since those of k and e are known, and it is quite near to the value experimentally determined from the conductivities.

But Drude's theory led to a difficulty. Our simple kinetic theory assigns to each electron the average energy $3kT/2$. If we assume only one free electron for each atom of the metal this amounts to $3RT/2$ ergs per gramme-atom, and we learned in discussing the law of Dulong and Petit that the vibrational energy of the atoms (e.g. copper atoms) in a gramme-atom is equal to $3RT$. Thus we have an atomic heat for copper of $9R/2$

* P. Drude: *Ann. der Phys.*, 1, p. 566 (1900); 3, p. 369 (1900).

ergs per 1°C., and when we divide by Joule's equivalent and by 63, the atomic weight of copper, we get about 0·143 calorie per gramme per 1°C. This is far too big—the specific heat of copper is about 0·095.

The difficulty has been solved by applying to the electron gas a new form of statistics—FERMI-DIRAC STATISTICS. The application of it to the thermal and electrical conductivities of metals has been successfully achieved by A. Sommerfeld.* It appears that the new statistics do not make any appreciable change in the calculated ratio of the conductivities, but the specific heat of the electron gas in a metal is zero or negligible.†

SUPERCONDUCTIVITY

The electrical conductivities of metals *increase* as the temperature is lowered, or, as it may be expressed, the resistance of a coil of wire becomes less as its temperature is lowered. Kamerlingh Onnes observed quite unexpectedly in 1911 that the resistance of mercury disappeared altogether when it was cooled to 4·12°K.‡ This SUPERCONDUCTIVITY was then observed in a number of metallic conductors, in lead below 7·26°K, in tin below 3·69°K, and in many others. It can be recognized by the persistent magnetic field of a superconducting ring of metal after the inducing electromotive force, or any other electromotive force, has ceased to act. The resistance of a metal in the superconducting state, according to the experiments of Onnes and his collaborators, is less than 10^{-15} times its resistance at room temperature. A remarkable feature of the phenomenon is that the resistance of a superconducting wire can be restored by a magnetic field parallel to the length of the wire, provided this field exceeds a certain critical value which depends on the temperature and on the particular conductor. Meissner and

* A. Sommerfeld: *Zeitsch. für Physik*, 47, pp. 1 and 43 (1928). Fermi: *Rend. Accad. Lincei*, 3, p. 145 (1926). P. Dirac: *Proc. Roy. Soc.*, 112, p. 661 (1926).

† An electron gas (Fermi-Dirac statistics) approximates in its character to the old-fashioned gas of classical physics at sufficiently high temperatures and low enough pressures. This explains Sir Owen Richardson's observation of the Maxwellian distribution in the electrons emitted from hot metals.

‡ Kamerlingh Onnes: *Leiden Comm.*, 122B (1911).

Ochsenfeld have found in the case of a pure metal, in a superconducting state, that the magnetic induction, B, is zero* (MEISSNER EFFECT).

FURTHER INVESTIGATION OF CATHODE RAYS

The suggestion that cathode rays consist of streams of negatively charged particles was proved experimentally by Jean Perrin.†

He simply allowed the cathode stream to impinge on a conductor connected to an electroscope (or electrometer) and observed that it charged up negatively. The anode of his tube was hollow (like a Faraday vessel). Its potential was kept constant, and the conductor, on which the cathode particles fell, was inside it and, as it were, protected from everything except the corpuscular stream.

Sir J. J. Thomson made one of the earliest determinations of the ratio of the charge on a cathode particle to its mass. He called these cathode particles CORPUSCLES, but the name ELECTRON has come to be universally adopted. A stream of charged bodies or particles has the magnetic properties of an electric current in a wire,‡ so that the cathode stream is deflected by a magnetic field directed at right angles to it. An electric field also deflects it, but when we apply both fields at the same time in suitable directions at right angles to one another, and to the direction of the flying electrons, and suitably adjust them, the forces they exert annul one another and the cathode stream is undeflected. If e is the charge on an electron, v its velocity and H the intensity of the magnetic field, the force exerted on it is Hev. The force due to the electric field is Ee, if E is the electric field intensity, and when both forces are equal

$$eE = Hev$$

so that

$$v = E/H$$

in electromagnetic units. Sir J. J. Thomson determined v in

* Meissner and Ochsenfeld: *Naturwiss.*, 21, p. 787 (1933). See also D. Shoenberg: *Superconductivity* (Cambridge University Press).

† *Comptes rendus*, 121, p. 1130 (1895).

‡ H. A. Rowland: "On the Magnetic Effect of Electric Convection," *Amer. Journ. of Science* (3), 13, pp. 30–38 (1878).

this way, and by measuring the deflection produced when only the magnetic field was applied he was able to find e/m.* The ratio was found by Kaufmann and by Wiechert about the same time.† It turned out eventually to be in the neighbourhood of $1{\cdot}77 \times 10^7$ in E.M. units, a little more than 1840 times that of a hydrogen ion as found by measurements of the electrochemical equivalent in electrolysis.

Electrons‡ were found to be identical whatever the material of the cathode from which they originated. Since they carry the same charge as a hydrogen ion and have a ratio e/m which is about 1845 times as big, their masses are all equal and much smaller than that of any atom—about 9×10^{-28} gramme. It therefore appeared to Thomson that electrons were actually constituents of atoms and he adopted the view that every atom consisted of a massive positively charged part and one or more negatively charged electrons, whose masses were negligible compared with that of the whole atom. His picture of an atom was not quite the same as Rutherford's. He thought of the positive charge as spread over a spherical region and the electrons as moving inside it.

* J. J. Thomson: *Phil. Mag.*, 44, p. 293 (1897).

† Kaufmann: *Wied. Ann.*, 62, p. 598 (1897); Wiechert: *Wied. Ann.* (Beiblätter), 21, p. 443 (1897).

‡ The name "electron" was first used by G. Johnstone Stoney (1880), not for what were later called "electrons", but for the natural unit of electric charge: the charge on a hydrogen ion, or on an electron as the name is now used.

CHAPTER FIFTEEN

MISCELLANEOUS SUBJECTS

ELECTRON THEORY

IONS and charged particles became very prominent in physical theory as it began to be appreciated that cathode rays consisted of charged particles and the view began to find acceptance that all electric currents consist of streams of charged particles. H. A. Lorentz modified the expression σE (conductivity $\times$ field intensity) for current density (e.g. in a metal) to ρv (electric density $\times$ velocity of convection) or nev where n and e are respectively the number of charged particles (electrons) per unit volume and the elementary charge on one of them. He amplified Maxwell's electromagnetic theory in this sense and found the now familiar expression for the force (due to a magnetic field) on a moving charged particle, namely, in our electromagnetic units for example,

Charge $\times$ *Vector product of* v *and* B,

where B is the magnetic induction, i.e. $\mu \times H$. We have already met with this in the form Hev which it assumes when H and v are at right angles to one another and $\mu = 1$.

As early as 1881 Sir J. J. Thomson pointed out that a charge of electricity has a mass (electromagnetic mass) associated with it, in addition to the mass peculiar to the material body that may carry the charge. It is very easy to see that this must be the case. A moving charged body, as Rowland showed experimentally, produces a magnetic field round about it—since it constitutes in fact a convection current. In order therefore to set such a charged body in motion work must be done on it to provide the energy of its magnetic field. This energy is in fact kinetic energy (energy of motion), and when we equate it to $mv^2/2$, we can calculate m, which turns out to be constant *for small velocities*. In fact, in the case of a spherical body with a charge uniformly spread over its surface and moving through air or *in vacuo* ($\mu = 1$ approximately) it is expressed by

$$m = 2e^2/3R$$

(in ordinary electromagnetic units), R being the radius of the spherical body and e the charge on it. Both the charge on an electron and its mass are known from experimental observations and from them the radius of such an electron is found to be

$$1{\cdot}9 \times 10^{-13} \text{ cm.}$$

We can scarcely wonder that, as the end of last century approached, the suggestion arose that perhaps all mass is of this electromagnetic kind and that it should give birth to the ELECTRON THEORY OF MATTER, according to which all material things are constituted of electricity, electrons having the most prominent place in it. In its most extreme form (Max Abraham, 1903) it presented electrons as charged spherical bodies and absolutely rigid.* They had no mass except that of the electromagnetic kind. One consequence of it, which lent it some support, was that the masses of bodies must increase with their velocity. Even this extreme form of theory could not dispense with something which was non-electrical—the rigid framework which held the electric charge together by some sort of force which was necessarily not electrical. Lorentz's theory was superior to that of Abraham. He anticipated much of the relativity theory which developed later, as we have already seen—including the dependence of mass on velocity, whether mass was of the electrical kind or not. His general attitude is set out in a little book called *Versuch einer Theorie der Electrischen und Optischen Erscheinungen in Bewegten Körpern* (Brill, Leiden, 1895; reprinted in 1906 by Teubner, Leipzig). The outlook of the theoretical physicists at the *fin de siècle* is fairly clearly presented in Joseph Larmor's *Aether and Matter* (Cambridge, 1900).

THE CLOUD CHAMBER

One of the most important pieces of apparatus for investigating atomic phenomena is the CLOUD CHAMBER devised by C. T. R. Wilson (1869–). It was not originally intended for any such purpose at all. When a young man Wilson, on holiday in the highlands of his native Scotland, climbed Ben Nevis and

* Max Abraham: "Prinzipien der Dynamik des Elektrons," *Ann. der Phys.*, 10, p. 105 (1903).

became interested in the wonderful colour phenomena (glories) due to sun and mist. On returning to Cambridge he determined to reproduce and study them in the laboratory. Previous investigations indicated that the fog produced when air, initially saturated with water vapour, is caused to expand adiabatically, is due to the presence of small dust particles.* Without them the vapour appeared to remain supersaturated (Coulier, 1875, and Aitken, 1888). Röntgen's discovery of X-rays became known about the time when Wilson was occupied with these experiments and he tried the effect of passing a beam of X-rays through the saturated water vapour. He now obtained a dense cloud with an expansion in the ratio of 1 : 1·26, even when all dust particles had previously been removed. The nuclei, which were now effective, he proved to be charged ions by demonstrating that they could be removed by an electric field. He also found that a bigger expansion ratio was needed to cause condensation on positive ions than on negative ones.†

The great importance of the cloud chamber is due to the fact that it enables *individual* (charged) particles such as α particles, electrons, protons, to be observed. One of these, as it travels through the cloud chamber, ionizes the molecules of air and water vapour, so that its track is indicated by a line of water drops.

EXPERIMENTAL MEASUREMENT OF THE ELEMENTARY CHARGE

There are many indirect ways of determining the charge on an ion. For example, a knowledge of Boltzmann's constant, k, or of Loschmidt's number, when combined with electrolytic data, enables this charge to be computed. The first *direct* experimental measurement of e (as I shall call it) was achieved by Sir John Townsend (1868–).‡ The gases which come off during electrolysis, especially with large currents, are charged and, when passed through water, form dense clouds. Townsend determined the charge on such a cloud and the amount of water

* *Vide* Chapter II.

† C. T. R. Wilson: *Phil. Trans. A.*, 189, p. 265 (1897); *A.*, 192, p. 403 (1899). *Proc. Roy. Soc. A.*, 85, p. 285 (1911).

‡ *Proc. Camb. Phil. Soc.*, 9, p. 224 (1897); *Phil. Mag.*, Feb. 1898.

in it. He found the size of an individual drop by observing the rate at which the cloud fell and using Stokes' formula:

$$v = 2(\rho - \rho') \times g \times r^2/9\eta.$$

This gave the average radius and thus the mass of a drop. The ionic charge e he got by dividing the total charge by the number of drops, which latter is given by dividing the total amount of water carrying the charge by the mass of a drop. He found about 3×10^{-10} electrostatic units at first. Making certain corrections later, he revised his estimate to 5×10^{-10} which is very close to what is now thought to be the best estimate, namely

$$4{\cdot}803 \times 10^{-10} \text{ E.S. units.}$$

Sir J. J. Thomson followed. His method consisted in producing a cloud of charged water drops in a Wilson cloud chamber.* Apart from this his method was almost the same as Townsend's. Thomson's earliest result was $6{\cdot}5 \times 10^{-10}$, but in his later experiments he obtained $3{\cdot}4 \times 10^{-10}$ E.S. units. Millikan's judgment is that Thomson's determinations contained all the *theoretical uncertainties* that attach to Townsend's, but that the *experimental uncertainties* are more serious in Thomson's type of experiment.†

H. A. Wilson's method was like Sir Joseph Thomson's with the added feature that he could subject the charged cloud to an electric field.

R. A. Millikan's method was a great improvement on those of his predecessors. Instead of a cloud he used charged oil drops. He made use, like H. A. Wilson, of an electric field and he was able to deal with *an individual drop*. Moreover in using Stokes' law he and his pupils made careful measurements of the viscosity of air and he also corrected the law for the inhomogeneity of the medium through which his oil drops moved.‡ Millikan's final result was

$$e = 4{\cdot}774 \times 10^{-10} \text{ E.S. units.}$$

* *Phil. Mag.*, 46, p. 528 (1898); *Phil. Mag.* 5, p. 354 (1903).

† *Vide* R. A. Millikan: *The Electron* (University of Chicago Press, 1917). This little book contains detailed information about the determinations of e up to 1917, including Millikan's own work.

‡ Air and gases are not homogeneous, being constituted of relatively widely spaced molecules, and Stokes' law holds only for a homogeneous medium.

I should like to make the observation here that in my view Millikan's experimental determination of e is the most impressive one that has been made. Slightly more accurate estimates have been made since, but they have been made by rather indirect methods which cannot be described here. They indicate, as we have seen, that the true value of e is a trifle greater than that obtained by Millikan.

RATIO OF CHARGE TO MASS FOR HIGH-VELOCITY ELECTRONS

The beta rays from radioactive materials are electrons and have a very wide range of velocities, approaching within one or two per cent of the velocity of light in free space, and the electron theories of Abraham and Lorentz, both of them, indicated a variation of mass with velocity which, though insensible for small velocities, becomes very appreciable with high velocities. So W. Kaufmann (1871–), A. H. Bucherer (1863–) and others measured the ratio e/m for beta particles, to decide the rival claims of Abraham's electron and that of Lorentz (which was the so-called contractile electron, experiencing the FitzGerald–Lorentz contraction, or, we may say, the electron of special relativistic theory). The results of these experiments were very definitely in favour of Lorentz's electron and of Einstein's relativity theory.*

Bucherer used a very elegant method which is illustrated in Fig. XV—1. He placed a speck of radium fluoride at the centre, O, of an arrangement like a parallel plate condenser, A. This is contained in a wide cylindrical vessel, B (Fig. XV—1 (a)). A photographic film extended round the cylindrical wall of this vessel and the particles shot out (alpha and beta particles) from the fluoride, in the absence of fields, produce the line $\alpha\alpha$ (Fig. XV—1 (b)). Electric and magnetic fields are applied at right angles to one another, the former by establishing a potential difference between the condenser plates, so that the field stopped fairly sharply at the edge of the condenser; the latter by large

* W. Kaufmann: "Ueber die Konstitution des Elektrons," *Ann. der Phys.*, 19, p. 487 (1906). A. H. Bucherer: "Die experimentelle Bestätigung des Relativitätsprinzips," *Ann. der Phys.*, 28, p. 513 (1908).

Helmholtz coils with their axis parallel to the condenser plates. The fields are so directed that the vertical forces they exert on the electrons are in opposite directions. Electrons (beta particles) can emerge and their directions depend on their velocities. In fact

$$v = E/H \sin\theta \text{ (E.M. units)},$$

where θ is the angle between the direction of the magnetic field, H, and that in which the electron is travelling within the

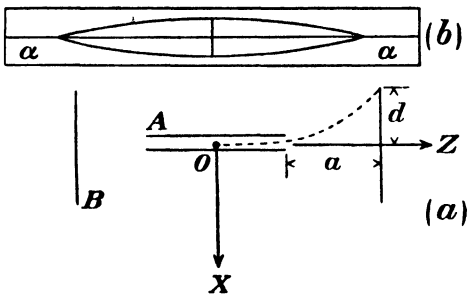


Fig. XV—1

condenser. The emerging electrons now produce one or other of the bow-shaped lines (Fig. XV—1 (*b*)), according to the directions of the fields. The theory is too long to give here;* suffice it to say that measurements made on the lines (Fig. XV—1 (*b*)) enable the ratio e/m to be computed and the associated velocity of the electron is that given by the above formula. It is interesting to note that the bow-shaped lines (Fig. XV—1 (*b*)) meet at points on the median line which correspond to c, the speed of light in free space, and that electrons do not reach the film beyond these points (except of course in the absence of fields). In fact it appears that c is the upper limit for the speed of an electron. Bucherer's experiments confirmed Lorentz's law (which is also that of Einstein's relativity theory) for the variation of mass with velocity. In fact experiment shows that

$$m = m_0\gamma,$$

where m_0 is the mass of the particle when at rest and m its mass

* *Vide* O. W. Richardson: *Electron Theory of Matter* (Cambridge), pp. 239 *et seq.*, or W. Wilson: *Theoretical Physics*, Vol. ii, p. 210 (Methuen).

when it has the velocity v. The factor γ is of course the same factor as that represented by γ in the chapter on the Aether—

$$\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}.$$

It has been assumed in the foregoing description that e, the charge on an electron, is independent of its velocity. It may suffice here to say that there are strong reasons for believing this to be the case.

BARKLA'S INVESTIGATION OF X-RAYS

The first to make any appreciable progress in elucidating the nature of X-rays was Charles Glover Barkla (1877–1944), Professor of Physics at King's College, London, and later in the University of Edinburgh. He definitely established that they were of the same nature as light, but of very short wave-length. He had no means of measuring wave-lengths, but was able to polarize X-rays. This combined with the fact that the refractive index of the radiation in any material is practically unity meant very short transverse waves.

The following illustration, which Barkla used himself, explains the experiments he did to produce polarization in X-rays.

Imagine a horizontal (inextensible) cord running north and south and attached at its northerly end to a point, A, which is not quite rigidly fixed. Transverse waves travelling along the cord will cause motions of A in a plane perpendicular to the cord. Another cord running from A in, say, a westerly direction will have transverse waves developed in it, but its vibrations will necessarily be up and down ones only. The wave along this second cord is *polarized*. Suppose it to end at some point, B, which, like A, is not quite rigidly fixed. It is easy to see that no wave at all can travel from B in a *vertical* direction along some third cord, but if the direction of this third cord is horizontal and north and south, polarized waves will travel along it. Now we can understand Barkla's experiment. The waves travelling northwards are waves of X-radiation. The X-radiation scattered from A (which in some of Barkla's experiments was a piece of carbon) in the westerly direction are (suspected to be) polarized. The test was that no radiation, or a very weak radiation, is scattered from the piece of carbon at B in a vertical direction, but

a moderately intense beam is scattered northwards or southwards from *B*. Barkla observed and measured the intensity of X-ray beams by means of an ionization chamber.*

When a primary beam of X-rays falls on some material this latter emits a secondary radiation consisting partly of a stream of electrons (photo-electric emission) and partly of a secondary beam of X-rays. Barkla investigated the secondary X-rays by observing their absorbability in aluminium plates. If, for example, one of his aluminium plates cut down the intensity of the X-ray beam to one-half, and if a second exactly similar plate cut down the intensity of what had passed through the first plate to one-half, and so on, he described the X-radiation as HOMOGENEOUS and assumed (correctly as we now believe) that it was "monochromatic," i.e. radiation of a very short range of wave-lengths. A portion of the secondary X-ray beam he found to be a replica, as it were, of the primary beam—it was absorbed in the same way as the primary by the sequence of aluminium plates. He concluded correctly that this was just the part of the primary beam which the irradiated material scattered. According to the classical theory this scattered radiation should be *exactly* like the primary beam in all respects, except its intensity, but Barkla noticed one difference: the scattered radiation was slightly more absorbable than the primary, indicating a very slightly longer wave-length. This was the earliest indication of the COMPTON EFFECT.

Barkla's observations were also the earliest which indicated that the number of electrons (extra-nuclear electrons) in an atom is identical with its ATOMIC NUMBER, i.e. the number which marks its place in the periodic system. It is impossible to give all the detail here,† but, briefly, he determined the MASS COEFFICIENT of scattering in the case of a number of light elements and found it to be near 0·2. Now the classical theory indicates that

$$\text{(Mass coefficient of scattering)} = 0{\cdot}4 \times \frac{\text{(Number of scattering electrons in one atom)}}{\text{(Atomic weight)}}$$

* *Phil. Trans. A.*, 204, p. 467 (1905); *Proc. Roy. Soc.*, 77, p. 247 (1906).

† *Vide* W. Wilson: *Theoretical Physics* (Methuen), Vol. ii, pp. 227, 228, 229.

Thus Barkla found the number of electrons per atom to be one-half the atomic weight for a number of elements of low atomic weight. But for such elements the ATOMIC NUMBER is also equal to one-half the atomic weight and there was thus a strong presumption that atomic number and number of electrons per atom are identical. We should now identify these scattering electrons with extra-nuclear electrons and since their total negative charge must be balanced by the positive charge on the nucleus we may restate Barkla's result by saying that the atomic number is identical with the number of elementary units of charge on the nucleus of the atom.*

Perhaps Barkla's greatest discovery was that of CHARACTERISTIC (or FLUORESCENT) X-radiation. In addition to that part of the secondary radiation which is scattered, he found one or more radiations which were *characteristic of the material itself* (e.g. silver) and independent of the nature of the primary beam, provided the latter was hard enough, or penetrating enough (in terms of later knowledge, had a short enough wave-length or high enough frequency). He labelled them *K*, *L*, *M*, and so on. Silver, for example, he found to emit two, *K* and *L*. Each of these he found to be homogeneous (in the sense described above). Indeed this characteristic led him to their recognition. Of course homogeneous, as was learned later and as Barkla believed, means "monochromatic."† The *K* radiation he found to be more penetrating (shorter wave-length or higher frequency) than the *L* radiation and the *L* radiation more penetrating than the *M*, and so on.‡ The importance of these discoveries for the subsequent inquiries into the structure of atoms was very great indeed.

THERMIONICS

If a charged body, e.g. a metallic body, is raised in temperature to 1000°C. or more it quickly loses its charge. This phenomenon has been known for a long time. It was first seriously

* C. G. Barkla: *Phil. Mag.*, 21, p. 648 (1911).

† This is only an approximate description. The *K* radiation from an element, e.g. silver, exhibits discrete spectral lines, but the extreme variation in wave-length is not great.

‡ C. G. Barkla and C. A. Sadler: *Phil. Mag.*, 16, p. 550 (1908). C. G. Barkla: *Phil. Mag.*, 22, p. 396 (1911).

examined by F. Guthrie,* but our knowledge of it is due almost wholly to Sir Owen Richardson, formerly professor at Princeton, U.S.A., and lately at King's College, London, who gave the name THERMIONICS to the subject and the name THERMIONS to the emitted ions.†

When a hot wire is maintained at a positive potential relatively to its surroundings it emits positive electricity, but this effect quickly decays and appears to be due to traces of volatile materials, salts, etc., on the surface of the wire. On the other hand, when maintained at a *negative* potential the emission settles down to a steady rate which lasts indefinitely. Richardson assumed that the electrons of which the negative emission consists‡ occupy the hot metal in the form of a gas (electron gas) and applied to it the old kinetic theory of gases. He thus deduced the following formula (RICHARDSON'S LAW) for the number of electrons emitted per unit area per second, or for the current per unit area (they differ only by a constant):

$$i = AT^{\frac{1}{2}}e^{-\varphi/kT}, \qquad \text{(XV—1)}$$

where A is a constant, φ the work that has to be done to bring an electron through the surface of the metal and the remaining symbols have their usual meaning.§ Later investigation, by a thermodynamical method, of the dependence of φ on temperature led Richardson to the result:

$$\varphi = \varphi_0 + 3kT/2,$$

where φ_0 is a constant and so to an improved formula

$$i = AT^2e^{-\varphi_0/kT}, \qquad \text{(XV—1A)}$$

where A is again a constant though somewhat different in value. It is impossible to distinguish between these two formulae by measurements of the thermionic current—because of the dominating character of the exponential, but it seems certain, on indirect grounds, that the latter is the more correct one.

Very simple thermodynamical reflections suggest that the

* *Phil. Mag.*, 46, p. 257 (1873).

† O. W. Richardson: *The Emission of Electricity from Hot Bodies* (Longmans, Green & Co., 1922). O. W. Richardson and W. Wilson: Article on Thermionics in the *Dictionary of Applied Physics*, Vol. ii.

‡ That it consists of electrons is proved by measurements of the ratio e/m. J. J. Thomson: *Phil. Mag.*, 48, p. 547 (1899).

§ O. W. Richardson: *Phil. Trans. A.*, 201, p. 497 (1903).

photo-electric emission (emission of electrons under the influence of light) due to black body radiation should follow Richardson's law and the author of this work suggested in 1912 that the thermionic emission is of this nature—due to the radiation within the hot metal.*

Richardson showed that the electrons emitted by a hot body conform to Maxwell's law of distribution of velocities. We appreciate to-day that this is only true because the electron gas outside the emitting metal has a very low density.

Thermionic emission has an immense range of practical application. The electron tubes (triodes, pentodes, etc.) used for the emission and reception of broadcasting waves are simply adaptations of the tubes used by Richardson and his pupils for the investigation of thermionic phenomena.

The modern X-ray tube, invented by W. Coolidge, is also such an adaptation. Instead of cathode rays produced by a current in a rarefied gas, electrons are produced thermionically by heating the cathode, which has the form of a coiled wire, with an electric current. The tube is exhausted as far as possible. With such a tube the intensity and hardness of the X-ray beam can be varied at will by varying the supply of electrons and the potential drop through which they are made to fall.

In the recent forms of the cathode ray oscillograph, the cathode rays (electrons) are produced by a hot filament. The electrons impinge on a fluorescent screen and produce a bright spot there. A transverse electric field applied across the tube through which the electrons are flying causes them to impinge at another point on the screen. The instrument played a great part in "radar" (radio detection and ranging) during the late war, since it could be used to determine the minute fraction of a second elapsing between the emission of a radio pulse and the reception of its echo and thus to estimate the distance of an enemy aeroplane for example.

PHOTO-ELECTRICITY

In the later years of last century Heinrich Hertz noticed that light in some way facilitated the passage of a spark across the gap in the secondary of an induction coil. The length of the gap

* W. Wilson: *Proc. Phys. Soc.* (1912); *Ann. der Phys.*, 42, p. 1154, (1913).

could be adjusted so that while no discharge passed across it in the dark it did so when the electrodes were illuminated by the light from another spark in the same neighbourhood. Metallic bodies emit electrons when illuminated by light of sufficiently short wave-length (sufficiently high frequency). The earliest

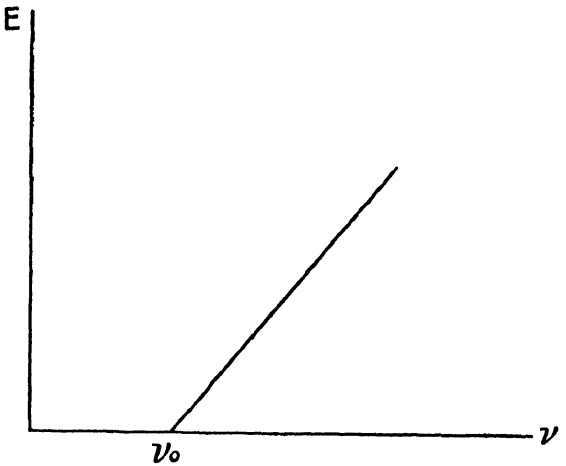


Fig. XV—2

notable investigation of the phenomenon was by W. Hallwachs, of Dresden. Indeed it has been called the Hallwachs phenomenon. Hallwachs made the important observation that, while a negatively charged body will lose its charge under the influence of ultra-violet light, this is not the case with a positively charged one.* The phenomenon has the following features:

(i) There is no emission of electrons at all† unless the frequency, ν, of the light is great enough. For most metals it is in the ultra-violet region. Each metal has a THRESHOLD FREQUENCY. Only when this frequency is exceeded does the effect occur.

(ii) The kinetic energy of the individual electrons is a simple function of the frequency of the light, illustrated by Fig. XV—2, in which the ordinates represent the kinetic energies of individual electrons and the abscissae the corresponding frequency

* W. Hallwachs: *Ann. der Phys.*, 33, p. 301 (1888).

† J. J. Thomson demonstrated that the emission consists of electrons. *Phil. Mag.*, 48, p. 517 (1899).

of the light. In fact the energy of an electron (its maximum energy at any rate) is proportional to

Frequency minus constant.

Of course the greater the intensity of the light the greater the rate of emission of electrons and their total energy, but the energy of *individual* electrons is quite independent of the intensity of the light. These facts about photo-electricity were very perplexing. However weak the intensity of the radiation, the ejected photo-electrons have the *same energy* (for radiation of a given frequency) on leaving the metallic surface as they have when the illumination is strong. It appeared to be impossible that electrons could be ejected in this way, since the energy in the irradiating beam of ultra-violet light, or X-rays, is uniformly spread over the metallic surface.

EINSTEIN'S THEORY OF THE PHOTO-ELECTRIC EFFECT

Einstein met the difficulty in characteristic fashion. He assumed that the energy in a beam of light was concentrated in small bundles and assigned to each bundle the amount of energy which Planck's quantum theory suggests, namely $h\nu$.* When the light bundle is absorbed by the metal, or other substance, an electron is liberated and a definite part, w, of the energy, $h\nu$, is used up in doing the work necessary to drag the electron out through the surface of the material. The rest of the $h\nu$ units of energy endows the electron with its kinetic energy. Thus

$$\tfrac{1}{2} mv^2 = h\nu - w.\dagger \qquad \text{(XV—2)}$$

An electron can only emerge when $h\nu$ is just equal to or bigger than w. The lowest possible frequency, ν_0 (threshold frequency), is therefore given by

$$h\nu_0 = w.$$

Consequently

$$\tfrac{1}{2} mv^2 = h\nu - h\nu_0 \qquad \text{(XV—2A)}$$

in conformity with the diagram (Fig. XV—2). This formula was

* A. Einstein: *Ann. der Phys.*, 17, p. 132 (1905).

† If ν is *very* big the relativistic expression for the kinetic energy must be used, namely, $mc^2 - m_0c^2$, where m is mass of the electron and m_0 its mass when at rest.

confirmed experimentally by A. L. Hughes* and a value for h obtained which, within the limits of experimental error, was in agreement with Planck's estimate.

Einstein's suggestion raised difficulties in another direction. If light has the constitution he suggested, i.e. if it consists of separate bundles of energy, as it were, one would expect a modification of the familiar interference phenomena, at any rate when the light is so weak that each light bundle travels in isolation from its fellows. Sir G. I. Taylor tested this experimentally with light so faint that he had to expose his photographic plate for 2000 hours, but the fringes duly appeared with unimpaired distinctness. The only inference seemed to be that light *could not* have this bundle constitution, but in that case photo-electric phenomena were inexplicable. The solution of the riddle and much more was revealed by the WAVE MECHANICS of Prince Louis de Broglie and its interpretation by his successors, especially Erwin Schroedinger and Max Born. Einstein's bundles are our PHOTONS, the strange particles or quasi-particles of which light and radiation seems to be constituted.

The excitation of X-rays by electron bombardment is a converse phenomenon to the photo-electric one. In falling through a potential drop, V, an electron acquires the additional energy eV (e is the electronic charge, namely $4{\cdot}8 \times 10^{-10}$ E.S. units) and, if its initial energy is negligible, as it is when the electron is liberated thermionically, the eV is practically the whole energy of the electron.

When now the electron with the energy eV strikes a piece of metal, e.g. the anti-cathode in an X-ray tube, it generates radiation, i.e. X-radiation, of frequency ν, in accordance with

$$eV = h\nu. \qquad \text{(XV—3)}$$

Since in the phenomenon we are considering both eV and $h\nu$ are very big we need not worry about any work, w, done on the electron at the metallic surface.

Of course the collision of an electron and an atom generally results, as we shall better appreciate later, in a rearrangement of the electrons in the atom with the consequent emission of

* *Phil. Trans. A.*, 212, p. 205 (1912); see also O. W. Richardson and K. Compton, *Phil. Mag.*, 24, p. 576 (1912).

X-ray photons (Barkla's CHARACTERISTIC RADIATION) of lower frequency than ν, as indeed experiment confirms. The frequency given in (XV—3) is an upper limit.

THE COMPTON EFFECT

When an X-ray photon—or for that matter any photon—collides with an electron it gives up some of its energy to it and then glances off with diminished energy. The photon is like a rapidly moving billiard ball and the electron like a slowly moving and much more massive ball. Before the collision the photon has the energy $h\nu$ and after being scattered the smaller energy $h\nu_1$. Reflections like these led A. H. Compton to predict that the frequency of scattered X-rays should be slightly lower than that of the primary beam, or, what amounts to the same thing, the wave-length of scattered X-rays (or indeed of scattered light in general) should be slightly longer,* as indeed Barkla had inferred from measurements of absorbability. Compton deduced the following formula and verified it experimentally:

$$\lambda_1 - \lambda = \frac{2h}{m_0 c} \sin^2 \frac{\theta}{2},\dagger$$

in which λ and λ_1 are the wave-lengths of the primary and scattered radiations respectively, m_0 is the mass of an electron at rest and θ is the angle between the directions of the primary and the scattered beams.

THE ELECTRON MICROSCOPE

In the chapter on wave-mechanics we shall learn that waves are associated even with such things as electrons, just as light waves are associated with photons. A beam of electrons exhibits the phenomena of interference and diffraction characteristic of waves, and a new subject, rather naturally but not quite correctly called ELECTRON OPTICS, has grown up in recent times. Electrons emerging from a point can be focussed at another point and an image of a structure irradiated by electrons can be formed and

* A. H. Compton: *Phys. Rev.*, 21, pp. 483, 715 (1923); 22, p. 409 (1923).

† *Vide* W. Wilson: *Theoretical Physics*, Vol. iii, pp. 164 and 165.

indirectly rendered visible by a fluorescent screen or photographically. A uniform magnetic field will cause electrons, all starting from the same point with the same component of velocity parallel to the field, to pass through another point common to them all. In fact the distance of this point from the starting point can be shown to be equal to

$$2\pi mv/H,$$

where v is the component of their velocity parallel to the magnetic field of intensity H and m is the mass of an electron. The focusing can also be achieved by suitably arranged electrostatic fields. This is the basis of the ELECTRON MICROSCOPE. It is a cumbersome and awkward piece of apparatus and probably no one would have taken the trouble to construct or use one, but for the high resolving power that can be achieved with it. I ought to explain that RESOLVING POWER, as applied to a microscope or telescope, is not quite the same thing as spectroscopic resolving power, but they are similar and related to one another. As applied to a microscope or telescope the term means its capacity for separating and making clear the details in the structure or appearance of the object being examined—in fact to reproduce in the image the fine detail of the object. The resolving power of a microscope is all the greater the greater its aperture and *the smaller the wave-length* of the undulations (light or electron waves) used with it. Now electrons can easily be given a momentum which is associated with a very much shorter wave-length than that of the light used with the common microscope, so that the electron microscope has a much higher resolving power, notwithstanding the fact that it cannot be endowed with a very great aperture.

CHAPTER SIXTEEN

ATOMIC STRUCTURE

SPECTRA

THE first to make successful calculations of the wave-lengths of spectral lines from an assumed model of the emitting atom appears to have been J. W. Nicholson, Professor of Mathematics at King's College, London. His work was published in a remarkable series of papers on the spectra of the light from certain nebulae and from the solar corona, in 1911 and 1912. These spectra were unlike any terrestrial spectra and he ascribed them to hypothetical elements which he named nebulium and protofluorine. Still earlier Rutherford (1871–1937) had come to the conclusion that all atoms consisted of a massive positively charged nucleus, a kind of sun, with planetary electrons revolving about it. This was suggested to him by the way in which metals scattered the alpha particles shot out by radioactive elements. The very wide angles through which some of the particles were deflected could only be accounted for on the supposition that the atom had a very small, but relatively very massive, repelling central nucleus. Under the influence of Rutherford's suggestion Nicholson imagined atoms with a positively charged central nucleus and 4 or 5 electrons revolving round it, all in the same ring. His mathematical methods were rather similar to those employed by Clerk Maxwell many years earlier in his theoretical investigation of Saturn's rings. He calculated the frequencies of vibration of the electrons perpendicular to and in the plane of the ring. He followed classical methods entirely at first, and since he had nothing which fixed the angular velocity of the electrons in the ring he could at first only calculate the *ratios* of the possible frequencies of vibration. Now according to classical notions the frequencies of the emitted radiation (light) are identical with the frequencies of vibration in the emitting system. By *assuming* one wave-length (as given by observation) Nicholson was able to calculate others quite correctly and even

to predict correctly the wave-lengths of spectral lines which had not yet been observed. Finally he made a most important discovery, namely, that the angular momentum of the atom was an integral multiple of $h/2\pi$. This enabled him to calculate wave-lengths absolutely.*

It was a great piece of work which has undeservedly fallen into oblivion and one cannot doubt that Bohr's successful theory of the hydrogen atom which followed soon afterwards owes much to it.

BALMER'S SERIES

The spectrum of hydrogen, caused to glow by sending a current through the gas at low pressure, is characterized by a prominent series of lines called BALMER'S SERIES after a Swiss schoolmaster of that name who, in 1885, discovered that the wave-lengths of the lines in the series could be represented by the following formula:

$$\lambda = 3645{\cdot}6 \times \frac{n^2}{n^2 - 4},$$

in which n is any integer greater than 2 and the unit of length is the Ångström (10^{-8} cm.).† He found this formula quite empirically, by trial and patience. Rydberg expressed it so as to give wave-numbers, i.e. $1/\lambda$, thus:

$$\nu' = \frac{1}{\lambda} = \frac{4}{3645{\cdot}6}\left(\frac{1}{2^2} - \frac{1}{n^2}\right),$$

or, on multiplying by 10^8,

$$\nu' = 109{,}720\left(\frac{1}{2^2} - \frac{1}{n^2}\right),$$

when lengths are measured *in centimetres.* This is usually written

$$\nu' = R\left(\frac{1}{2^2} - \frac{1}{n^2}\right)$$

The constant 109,720 is called RYDBERG'S CONSTANT, but this term is also used for its product with c, the velocity of light, i.e. for

$$109{,}720 \times 3 \times 10^{10},$$

* J. W. Nicholson: *Monthly Notices, R.A.S.*, 72, p. 679 (1912).

† It will be noticed that the wave-lengths have the lower limit 3645·6 Å.U.

which must replace 109,720 when we wish to calculate FREQUENCIES instead of wave-numbers.

W. Ritz, who developed a spectral theory of his own,* conjectured that the wave-numbers (and consequently also frequencies) of all spectral lines can be expressed as differences of terms which are characteristic of the emitting atom. This conjecture, which is suggested by Rydberg's expressions for wave-numbers, has been confirmed by all subsequent observation and is now known as Ritz's COMBINATION PRINCIPLE. In the case of hydrogen these SPECTRAL TERMS take the simple form:

$$R/n^2,$$

where R is Rydberg's constant and n is a positive whole number.

BOHR'S THEORY OF THE HYDROGEN ATOM

Niels Bohr (1885–), Professor of Theoretical Physics in Copenhagen, like Nicholson before him, assumed a model of the emitting atom which was suggested by Rutherford's general notion of atomic structure. He assumed for hydrogen a massive central nucleus, now called the PROTON, with the elementary positive charge and a single electron travelling round it in a circular orbit (Fig. XVI—1).

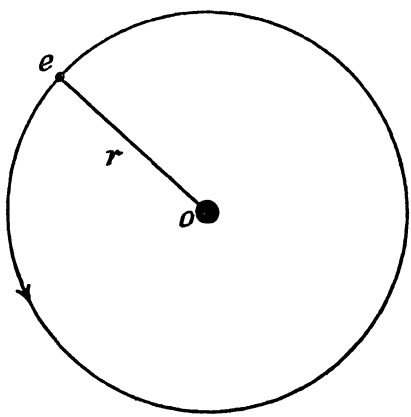


Fig. XVI—1

Before describing this theory of Bohr's, some simple general considerations may be helpful. We shall, for simplicity, confine our attention to systems with one degree of freedom, though the important features which I shall try to illustrate by reference to systems of *one* degree of freedom can be generalized to apply to many systems with several degrees of

* W. Ritz: *Ann. der Phys.*, 25, p. 660 (1908).

freedom.* In the case of simple harmonic motion we have seen in Chapter XIII that

$$E = J \times \nu,$$

E is the energy of the simple harmonic oscillator, ν its frequency and J is the PHASE INTEGRAL, $\int pdq$, equal in this case to the area of the ellipse of Fig. XIII—3. Hence

$$\frac{dE}{dJ} = \nu.$$

This applies generally to conservative periodic systems of one degree of freedom. We are now speaking according to the classical (pre-quantum) theory and according to this the waves emitted by such a system have this frequency, or, it may be, integral multiples of it. So the frequency of the emitted waves is given by

$$\nu = \frac{dE}{dJ} \times s, \qquad \text{(XVI—1)}$$

where s is an integer. Now let us examine the modification which the simplest form of quantum theory introduces. We may write

$$\Delta E = \frac{\Delta E}{\Delta J} \times \Delta J$$

where ΔE is the energy emitted by a system in consequence of a "quantum jump". But since J is always equal (in the older quantum theory) to

$$Integer \times h,$$

so ΔJ, the *change* of J corresponding to ΔE, must be equal to

$$s \times h$$

where s is some integer—positive or negative. Therefore

$$\Delta E = \frac{\Delta E}{\Delta J} \times s \times h.$$

Now Bohr adopted the hypothesis that the energy, ΔE, emitted in the form of light waves in consequence of a quantum jump, must be equal to

$$h\nu,$$

where ν is the frequency in the emitted light wave. So we may write

* *Vide* section on Wilson-Sommerfeld quantum conditions.

$$h\nu = \frac{\Delta E}{\Delta J} \times s \times h,$$

or

$$\nu = \frac{\Delta E}{\Delta J} \times s. \qquad \text{(XVI—1A)}$$

Compare this statement with (XVI—1) and observe the close correspondence between the classical and the older quantum theory. It constitutes the basis of Bohr's CORRESPONDENCE PRINCIPLE to which I shall refer again later.

Now turning to Bohr's model (Fig. XVI—1), the kinetic energy of the electron can be shown, indeed very simply, to be equal to

$$\frac{2\pi^2 me^4}{J^2}$$

and in this case J has a very simple form: it is equal to

$$mv \times 2\pi r,$$

and it is immaterial whether we regard it as momentum times circumference or as $2\pi \times$ angular momentum.* Moreover the *whole* energy is *numerically* equal to the kinetic energy, if we ignore a conventional constant. Therefore we may write for the whole energy:

$$E = \frac{2\pi^2 me^4}{J^2} \qquad \text{(XVI—2)}$$

numerically.

There are two kinds of physical constants: (*a*) those like m and e in (XVI—2) which characterize the parts or the structure of an atom or other physical system, and (*b*) those like energy and momentum which may be changed by contact with external systems. Now it was part of the method of the old quantum theory to express the constants (*b*) in terms of the phase integrals J (of which there are as many as the system has degrees of freedom). Formula (XVI—2), for example, expresses the energy of Bohr's hydrogen atom in this way and its angular momentum is simply $J/2\pi$. Now the quantum theory makes $J = nh$ (n is a positive whole number).† Thus (XVI—2) becomes

* In this calculation I have assumed, as did Bohr in his first paper, for simplicity, that the centre of mass of the system is actually in the nucleus.

† Note that the condition $J = nh$, as applied here, is simply that of Nicholson, namely, angular momentum is equal to $nh/2\pi$.

$$E = \frac{2\pi^2 me^4}{n^2 h^2};$$

therefore E can only change discontinuously, due to the sudden changes of n from one whole number to another (quantum jumps). Now Bohr, as we have seen, adopted the hypothesis that any such discontinuous change of the energy, E, is equal to $h\nu$, where ν is the frequency of the emitted (or absorbed, as the case may be) radiation. Consequently

$$h\nu = \frac{2\pi^2 me^4}{h^2}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

or

$$\nu = \frac{2\pi^2 me^4}{h^3}\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right). \qquad \text{(XVI—3)}$$

We notice that this formula becomes identical with Rydberg's form of Balmer's formula when we replace n_1 by 2—provided of course that $2\pi^2 me^4/h^3$ has a numerical magnitude which enables us to identify it with Rydberg's constant, R. That this is actually the case was one of Bohr's greatest triumphs. When we substitute in it

$$m = 9 \times 10^{-28} \text{ gramme},$$
$$e = 4{\cdot}8 \times 10^{-10} \text{ E.S. units},$$
$$\text{and } h = 6{\cdot}6 \times 10^{-27} \text{ erg} \times \text{sec.},$$

we get a result which is miraculously close to the observational value, that is, approximately,

$$109{,}700 \times 3 \times 10^{10}$$
$$\text{or } 3{\cdot}29 \times 10^{15}.$$

We are dealing now, of course, with frequencies and not with wave-numbers.* The spectral terms, R/n^2, of hydrogen, discovered of course before Bohr's theory was developed, indicated —as does also the formula (XVI—3)—other series of lines which were to be expected and which were indeed observed in the hydrogen spectrum:

Lyman's series, represented by

$$\nu = R\left(\frac{1}{1^2} - \frac{1}{n^2}\right); n = 2, 3, 4, \ldots$$

* N. Bohr: *Phil. Mag.*, 26, pp. 1, 476, 857 (1913).

Paschen's series,

$$\nu = R\left(\frac{1}{3^2} - \frac{1}{n^2}\right); n = 4, 5, 6, \ldots$$

Brackett's series,

$$\nu = R\left(\frac{1}{4^2} - \frac{1}{n^2}\right); n = 5, 6, 7 \ldots$$

There is no end to them, except what is determined by the practical questions of suitable conditions for their excitation and means of observing them.*

THE HELIUM ATOM

Naturally Bohr turned next to helium and assigned to its nucleus a double positive charge, while assuming two planetary electrons. Since the atomic weight of helium is approximately four times that of hydrogen, Bohr assumed its nucleus to consist of four protons and two electrons; the latter have negligible mass and bring the net positive charge down to two units. This view of the helium nucleus has been slightly modified in recent times as we shall see. The case of an ionized helium atom—ionized in such a way that it retains one of its two planetary electrons—is of special interest. The atom is now very like a hydrogen atom and we can calculate the kinetic energy of the planetary electron in the same way as that of the hydrogen electron. The difference lies in the fact that the nuclear charge is now two units, so that the e^4 of (XVI—2) is replaced by $2 \times 2 \times e^4$ or $4 \times e^4$ and the energy of the atom is expressed by

$$E = \frac{2\pi^2 m 2^2 e^4}{J^2}, \qquad \text{(XVI—4)}$$

or

$$E = \frac{2^2 Rh}{n^2},$$

when we replace J by nh and introduce Rydberg's constant. More generally, when the nuclear positive charge (atomic number) is Z times that on a proton, and the atom is ionized to

* Lyman: *Astrophys. Journ.*, 19, p. 667 (1908); Paschen: *Ann. der Phys.*, 50, p. 935 (1916).

such an extent that only one planetary electron remains, then numerically

$$E = \frac{Z^2 Rh}{n^2}. \qquad \text{(XVI—4A)}$$

Thus in the case of the ionized helium, it is obvious that we may expect spectral series resembling those of hydrogen. Thus

$$\nu = R \times 4\left(\frac{1}{n_1^2} - \frac{1}{n_2^2}\right)$$

or, as we may write it,

$$\nu = R\left(\frac{1}{(\frac{n_1}{2})^2} - \frac{1}{(\frac{n_2}{2})^2}\right)$$

The particular case $n_1 = 4$ yields a series some of the lines of which are identical (apparently—a certain very small difference, with an important significance, will be discussed later) with those of Balmer's series. In fact

$$\nu = R\left(\frac{1}{2^2} - \frac{1}{(\frac{n}{2})^2}\right)$$

and when $n = 6, 8, 10, \ldots$ we have in fact the Balmer lines; but there are other lines corresponding to $n = 5, 7, 9, \ldots$. These latter, along with the Balmer lines, were actually observed by Pickering in the spectrum of the star ζ *Puppis* in 1896, and they were all naturally ascribed to hydrogen at that time.*

A. Fowler (1868–1940), a distinguished astronomer and spectroscopist and, in his youth, assistant to Sir Norman Lockyer, pointed out that Rydberg's constant had a slightly different value for different elements. This† resulted in another great triumph for Bohr. It will be remembered that, as indicated in Fig. XVI—1, the mass of the nucleus has been assumed to be so enormous by comparison with that of the electron that it can be regarded as remaining at rest in the centre of the circular orbit of the electron. Actually it rotates in a very small circular orbit and Rydberg's constant, as Bohr clearly appreciated, requires a correction.

The value we have ascribed to it should be multiplied by

$$M/(M + m),$$

* They were observed to have the same limit as the Balmer lines of hydrogen.

† A. Fowler: Bakerian Lecture, 1914.

where M and m are, respectively, the masses of the nucleus and the electron. When the observational values R_{hydrogen} and R_{helium} are compared, it is easy to get the ratio of the masses of the proton and electron. The spectroscopic values of the R's yield about 1845 for this ratio, which is quite near to what is given by the e/m measurements for the electron and for a hydrogen ion.

X-RAY SPECTRA AND ATOMIC STRUCTURE

In 1912 Max von Laue (1873–) suggested that the fine grained regularities in a crystal might serve for X-rays as an ordinary grating does for light. It seemed clear that if X-rays were an undulatory phenomenon like light, the wave-lengths involved would be very short and would require a regular structure whose spacings were much more minute than any artificially constructed grating could provide. His suggestion was confirmed experimentally by his pupils Friedrich and Knipping, and was followed up by Sir William H. Bragg (1862–1942) and his son, W. L. Bragg (now Sir Lawrence Bragg (1890–), who constructed the earliest X-ray spectrometer. At first they used an ionization chamber for detecting diffracted X-ray beams, and X-ray spectral "lines" were indicated by sharp increases in the ionization current. It is rather remarkable that the elder Bragg, who had spent several years over experiments which seemed to demonstrate the corpuscular character of X-rays, should be the first to measure the wave-length of an X-ray beam. But Sir William Bragg turned his later attention rather to the use of X-rays for the investigation of crystal structure than to the examination of X-ray spectra, or their use for elucidating atomic structure.

It was naturally expected that the examination of Barkla's K, L, M, etc., radiations would throw light on atomic structure and this kind of investigation was first undertaken by H. G. J. Moseley, who produced the earliest photograph, I believe, of an X-ray spectrum which exhibited *lines* and looked somewhat like the spectra that the old spectroscopes produced with visible light.* He began with the K radiation of a number of different

* Moseley: *Phil. Mag.*, 26, p. 1024 (1913); 27, p. 703 (1914). Moseley lost his life in the Dardanelles campaign during the first world war.

elements and found that the first line, called K_α, of a large number of elements obeyed the following formula:

$$\nu = \tfrac{3}{4}R(Z - s)^2,$$

ν being the frequency, i.e. c divided by the measured wavelength, Z is in fact the atomic number of the element, and s a small number near unity. It is easy to interpret this formula. If we make $Z = 1$ and ignore the s, it is identical with the first line of Lyman's series for hydrogen. In fact the first line of Lyman's series is expressed by

$$\nu = R\left(\frac{1}{1^2} - \frac{1}{2^2}\right)$$
$$= \tfrac{3}{4}R.$$

The lines of Lyman's series are due to electron jumps from orbits with quantum numbers 2, 3, 4, etc., to the innermost orbit of quantum number 1. Barkla's L radiation is like Balmer's series; the first line is due to a jump from $n = 3$ to $n = 2$. The differences are due to the fact that $Z = 1$ for hydrogen, while for copper, for example, $Z = 29$; and of course, in the case of hydrogen, $s = 0$. Turning again to the example of copper, its first K line, or K_α as it is labelled, is due to the fact that an electron in an orbit for which $n = 1$ has been removed by some kind of violence—e.g. a primary beam of X-radiation—and in consequence an electron jumps from an orbit for which $n = 2$ into the $n = 1$ orbit. All the rest of the electrons have hardly any influence on this happening. They resemble the charge on the outside of a sphere which produces no field in the interior. The jumping electron is affected only by the charge Ze on the nucleus, or, to be exact, it is only very slightly influenced by the 27 remaining electrons of the copper atom. It is this *slight* influence which the SCREENING CONSTANT s represents.

THE IMPACT ON BOHR'S THEORY OF THE WILSON–SOMMERFELD QUANTUM CONDITIONS

In the simple case of an electron travelling round a nucleus the orbit is in general elliptical and there are two phase integrals, J_1 and J_2; the former is the product (or the sum of products):

(Radial momentum) × (Increment of radius vector),

the latter

(Angular momentum) × (Increment of angle),

and (XVI—2) is just a particular case of the more general

$$E = \frac{2\pi^2 m Z^2 e^4}{(J_1 + J_2)^2}.$$

The eccentricity of an orbit—let us call it ϵ—conforms to the rule

$$1 - \epsilon^2 = J_2^2/(J_1 + J_2)^2$$

and, as Sommerfeld and the present writer showed,

$$1 - \epsilon^2 = n_2^2/(n_1 + n_2)^2, \qquad \text{(XVI—5)}$$

since J_1 and J_2 are replaced by $n_1 h$ and $n_2 h$.* We might describe an orbit for which the two quantum numbers are n_1 and n_2 as an n_1, n_2 orbit. Bohr preferred to label such an orbit in rather a different way: he used the letter n for the sum $n_1 + n_2$ and the letter k for n_2 and called the orbit an n_k orbit. Thus an orbit which was originally described as a 3, 2 orbit—radial quantum number $n_1 = 3$, and angular quantum number $n_2 = 2$—Bohr described as a 5_2 orbit. It is a convenient mode of description, since the energy is practically determined by the sum $n_1 + n_2$ and not very much influenced by the individual values of n_1 and n_2; indeed, in the case where there is only one electron orbit (hydrogen) the energy, as we have seen, is fixed exactly by the sum $n_1 + n_2$.

GENERAL DESCRIPTION OF ATOMIC STRUCTURE

Perhaps the best way to describe the general scheme of atomic structure, the discovery of which is mainly due to Bohr, is to begin by studying a few of the atoms in the order of their atomic numbers. An undisturbed hydrogen atom has a single 1_1 orbit, i.e. $n_1 = 0$ and $n_2 = 1$. Its nucleus is the proton, a particle with a single unit of positive charge. A helium atom has a nucleus approximately four times as massive as the proton; but a net positive

* W. Wilson: *Phil. Mag.*, 31, p. 156 (1916).

charge of two units. This has been established not only by the concordance of the lines in the helium spectrum with what Bohr's theory predicted, but also by direct experiment. Rutherford and his pupil Geiger counted the α particles (Sir William Ramsay and F. Soddy demonstrated that they are nuclei of helium atoms) passing through a small aperture into an ionization chamber and measured the total charge they carried. Their apparatus was the first GEIGER COUNTER. To account for the mass 4 and the charge 2 it was formerly thought that the helium nucleus was made of 4 protons and 2 electrons. The subsequent development of quantum mechanics made it difficult to understand how an electron could exist in a nucleus, and when another particle, almost exactly like a proton, but with no charge, the NEUTRON, was discovered by Sir James Chadwick in 1932, it seemed obvious that the helium nucleus must consist of 2 protons and 2 neutrons. Round this are two 1_1 electron orbits in the undisturbed or unexcited helium atom. The next element in order is lithium, atomic number 3. Its nucleus has 3 units of positive charge and it has 3 electrons in orbits, two of which are like the helium orbits, i.e. they are 1_1 orbits, while the third orbit is a 2_1 orbit. Helium, first noticed as present in the sun, was discovered as a terrestrial element by Sir William Ramsay and Sir William Crookes in the mineral cleveite. It is a very inert gas and does not take part, so far as I know, in any chemical reaction. The significance of this, in terms of our atomic theory, is that the two 1_1 orbits are an exceedingly stable structure. Exceedingly stable atomic structures, comparable with that of helium, are repeated among the atoms of elements of higher atomic number, the next one being the inert gas *neon*. This element is followed by sodium, which like lithium has an electron orbit, in this case a 3_1 orbit, outside the stable structure of ten electron orbits which characterize the inert gas neon, and so we get some insight into the significance of the fact that the two elements, lithium and sodium, are extraordinarily similar. As we proceed further we find another stable configuration of electron orbits characterizing the atom of the inert gas *argon* and the following element, potassium, has an electron orbit, a 4_1 orbit, thus resembling lithium and sodium.

Originally, when it was still not known how the various electron orbits were distributed in the atoms Bohr made the tentative

supposition that the two helium orbits were followed by four 2_1 orbits and four 2_2 orbits, whereas facts discovered later required, as shown in the table, two 2_1 orbits and six 2_2 in the completed system of L orbits. This was first suggested by E. C. Stonor, now Professor of Theoretical Physics in the University of Leeds, mainly to account for the peculiarities of X-ray phenomena, and independently by the American Main Smith, on chemical grounds. It is not possible to explain fully here just why the eight L orbits should be made up of two 2_1 orbits and six 2_2 orbits, but part of the explanation may be indicated. It is based on a principle due to W. Pauli,* according to which only one electron orbit in the atom can have the same set of quantum numbers. Associated with this is a notion called the SPIN of the electron. The possibility that an electron is a rotating structure has been suggested by many of the investigators of atomic problems in this century, by Voigt, Max Abraham, Ritz and Nicholson. It is also of special interest to-day that H. S. Allen, Professor of Natural Philosophy at the University of St. Andrews, made the suggestion in 1914, in a discussion of atomic structure at a meeting of the Royal Society, that the core of the atom might be a rotating sphere of positive electricity, or possibly the electron itself might be a rotating sphere of negative electricity.† The suggestion in regard to the electron had to be taken seriously when the Dutch physicists, Uhlenbeck and Goudsmit, showed that it appeared to resolve difficulties connected with the fine structure of spectral lines and promised to solve the problem of the anomalous Zeeman phenomenon.‡

Broadly speaking, the need for assuming a spinning electron is that it may have the magnetic moment $eh/4\pi mc$ (e expressed in E.S. units). With this assumption an electron orbit has *four* quantum numbers associated with it:

$$n,\ l,\ m,\ \text{and}\ s,$$

where n is Bohr's total quantum number, $l = k - 1$, m is the component of l in the direction of s, and s measures the spin

* Pauli: *Zeits. für Physik.*, 31, p. 765 (1925).

† H. S. Allen: *Nature*, 92, p. 713 (1914); *Phil. Mag.*, 29, p. 714 (1915).

‡Uhlenbeck and Goudsmit: *Naturw.*, 13, p. 953 (1925); *Nature*, 117, p. 264 (1926); *Zeits. für Physik.*, 35, p. 618 (1926).

momentum (spin momentum $= sh/2\pi$). Strangely enough s has the values $+\frac{1}{2}$ or $-\frac{1}{2}$.

Let us combine this with Pauli's principle—called the EXCLUSION PRINCIPLE. We ask how many 1_1 orbits are possible. In this case we have

n	l	m	s
1	0	0	$+\frac{1}{2}$ or $-\frac{1}{2}$.

There are only two of these orbits therefore, since the exclusion principle requires that only *one* electron orbit may be associated with the same set of numbers. Now consider the 2_1 and 2_2 orbits. For the former we have the same result again, namely two orbits; but with the 2_2 orbits $k = 2$, equivalent to $l = 1$. The number m is necessarily an integer and may therefore have the values $+1, 0, -1$ only, and combined with each of these is $s = +\frac{1}{2}$ or $s = -\frac{1}{2}$. There are therefore 3×2 possibilities in all. When we ask about the orbits for which $n = 3$, we have as before *two* 3_1 orbits and *six* 3_2 orbits, but in the case of the 3_3 orbits, $k = 3$, therefore $l = 2$, and we have for m the possibilities $m = +2, +1, 0, -1, -2$, five of them, and each of the five may be associated with $s = +\frac{1}{2}$ or $s = -\frac{1}{2}$: *ten* possibilities.

ELECTRON ORBITS OF ATOMS

Taking all the orbits for which $n = 3$, the complete set is made up of two 3_1 orbits, six 3_2 orbits and ten 3_3 orbits. The accompanying table gives the electron configuration of the first thirty-six elements. The numbers in the vertical columns represent the number of orbits in the atom of the specified type. It will be noticed that sometimes, as happens in the case of potassium, atomic number 19, that the conceivably possible electron orbits are not occupied in succession as we follow the atomic numbers. Thus the electron which is distinctive of potassium is not in a 3_3 orbit, but in a 4_1 orbit. The 3_3 orbits do not begin to get filled up till we come to scandium (21). Evidently the potassium configuration, namely two 1_1, two 2_1, six 2_2, two 3_1, six 3_2 and one 4_1, is more stable than one with the last orbit replaced by a 3_3 one. In fact there is no atom with the configuration two 1_1, two 2_1, six 2_2, two 3_1, six 3_2, one 3_3. Similarly when we pass beyond the inert gas *krypton*, in whose atom the 4_1 and 4_2 orbits

Element and Atomic Number			K	L		M			N	
			1_1	2_1	2_2	3_1	3_2	3_3	4_1	4_2
H	1	..	1							
He	2	..	2							
Li	3	..	2	1						
Be	4	..	2	2						
B	5	..	2	2	1					
C	6	..	2	2	2					
N	7	..	2	2	3					
O	8	..	2	2	4					
F	9	..	2	2	5					
Ne	10	..	2	2	6					
Na	11	..	2	2	6	1				
Mg	12	..	2	2	6	2				
Al	13	..	2	2	6	2	1			
Si	14	..	2	2	6	2	2			
P	15	..	2	2	6	2	3			
S	16	..	2	2	6	2	4			
Cl	17	..	2	2	6	2	5			
A	18	..	2	2	6	2	6			
K	19	..	2	2	6	2	6		1	
Ca	20	..	2	2	6	2	6		2	
Sc	21	..	2	2	6	2	6	1	2	
Ti	22	..	2	2	6	2	6	2	2	
V	23	..	2	2	6	2	6	3	2	
Cr	24	..	2	2	6	2	6	5	1	
Mn	25	..	2	2	6	2	6	5	2	
Fe	26	..	2	2	6	2	6	6	2	
Co	27	..	2	2	6	2	6	7	2	
Ni	28	..	2	2	6	2	6	8	2	
Cu	29	..	2	2	6	2	6	10	1	
Zn	30	..	2	2	6	2	6	10	2	
Ga	31	..	2	2	6	2	6	10	2	1
Ge	32	..	2	2	6	2	6	10	2	2
As	33	..	2	2	6	2	6	10	2	3
Se	34	..	2	2	6	2	6	10	2	4
Br	35	..	2	2	6	2	6	10	2	5
Kr	36	..	2	2	6	2	6	10	2	6

are completed, to rubidium, this latter has its outermost electron in a 5_1 orbit and, as we can understand, its properties strongly resemble those of lithium, sodium and potassium.*

* Much more detailed information will be found in Professor Max Born's *Atomic Physics* (Blackie & Son). He uses a different convention in describing electron orbits, preferring l to Bohr's k, so that, for example, an orbit which I have described by 3_2 (as Bohr did), Born describes by 3, 1.

CHAPTER SEVENTEEN

THE CORRESPONDENCE PRINCIPLE AND ITS APPLICATIONS

SPECTRAL SERIES AND SPECTRAL TERMS

SPECTRAL series were first observed almost a century ago and also series of distinctive types—sharp series, principal series, diffuse series, and so on. Some kind of order began to be recognized in them when Ritz published his COMBINATION PRINCIPLE, namely that the associated frequencies, or wave-numbers, are differences of terms which are characteristic of the emitting atom or molecule. On the other hand, if we take any two spectral terms of an element and subtract one from the other, the result *does not always* represent the wave-number (or frequency) of an observed spectral line and certain SELECTION RULES were developed which indicated what pairs of spectral terms yielded observed frequencies, and the atomic terms were classified as *s*, *p*, *d*, etc., terms. One of Bohr's most important discoveries was that these various types of term were to be associated with specific values of the quantum number k—later replaced by $l + 1$. For *s* terms $k = 1$ ($l = 0$), for *p* terms $k = 2$, and so on. Differences between two different *s* terms, or between two different *p* terms, do not for example represent observed wave-numbers. Bohr introduced order into these perplexing things by his CORRESPONDENCE PRINCIPLE. It is based on the equations (XVI—1) and (XVI—1A) and their generalization. Each frequency as it would be given by the classical theory can be associated with a similar quantum formula. Now one weakness of the old quantum theory lay in the fact that it gave no indications about the possible values of the whole numbers, s in (XVI—1A), and Δk, the change in k during a quantum jump, is perhaps the most important of these; whereas the classical theory is quite definite about them. It requires that Δk *must be* $+ 1$ or $- 1$. Now Bohr adopted in the quantum theory (where this proved insufficient) whatever the classical theory laid down for the *corresponding* frequency. He

was thus led to suppose that s terms are associated with orbits for which $k = 1$, p terms with orbits for which $k = 2$, d terms with orbits for which $k = 3$, and so on, and the correspondence principle allowed only those quantum transitions for which, as we have seen, $\Delta k = +1$ or -1, for instance from a p orbit to an s orbit.

THE STARK EFFECT

When emitting hydrogen atoms are subjected to a strong electric field each spectral line of the ordinary spectrum appears split up into a number of lines which are symmetrically situated with respect to the position of the original line. This was discovered by Johannes Stark (1874–) and independently by Lo Surdo.* Its explanation was one of the great triumphs of the quantum theory. It will be remembered that the general method of the old quantum theory was to proceed according to classical principles, using suitable co-ordinates, and express the energy in terms of the phase integrals J, of which in this case there are three. Finally each J is replaced by the product of an integer and h. The mathematical problem involved is a particular case of the problem of a particle attracted to two fixed points (under the inverse square law of force) and was solved by the great Königsberger mathematician C. G. J. Jacobi more than a century ago.† The particular case is that in which one of the fixed attracting centres is at infinity. P. Epstein (1873–) successfully calculated the frequencies by expressing the energy of the system atom plus external field in terms of the J's of (XIII—10A) and substituting for each of them the product

$$\textit{Integer} \times h.$$

Karl Schwarzschild (1873–1916) also achieved the same result as Epstein independently.‡ His solution of the problem has a certain distinction: he showed that the J's may be made to

* Stark: *Ber. Akad.*, 40, p. 932 (1913); Lo Surdo: *Accad. Lincei*, 22, p. 664 (1913), 32, p. 82 (1914).

† Jacobi: *Vorlesungen über Dynamik*, (Berlin Georg Reimer). It will be found on page 221 of the edition of 1866.

‡ K. Schwarzschild: *Ber. Akad.* (1916); P. Epstein: *Ann. der Phys.*, 50, p. 498 (1916).

function as p's and that instead of the q's a simpler type of variables, called ANGLE VARIABLES, can be used.*

Kramers has shown that the intensities of the components into which a spectral line is resolved in the Stark effect are well given by the correspondence principle while their state of polarization is exactly predicted by it.

THE ZEEMAN EFFECT

Faraday (1791–1867) was aware of the existence of some relationship between light and electromagnetism and tried to observe (about 1862) a conjectured influence of a magnetic field on the spectral lines of a source of light. He was unsuccessful; no doubt because his magnetic field was not strong enough and the resolving power of his spectroscope too low; but such an effect was observed in 1896 by P. Zeeman (1865–1943), a pupil of the great Lorentz.† Zeeman was able to observe a broadening of the D lines of sodium from a sodium flame in a strong magnetic field. Lorentz then worked out a theory of the phenomenon—indeed he possibly did this before Zeeman began his experiments—which predicted the details of the effect quite accurately in the simplest case. He imagined the light forming the spectral line to be due to a simple harmonically vibrating ion and calculated that it would split up, in a magnetic field, into *three* lines, all three being visible to an observer receiving the light along a direction perpendicular to that of the field, but only the two outer lines when the observer is looking along the line of the field. He also calculated correctly the frequency differences and the state of polarization of the three components. The middle line of the three, which can only be seen transversely, is in the position of the original line and polarized as if the vibrations of the ion were along the line of the field. The remaining two were polarized *as if* the vibrations were perpendicular to the field, but circularly polarized when seen longitudinally. Zeeman confirmed this by observing that the broadened D lines were circularly polarized at their edges. Assuming that the broadened line really consisted of two (or three, according to the direction along which the light

* *Vide* W. Wilson, *Theoretical Physics*, Vol. iii (1940), pp. 177, 178.

† Zeeman: *Amsterdam Akad.*, 5, pp. 181, 242 (1896), and *Phil. Mag.*, 43, p. 226 (1897).

was received) he estimated the value of the ratio e/m for the vibrating ion, using Lorentz's formula, and found it to be about $1{\cdot}6 \times 10^7$ E.M. units and *negative*. Clearly it was an electron. Lorentz found that each of the two outer components differed in frequency from that of the original line by the amount $eH/4\pi mc$, where H is the intensity of the magnetic field, e is the charge on the ion (electron) in E.S. units, m is its mass and c as usual the speed of light in free space.

Sir Joseph Larmor, after Lorentz, was also able to account for the Zeeman effect, though, as later investigations showed, it was only the simplest form of Zeeman effect which they were able to explain. At that time (in the late nineties) nobody seems to have contemplated that it might be necessary to modify the classical principles on which they based their work. Let us for a moment think of an electron revolving round a nucleus in an elliptical orbit and imagine a straight line through the nucleus perpendicular to the plane of the orbit. So long as there is no other force but the attraction between the nucleus and the electron the position of this line does not vary. When however we apply a magnetic field the electron orbit precesses round a straight line through the nucleus in the direction of the magnetic field. The direction of the former straight line always makes the same constant angle with that of the latter, but precesses round it with the angular velocity $eH/2mc$; where H means the intensity of the magnetic field, measured in the usual units, and e is the electronic charge in E.S. units. This angular velocity is called the LARMOR PRECESSION. When divided by 2π an angular velocity always gives a *frequency*, so we expect from this a change in frequency, due to the magnetic field, equal to

$$eH/2mc \div 2\pi$$

which is exactly what Lorentz found.

In applying the quantum theory to the problem we note that —though it is not possible to give the full reasons here—the classical expression for the frequency has not the form corresponding to (XVI—1), but very happily when we introduce a more general momentum, which I represent by the symbol Π, and which I defined* as the vector sum

* W. Wilson: *Proc. Roy. Soc. A.*, 102, p. 478 (1922).

(Ordinary momentum) + (Product of charge on electron and electromagnetic vector potential)

or

$$\Pi = p + eK, \qquad \text{(XVII—1)}$$

and define a new J by

$$J = \int \Pi dq, \qquad \text{(XVII—1A)}$$

then the old quantum rules can be applied and it turns out that the particular s that occurs in that J which is associated with the Larmor precession must have one or other of the three values $+1$, 0, -1 in virtue of the correspondence principle. The quantum theory when applied in this way gives precisely the same result as does the classical theory.

The vector Π is prominent in Dirac's form of electron theory. It is also remarkable that it gives the equations of motion of a charged particle, in an electromagnetic field, the form of a GEODESIC and thus simplifies the expression of Kaluza's form of relativity and indeed suggests it.*

The more complex forms of the Zeeman effect are now accounted for in terms of the rather strange electron spin which we have encountered in the description of the electron orbits of atoms. Paschen and Back have shown that in sufficiently intense magnetic fields the Zeeman effect approximates to the simple phenomenon which the classical theory of Lorentz and Larmor is able to account for.† This is known as the PASCHEN–BACK EFFECT.

* W. Wilson: *Proc. Roy. Soc. A.*, 118, p. 441 (1928).
† Paschen and Back: *Ann. der Phys.*, 39, p. 897 (1912).

CHAPTER EIGHTEEN

QUANTUM MECHANICS

SOME OBSERVATIONS ON PLANCK'S THEORY

THE quantum theory, as it appeared in 1900 and as it continued to be till about 1923, might be described, at least approximately, as the mechanics of Hamilton with which certain liberties had been taken; but while Hamilton's mechanics was logically coherent and complete, the modified mechanics was far from having this character. The former rests on a principle, known as HAMILTON'S PRINCIPLE, which is a generalization of the still older PRINCIPLE OF LEAST ACTION of Moreau de Maupertuis (1698–1759). The thing called ACTION is the most fundamental thing in mechanics, old or new, and Planck's theory may be said to have given it an atomic structure, the atoms of action having the magnitude h.

THE NEW MECHANICS

A coherent new mechanics began to develop about 1923 from the initial efforts of Werner Heisenberg in Germany and Prince Louis de Broglie (1892–) in France. The outward aspect of the new mechanics, as presented by Heisenberg on the one hand and by de Broglie and Schrödinger on the other, was so different that a new problem was raised: the problem of accounting for the strange fact that two such different mathematical schemes should lead to such similar, indeed identical, consequences. This problem was solved by Schrödinger, who exposed the intimate inner relationship between the two forms of mechanics.* He compared this relationship very aptly to that between electrostatics in its Faraday–Maxwell field form on the one hand and the older action at a distance form on the other. In the former the energy is regarded as continuously distributed through the dielectric medium; in the latter it is associated discontinuously, as it were,

* E. Schrödinger: *Ann. der Phys.*, 79, p. 734 (1926).

with the individual conductors. But both lead to identical consequences. Schrödinger's theory resembles the Faraday–Maxwell form of electrostatics and Heisenberg's the other form. The latter theory was named MATRIX MECHANICS, because the algebraical things called MATRICES were so conspicuous in it, and the former, with the study of which it is preferable to begin, WAVE MECHANICS.

It was the fact that light behaves sometimes (photo-electric and Compton effects) as if it were corpuscular and sometimes (in diffraction and interference phenomena) as if it were undulatory in its nature, that led de Broglie to suspect that a beam of light consisted of both waves and corpuscles (photons), the energy of the beam being carried by the corpuscles, while the wave itself, de Broglie thought, in some way or other guided the photons. But he went further and made the brilliant suggestion which is best given in his own words: "But if for a century we have neglected too much the corpuscular aspect in the theory of light, in our exclusive attachment to waves, have we not erred in the opposite direction in the theory of matter? Have we not wrongly neglected the point of view of waves and thought only of corpuscles?"* He thus predicted what was later observed by Davisson and Germer, G. P. Thomson, Rupp and others, namely, the extraordinary phenomenon of the diffraction of electrons.

STATIONARY PRINCIPLES

In order to get some insight into the nature of the wave mechanics which we owe to the brilliant work of de Broglie and Schrödinger it is very helpful, if not essential, to understand the principle of the STATIONARY PATH in geometrical optics and those of ACTION in classical mechanics and the remarkable parallelism, or analogy, between them. So we go back to the seventeenth century and look carefully at the principle of P. Fermat (1601–1665), returning to the present century after halting for a short while in the middle of the eighteenth century to examine de Maupertuis' principle and in the early nineteenth century to examine that of Hamilton. These principles have all of them one thing in common: they assign a stationary value to something. I

* L. de Broglie: *Wave Mechanics*, p. 3 (translated by H. T. Flint (Methuen)).

have tried to indicate the meaning of "stationary" in the chapter on general relativity. Usually, but not always, it means greatest or least, i.e. maximum or minimum. Fermat's principle assigns a stationary value to the length of an optical path, *assuming this to be measured by using the length of the wave in the medium as a unit*; so that, when the ray traverses two or more media, the unit will usually be different in different media. In geometrical optics with which Fermat's principle is concerned, we have to do with a certain limiting case of the simple groups described in the section on characteristics of waves in Chapter VII. Such a group is practically monochromatic and the wave-length in it is very small by comparison with the dimensions of the group. Now the limiting case of geometrical optics is that in which a portion of a beam of light, though so small as to be practically a point, is still a simple group and the path it traces is a ray of light. The wave-length is a *second order small quantity* and diffraction patterns are too small to be observable. The length of the path (in our special sense) traced by such a punctual group, in travelling from some point A to another B, is equal to

$$\nu' \times q,$$

where ν' is the wave number, i e. the number of waves in one centimetre, and q is the length of the path *in centimetres*. Of course when ν' varies from point to point we multiply each short bit, Δq or dq, of the path by the appropriate value of ν' and add up the products to get

$$\nu_1' dq_1 + \nu_2' dq_2 + \nu'_3 dq_3 + \ldots,$$

which is usually written

$$\int_A^B \nu' dq. \qquad \text{(XVIII—1)}$$

To get the stationary value of this sum we inquire how it changes as we pass from a path to a neighbouring one. When it is stationary the neighbouring paths have the same sum. This condition of being stationary is usually shortly expressed by

$$\delta \int_A^B \nu' dq = 0. \qquad \text{(XVIII–1A)}$$

The δ is a symbol which is often used to express the difference between two neighbouring things, and STATIONARINESS is determined by the vanishing of such a difference.

Now turning to the case of the motion of a particle, de Maupertuis laid down the principle which he called the PRINCIPLE OF THE LEAST QUANTITY OF ACTION (principle of least action). If p means the momentum of the particle, i.e. $m \times v$, the principle may be stated in the form:

$$\delta \int_A^B p dq = 0. \qquad \text{(XVIII—2)}$$

I am assuming for simplicity that the position of the particle is given by one co-ordinate, q, only. It can of course be stated in such a form as to apply to more complicated systems. We notice that ν' in Fermat's principle and p in de Maupertuis' are analogous. But there is something more: in applying Fermat's principle we assume the constancy of the *frequency*, ν. Similarly in applying de Maupertuis' principle we have to assume the constancy of the *energy*, E, of the system. So ν in geometrical optics is analogous to E in mechanics.

The following passage from de Maupertuis' *Essay de Cosmologie* (1751) is of great interest: *Notre principe . . . laisse le monde dans le besoin continuel de la puissance du Créatur, & est une suite nécessaire de l'emploi le plus sage de cette puissance.* He was clearly aware of its great importance. It may correctly be described as a special case of a much wider principle which embraces those of Fermat and Hamilton and includes within its scope the geodesics of Einstein's gravitational theory. The Germans have claimed that it was known much earlier to Gottfried Wilhelm Leibniz.

Sir William Hamilton's (1805–1865) generalization of it in the earlier half of last century is really a remarkable prevision of Minkowski's form of relativity as well as of wave mechanics. Our study of Minkowski's space-time continuum suggests that we should add to

$$pdq$$

the additional term

$$p_w dw,$$

which is the product of the W component of momentum, p_w, and dw. Now

$$p_w = m\frac{dw}{dt},$$

so that $p_w = mic$, because $dw = icdt$. Thus $p_w\, dw = mic \times icdt$

or $p_w\, dw = -mc^2 dt$. Now mc^2 is the energy of the particle we are dealing with, Therefore

$$pdq + p_w\, dw = pdq - Edt$$

and the statement

$$\delta\int_A^B (pdq - Edt) = 0^* \qquad \text{(XVIII—3)}$$

is HAMILTON'S PRINCIPLE. The important condition attaches to it that when we compare neighbouring paths they must start simultaneously from A and end simultaneously at B. Hamilton derived his form of mechanics by consciously using the analogy between geometrical optics and mechanics and we may note here that the statement of Fermat's principle may be widened as follows:

$$\delta\int_A^B (\nu' dq - \nu dt) = 0. \qquad \text{(XVIII—3A)}$$

This is quite analogous to Hamilton's principle. The mechanical equations which Hamilton deduced from his principle (the CANONICAL EQUATIONS) include among others

$$Velocity = \frac{dq}{dt} = \frac{\partial E}{\partial p}, \qquad \text{(XVIII—4)}$$

which we can verify by a simple example A body moving along a straight line under no forces has the (kinetic) energy

$$E = \tfrac{1}{2}\, mass \times (velocity)^2,$$

or
$$E = \tfrac{1}{2}\frac{(mv)^2}{m},$$

i.e.
$$E = \tfrac{1}{2}\frac{p^2}{m}$$

and
$$\frac{\partial E}{\partial p} = \frac{p}{m} = \frac{mv}{m} = v,$$

* Conservatism is, perhaps fortunately, deeply entrenched in human nature and Hamilton's principle is still written in the form

$$\delta\int_A^B (T - V)dt = 0,$$

though (XVIII—3) is much more illuminating.

in accordance with (XVIII—4). A strictly analogous equation can be obtained from (XVIII—3A) namely,

$$Velocity = \frac{dq}{dt} = \frac{\partial \nu}{\partial \nu'}. \qquad \text{(XVIII—4A)}$$

This is obviously the group velocity of the small point group which traces out a ray of light.* Perhaps it is needless to remark here that, instead of speaking of an analogy between mechanics and geometrical optics, we may alternatively speak of an analogy between mechanics and any wave motion of *very short* wave-length (or very big wave-number).

CONCISE EXPRESSION OF THE ANALOGY

If in a geometrical optical problem we multiply both wave-number, ν', and frequency, ν, by the same constant, which we may call h, and make the identifications

$$p = h\nu',$$
$$E = h\nu,$$

we have a problem in mechanics (that is the old mechanics of Hamilton). *It does not matter what precise value is chosen for h.* The reason for this is that in a problem in geometrical optics—in the sense in which I have been using the term—the precise value of the frequency, ν, or of the wave-number, ν' $(= 1/\lambda)$, is of no account. Indeed as we have seen, diffraction patterns are too small to be observed and wave-lengths or wave-numbers cannot be known.

WAVE MECHANICS

When wave-numbers and frequencies are not big enough or, to put it in another way, when wave-lengths and periods are no longer small, Fermat's principle and the simple laws of geometrical optics fail us. We now have recourse to the differential equation which describes a light wave, and its solution solves our optical problem. As we have seen, the laws of classical mechanics too have turned out to be inadequate for *small systems* (systems in which the momenta, p, and the energy, E,

* *Vide* "Origin and Nature of Wave Mechanics," *Science Progress*, 32, p. 209 (1937).

are small), and the analogy or parallelism just described suggests strongly that we should proceed as we do with optics when wave-numbers and frequencies are small. The suggestion is that we widen the scope of the analogy so that it becomes an analogy between mechanics and optics in the widest sense of optics. It is thus that WAVE MECHANICS arises.

DE BROGLIE WAVES

In the case of a beam of electrons, for example, all travelling in the same direction with the same velocity, which, as we have seen, can be measured, there is a perfectly definite wave-length which can be measured by using a crystal to perform the function of the grating used to measure optical wave-lengths. The wave-length being known, its reciprocal ν', the wave number, is also known, and in the relation

$$Momentum = mv = h\nu',$$

both m and v can also be determined. Thus a definite value can be assigned to h. It turns out to be identical with Planck's h. These waves have a certain phase velocity, which I call u. De Broglie found the relationship between it and the associated group velocity, which may be called v. We may derive it in the following way:

$$Phase\ velocity,\ u = \nu/\nu'$$

and our Hamiltonian analogy requires that

$$(Energy)/(Momentum) = h\nu/h\nu'.$$

So when we substitute Einstein's expression ($mass \times c^2$) for energy, we find

$$(mc^2)/(mv) = \nu/\nu' = u.$$

Thus

$$u = c^2/v$$

or

$$uv = c^2 \qquad \text{(XVIII—5)}$$

This is de Broglie's relation between the velocity of a particle and the phase velocity of the associated wave. Such waves are now known as DE BROGLIE WAVES. The velocity of the electron must of course be identified with the GROUP VELOCITY* of the

* L. de Broglie: *Ann. de Physique*, 3, p. 22 (1925), Thèses, Paris (1924).

simple group which we associate with it, since it is equal to

$$\partial E/\partial p = \partial(h\nu)/\partial(h\nu') = \partial\nu/\partial\nu'.$$

So far we have thought of a particle, e.g. an electron, as in some way associated with a simple wave group. It is travelling with the same speed, v, as the group and we must therefore think of it as being at some point in the group and travelling with it. We are dealing now with a state of affairs in which we know the velocity of the electron; but not where it is situated, except of course that it is somewhere within the limits of the group and even the outline or boundary of the group is not sharply defined. If we think of this group approaching perpendicularly to a transmission grating (for example) it will be partly reflected and partly transmitted along directions given by (VII—5). In fact it will now have split up into several groups, and all we know, so far as the position of the electron is concerned, is that it is in *one* of these groups. It may be in that one proceeding in the direction θ of (VII—5) and we may represent the distance $n\lambda$ by q, so that

$$q = n\lambda$$

or

$$\nu' q = n.$$

If we multiply both sides of this equation by h we get

$$h\nu' \times q = nh,$$

i.e.

$$p \times q = nh.$$

This is a simple case of the familiar quantum condition of the old quantum theory, i.e. a special case of $J = nh$ (XIII—10 and XIII—10A).

We may of course carry out some experiment to locate the particle, or electron, at some instant, with great precision. When this has been done we are forced to regard the wave group in which it is situated as very small—almost a point. But a very little reflection indicates that it cannot now be a *simple* group. We have in mind a particle of very small energy and momentum and consequently the frequency and wave-number (or frequencies and wave-numbers) associated with it are very small and a simple group which is almost a point is necessarily characterized by an enormously big wave-number. In fact the very small group (sometimes called a wave-packet) representing a very precisely located particle is a superposition of an immense range of wavelengths, i.e. an immense range of wave-numbers associated with

a great range of frequencies. We cannot now assign a unique momentum or energy to the particle. This rather cumbersome argument has led us to two things. It appears that when we know the momentum (or velocity) of the particle very accurately, we cannot know its position very accurately, and conversely, when we know its position very accurately, we cannot have accurate information about its velocity. The other thing is that the wave associated with a particle seems to be determined by the experiments that have been carried out on the particle.

Turning back to the simple formula (VII—3A) and replacing the length L by Δq, we have

$$\Delta \nu' \times \Delta q \sim 1$$

and Δq (or L) means the distance across the group in the direction in which it is travelling. In other words it means the range of uncertainty of this co-ordinate of the particle the group represents, and $\Delta \nu'$ is the associated uncertainty in the wave-number, which lies between ν' and $\nu' + \Delta \nu'$. Multiplying both sides of the last equation by h, we have

$$\Delta p \times \Delta q \sim h. \qquad \text{(XVIII—6)}$$

This is Heisenberg's UNCERTAINTY RELATION. It simply means that if we do experiments to determine simultaneously the momentum and position of a small particle (electron) the product of the two errors (or, better expressed, uncertainties) is of the order of magnitude of h. There is of course a corresponding relation, namely

$$\Delta E \times \Delta t \sim h,$$

connecting the uncertainties in the measured values of the energy and the time.

SCHRÖDINGER'S WAVE EQUATION

The simplest equation representing a wave of any kind may be written in the form

$$\psi = A \cos 2\pi (\nu' q - \nu t)$$

in which ν' and ν, the wave-number and frequency, are constants. This equation has a mechanical interpretation when we replace ν' by p/h and ν by E/h, so that we now have

$$\psi = A \cos \frac{2\pi}{h} (pq - Et) \qquad \text{(XVIII—7)}$$

This is the simplest kind of equation we can meet with in wave mechanics. It might, for example, represent a particle moving under the influence of no forces with constant momentum, p, and constant energy, E. By differentiating it *twice* with respect to q we get.

$$\frac{\partial^2 \psi}{\partial q^2} = -\frac{4\pi^2}{h^2} p^2 \psi ;$$

so that

$$\frac{\partial^2}{\partial q^2} \text{ is, in some sense, equivalent to } -\frac{4\pi^2}{h^2} \times p^2$$

and

$$\frac{\partial}{\partial q} \equiv \frac{2\pi i}{h} \times p, \qquad \text{(XVIII—7A)}$$

if we use the symbol $\equiv$ for "equivalent to". Similarly

$$\frac{\partial}{\partial t} \equiv \frac{2\pi i}{h} \times E \qquad \text{(XVIII—7B)}$$

Now Schrödinger's differential wave equation is what we get when we replace p and E in the classical energy equation by the aid of (XVIII—7A) and (XVIII—7B).*

The significance of the amplitude A of the wave (XVIII—7), assuming the case where it is representing an electron, may be described by saying that the probability that the electron is in some small volume, Ω, is proportional to

$$A^2 \times \Omega.$$

Since only relative values of A are given, we may fix its absolute value so that the above expression is actually *equal to the probability*.† In other words, we fix the absolute value of A by equating the sum of all the $A^2 \times \Omega$, over the whole volume occupied by the wave, to unity.

In general the problems of wave mechanics are not so simple that we can detect immediately, as it were, some expression like, or corresponding to, (XVIII—7) for the wave function ψ. As in optical problems, we have to start out from the Schrödinger differential equation, already mentioned, which represents the

* Erwin Schrödinger's most important papers are contained in the volume entitled *Abhandlungen zur Wellen Mechanik* (Barth, Leipzig). An English translation is published by Blackie & Son.

† This interpretation is due to Max Born.

problem. This equation is of the type the mathematicians call LINEAR. Such equations have the peculiarity that when we have found particular solutions of them (i.e. particular expressions for ψ) the sum of these is also a solution. The simplest of these particular solutions are sometimes called PROPER FUNCTIONS (sometimes, such seems to be the poverty of our language, *eigen* functions). In some of their properties they strongly resemble vectors and indeed the mathematicians in recent times have been accustomed to think of a special kind of space called a HILBERT SPACE, in which these proper functions play a part resembling, or related to, that of vectors in Euclidean or Minkowskian space. Each proper function is associated with PROPER VALUES which determine the possible values of the energy or momentum of the system to which they refer and thus replace the rules expressed by the Wilson–Sommerfeld formulae (XIII—10) or (XIII—10A).

Wave mechanics is superior in various ways to the older quantum theory. It is logically complete and does not require the assistance of *ad hoc* devices, such as the correspondence principle, to answer the questions that are put to it. In one respect the form of wave mechanics developed by Schrödinger requires amplification: it retains its form only under the transformations of Newtonian mechanics, e.g. (XI—4), and not under the Lorentz transformation (XI—6) or (XI—8). In this respect it has been improved by P. Dirac, at least in so far as the theory of the electron is concerned. He makes use of the suggestion (XVII—1) namely that the ordinary momentum, $p = mv$, should be replaced by the more general momentum $\Pi = p + eK$, in which e is the charge on the electron and K is the vector potential.* His theory has at least two merits—it conforms to the special principle of relativity and it gives a reason for the existence of the strange spin which characterizes the electron. Moreover it predicted the existence of the counterpart of the electron—the *positively* charged electron, or POSITRON† as it has been named

* P. A. M. Dirac: *The Principles of Quantum Mechanics* (Clarendon Press, 1935). See also Max Born: *Atomic Physics*, p. 182 (Blackie & Son, 1947).

† P. A. M. Dirac: "The Quantum Theory of the Electron," *Proc. Roy. Soc. A.*, 117, p. 610 (1928), and 118, p. 351 (1928). Dirac originally identified the positively charged particle, which appeared in his theory, with the proton; but in fact his theory predicted the exact counterpart of the electron namely the positron, as it is now called.

since and which was discovered later by the American physicist Anderson in cosmic radiation.

MATRIX MECHANICS

This form of quantum mechanics was discovered by W. Heisenberg and appeared at first to be totally different from Schrödinger's mechanics until the latter showed that they were equivalent to one another.* Important contributions have been made to it by Max Born, P. Jordan and P. Dirac. Its nature can only be indicated here. The proper functions of wave mechanics are now replaced by vectors in a space appropriate to the particular problem.† To take a very simple case let us think of a two-dimensional space, like a Euclidean plane and a *vector*, ψ, referred to rectangular axes of co-ordinates. Its components are, we may suppose, ψ_x and ψ_y. We may form from it another vector ψ' by some linear operation or other; for example

$$\begin{aligned} \psi'_x &= a\psi_x + b\psi_y \\ \psi'_y &= c\psi_x + d\psi_y \end{aligned} \qquad \text{(XVIII—8)}$$

where a, b, c and d are numerical constants. Such an operation may be written, as it was originally, for brevity only, in the following way:

$$\psi' = \begin{bmatrix} a, & b \\ c, & d \end{bmatrix} \psi \qquad \text{(XVIII—8A)}$$

The square bracket symbol is called a MATRIX. We may go further and get a third vector ψ'' by a similar operation on ψ', thus shortly:

$$\psi'' = \begin{bmatrix} e, & f \\ g, & h \end{bmatrix} \psi' \qquad \text{(XVIII—8B)}$$

and obviously we may write

$$\psi'' = \begin{bmatrix} e, & f \\ g, & h \end{bmatrix} \begin{bmatrix} a, & b \\ c, & d \end{bmatrix} \psi. \qquad \text{(XVIII—8C)}$$

The multiplication of matrices is done by the same rule, which indeed the reader might find for himself, as in the case of determinants.

* Heisenberg: *Zeits. für Physik*, 33, p. 879 (1925). Schrödinger: *Ann. der Phys.*, 79, p. 734 (1926).

† That is to say, in the Hilbert space appropriate to the particular problem.

For example,

$$\psi''_x = e\psi'_x + f\psi'_y,$$

as indicated in (XVIII—8B). Therefore

$$\psi''_x = e(a\psi_x + b\psi_y) + f(c\psi_x + d\psi_y),$$

on substituting from (XVIII—8). Therefore

$$\psi''_x = (ea + fc)\psi_x + (eb + fd)\psi_y,$$

Similarly

$$\psi''_y = (ga + hc)\psi_x + (gb + hd)\psi_y.$$

So we get

$$\psi'' = \begin{bmatrix} ea + fc, & eb + fd \\ ga + hc, & gb + hd \end{bmatrix}\psi$$

This last matrix is the product of the two matrices in (XVIII—8c). It is important to notice that the product would, in general, be different if the two factor matrices were in the other order.

Now in matrix mechanics, momentum, energy and observable things in general are represented by matrices. They are also associated with PROPER VALUES, as in wave mechanics, which represent just what we can observe. When the co-ordinates are suitably chosen, a matrix like

$$\begin{bmatrix} a, & b \\ c, & d \end{bmatrix} \text{ becomes } \begin{bmatrix} A, & 0 \\ 0, & B \end{bmatrix}$$

and A and B are its proper values.

To bring out the intimate relationship between this form of quantum mechanics and that of Schrödinger, let us consider the following sequence of operations. We start with Schrödinger's wave function ψ as illustrated by (XVIII—7) and multiply it by any function (of p or q or both), thus:

$$z\,\psi$$

where z means the particular function. We then differentiate the result by q thus:

$$\frac{\partial}{\partial q} z\,\psi$$

It is important to appreciate that this last statement means the result of *first* multiplying ψ by z and then differentiating the *result* by q. Now

$$\frac{\partial}{\partial q}\, z\,\psi = \left(\frac{\partial z}{\partial q}\right)\psi + z\frac{\partial \psi}{\partial q},$$

in which the brackets are meant to indicate that in $\partial z/\partial q$ the differentiation is carried out on z only. Thus

$$\frac{\partial}{\partial q} z\psi - z\frac{\partial}{\partial q}\psi = \left(\frac{\partial z}{\partial q}\right)\psi$$

and so

$$\frac{\partial}{\partial q} z - z\frac{\partial}{\partial q} \equiv \left(\frac{\partial z}{\partial q}\right),$$

in which the symbol $\equiv$ means "is equivalent to".

Now if we replace $\partial/\partial q$ by (XVIII—7A) we get

$$\frac{2\pi i}{h}(pz - zp) \equiv \left(\frac{\partial z}{\partial q}\right) \qquad \text{(XVIII—9)}$$

In matrix mechanics p and z are replaced on the left by suitable matrices, and differential quotients like $\partial z/\partial q$, wherever they occur in classical mechanics, are replaced by this matrix expression on the left. In particular when $z = q$ we have

$$\frac{2\pi i}{h}(pq - qp) \equiv 1, \qquad \text{(XVIII—9A)}$$

which is one of the fundamental statements of matrix mechanics. The unity, 1, on the right of (XVIII—9A) means of course the matrix, 1, with its diagonal elements each equal to 1 and all the rest equal to zero. Its essential characteristic is that when it multiplies another matrix, or is multiplied by it, the latter matrix is unchanged, just as with ordinary numbers $a \times 1 = 1 \times a = a$.

CHAPTER NINETEEN

RADIOACTIVITY AND COSMIC RADIATION

INTRODUCTORY OBSERVATIONS

SOME information about atomic nuclei was revealed by the study of spectra in the light of the older quantum theory, and, as we have seen, by Rutherford's experiments on the scattering of α particles from which it appeared that there is in every atom a massive positively charged central structure of relatively very small dimensions. This central structure, or nucleus, is rather like a sun with planetary electrons revolving about it. It is simplest in the hydrogen atom whose nucleus is called the PROTON and carries the positive charge characteristic of hydrogen ions in electrolysis. The nucleus of any atom has a positive charge equal to a whole number times the charge on the proton. This whole number represents the place of the particular element in the periodic system of Mendeléeff and his contemporaries. Theoretical investigations, more especially those of Bohr, have suggested that each atomic nucleus consists of a number of protons equal to the atomic weight of the element, or nearly so, with just so many electrons that the net positive charge on it is equal to the product of the atomic number and the protonic charge. This suggestion—reminiscent of Dr. William Prout's view—has had to be modified, as we shall see, in consequence of the experimental results of radioactive investigations.

RADIOACTIVITY

The phenomena of radioactivity were first noticed by Antoine Henri Becquerel (1852–1908) who observed that salts of uranium caused a photographic plate to be blackened when it was not exposed to light at all. Moreover he observed that the radiation from the uranium salt, which caused this blackening, was able to pass through very considerable thicknesses of wood and other

materials opaque to ordinary light, just as X-rays would do.* He was followed by Marya Sklodowska, a native of Warsaw, who, on the advice of her husband, Professor Pierre Curie, studied this uranium radiation and tracked it down to the uranium atom. Becquerel had observed that his uranium radiation, like X-radiation, renders air conducting—produces ions in it—and Pierre Curie devised a simple apparatus for measuring such ionization currents. With this apparatus Madame Curie examined the pitchblende from Joachimstal, Bohemia (later Czechoslovakia), and found that its radioactivity was greater than could be accounted for by the quantity of uranium it contained. This led her to the discovery and eventual isolation of the element RADIUM (1898). It is chemically related to the group of metals beryllium, magnesium, calcium, strontium and barium. To give some idea of the laboriousness of this piece of work it may be pointed out that a ton of the pitchblende yields less than a gramme of radium.

ALPHA, BETA AND GAMMA RAYS

Examination of the radiation from radioactive materials revealed three distinct types of rays. A thin sheet of tin-foil absorbs most of the radiation, as measured by its power to ionize air. What succeeds in passing through the tin-foil is much more penetrating and can be recognized to consist of two distinct types of radiation by absorption experiments. Rutherford named the three types α rays (the most easily absorbable), β and γ rays, the last-named being much the most penetrating. They are most effectively distinguished from one another by the way in which they behave in a magnetic field. The α rays are very slightly deflected in a strong magnetic field and in a sense indicating them to be positively charged particles. The β rays are strongly deflected in a magnetic field in the sense which indicates them to be negatively charged particles. Indeed they were easily recognized to be electrons. The γ rays are not deflected at all.† The ratio e/m for α particles turned out to be nearly one half that of a hydrogen ion. They have turned out to be the

* *Compt. rend.*, 122, pp. 420, 501, 689, 1086 (1896).
† Rutherford: *Phil. Mag.*, 5, p. 177 (1903).

nuclei of helium atoms, as Ramsay and Soddy proved.* They carry twice the charge of a proton, but have approximately four times its mass. Rutherford and Geiger actually counted the number of α particles emitted by radioactive materials. Their apparatus, devised by Geiger, was the prototype of the Geiger counters now in use.† By dividing the total charge by the number of particles they estimated the charge on a single particle.

The γ rays emitted by radioactive materials are simply X-rays and consist to some extent of the characteristic X-radiations of the radioactive element,‡ as indeed we might expect, since α and β particles ejected from the nucleus are likely to knock out the element's *K*, *L*, etc., electrons. Of course not all of the γ radiation consists of this characteristic radiation; some of it originates in the nucleus itself.

RADIOACTIVE DECAY—ISOTOPES

Sir William Crookes carried out some simple and illuminating experiments in the early days of radioactive investigation.§ He noticed that, when he precipitated uranium from solution by means of ammonium carbonate and redissolved it in excess of the carbonate, a precipitate remained over which was photographically very very radioactive. As we now know, this meant that it emitted β rays copiously. Crookes called it uranium X (Ur X). The original uranium had lost its photographic activity. Further investigation revealed that uranium (atomic mass 238, atomic number 92) is constantly emitting α particles and becoming a new kind of element (Rutherford and Soddy). This new element is the uranium X of Crookes. It emits β particles copiously. As Soddy pointed out,‖ when an atom emits an α particle (which is simply the nucleus of a helium atom with the positive charge 2 corresponding to its atomic number) its charge

* *Proc. Roy. Soc.*, 72, p. 204 (1903); 73, p. 341 (1904).

† E. Rutherford and H. Geiger: *Proc. Roy. Soc. A.*, 81, pp. 141, 162 (1908).

‡ E. Rutherford and E. N. da C. Andrade: *Phil. Mag.*, 27, p. 854; 28, p. 263 (1914).

§ W. Crookes: *Proc. Roy. Soc.*, 66, p. 409 (1900).

‖ F. Soddy: *Chemical News*, 107, p. 97 (1913).

is reduced by 2. In fact it becomes the atom of an element of atomic number 2 less than the original atomic number, and its atomic weight is reduced by 4 approximately. When an atom emits a β particle (charge minus 1) its atomic charge is increased by 1; the element becomes a new element whose atomic number is 1 higher than the original value. In this case, however, the new element has the same atomic mass (or atomic weight, in the old terminology) as the original one from which it has been formed by the β ray emission, since the mass of a β particle is negligible. Soddy named elements which have the same atomic number (and consequently the same chemical properties); but different atomic masses (atomic weights) ISOTOPES. Elements with the same atomic mass and different atomic numbers—so that they are in fact different chemical elements—are named ISOBARS. Relative atomic masses, experiment has shown, are very nearly *whole numbers* when oxygen—one ought to say *ordinary* oxygen—is taken to be 16. The reader will suspect already how it happens that the chemically determined atomic weight of the chlorine of the chemical laboratories is not a whole number, but 35·46. In fact it is a mixture of two isotopes whose atomic masses are 35 and 37 in the proportion of 3 of the former to 1 of the latter.

This had been shown directly by F. W. Aston (1877–1945), who devised a very effective method of separating isotopes by subjecting their nuclei to the influence of an electric and a magnetic field.* His apparatus is called the MASS SPECTROGRAPH. The existence and proportions of the isotopes of chlorine are clearly indicated in the infra-red band spectrum of HCl. The two sorts of chlorine atom, having different masses, vibrate in the HCl molecule with slightly different frequencies and consequently the spectral lines in the band spectrum are double, their relative intensities indicating the proportion of Cl^{35} to Cl^{37} to be 3 : 1.

One of the most interesting isotopes is known as DEUTERIUM. It is an isotope of hydrogen of atomic mass 2.† There is about one part of it in 4000 parts of the hydrogen in ordinary water.

* F. W. Aston: *Phil. Mag.*, 38, p. 709 (1919); *Mass Spectra and Isotopes* (Arnold).

† H. C. Urey, F. G. Brickwedde and G. M. Murphy: *Phys. Rev.*, 40, p. 1 (1932).

It can be recognized by the doubling of the lines in the band spectra of hydrogen compounds. In ordinary water nearly all the molecules are H_2O molecules, but some are HDO molecules (D represents the deuterium atom) and some D_2O. The mass spectrograph has revealed the existence of still another isotope of hydrogen of mass 3. The chemical properties of HEAVY WATER (water consisting of D_2O molecules) are just like those of ordinary water, but its physical properties—density, freezing point, etc.—are so different that deuterium seems almost like another element.

One of the pioneers in the separation of isotopes by the application of electric and magnetic fields, it should be mentioned, was Sir Joseph Thomson, to whose inspiration the successful work of Aston, as of many others in the Cavendish Laboratory, including Rutherford himself, was largely due.

The atoms of uranium or any other radioactive element are constantly breaking down. This is better expressed perhaps as follows. In a mass of a uranium compound many atoms break down into Ur X during every second, but of course most of the atoms endure for a very long time. In the particular case of uranium most of its atoms endure for thousands of millions of years. One might infer this from the fact that there is so much uranium in the earth's crust. The activity of uranium, as of other elements, is due to the occasional and seemingly accidental breaking down of individual atoms. The frequency with which this happens varies from one element to another. If we use the letter M for a quantity of a radioactive element and dM for the amount of it that decays in the short time dt, then

$$dM = \lambda M dt,$$

where λ is a constant characteristic of the element and called its TRANSFORMATION CONSTANT.*

In other terms the rate of decay (expressed, for example, in

* The formula is more usually expressed as

$$-dM = \lambda M dt$$

because mathematicians and those with mathematical instincts prefer to use the symbol d to represent only increments, a decrement being regarded as a negative increment. It easily follows that

$$M = M_0 e^{-\lambda t}$$

where M_0 is the mass of the radioactive element at a certain moment and M is the mass of it still enduring when the time t has elapsed.

grammes per second) is proportional to the mass of the radioactive material. This is like the law of mass action in chemistry. The time taken for the mass of a radioactive element to diminish to one-half of its original value is called its HALF-VALUE PERIOD. It varies enormously for different elements. For uranium it is about five thousand million years; for radon, the immediate product of the disintegration of radium, it is only 3·85 days. The end result of the disintegration of uranium and of the resulting products appears to be lead (of which there are several isotopes) in one or other of its isotopic forms. Lead is always found in the neighbourhood of minerals containing uranium.

Radioactive disintegration is always associated with generation of heat. A quantity of a radium compound, for example, constantly maintains a higher temperature than its surroundings, so that heat is constantly flowing out from it. This energy is not created. It represents simply the difference between the energy of the original radium, before its disintegration, and that of the atom of radon or its successors which have resulted from the demolition of the radium aton.

ARTIFICIAL TRANSMUTATION OF ATOMS

The α particles emitted by a particular radioactive substance, e.g. by radium, are characterized by having a definite range (for example, in air at normal pressure and temperature). This was first noticed by Sir William H. Bragg and R. Kleeman.* They measured the ionization produced by α particles after travelling various distances and found it to cease rather abruptly when the α particles had travelled a characteristic distance. This distance they called the RANGE of the particles. The range, in this sense, is not the actual distance they travel, but the distance they travel while still having enough energy to ionize the molecules of the air.

Rutherford noticed the presence in air and in nitrogen gas, subjected to bombardment by α particles, a few particles of much longer range than that of the α particles. He had some reason to believe they were protons—the charged nuclei of hydrogen atoms—and naturally came to the conclusion that these had been knocked out of the nitrogen atoms. He could definitely associate

* W. H. Bragg and R. Kleeman: *Phil. Mag.*, 10, p. 318 (1905).

them with nitrogen.* P. M. S. Blackett subsequently confirmed this with the aid of Wilson's cloud chamber. In fact he produced photographs of individual collisions.† What actually happens to a nitrogen nucleus may be expressed as follows:

$$_7N^{14} + {}_2He^4 = {}_8O^{17} + {}_1H^1.$$

A nitrogen nucleus, mass number 14 and nuclear charge 7, is struck by a helium nucleus (α particle) mass number 4 and nuclear charge 2, with the result that an oxygen atom of mass number 17 and nuclear charge 8 and a proton of mass number 1 and charge 1 are produced. Clearly the oxygen produced is an ISOTOPE of the ordinary oxygen.

Since these remarkable experiments were carried out, the atoms of many elements have been broken down by bombarding them with rapidly moving particles and in many instances the resulting products are radioactive—induced radioactivity. "Broken down" is not always the appropriate expression. In the case of nitrogen, as we have seen, a heavier atom, $_8O^{17}$, was built up from it.

DISCOVERY OF THE NEUTRON

About 1930 a very penetrating radiation, at first thought to be very hard γ or X-rays, was obtained by bombarding beryllium with the α rays from polonium.‡

Irene Joliot, daughter of Madame Curie, and her husband, M. F. Joliot, noticed that, while this radiation passed easily enough through thick sheets of metal, it was strongly absorbed by substances like paraffin, which contains a lot of hydrogen, with the expulsion of protons. These peculiarities of the radiation proved that it was not γ radiation.§ Investigations by J. Chad-

* E. Rutherford: "Collision of α Particles with Light Atoms," *Phil. Mag.*, 37, p. 537 (1919), and 41, p. 307 (1921). "Artificial Disintegration of the Elements," *Nature*, 109, p. 614 (1922). E. Rutherford and J. Chadwick: "Artificial Disintegration of Light Elements," *Phil. Mag.*, 42, p. 809 (1921); "Disintegration of Elements by α Particles," *Phil. Mag.*, 44, p. 417 (1922).

† P. M. S. Blackett: "The Ejection of Protons from Nitrogen Nuclei. Photographed by the Wilson Method," *Proc. Roy. Soc.*, 107, p. 349 (1925).

‡ Bothe and Becker: *Naturwiss*, 18, p. 705 (1930).

§ I. Curie and M. F. Joliot: *Comptes Rendus*, 194, pp. 273, 708 (1932).

wick (now Sir James Chadwick) indicated that this penetrating radiation consisted of *uncharged* comparatively heavy particles* and this was confirmed by Norman Feather, now professor in the University of Edinburgh, with the aid of the cloud chamber. Later Chadwick succeeded in making an estimate of the mass of the new particle and found it to be almost indistinguishable from that of the proton. It is in fact a trifle greater than that of the proton. On the scale of oxygen = 16, the proton has a mass very near to 1·0076 while that of the NEUTRON, as the new particle was named, is almost 1·009. Its discovery has modified our view of the constitution of atomic nuclei. Thus we now think of a helium nucleus as made up of 2 protons and 2 neutrons; not 4 protons and 2 electrons. A carbon nucleus we believe to consist of 6 protons and 6 neutrons, not of 12 protons and 6 electrons. Quantum mechanics raises difficulties about electrons as nuclear constituents. The ejection of neutrons from beryllium under bombardment with α particles is represented by

$$ {}_4\mathrm{Be}^9 + {}_2\mathrm{He}^4 = {}_6\mathrm{C}^{12} + {}_0\mathrm{n}^1. $$

The symbol ${}_0\mathrm{n}^1$ represents the neutron—mass number 1 and charge 0. Neutrons are produced, together with nitrogen atoms, when boron is bombarded with α particles, thus

$$ {}_5\mathrm{B}^{11} + {}_2\mathrm{He}^4 = {}_7\mathrm{N}^{14} + {}_0\mathrm{n}^1. $$

BETA RAY EMISSION AND THE NEUTRINO

Since there are, it is believed, no electrons in atomic nuclei the emission of β rays, i.e. the emission of electrons, must be a process in which they are, as it were, manufactured. It appears as if in the emission of an electron a neutron in the nucleus splits up into a proton and an electron.

Measurements of the energy of ejected β particles which cannot be described here suggested rather strongly a conflict with the fundamental principles of conservation of energy and momentum. To save these principles Fermi (1934) suggested that the emission of a β particle is accompanied in general by the simultaneous ejection of a neutral particle of very small mass which had not hitherto been observed. The existence of this particle, called the NEUTRINO, has seemed doubtful; but quite recently

* J. Chadwick: *Nature*, 129, p. 312 (1932).

(1942) experiments by J. F. Allen in the Cavendish Laboratory have confirmed it. He came to the conclusion that the beryllium isotope ${}_4Be^7$, which is formed in consequence of bombardment of a lithium isotope with deuterium nuclei (deuterons) in accordance with

$$ {}_3Li^6 + {}_1D^2 = {}_4Be^7 + {}_0n^1, $$

is transmuted to ${}_3Li^7$ by capturing *one of its own K electrons*, thus:

$$ {}_4Be^7 + {}_{-1}e^0 = {}_3Li^7. $$

The total momentum is unaffected by this process. If, for example, the moon were to fall to the earth their common centre of mass would not thereby be affected. So with ${}_4Be^7$ when the circumnuclear electron ${}_{-1}e^0$ falls into it; but Allen noticed that the atom recoiled, thus indicating the ejection of a particle. This is probably Fermi's elusive neutrino. Uncharged particles are necessarily difficult to observe. They are not attracted or repelled by atomic nuclei, or electrons—so long at any rate as they do not get so extremely near to them that forces of the mesonic order come into play. They are therefore very penetrating. Moreover they ionize very little and are consequently difficult to observe by cloud-chamber methods.

THE CYCLOTRON

The artificial transmutation of atoms was achieved at first, as we have seen, by the use of naturally occurring high-velocity particles, namely α particles. The next step was the development of methods of giving high velocities (and energies) to charged particles of various kinds. The most successful of the devices for this purpose is the CYCLOTRON invented by E. O. Lawrence, of California.* It is easy to describe how it functions. Think of a circular metal cylinder the height of which is not very great compared with its diameter. Imagine it to be cut in half to produce two D-shaped portions—they are called "dees"—as indicated in Fig. XIX—1. Perpendicular to the semicircular faces is a strong uniform magnetic field, produced by an enormous electromagnet, weighing, in the latest instrument, not

* E. O. Lawrence and N. E. Edlefsen: *Science*, 72, p. 376 (1930). Lawrence was awarded the Nobel Prize in 1940 for his work in developing the cyclotron.

far short of 4000 tons, I believe. The dees are connected (see *a* and *b* in the figure) to the terminals of a high-frequency oscillator with an amplitude of five to ten thousand volts. Consider what will happen to a charged particle which is projected from

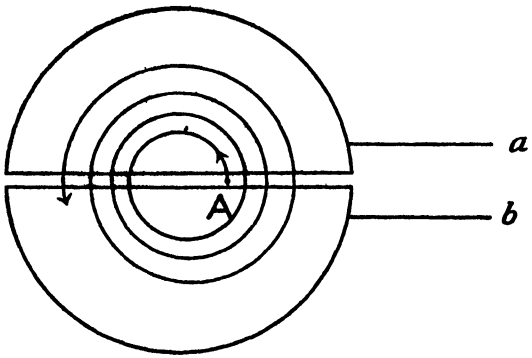


Fig. XIX—1

a point *A* near the middle with a suitable initial velocity, say perpendicular to the edges of the dees. It will travel in a circle and the line of separation of the dees cuts this circle along, or almost along, its diameter. The force on the charged particle, due to the field, is equal to Hev where $H \equiv$ intensity of field, $e \equiv$ charge on particle and $v \equiv$ particle velocity. This force is balanced by the centrifugal force, mv^2/r, r being the radius of the circle along which the charged particle travels. Therefore

$$Hev = mv^2/r$$

or

$$r/v = m/eH.$$

Multiply both sides by 2π, thus

$$2\pi r/v = 2\pi m/eH.$$

The left-hand side of this statement means the time taken by the particle to travel once round its circular path. Call it T. Therefore

$$T = 2\pi m/eH \qquad \text{(XIX—1)}$$

Notice that this is quite independent of the speed of the particle —unless indeed its speed is so great that the relativistic increase of mass with velocity becomes appreciable. Now let us suppose that the field intensity, H, or, for that matter, the period of the

oscillator, to be adjusted so that this period is that of the oscillator. Then the charged particle will complete a circle in the same time as the oscillator's period. If then it crosses the gap between the dees while the electric field is accelerating it, it will acquire a higher speed. The isochronism will not be affected by this and it will again reach the gap between the dees when the conditions are conducive to a further increase in its velocity. Consequently the charged particle, proton, deuteron or whatever it may be, has its speed progressively augmented.

The cyclotron provides the most effective way of giving enormously high speeds to such things as portons, deuterons, helium nuclei (α particles), etc., speeds corresponding to many millions of ELECTRON VOLTS. An electron volt is the energy acquired by an electron (or any free particle with the electronic charge of $4{\cdot}8 \times 10^{-10}$ E.S.U.) in moving under the influence of an electric field from one point to another whose potential differs by a volt. It is easy to estimate it in terms of the erg. It is in fact

$$4{\cdot}8 \times 10^{-10} \times 1/300,$$

since 1 E.S.U. of potential difference is equal to 300 volts. Therefore

$$1 \text{ electron volt} = 1{\cdot}6 \times 10^{-12} \text{ erg.}$$

This is a very small amount of energy. Even 100 million electron volts is only

$$1{\cdot}6 \times 10^{-4} \text{ erg,}$$

but a deuteron with so much kinetic energy will reach a velocity approaching 10^{10} cm./sec. At this speed the relativistic dependence of mass on velocity is beginning to manifest itself.

The cyclotron cannot be used for generating velocities which are so great as to make the relativistic factor $\gamma = (1 - v^2/c^2)^{-\frac{1}{2}}$, appreciably greater than unity. Another type of apparatus has therefore been devised (Kerst and Serber, 1941) which can generate enormous velocities in electrons. It is called the BETATRON. In some respects it is like the cyclotron. There is the same flat cylinder and perpendicular magnetic field, but the cylinder is not divided into two dees and no electric field is applied. The electrons move in circular paths, their centrifugal force being balanced by *Hev*, while their enormous speed is given them by the electromotive force induced by the increasing magnetic

flux (Faraday's law). In this instrument the magnetic field intensity is *not* uniform. In order to keep the electron on a circle, the field intensity on the circle must be half the mean value of the inducing field intensity.

Van de Graaff has invented an interesting method of producing high-velocity particles. It is based on the generation of an enormously high potential by charging up a large hollow metal sphere. Charges are conveyed on a belt into the interior of the sphere and contact made with it. The high voltage developed can be applied to accelerate charged particles.

COSMIC RADIATION

A very penetrating radiation comes to us from interstellar space. It was first noticed about the beginning of the century. C. T. R. Wilson observed that the leaves of a charged electroscope gradually collapsed, even when the strictest precautions were taken to ensure insulation. It was naturally thought at first that the air in the electroscope was ionized by radiations from the earth due to radioactive materials. The investigations of Hess (1912), Kohlhörster (1914) and Millikan, however, gradually revealed that this radiation came from outer space and apparently from all directions. Its intensity was found to be the same during the night as during the day, so that the sun had no appreciable share in its production. It has been investigated by balloon ascents in which the observers carried electroscopes with them and also by sending up small unmanned balloons carrying suitable recording apparatus. These have shown that the intensity of the radiation—as measured by its ionizing power—at first diminishes as the height increases, reaches a minimum and then increases with the height. The primary radiation appears to consist of positively charged particles of great energy and penetrating power—they are probably protons. Their positive charge can be inferred from the effect on them of the earth's magnetic field. The ionization due to it is weak at the equator and in low latitudes, but increases rather suddenly about latitude 49°.

The primary particles, it has been shown, produce secondary particles: electrons, positrons and MESONS ($\mu\acute{\epsilon}\sigma o\varsigma \equiv$ middle), the last-named having a mass somewhere between that of an

electron (or positron) and that of a proton. Both the positron and the meson were predicted before they were actually observed, the former, as we have already seen, by Dirac and the latter by a Japanese, Yukawa, to whom I shall refer again. A charged cosmic particle can be recognized by the curvature of its track in a magnetic field, but unless it is known in which direction it is moving it is impossible to know what is the sign of its charge. Anderson thought of a simple and ingenious way of finding out the direction in which a charged cosmic particle moved through his cloud chamber. He placed a sheet of lead across the middle of it and observed cloud tracks indicating paths which traversed the lead sheet. These were more curved, in a magnetic field, on one side of the sheet than on the other—the greater curvature indicating smaller velocity. And the smaller velocity could only be due, of course, to the reduction of the velocity of the particle in consequence of passing through the lead. Thus the *direction* of motion of the particle was indicated. In this way Anderson was able to recognize that some cloud tracks, which were exactly like those of electrons, were associated with a *positive* charge. Thus the positive electron, or POSITRON, was discovered. The MESON is a particle of rather uncertain mass—perhaps 200 times that of an electron—which has been observed in cosmic radiation. Both positive and negative mesons have been observed and there is reason to believe that mesons are sometimes neutral.

A remarkable phenomenon due to cosmic radiation was first observed by Blackett and Occhialini (1933). They noticed that occasionally a shower of tracks appeared which all started from the same point on or in the wall of the cloud chamber. These cloud tracks appear to consist of equal numbers of positrons and electrons. It is believed that their formation is effected in the following way. A photon of enormously high frequency in the cosmic radiation—Max Born speaks of it as an ultra γ ray—strikes an atom and a very high energy pair, positron and electron, is produced in consequence. Each of these in striking an atom generates an ultra γ photon and each photon in its turn generates a pair (positron and electron), and so on, the energy of the pairs becoming of course progressively smaller.

The origin of cosmic radiation is still unknown.

CHAPTER TWENTY

THE ATOMIC NUCLEUS

MESON THEORY

THE positive charge on the protons in a nucleus must obviously tend to disrupt it. What is it then that prevents a complex nucleus, e.g. a carbon nucleus, from flying apart under the mutual repulsion of its protons? An explanation was suggested some years ago by the Japanese physicist, Yukawa.* In broad outline his theory is as follows. Both protons and neutrons carry a kind of charge—not to be identified with an electric charge, but analogous. In virtue of this they attract one another and we might tentatively suggest that the force holding together a proton and a neutron is equal to

$$g^2/r^2$$

where g is the "charge" on the proton and on the neutron. But the new kind of force is limited to the region occupied by the nucleus, so Yukawa made what I think was originally an *ad hoc* amendment: he assumed for the mutual potential energy of the particles

$$U = g^2 e^{-\lambda r}/r, \qquad \text{(XX—1)}$$

λ being very big, 10^{12} or 10^{13} cm.$^{-1}$. The exponential makes the expression diminish very rapidly with increasing r, so that this new field of force is effectively confined to the nucleus.

In a changing electromagnetic field electromagnetic waves are constantly being generated, in other words *photons* are being generated. In Yukawa's field a new type of particle plays the part of the photon. It is called the MESON and is effectively distinguished from the photon by having a REST MASS and also by the possibility of carrying a positive or a negative charge of electricity, though neutral mesons are also assumed. When a positively charged meson leaves a proton the latter becomes a neutron, and when a negatively charged meson leaves a neutron

* H. Yukawa: *Proc. Phys.-Math. Soc. Japan*, 17, p. 48 (1935).

the latter becomes a proton. The forces holding protons and neutrons together spring out of this exchange of their rôles by the protons and neutrons (exchange forces).

The meson is a very short-lived thing. It has the spin momentum $h/2\pi$ (an electron, it will be remembered, has a spin momentum equal to one-half of $h/2\pi$) and a meson probably decays into an electron (or positron) and a neutrino, each of these having the half-unit of spin momentum. Both the electron and the neutrino have relatively enormous kinetic energy, since the greater part of the mass of the meson is used up to produce it.

There is a term in the wave equation of the meson which indicates that Yukawa's constant, λ, must conform to

$$\lambda = \frac{2\pi m_0 c}{h},$$

where m_0 is the mass of the meson. Yukawa's estimate of λ was 5×10^{12} cm.$^{-1}$, which yields a value for m_0, the mass of the meson, about 200 times that of the electron.

MASS DEFECT AND NUCLEAR BINDING ENERGY

In ordinary mechanical systems stability is associated with low energy. As an illustration the case of a massive body supported at a point on it, well separated from its centre of gravity, will serve. One position of equilibrium is that in which its centre of gravity is vertically *above* the point of support and the other is that in which the centre of gravity is vertically *below* the point of support. The latter is the stable case. It is associated with *lower energy*, the difference between the corresponding energy values being the product of the vertical distance between the two positions of the centre of gravity and the weight of the body.

So with the formation of a stable atomic nucleus from its constituent protons and neutrons we expect a reduction in energy (or the equivalent mass). The rate of exchange between mass and energy, i.e. grammes and ergs, is, as Einstein showed,

$$E = mc^2$$

or

$$\text{Ergs} = \text{grammes} \times 9 \times 10^{20}.$$

Now calling the mass of an oxygen atom 16, that of a proton is approximately 1·0076 and that of a neutron 1·009, and in what

appears to be the stablest of all nuclei, the helium nucleus, there are two of each and their total mass is equal to

$$4{\cdot}0332.$$

But a helium nucleus is estimated to have, in terms of the same unit, a mass of

$$4{\cdot}003,$$

so that in its formation there has been a loss of mass equal to

$$0{\cdot}0302$$

in terms of this unit. It amounts to this: when four grammes (approximately) of helium are generated from protons and neutrons—or we might say from hydrogen—there is a loss of mass of a trifle over 0·03 gramme and the mass-energy law requires an emission of energy equal to

$$0{\cdot}03 \times 9 \times 10^{20} \text{ ergs}$$
$$= 2{\cdot}7 \times 10^{19} \text{ ergs.}$$

It is enormous. This diminution of mass is usually called the MASS DEFECT and the equivalent energy the BINDING ENERGY of the nucleus.

In nuclei of small mass number—mass number means the number of protons plus the number of neutrons in the nucleus—the number of protons is equal, or nearly equal, to the number of neutrons; in the case of helium two and two; in the case of carbon six and six. In the case of more massive nuclei the proportion of neutrons increases. This is easily comprehensible. The more protons the greater the electrostatic repulsion between them and the binding influence of the extra neutrons, which is of course not offset by repulsion, is needed to hold the nucleus together.

ARTIFICIAL ISOTOPES AND INDUCED RADIOACTIVITY

The first instance of the artificial production of an isotope was that of ${}_8O^{17}$ which emerged as the result of bombarding ordinary nitrogen with α particles (Rutherford). The lighter of the two well-known isotopes of lithium has been transformed into the heavier one by bombarding it with high-velocity deuterons (Cockcroft and Walton, 1934) in accordance with

$${}_3Li^6 + {}_1D^2 = {}_3Li^7 + {}_1H^1,$$

a proton being thereby produced. The bombardment of ordinary phosphorus with high-velocity deuterons gives rise to the isotope ${}_{15}P^{32}$ with emission of protons, thus:

$$ {}_{15}P^{31} + {}_{1}D^{2} = {}_{15}P^{32} + {}_{1}H^{1}. $$

This isotope of phosphorus, like many other artificially produced isotopes, is radioactive, with a half-period of about a fortnight. It is one of a number of isotopes with artificially induced radioactivity which are now being used by the biologists to assist them in investigating the metabolic processes in the bodies of animals and plants. The course of such a thing can easily be traced through the body of the organism which has assimilated it, since its presence anywhere in the organism can be detected by its radioactivity—by a Geiger counter for example.

CHARACTERISTICS OF NEUTRONS

Neutrons were detected by their great penetrating power and were in fact first suspected to be γ radiation until it was noticed that they were strongly absorbed by materials containing a lot of hydrogen. This alone suggested that they might be neutral particles of the same, or nearly the same, mass as hydrogen nuclei. When an elastic body strikes another, initially at rest and of very much greater mass, it rebounds with its energy only very slightly diminished, but when it makes a head-on collision with an elastic body of the same mass and initially at rest, it transfers all its energy to the second body. This kind of thing can be observed in the collisions of billiard balls. Being uncharged, neutrons can approach, even when their velocities are quite small, right up to the nuclei of atoms, since they are not hindered by the positive nuclear charge which would repel α particles or protons for instance. Neutrons flying about among the molecules and atoms of a substance, especially if these latter are comparatively light; rapidly reach statistical or temperature equilibrium. They have become THERMAL NEUTRONS. Lastly, neutrons which collide with the nucleus of an element may become attached to it, or enter into it. Sometimes the nucleus breaks down with the emission of perhaps a proton or α particle.

Naturally the effect of bombarding very massive nuclei, those of uranium in particular, with neutrons was investigated soon

after neutrons were discovered. It was hoped to produce in this way elements of still higher atomic number than uranium (92), which before the discovery of neutrons was the element of highest known atomic number. Its commonest isotope has the mass number 238. When this is bombarded with neutrons the immediate result is usually another isotope with the mass number 239. Thus

$$_{92}U^{238} + {}_{0}n^{1} = {}_{92}U^{239} + \gamma.$$

That is to say, the neutron (mass number 1 and charge zero) unites with the uranium nucleus (mass number 238 and charge 92) to form $_{92}U^{239}$ with the emission of a γ photon. The isotope thus formed is radioactive. It emits β particles and becomes a new element of mass number 239 and atomic number 93, thus:

$$_{92}U^{239} \text{ becomes } {}_{93}Np^{239} + {}_{-1}e^{0},$$

the new element being called NEPTUNIUM. Evidently the guiding principle of this nomenclature is the fact that Neptune is the next planet beyond Uranus, and on the same principle the next isotope which emerges from the break-down of the neptunium nucleus is called PLUTONIUM. Thus

$$_{93}Np^{239} \text{ becomes } {}_{94}Pu^{239} + {}_{-1}e^{0}.$$

The extraordinary appropriateness of the last name appears to be due entirely to chance.

FISSION

In 1938 Hahn and Strassmann noticed that one of the products of the bombardment of uranium with neutrons was like barium; indeed it was barium. The uranium atom had in fact split into two parts of comparable mass with the release of a relatively enormous amount of energy, the fragments of the nucleus flying apart with the energy equivalent of millions of electron volts. The uranium nucleus owes its instability to the repulsive force of its 92 protons, which is not at best very securely balanced by the mesonic binding forces of its neutrons and protons, and probably the entrance of the neutron from outside sets up oscillations which separate two portions of the nucleus just far enough for the repulsive force to predominate over the attractive exchange force.

This phenomenon of fission is part of the basis of the devastating bomb used near the close of the late war. What the makers of the bomb achieved was the splitting, within a small fraction of a second, of all, or nearly all, of the atoms in a lump of uranium (and later plutonium). The uranium isotope which was found to be suitable is ${}_{92}U^{235}$. It is split up by slow-moving neutrons into barium and krypton nuclei. When this happens several fresh neutrons fly out, the barium and krypton nuclei having altogether appreciably fewer (at least 11 fewer) neutrons than the original uranium atom. These in their turn send off other uranium nuclei and thus a so-called CHAIN REACTION is started—provided of course that the lump of fissible material is not so small that most of the neutrons get right out of it before they get a chance to do any damage, and provided further that the velocities of the ejected neutrons are rather small. Now it happens that the neutrons ejected when the uranium 235 atom splits have quite big velocities and therefore a so-called moderator was used to bring down their velocities. Carbon in the form of very pure graphite was chosen for this purpose. Being rather light the carbon atoms are very effective in slowing down neutrons, while they do not capture them—though impurities in the carbon (graphite) might be very effective in absorbing them.

CHAPTER TWENTY-ONE

ASTROPHYSICS AND COSMOLOGICAL SPECULATION

COSMICAL PHYSICS

ASTRONOMY is very largely physics. The investigation of the motions of stellar and planetary bodies is an application of mechanics. The examination of the spectra of stars and nebulae is the main part of the routine work of many astronomers. The study of such spectra and the inferences drawn from them constitute the subject of ASTROPHYSICS. It goes back to the time of Wollaston and Fraunhofer. Among the earlier contributors to the subject was Sir William Huggins, who had a small observatory built for himself in the neighbourhood of London in 1855. About the same time Father Secchi began his observations of stellar spectra at the Vatican Observatory in Rome and appears to have been the first to classify stars according to their spectral type (1867). Norman Lockyer was active in the eighteen-seventies and later. He became a professor in the Normal School of Science, later called the Royal College of Science, which had an observatory in Exhibition Road, South Kensington. Part of its equipment included a very respectable reflecting telescope.

STELLAR TEMPERATURES AND DIMENSIONS

The radiation from the sun and stars approximates fairly closely to black body radiation. This fact has made it possible to make moderately good determinations of their surface temperatures. For this purpose a spectroscope is attached to a reflecting telescope and observations are made of the distribution of energy in the spectrum of the light from the star whose temperature is being sought; or this distribution is estimated from the relative brightness of the light of different wave-lengths. The temperature may then be determined by Wien's displacement law, or by Planck's law. When the temperature at the surface of the star

has been found and the total rate at which it radiates heat, it is a simple matter to calculate its diameter. This method of getting stellar diameters appears to have been adopted first by Professor E. Hertzsprung, of Leyden, in 1913 or earlier. The rate of emission of heat by the star can be found by focusing its image on a suitable radiometric device. The immediate observation gives the rate at which heat enters the telescope, but a knowledge of the aperture of the instrument and of the distance of the star enables its total rate of emission to be computed. Since the star emits, at least approximately, like a black body, its rate of emission per unit area is of course equal to

$$\sigma T^4 = 5{\cdot}75 \times 10^{-5} T^4$$

in ergs per cm.^{2} (Chapter XIII) and division of the total rate by the rate per unit area gives the area of the star's surface and thus its diameter and volume.

The determinations of stellar dimensions by Hertzsprung's method received a remarkable confirmation from the observations made with the aid of the stellar interferometer invented by Michelson (1920). This was not of course the famous apparatus of the aether drift experiments—though equally ingenious.* It was an instrument which, when mounted on a telescope, virtually enlarged its aperture and consequently its resolving power. One of the first stars, if not the first, on which it was tried was Betelgeuse (α *Orionis*) the *angular* diameter of which was found to be 0·046 of a second, and since its distance is about 190 light years its *linear* diameter turned out to be about 250 million miles, about 300 times that of the sun, amply big enough to include the earth's orbit within it. It is not the largest star: the diameter of Antares is half as big again.

The surface temperatures of stars range from about 2000°K. to perhaps 70,000°K. Capella is a double star. Capella A has a temperature about equal to that of the sun, while Capella B has a somewhat higher temperature. That of Arcturus is a trifle more than 4000°K. The blueness of Sirius, that is of Sirius A, indicates its much higher temperature, nearly 14,000°K.

Hertzsprung noticed that red stars—i.e. those with a comparatively low surface temperature—had one or other of two

* For a description of it see *An Introduction to the Theory of Optics*, by A. Schuster and J. W. Nicholson, p. 167E (Arnold, 1924), or *Theoretical Physics*, by W. Wilson, Vol. ii, pp. 265 and 267 (Methuen).

fairly definite sizes, provided they had the same surface temperatures. They were either giants or dwarfs—it was he who introduced these now familiar terms—with nothing in between.

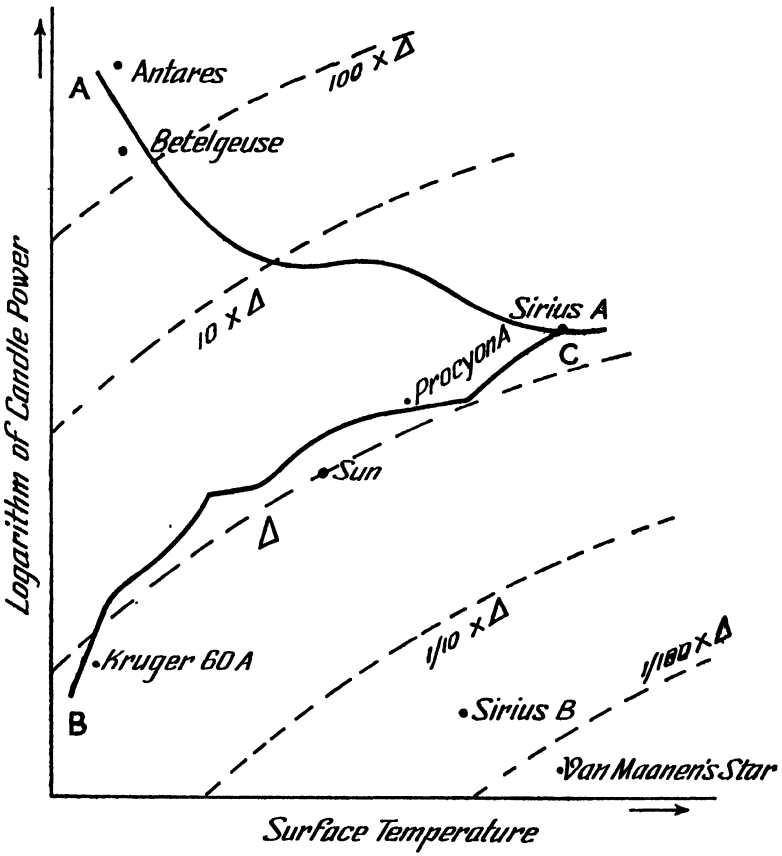


Fig. XXI—1

As we proceed to stars of higher surface temperature the two groups differ less and less in size and in candle-power until at last, at the temperature of 15,000°K. or thereabouts, the two groups fall together. The intrinsic brightness (candle-power) of a star of given surface temperature is a measure of the area of its surface. The American astronomer, H. N. Russell, made a diagram on which he arranged the stars according to their candle-power and surface temperature. The accompanying diagram

(Fig. XXI—1) is meant to indicate what Russell's diagram is like. It is not drawn exactly to scale, though I believe it does represent, if rather roughly, the magnitudes involved. The broken contour lines indicate the diameters of stars which lie near them on the diagram, Δ meaning the Sun's diameter. Russell found the stars to lie, with some exceptions, on or near the lines *AC* and *BC*. The exceptions are the white dwarfs, such as the companion of Sirius and van Maanen's star, which lie in the right-hand lower corner of the diagram (Fig. XXI—1). As we pass across the diagram from the right lower corner to the left-hand upper one the diameters of the stars on the diagram increase roughly in geometrical progression, those on or near the line *BC*, so-called MAIN SEQUENCE stars, such as Alpha Centauri, Altair, Procyon, Sirius and Vega, having dimensions of the order of that of the sun.

STELLAR INTERIORS

No doubt every star was at some time in its history a huge mass of attenuated gas contracting under the gravitational attraction of its parts and developing heat in its interior at a great rate. The giant red stars like Betelgeuse are passing through this phase of their life history now and probably it is a relatively short one. Russell supposed that, as the star aged, it passed along the upper line of the diagram (Fig. XXI—1) from *A* to *C* and then to *B*. In this he was no doubt right. In recent times A. S. Eddington (Sir Arthur Eddington, 1879–1944) and E. A. Milne have also occupied themselves with the problems of the internal condition and life histories of stars. The general outcome of their efforts may be described as follows. At first it seemed that while giants like Antares and Betelgeuse were entirely gaseous, smaller stars, like the Sun, could hardly be in the gas state throughout their interiors, until Eddington noticed that the luminosities of both types of star depend on their masses in the same way. It appeared then that the Sun and similar stars were also gaseous throughout their interiors. The question naturally arose as to how this could be when the densities at the centres of such stars are so great. The simple answer, as Eddington saw, is that at the enormous temperatures developed in the interior of these stars not only are all compounds disrupted into their atomic constituents, but the atoms themselves have their

outer electrons more or less completely stripped off. We have seen that an atom, with its system of electrons, *K*, *L*, *M*, etc., intact, has a diameter of the order of 10^{-8} centimetre, i.e. a hundredth of a millionth of a centimetre, and it excludes a corresponding volume from occupation by other atoms. But when the atoms are stripped of their electrons we have a set of particles: nuclei and electrons whose linear dimensions are only of the order of 10^{-13} cm. or, at most, perhaps 10^{-12} cm., that is between a *ten thousandth* of that of an undamaged atom (at most) and a *hundred thousandth* of it. Collectively a relatively immense volume is thus provided for these small particles to fly about in, as the molecules of a gas fly about in the containing vessel. We can therefore understand how it is that the material of a star may still behave like a gas even when compressed to such a degree that this would be quite impossible under terrestrial conditions. There is one other very important difference between such a gas and one that can be observed in a terrestrial laboratory: the temperatures in the interiors of stars like the Sun are estimated to be of the order of 20 million degrees and consequently the pressure of radiation in such a star becomes enormous. It will be remembered that this pressure is equal to one-third of the density of the radiant heat, which means that it is equal to about

$$2{\cdot}56 \times 10^{-15} \times T^4$$

dynes per square centimetre. When 20,000,000 is substituted for T we easily find a trifle more than

$$4 \times 10^{14} \text{ dynes/cm.}^2$$

This is approximately 400 million atmospheres.

The behaviour of this super gas is very largely determined by the percentage of hydrogen it contains. The reason is that the average mass of its "molecules" or ultimate particles would be one-half (calling a hydrogen atom unity) if it were wholly hydrogen. Each atom gives rise to two "molecules" or particles, a proton and an electron whose combined mass is approximately 1, and 1 divided by 2 gives the average. But for nearly all the other elements this average is in the neighbourhood of 2. Even helium is $1\frac{1}{3}$. Its atom has a mass of 4 and splits up into a nucleus and 2 electrons = 3 particles.

The lithium atom has a mass 7 and it splits up into a nucleus

and 3 electrons so that the average mass of the particles is $7/4 = 1\frac{3}{4}$. The carbon atom has a mass 12 and splits up into a nucleus and 6 electrons—7 particles; average mass equal to 12/7. The copper atom has a mass 63 and splits up into a nucleus and 29 electrons, yielding the average 63/30. For uranium we find $238/(1 + 92)$, roughly $2\frac{1}{2}$. The general formula is

$$(\text{Atomic mass number})/(1 + \text{Atomic number}).$$

The amount of hydrogen in a star has been estimated to be about one-third of its mass. It is this proportion which gives the right relation between the mass and the candle-power of the star.

The life history of a star appears to be somewhat as follows. It passes quickly ("quickly" means occupying a period of time which is a rather small fraction of the time during which it can be identified as a star—still many million years) through states represented by the line from *A* to *C* in Fig. XXI—1. During this phase it is contracting under gravity and the heat it radiates is mainly derived from the lost gravitational energy. Thus far the picture is that which Russell gave and it is very like that given by those earlier worthies, Kelvin and Helmholtz, but more recent knowledge and inquiry, and perhaps especially Eddington's work, have suggested something quite different for the further progress of the star. At *C* its radius has got down to the size indicated by the position of *C* on the diagram. Its internal and surface temperatures are much higher, especially the former. Not only are its atoms beginning to get stripped, but a new source of energy is being drawn on for the heat radiation of the star. Protons are being converted into deuterons, thus

$$_1H^1 + {}_1H^1 = {}_1D^2 + {}_1e^0,$$

with emission of positrons and γ radiation and as the internal temperature rises still further helium nuclei are produced from protons. H. A. Bethe and others believe that this occurs in the following way:

$_6C^{12} + {}_1H^1$ yields $_7N^{13}$ and γ radiation;
$_7N^{13}$ disintegrates into $_6C^{13}$ and $_1e^0$;
$_6C^{13} + {}_1H^1$ yields $_7N^{14}$ and γ radiation;
$_7N^{14} + {}_1H^1$ yields $_8O^{15}$ and γ radiation;
$_8O^{15}$ disintegrates into $_7N^{15} + {}_1e^0$,

and finally

$$_7N^{15} + {}_1H^1 \text{ yields } {}_6C^{12} + {}_2He^4,$$

the final result being the production of helium nuclei and positrons from protons with emission of γ radiation. The star now practically ceases to contract because of the immense pressure of the radiation. This state of affairs continues during the greater part of the star's life while it is passing roughly speaking along the line *CB*. After this it is conjectured that as the protons get used up the generation of radiation goes on at a slower rate, the temperature drops and the pressure of radiation, depending as it does on the *fourth power* of the temperature, drops in *catastrophic* fashion. The vast gravitational pressure is no longer counterbalanced by radiation pressure and a collapse of the star occurs on a gigantic scale. There is an immense and sudden development of heat at the expense of gravitational energy and the star becomes a *nova*, so called because it gives to the observers the impression of a new star coming into being. After this it probably settles down into the condition exemplified by Sirius *comes* or van Maanen's star, which have densities of the order of tons per cubic inch.

Stable conditions in any kind of system whatsoever are only reached by loss of energy and, if I may speculate here, I would suggest that, where a vast mass of atomic nuclei are jammed together, as they must be in the companions of Sirius and Procyon, van Maanen's star and others like them, their combined energy is far above its minimum value. So possibly a star becomes a nova at least once again before it reaches the death which is represented by an absolute minimum of energy.

One last word about the life of a star: the continual radiation from it is necessarily associated with a diminution in its mass, in accordance with Einstein's mass–energy law. It is not difficult to make the estimate that the sun is now losing its mass at the rate of some four million tons a second, and it was, there is no doubt, losing it at a much greater rate when it was younger.

ABOUT THE SPECTRA OF STARS AND NEBULAE

The stars have been classified according to their spectral type into a number of groups (by Harvard Observatory) labelled *O*, *B*, *A*, *F*, *G*, *K*, *M*, *R* and *N*, which follow the order of their surface temperatures, those in the class *O* having the very highest surface temperatures and those in the class *N* the very lowest.

M. N. Saha showed how to calculate the way in which the ionization of an element depends on its temperature. The higher the temperature the more violent are the collisions between the atoms and the greater the extent to which they are denuded of their circumnuclear electrons. We thus get an explanation, as Saha showed,* for the spectral types associated with various surface temperatures.

Huggins was able to distinguish (1862) between two types of nebulae. One of these had a greenish appearance, due to two green lines which had not been observed in the spectrum of any terrestrial source. They were ascribed to a hypothetical element, appropriately named *nebulium* (*vide* Chapter XVI), but are now known to be due to oxygen atoms which have lost two of their complement of electrons. These green nebulae consist of masses of widely diffused gas and the oxygen atoms in them have been robbed of two of their electrons by the light from a neighbouring very hot star—light the photons in which have a high frequency and correspondingly great energy.

The other type of nebulae had been observed, before Huggins started his work, by Sir William Herschell about the end of the eighteenth century, and later by Lord Rosse, who had constructed and erected at Parsonstown in Ireland what was then the finest telescope in the world. He described many of these WHITE NEBULAE (also called SPIRAL NEBULAE because of their appearance).† Huggins noticed that their spectra were just like the spectrum of the light from the Milky Way. None of these—Herschell, Rosse or Huggins—had any means of estimating how far away these nebulae of either sort were.

THE SPIRAL NEBULAE

In Victorian days many romances of science were written, but not one of them rivals the story of the investigation of the spiral or white nebulae since about 1917. The happy discovery, made by Miss Henrietta Leavitt, of Harvard Observatory (1912), that all those variable stars called cepheid variables (after δ *Cepheus*, the first one to be noticed), which have the same period of

* M. N. Saha: *Phil. Mag.*, pp. 472, 809 (1920); 41, p. 267 (1921); *Proc. Roy. Soc.*, 99, p. 135 (1921).

† *Phil. Trans.* (1862).

variation, also have the same candle-power, has made it possible to determine distances which are enormously greater than any previously measured. It may be explained here, parenthetically, that Miss Leavitt noticed that all the cepheids in the Lesser Magellanic Cloud which had the same apparent brightness, also had the same period of variation. Now it could be assumed with some confidence that these cepheids were practically equally distant, so that it was almost certain their candle-powers were equal. If therefore the distances of comparatively near cepheids are measured by the old parallax method, it is simply a matter of comparing the apparent brightness of a distant cepheid with that of a near one of known distance and of the same period to get the distance of the former. In this way the dimensions of the Galaxy have been estimated. Its diameter exceeds 100,000 light years. E. P. Hubble, of Mount Wilson Observatory, noticed cepheid variables in some of the spiral nebulae, among others in the Great Nebula in Andromeda (1924). He was thus enabled to estimate their distances away. It turned out that they are all of them beyond our Galaxy (that in Andromeda being 890,000 light years away) and that in fact they are galaxies similar to and comparable in size with ours. They appear to be uniformly distributed in space. This means that the number per unit volume is constant (approximately) provided the unit volume is not too small. The average separation between neighbouring spiral nebulae or galaxies is about two million light years.

There is no doubt that the stars in the Galaxy are revolving about a centre situated in the direction of the constellation Sagittarius and between 30 and 40 thousand light years from the Sun. The flattened shape of the Galaxy, and also of the spiral nebulae, suggests that they are in rotation. The observations of J. C. Kapteyn, of Gröningen, indicated (1905) that the motions of the stars in the Galaxy were not random motions.

In 1913 the Swedish astronomer, Charlier, had come to the conclusion, on the ground of observation, that the plane of the ecliptic was moving or rotating relatively to the stars in the Milky Way with a period which he estimated to be 370 million years. We may, quite roughly, think of the earth, in motion in its orbit, as a gyroscope. Its axis, or a straight line through the Sun perpendicular to the plane of the ecliptic, should maintain an invariable direction relatively to an inertial reference system.

So we understand how Charlier's observations led to the inference that it is the stars in the Milky Way which are rotating.

The Dutch astronomer, W. de Sitter, predicted that observation would probably reveal that the spiral nebulae were receding from us.* I believe de Sitter's theory predicts recession *or approach* and that he thought recession to be the more probable of the two possibilities. His theory indicated a velocity of recession proportional to the distance of the nebula. The prediction has been triumphantly vindicated, chiefly by Hubble at Mount Wilson, by observing and measuring the slight difference between the positions of the nebular spectral lines and those of the corresponding lines in a terrestrial source, or in the Sun's spectrum. The nebular lines are shifted a little towards the red, and Doppler's principle enables the associated velocity of recession to be calculated. It might of course be argued that the spectral shift is due to some other cause than a recessional motion, and indeed it is probably true that a very small part of it has a cause which is precisely analogous to that of the red shift predicted by Einstein and which I have already described. We may, however, feel confident that all except a quite negligible part of the shift is a measure (Doppler's principle) of a velocity of recession.

The most distant of these spiral nebulae that I have heard of has been estimated to be 500 million light years away from us. Naturally there is some uncertainty in this estimate, but probably an allowance of 20 per cent for this is more than ample and it is not a wild statement when one says that the light from the nebula which formed the photographic negative in the 100-inch telescope at Mount Wilson started off on its journey 500 million years ago, and its velocity of recession is, or more certainly was, 500 million years ago, about 50,000 miles per second. In fact the velocity of recession in miles per second is equal, approximately, to

$$100 \times \text{Distance}$$

when the unit of distance is a million light years. Meanwhile a 200-inch telescope is being completed and mounted in California.†

* W. de Sitter: "On Einstein's Theory of Gravitation and its Astronomical Consequences," *Monthly Notices of the Royal Astronomical Society*, 76, p. 699; 77, p. 155; 78, p. 3 (1916–1917).

† Since writing this sentence I have read of still more distant nebulae which have been observed with it.

THE EXPANDING UNIVERSE

In 1917 Einstein published a paper* in which he assigned finite dimensions to the spatial world. Though the universe as he imagined it was not an expanding one, it is very relevant in the study of any cosmical speculations based on his type of relativity theory. Einstein's paper has three outstanding characteristics: the greatness of the result achieved in it; its lucidity; and the pleasing elegance of its form. It begins by a discussion of the formidable difficulties in the way of the assumption that the spatial world is infinite. I cannot deal with that here, but I can, I think, give a correct indication of the finite world as Einstein imagined it. According to the gravitational theory which he had published some two years earlier, gravitation is bound up with the curvature of the Riemannian continuum (Chapter XII) and this in its turn is conditioned by masses. If we therefore assume that material masses are uniformly spread through the universe—this indeed appears to be the case if we ignore local irregularities and measure density with a large unit of volume—we are led to assume a uniform curvature and a finite volume for the universe. First of all let us study the meaning of curvature, more particularly uniform curvature. We may exemplify it by a spherical surface. Any point on a *surface* is determined by *two independent co-ordinates*. Think, for example, of a plane surface and rectangular axes of co-ordinates. In the case of a spherical surface, for example, these co-ordinates might be latitude and longitude, but quite obviously another way of fixing or defining a point on a spherical surface has its peculiar merits: we may use three-dimensional rectangular axes of co-ordinates, X, Y, Z, with the origin at the centre of the sphere. Naturally, when our concern is confined to the *surface* of the sphere, the three co-ordinates, x, y and z, of a point are *not independent*. There is the relation between them

$$x^2 + y^2 + z^2 = R^2, \qquad \text{(XXI—1)}$$

R being a constant, namely the radius of the sphere. Or, to put

* "Kosmologische Betrachtungen znr Allgemeinen Relativitätstheorie," *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (1917). This paper has been translated by W. Perrett and G. B. Jeffery and appears in a volume entitled *The Principle of Relativity* (Methuen).

it in another way: $1/R$ is a constant, namely the CURVATURE of the surface. Note, too, that the *area* of the spherical surface is *finite*. Now the points in Einstein's three-dimensional curved space are determined by *four* co-ordinates between which the following relation holds:

$$x^2 + y^2 + z^2 + \xi^2 = R^2, \qquad \text{(XXI—1A)}$$

R being a constant, namely the radius of Einstein's spatial world. His own words, as translated by Perrett and Jeffery are: "The points of this hyper-surface form a three-dimensional continuum, a spherical space of radius of curvature R." The volume of this spherical space is limited in precisely the same kind of way as the area of a spherical surface is limited. In fact it is not terribly difficult to show that the

$$\text{Volume of Einstein's World} = 2\pi^2 R^3.$$

In 1917 not only did it seem likely that the material in the universe was uniformly distributed when relatively small local irregularities were ignored, but it also seemed likely that the motion in it might be ignored. The former of these two possibilities has been confirmed since, but not the latter. Einstein assumed them both and was thus inevitably led to a relation between the radius of curvature of his world and the (assumed uniform) density, ρ, of the material in it. As I write it, this relation is

$$\frac{1}{R^2} = \frac{4\pi\kappa\rho}{c^2},$$

in which κ means the Newtonian constant of gravitation.*

The uniform curvature of the world, as Einstein realized, is not compatible with his original gravitational equations; but he found a slight modification of these latter which did conform to the uniform curvature condition and yet left the important results which had already been achieved unaffected, provided R is sufficiently big. Indeed other reasons, quite independent of the suggestion of a spherical world, would have suggested the generalization.

* Einstein's formula, as given in Perrett and Jeffery's translation, is

$$\frac{1}{R^2} = \frac{\kappa\rho}{2};$$

but his units assume $c = 1$ and his κ means constant of gravitation multiplied by $8\pi/c^2$.

The astronomer H. Shapley, of Harvard, estimated the density ρ to be equal to 10^{-30} gm./cm.^{3} When we substitute this value in the formula for R we find

$$R = 3{\cdot}26 \times 10^{28} \text{ cm.}$$

or

$$R = 3{\cdot}45 \times 10^{10} \text{ light years.}$$

Furthermore, the mass in Einstein's world must be expressed by

$$Mass = Volume \times Density$$

or

$$Mass = 2\pi^2 R^3 \times \rho,$$

and consequently $Mass = \dfrac{\pi R c^2}{2\kappa}$

and when we substitute for R, κ and c we find

$$Mass\ of\ Universe = 6{\cdot}8 \times 10^{56} \text{ grammes}$$
$$= 6{\cdot}8 \times 10^{50} \text{ tons.}$$

Very likely Shapley's estimate of ρ is a bit too small.

Einstein's world, or rather his space-time, is often appropriately described as cylindrical. He gave the uniform curvature to space only and did not interfere with the co-ordinate which I represent by w (the time). As an illustration let us think of a plane rectangular sheet of paper on which two sets of parallel straight lines are drawn: one set (a) parallel to one edge of the sheet and the other (b) parallel to the other. If now we bend the sheet so that the set (a) remain straight lines while the set (b) are curved, we have a cylinder.

It is now appreciated that this cylindrical world is *unstable*, but, as we shall see after we have studied the cosmological speculations of W. de Sitter, this is perhaps its most interesting feature. In de Sitter's world the *space-time* continuum has a uniform curvature. Thus

$$x^2 + y^2 + z^2 + w^2 + \xi^2 = R_s^2, \qquad \text{(XXI—2)}$$

in which R_s is the constant radius of curvature and x, y, z and w have the meaning ascribed to them in the description of Minkowski's space-time continuum. One consequence of the hypothesis (XXI—2) is of great interest; it follows from it, as I shall show, that the spatial world is expanding (or contracting) and it led de Sitter to suggest that this might manifest itself by a recession (or approach) of the spiral nebulae. A number of marks or dots on the surface of a toy balloon will separate from one another when

the balloon is blown up, i.e. when it expands. In order to appreciate the nature of de Sitter's world, let us begin (as we have done with Einstein's world) with an ordinary sphere. Imagine a section of the sphere by a plane perpendicular to, shall we say, the Z axis. The radius of this section, which is of course a circle, will vary from one value of the co-ordinate, z, of the plane, to another. So too when we consider the generalization of a sphere which (XXI—2) represents and imagine a locus for which w has a particular value. This locus, or "section", will vary in dimensions from one value of w to another, i.e. from one time to another, since w is a measure of the time. Let us write (XXI—2) in the form

$$x^2 + y^2 + z^2 + \xi^2 = R_s^2 - w^2.$$

At a given instant of time, which means a given value of w, the right-hand side has a particular value whatever may be x, y, z or ξ individually. In fact we may replace $R_s^2 - w^2$ by R^2 and R will be the radius of a spherical spatial world rather like Einstein's. Notice that since

$$R^2 = R_s^2 - w^2$$

we have

$$R^2 = R_s^2 + c^2t^2.$$

R_s is a constant, consequently R^2 increases as t^2 increases. A small increment of t^2, on the right, is associated with a corresponding increment of R^2 on the left; thus

$$\Delta(R^2) = c^2\Delta(t^2).$$

This is equivalent to

$$\frac{\Delta R}{\Delta t} = c^2\,\frac{t}{R}, \qquad \text{(XXI—3)}$$

or in words: the rate of change of the radius of curvature of the spatial world is proportional to t/R, t being the time as measured from the instant which is determined by the centre of the de Sitter hypersphere and R being the radius of curvature of the spatial world at the instant, t. R is necessarily positive and it is therefore clear that the rate of change expressed by (XXI—3) would be negative if the instant, $t = 0$, has not yet arrived, for in that case t would be negative. The fact that the distant spiral nebulae are actually receding indicates that the velocity (XXI—3) is positive.

It would take me too far afield to derive an expression for the

velocity of a receding spiral nebula or to discuss adequately the law of gravitation in de Sitter's world. If we use the PROPER TIME the velocity of recession of a spiral nebula whose distance from us is r is proportional to r/R_s. It is strictly proportional to r, but if we use time in the usual meaning of the term the velocity of recession appears to be

$$c\frac{r}{R_s}\left(1-\frac{r^2}{R_s^2}\right)$$

and is proportional to r so long as r^2/R_s^2 is small. For reasons that cannot be given here de Sitter's world must be an *empty* world. Notwithstanding this it may very well approximate closely to the actual world in which the average density of matter is negligibly small by comparison with the densities which may be observed locally. The constant $3/R_s^2$ is usually represented by the letter λ and called the COSMIC CONSTANT.

It is not unreasonable to indulge in speculation about the cosmos. Attempts have been made by A. Friedmann (1922) and by the Abbé G. Lemaitre (1927) and more recently by many others, Eddington, McVittie, McCrea and Sen, to generalize the efforts of Einstein and de Sitter. The picture of the universe which Eddington favoured, and which appeals to me, is a superposition of those of Einstein and de Sitter. A long time ago, many thousands of millions of years, the world may have been in the condition described in Einstein's picture. The familiar gravitational attraction in it is balanced by the "gravitational" repulsion characteristic of the pure de Sitter world. This state of affairs is unstable and one or the other, attraction or repulsion, will predominate. The latter has happened and the universe is approaching more and more to the de Sitter type—the empty world.*

This all too brief account of the expanding universe must not close without reference to Professor E. A. Milne's *Kinematic Relativity*. Milne's investigations were suggested, I believe, by the fact of the recession of the extra-galactic nebulae. He makes few initial assumptions about the nature of space-time. In his own words: "The leading idea in our work is not that of transformations of *co-ordinates* but of transformations from *observer*

* A. S. Eddington: *The Mathematical Theory of Relativity* (Cambridge); W. Wilson: *Theoretical Physics*, Vol. iii (Methuen).

to *equivalent observer*. . . ."* He makes use of a principle which he terms appropriately the COSMOLOGICAL PRINCIPLE; stated by Einstein in the form "Alle Stellen des Universums sind gleichwertig,"† which may perhaps be freely translated by "One place in the universe is just as good as another". In brief Milne imagines equivalent observers spread through the universe. They are equivalent in the sense that "the totality of observations A can make on B can be described by A in the same way as the totality of observations which B can make on A can be described by B". He idealizes the universe to a system of particles. From this starting point, combined with the conventional fixing of the velocity of light to be a constant, he deduces a remarkable theory. If I have understood it correctly each equivalent particle is an idealization of a spiral nebula. It has always had its unique velocity of recession from a given equivalent observer either attached to it (in which case it is zero) or attached to some other equivalent particle. For Milne each extra-galactic nebula has always had the same velocity of recession (at least approximately). It is as if the universe began with all its mass concentrated in some relatively small region and a vast explosion occurred. Those fragments with the biggest velocities have travelled farthest. A simple computation indicates that this "explosion" must have happened about 2000 million years ago.

The vast majority of theoretical physicists incline to the Einsteinian type of relativity theory, but Milne's theory is rather profound and the significance of the distinction he makes between the transformations of the earlier relativity and those from observer to equivalent observer deserves to be very carefully studied.

* E. A. Milne: *Relativity Gravitation and World-Structure* (Oxford, Clarendon Press).

† *Sitzungsberichte d. Berliner Akad.* (1931). Milne would, I believe, regard this as an inadequate statement of the principle.

EPILOGUE

EVERY science of the experimental and observational kind is something more, indeed very much more, than a mere accumulation of facts. An essential part of it is its theory, the intention of which is to make of it a progressively more unified and coherent structure within which the relationship between one fact and another is revealed, and in this record of a hundred years I have tried to show not only how many remarkable new phenomena were discovered, but also how the unifying theoretical framework has gradually been built up.

The old ideal, as we have seen, was the mechanical one. The phenomena of physics and chemistry were believed to be matter and motion phenomena, conforming to mechanical principles laid down by Sir Isaac Newton. At the beginning of the period no serious difficulty was encountered in interpreting thermal phenomena in this way, and Clausius, Thomson, Neumann and other scientific worthies of that time called the newly developing science of heat (our thermodynamics) the *mechanical* theory of heat. Till 1860 the theoretical unification of the phenomena of light was like a triumphal progress of the old mechanics. It seemed to be finally settled that light consisted of waves in a quasi-material medium (the aether) not very different in their nature from distortional (transverse) waves in an elastic solid. Indeed the aether was assumed to have rigidity and inertia just as an elastic solid has. Students are still (very properly) introduced to the theoretical aspects of light by way of the old elastic solid theory. It represented important features of light correctly, and the phenomena of radiant heat, as they were known at the time when Planck was born, constituted a very natural and obvious link between optical and thermal phenomena.

Electromagnetic phenomena refused to be nicely linked up with those of light and the rest of physics in the old mechanical way. They conformed to such fundamental principles as the conservation of energy, but no expressions in mechanical terms could be found for characteristic electromagnetic quantities like

electric charge, magnetic induction, and so on; and Maxwell's electromagnetic theory did not seem to his contemporaries, Kelvin and others, to promise anything of the kind. Naturally it was slowly and reluctantly accepted, but Hertz's experiments and the impressive feature of the theory—the requirement that electromagnetic waves in free space should travel with the same velocity as light—began to convince the physicists that it represented a real advance. It made it certain that optical phenomena were actually a small part of the comprehensive group of electromagnetic phenomena. Visible light waves, we now appreciate, occupy roughly the octave from 4000 to 8000 Ångström units in the vast range of wave-lengths of electromagnetic waves extending from the broadcasting waves of hundreds of metres in wave-length on one side to the shortest gamma-ray wave-lengths on the other.

Maxwell's wonderful synthesis assigned to electromagnetic waves, including light waves, a character never before contemplated for waves. They were indeed still regarded as waves in the aether, but the displacements associated with them were not displacements of the strange substance of the luminiferous medium—as, for example, the displacements in the transverse (or longitudinal) waves in an elastic solid are displacements of the material of the solid—but DISPLACEMENTS OF ELECTRICITY.

Arthur Schuster, a distinguished physicist, wrote in 1904, "There is at present no theory of Optics in the sense that the elastic solid theory was accepted fifty years ago. We have abandoned that theory, and learned that the undulations of light are electromagnetic waves differing only in linear dimensions from the disturbances which are generated by oscillating electric currents or moving magnets. *But so long as the character of the displacements which constitute the waves remains undefined*, we cannot pretend to have established a theory of light."* This passage shows how strongly entrenched was the old ideal, even so recently as 1904. It was difficult to imagine any physical happening that could not be described in mechanical terms. Maxwell himself believed that his displacement would sooner or later be revealed as a clearly defined mechanical thing, and I have a little impatience with Eddington's caricature of the

* Preface to the first edition of Sir Arthur Schuster's *An Introduction to the Theory of Optics* (Arnold, London, 1904). The italics are mine.

intellectual level of the nineteenth century physicist, in which he describes him as saying to himself, on having got on the track of an explanatory piece of mechanism: "Here we are getting to bedrock. . . . This is ultimate reality."* The old confidence in mechanical "bedrock" foundations was of course misplaced, but only in the mistaken belief that the Newtonian scheme was sufficient. The ideal of the old natural philosopher was not *essentially* different from that of the present-day physicist—he was striving to incorporate the phenomena known to him in the grand and imposing unity of Euclid's geometry and Newton's mechanics. The present-day physicist is also striving after such a unity, but he has ceased to be dominated by the belief that Euclidean geometry and Newtonian mechanical principles are established for all time.

After Maxwell's day, when it had been definitely proved (Rowland) that the motion of a charged body is accompanied by a magnetic field—it is in fact equivalent to an electric current—J. J. Thomson and his contemporaries appreciated that the electric charge on a body endows it with an additional mass or inertia. This led to an effort—an unsuccessful one—to account for all mass by assuming for it an electromagnetic origin, and to replace the old mechanical ideal by another, in which electricity, rather than "matter", became the "building material" of which the physical world was made. The effort reached its extreme elaboration in Max Abraham's electron theory (1903). It had to be rejected on several grounds, but mainly because it did not give the correct law for the dependence of mass on velocity and because of the great promise of the theory of Lorentz, which in Einstein's hands developed into the special theory of relativity.

The magnificent generalization of Einstein, which removed the peculiar mystery that had enshrouded gravitation, *is still mechanics*, though of a more general kind which includes Newton's as a limiting case. Even this did not incorporate electromagnetic phenomena. They conform to it—so far as their macrophysical manifestations are concerned—but they are not an organic part of it. There is at present no general agreement as to which of the various suggested solutions of the problem of unifying electromagnetic and other phenomena in a still more general relativistic synthesis is the correct one. The most

* A. S. Eddington: *Science and the Unseen World.*

impressive is Kaluza's form of relativity in which electric charge appears as the fifth component of momentum in the five-dimensional continuum on which it is based.

When we turn to small-scale, or microphysical, phenomena we get an impression of impenetrable obscurity. We have indeed learned much about atoms, protons, neutrons, electrons, etc., and the laws which describe their behaviour; but these laws are statistical in their nature—of the same type as laws based on chance and large numbers, such for example as Maxwell's law of distribution in the kinetic theory of gases. They have been unified almost completely in the beautiful scheme of quantum or wave mechanics. But there is a profound difference between the type of picture presented by the kinetic theory as Maxwell and Boltzmann thought of it, and that presented by quantum mechanics. The former is CAUSAL, while the latter is, or seems to be, ACAUSAL. By way of illustration, the phenomenon of the alternation of day and night is causal. It is in fact caused by the rotation of the earth relative to the sun, but the small-scale phenomena of the physical world do not reveal the causality, if it exists, to which the statistical laws of quantum mechanics owe their origin, and some great authorities, like Johann von Neumann,* maintain that there cannot be any such causal laws. There are, he contends, no "hidden parameters" from which the statistical laws can be inferred. The meaning of "hidden parameters" is illustrated by the kinetic theory. The gas laws such as Boyle's or van der Waals' rest on the assumption that a gas consists of molecules, each having at any instant a definite position and momentum—these are hidden parameters—which determine causally, according to the laws of mechanics, the state of affairs in the gas and its behaviour at subsequent times. Von Neumann, however, claims to have proved that the existence of hidden parameters and causal laws is incompatible with the known and accepted laws of quantum mechanics.

However this may be, there is one remarkable causal feature in wave mechanics: the de Broglie wave, representing an electron for example, changes and behaves according to strictly causal laws (see equation XVIII—7). On the other hand it does

* J. von Neumann: *Mathematische Grundlagen der Quantenmechanik* (Dover Publications, New York). A French translation is published by the Librarie Alcan (Paris).

not appear to be identifiable with the *physical reality*, but is merely something which informs us about the *probabilities* of the results of measurements before they have been made.

No doubt future discoveries will throw light on the question as to whether there are causal laws (hidden from us now) behind the statistical ones that we know, and we may leave it and turn to the possibilities of physical inquiry in the immediate future. First of all it seems certain that the investigation of nuclear phenomena will be vigorously continued. Probably new kinds of atoms (new elements) will be produced artificially as have been already the elements neptunium and plutonium.* Some may be capable of fission and may be used in the manufacture of even more formidable bombs than those that threaten humanity now. It is certain that we shall learn a great deal more about the utilization of atomic energy for industrial and beneficent purposes and about the risks attached to it. And generally the technological utilization of those physical phenomena that are already well known will be developed in directions and to an extent far beyond anything we can foresee.

What we may call the middle regions of physics have been fairly exhaustively explored, and the scope for those whose interest lies in pure science, apart from its utilization for practical ends—and it is to be hoped that there will always be inquirers who will pursue physics and other sciences for their own sake—now seems to be at the extremes, e.g. in the study of extremely low-temperature phenomena and of the opposite extremes of nuclear physics and cosmical phenomena, in the investigation of cosmic radiation, and so on.

There is also another direction in which physical science may conceivably expand. There has long been a question as to whether the phenomena in a living organism are exclusively physico-chemical or whether—to use a perhaps obsolete nineteenth-century expression—there is operative in it a VITAL FORCE. If we had to decide *now*, i.e. on the basis of our present knowledge, in favour of the one or the other of these alternatives we should have to adopt the former. There appears to be no

* Americium and Curium, atomic numbers 95 and 96, have recently been discovered. It has been maintained (H. T. Flint) on theoretical grounds that atoms with a higher atomic number than *ca.* 97 cannot exist.

evidence of any phenomena, in a living organism, which cannot be classed as physical in the broad sense of the term.*

New types of law, different from those with which we are now acquainted in the inorganic world, may indeed be expected and discovered in living organisms in the future, but it seems likely that they will be found to belong to the same category as the laws operating in the inorganic world. So we may be concerned in biology, not only with biophysics in the sense of physics applied to biological phenomena, but possibly with new groups of purely physical phenomena.

There are "quantum jumps" in living cells. They are the MUTATIONS of the geneticist, discovered by the Dutchman de Vries, about two years after Planck published his quantum theory, but not then suspected to have any connection with a revolutionary physical theory. Even the impressive orderliness manifested in a living organism—one thinks of that superb structure, the human brain—appears to have the same type of statistical foundation as we find in the inorganic world. The familiar Maxwell–Boltzmann exponential

$$e^{-\frac{E}{kT}}$$

would seem to be as dominant in the organic as in the inorganic world. If we regard E as the energy requisite to effect some "quantum jump" or other, e.g. some isomeric change,† then the bigger E/kT the more seldom will such a change occur. Low temperature is associated with solidity and permanence and the measure of the stability of a structure is the bigness of the amount of energy needed to make it topple over. In fact the statistical basis of physics is compatible with extreme stability and orderliness.

I am not ignoring the fact that, in living organisms certainly, and possibly even in the inorganic world, there is a boundary where physical science ends and where we are confronted with purpose, volition, sense-perception, and so on. These things

* *Vide* E. Schrödinger's *What is Life?* (Cambridge University Press) and F. G. Donnan: "Zur Frage der Eigengesetzlichkeit der theoretischen Biologie," *Angewandte Chemie*, 52, p. 469 (1939).

† Ethyl alcohol and di-methyl ether furnish a simple instance of ISOMERISM (ἴσος, same; μέρος, part). Both are represented by C_2H_6O; but with their atoms differently combined.

appear to be outside the physical world and I find it impossible to distinguish the rôle of the physicists' percepts from that of symbols used in a code: the code in terms of which his information about the physical world is conveyed to him. He starts out, as Max Planck has said, from the basic METAPHYSICAL ASSUMPTION that a world exists which is independent of him and his affairs and of his investigation of it. This is his physical world; and his interpretation of the results of all his experimental activity is made to conform to the fundamental postulate of its independent existence.

Physic of Metaphysic begs defence.

BIBLIOGRAPHY

THOSE works in the following two lists which are calculated to be most useful to students of physics are marked with an asterisk, and two asterisks indicate that the particular book is easy to read or suitable for beginners. List I contains works referred to in the text in the order in which they are first mentioned. Many of these, but not all, are mainly of historical interest. List II contains books of all degrees of difficulty from the most elementary kind to works of a highly advanced character. Where a date is given it indicates, to the best of my belief, when the book first appeared.

I

SIR ISAAC NEWTON: *Philosophiae naturalis principia mathematica* (1687).

SADI CARNOT: ***Réflexions sur la Puissance Motrice du Feu* (1824).

JULIUS ROBERT MAYER: ***Die Mechanik der Wärme* (Stuttgart, 1867).

JAMES PRESCOTT JOULE: *Collected Papers* (Physical Society, Vol. 1, 1884; Vol. 2, 1887).

CARL NEUMANN: *Vorlesungen über die mechanische Theorie der Wärme* (Leipzig, 1875).

A. W. BARTON: ***A Text Book on Heat* (Longmans, Green & Co., 1933).

W. WILSON: **Theoretical Physics*, Vol. 1 (Methuen & Co., 1931).

R. CLAUSIUS: **The Mechanical Theory of Heat* (W. R. Browne's translation, Macmillan & Co., 1879).

S. G. STARLING: **Electricity and Magnetism for Advanced Students* (Longmans, Green & Co., 1912).

J. FOURIER: *Théorie Analytique de la Chaleur* (1822), translated by A. Freeman (Cambridge).

J. K. ROBERTS: **Heat and Thermodynamics* (Blackie & Son, 1928).

A. EINSTEIN: **Investigations on the Theory of the Brownian Movement* (Methuen & Co., 1926).

J. PERRIN: **Les Atomes* (Paris, Alcan).

M. KNUDSEN: **Radiometer Pressure and Coefficient of Accommodation* (Copenhagen, 1930).

L. BOLTZMANN: *Gastheorie* (Barth, Leipzig, 1896).

J. WILLARD GIBBS: *Elementary Principles in Statistical Mechanics* (1901).

N. ERNEST DORSEY: **The Velocity of Light* (American Philosophical Society, Philadelphia, 1944).

J. LARMOR: *Aether and Matter* (Cambridge, 1900).

SIR ISAAC NEWTON: *Opticks* (1704). Messrs. G. Bell & Sons have issued a reprint of the fourth edition (1931).

E. EDSER: ***Light* (Macmillan & Co.).

R. W. WOOD: **Physical Optics* (The Macmillan Co., 1905).

MAURICE DE BROGLIE: **Les Rayons X* (1922). English translation published by Methuen & Co.

P. DRUDE: **Lehrbuch der Optik* (1900). English translation published by Longmans, Green & Co.

W. WILSON: **Theoretical Physics*, Vol. 2 (Methuen & Co., 1933).

H. A. ROWLAND: *Physical Papers* (Johns Hopkins Press, Baltimore, 1902).

W. E. WILLIAMS: ***Applications of Interferometry* (Methuen's Monographs on Physical Subjects).

M. JAMIN: *Cours de Physique* (Gauthier-Villars, Paris).

A. SCHUSTER and J. W. NICHOLSON: **Introduction to the Theory of Optics* (Arnold, 1904).

R. GLAZEBROOK (ed.): *Dictionary of Applied Physics* (Macmillan & Co.).

J. C. MAXWELL: *A Treatise on Electricity and Magnetism* (Clarendon Press, 1873).

O. W. RICHARDSON: *The Electron Theory of Matter* (Cambridge, 1914).:

E. C. STONER **Magnetism and Atomic Structure* (Methuen & Co.).

H. A. LORENTZ: *Versuch einer Theorie der Elektrischen und Optischen Erscheinungen in Bewegten Körpern* (Leiden, 1895).

CARL NEUMANN: ***Ueber die Principien der Galilei-Newtonschen Theorie* (Teubner, Leipzig, 1870).

H. G. WELLS: ***The Time Machine* (Ernest Benn).

W. PERRETT and G. B. JEFFERY: *The Principle of Relativity* (Methuen & Co.). This is a collection of original papers on relativity translated from the German. It first appeared under the title *Das Relativitätsprinzip* (Teubner, Leipzig, 1913).

W. WILSON: **Theoretical Physics*, Vol. 3 (Methuen & Co., 1940).

A. S. EDDINGTON: **The Mathematical Theory of Relativity* (Cambridge, 1923).

J. C. MAXWELL: ***Theory of Heat* (Longmans, Green & Co.).

W. NERNST: **The New Heat Theorem* (Methuen & Co., 1926).

J. R. PARTINGTON: **A Text Book of Thermodynamics* (Constable, 1913).

A. SOMMERFELD: *Atombau und Spektrallinien* (Vieweg, Braunschweig, 1919). An English translation has been published by Methuen & Co.

C. G. J. JACOBI: *Vorlesungen über Dynamik* (Reimer, Berlin, 1866).

D. SHOENBERG: **Superconductivity* (Cambridge, 1938).

R. A. MILLIKAN: **The Electron* (University of Chicago Press, 1917).

O. W. RICHARDSON: **The Emission of Electricity from Hot Bodies* (Longmans, Green & Co., 1916).

MAX BORN: **Atomic Physics* (Blackie & Son).

L. DE BROGLIE: **Wave Mechanics*, translated by H. T. Flint (Methuen & Co.).

M. DE MAUPERTUIS: ***Essay de Cosmologie* (Elie Luzac, Leide, 1751).

E. SCHRÖDINGER: *Abhandlungen zur Wellenmechanik* (Barth, Leipzig). A collection of original papers which have appeared in English translation (Blackie & Son).

P. DIRAC: **The Principles of Quantum Mechanics* (Clarendon Press, 1930).

F. W. ASTON: **Mass Spectra and Isotopes* (Arnold & Co.).

E. A. MILNE: *Relativity, Gravitation and World Structure* (Clarendon Press, 1935).

J. VON NEUMANN: *Mathematische Grundlagen der Quantenmechanik* (Dover Publications, New York). Also in French (Librairie Alcan, Paris).

E. SCHRÖDINGER: *What is Life?* (Cambridge, 1948).

II

H. S. ALLEN: **The Quantum and its Interpretation* (Methuen & Co., 1928).

E. N. DA C. ANDRADE: **The Structure of the Atom* (G. Bell & Sons, 1923).

A. W. BARTON: ***A Text Book on Light* (Longmans, Green & Co., 1939).

J. A. CROWTHER: **Ions, Electrons and Ionizing Radiations* (Arnold, 1919).

A. W. DUFF: ***A Text Book of Physics* (J. & A. Churchill).

A. S. EDDINGTON: *Space, Time and Gravitation* (Cambridge, 1920).

A. EINSTEIN: ***The Theory of Relativity* (Methuen & Co., 1920).

H. E. HADLEY: ***Magnetism and Electricity* (Macmillan & Co., 1906).

E. MACH: *Die Mechanik in ihrer Entwickelung* (Brockhaus, Leipzig, 1883). *Principien der Wärmelehre* (Barth, Leipzig, 1896).

F. H. NEWMANN and V. H. L. SEARLE: **The General Properties of Matter* (Ernest Benn, 1928).

MAX PLANCK: **Vorlesungen über Thermodynamik* (Veit, Leipzig, 1897). An English translation exists.

B. RUSSELL: ***The ABC of Atoms* (1923). *The ABC of Relativity* (1926). Both published by Kegan, Paul.

E. RUTHERFORD, J. CHADWICK and C. ELLIS: **Radiations from Radioactive Substances* (Cambridge, 1930).

E. SCHRÖDINGER: *Wave Mechanics.* Four lectures at the Royal Institution (Blackie & Son, 1928).

A. SOMMERFELD: *Three Lectures on Atomic Physics* (Methuen & Co., 1926).

H. SPENCER JONES: ***Worlds without End* (Hodder & Stoughton, 1935).

S. G. STARLING: ***Electricity and Magnetism* (Longmans, Green & Co., 1931). ***Mechanical Properties of Matter* (Macmillan & Co., 1935).

G. P. THOMSON: ***The Atom* (Oxford University Press, 1930).

G. J. L. WAGSTAFF: ***Properties of Matter* (University Tutorial Press, 1906).

R. L. WATERFIELD: ***A Hundred Years of Astronomy* (Duckworth, 1938).

H. WEYL: *Raum, Zeit, Materie* (Julius Springer, Berlin, 1921).

W. H. WHITE: ***A Complete Physics* (Richard Clay & Sons, 1935).

E. J. WHITTAKER: *A History of the Theories of Aether and Electricity* (London, 1910).

INDEX

ABBÉ, 117
Aberration of light, 172
Absolute electrometer, 130
— manometer, 78, 79
— measurement, 118, 121, 125
— — of electric currents, 128
— space, 157, 166
— zero, 37
Absorbing power, 184
Adams, discovery of Neptune, 14
Aether, 11, 87, 93, 94, 95, 148, 167, 168
AIRY, 96
Airy's experiment, 172
AITKEN, 214
ALLEN, H. S., 240
— J. F., 271
Alpha, beta and gamma rays, 264
— body, 156, 157, 158
— particles, counting of, 239
AMAGAT, 49
Americium, 302
AMPÈRE, 145
— balance, 129
— (unit), 129
Amplitude and probability, 258
Analogy between mechanics and optics, 254
ANDERSON, 91, 275
ANDRADE, 265, 307
ANDREWS, 13, 24, 46, 47, 71
Andromeda, Great Nebula in, 290
Angle variables, 245
ÅNGSTRÖM, 60, 112
— (unit), 116
Angular momentum, 229
Anions, 203
Anomalous Zeeman Phenomenon, 240
Antares, 283
APPLETON, 147
Appleton layer, 147
ARAGO, 92, 96
ARISTARCHUS, 155
ARRHENIUS, 204
Artificial isotopes, 278
— transmutations of atoms, 268
Assemblage, 82, 84
ASTON, 266, 267
Atom bomb, 281
Atomic heat, 197
— nucleus, 276
— number, 219
— structure, 228, 238
AVOGADRO, 13
Avogadro's law, 52, 65
— number, 73
Axiom, of Planck, 39, 40
— of Thomson, 39

BALFOUR STEWART, 127
Ballistic galvanometer, 131
BALMER, 229
Balmer's formula, 229, 233
— series, 229, 237
BARKLA, 218, 219, 220, 256
BARNES, 52
BARTON, 33, 49, 54, 182
BECQUEREL, 263
BELL, 156
BENOIT, 116
BÉRARD, 16
BERNOULLI, D., 13, 62
BESSEL, 14
Beta ray emission, 270
Betatron, 273
Betelgeuse, 283
Bibliography, 305
Black body radiation, 85, 103, 182, 188
— — — problem of, 189
BLACK, JOSEPH, 50

Black surface, 185
BLACKETT, 269
— and OCCHIALINI, 275
BOHR, 230, 233, 234, 244, 263
Bohr's postulate, 233
Bolometer, 187, 188
BOLTZMANN, 62, 80, 81, 85, 135, 194
Boltzmann's constant, 64, 75, 191
— H theorem, 80
BORN, 225, 243, 258, 275
— and KÁRMÁN, 199
BOSE and EINSTEIN, 194
BOTHE and BECKER, 269
Boyle's law, 13, 22, 25, 62
— — deviations from, 48
BRACE, 154
Brackett's series, 234
BRADLEY, 11
BRAGG, W. H., 236, 268
— W. L., 236
BRANLEY, 146
BRINKWORTH, 52
British Association, 127, 130
— — (Standards Committee), 130
BROGLIE, LOUIS DE, 10, 168, 194, 225, 249, 250
— MAURICE, DUC DE, 106
Brownian movement, 75, 76
BUCHERER, 216, 217
BUNSEN, 51, 102, 103, 104

CAGNIARD DE LA TOUR, 13
CAILLETET, 49
CALLENDAR, 52, 73, 187
Caloric, 12
— conservation of, 12, 31, 33, 145
Calorimeters, 50, 51
Canal rays, 206, 207
Canonical distribution, 85
— equations, 10, 253
Carey Foster's bridge, 127
CARNOT, H., 13
— S., 12, 15, 31, 38, 40, 45, 183
— engine, 37
Carnot's function, 35, 37
— principle, 13, 33, 34, 36, 40
— proof, 32
Cascade process, 56
Cathode-ray oscillograph, 222
Cathode rays, 206, 207, 210
Cations, 203
Causality, 301
CAVENDISH, 11
Cavendish Laboratory, 267
Cepheid variables, 289
CHADWICK, 239, 267, 270
Characteristic radiation, 220, 226
Charles' law, 13, 23
CHARLIER, 290, 291
CHRISTIE, 12
CHRISTOFFEL, 179
Circuit, oscillating, 132
Clapeyron's formula, 35, 41
CLAUDE, 57
CLAUSIUS, 20, 22, 25, 31, 37, 38, 41, 42, 43, 44, 62, 63, 68, 69, 73, 204
Cloud chambers, 43, 213, 214, 269
Coblentz, 187
COCKCROFT and WALTON, 278
Coefficient of scattering, 219
Coherer, 146
COLDING, 19
Colloids, 76
Colour photography, 117
Combination principle, 230, 244
COMPTON, A. H., 219, 226
— effect, 219
— K., 225
Conduction of electricity, 203
— — — in gases, 205, 206
— — — in metals, 207
— of heat, 182
Conservation of caloric, 12, 31, 33, 45
— — electric charge, 181
Continuous flow calorimeter, 52
Contractile electron, 216
Contraction hypothesis, 152, 154, 165, 172, 175
Convection coefficient, 93
— of heat, 182
— of light, 92
COOK, CAPTAIN JAMES, 11
COOLIDGE, 222

COPERNICUS, 155
CORNU, 90
Correspondence principle, 232
— — applications of, 244
Corresponding states, theorem of, 73
Cosmical physics, 282
Cosmic constant, 296
— radiation, 274
Cosmological principle, 297
COULIER, 214
Coulomb (unit), 130
Covariance, 159
Critical data, 48
— point, 48
— pressure, 48
— temperature, 47
— volume, 48
Crookes, 79, 105, 239, 265
— dark space, 206
Crowther, 307
Crystalloids, 76
CURIE, MADAME, 264, 269
— P., 59, 60
Curie's constant, 59
Curium, 302
Current balance, 129
Cyclotron, 271, 272, 273

D'ALEMBERT, 9, 19
DALTON, 9, 23
Daniell's cell, 12
DARWIN, C., 182
DAVISSON and GERMER, 250
DAVY, 12
De Broglie waves, 255
— — causal nature of, 301
DEBYE, 52, 58, 60, 198, 205
Debye's law, 199
Degree of dissociation, 204
Degrees of freedom, 65, 197, 230, 231
DELAROCHE, 16
DELSAUX, 75
Densities of gases and vapours, 52
DE SITTER, 291, 296
De Sitter's world, 294, 295
Deuterium, 266
DEWAR, 51, 57
Diamagnetism, 135
Dielectric constant, 12, 120
— media, 137
DIETERICI, 73
Diffraction, 109
Diffusion in gases, 64, 65
Dimensions of units, 121
DIRAC, 248, 259, 275
Dispersion, 97
— and spectroscopy, 102
— anomalous, 97
Displacement hypothesis, 142
Distribution, canonical, 85
Doppler effect, 104, 291
DORSEY, 87
DRUDE, 111, 208
DUFOUR, 48
DULONG, 13, 16, 52
Dulong and Petit, law of, 197, 208
DUMAS, 53
Dyne (unit), 118

Echelon grating, 113, 114
EDDINGTON, 155, 178, 179, 181, 285, 296
EDLEFSEN, 271
EDSER, 98
Efficiency of engines, 31
— scale, 37
EINSTEIN, 75, 76, 167, 174, 177, 178, 179, 198, 216, 291, 296, 297, 300
Einstein's kinematical equations, 165
— mass-energy law, 200, 225
— theory, consequences of, 180
— world, finiteness of, 292
— — instability of, 294
— — mass of, 294
— — radius of, 293
— — volume of, 293
Electrical resistance, absolute measurement of, 123, 125, 126
Electrodynamometer, Weber's, 128
Electrolysis, 203

Electrolysis, laws of, 12
Electromagnetic mass, 212, 213
— phenomena, 181
— theory (Maxwell), 137
— waves, velocity of, 143, 144, 145
Electromagnetism, integration of, 118
Electromotive force, 123
Electron, 210
— microscope, 226, 227
— optics, 226
— orbits of atoms, 241
— theory, 212, 213
— volt, 273
Electrons, diffraction of, 250
Elementary charge, measurement of, 214
Emitting power, 185
Energy, 19
— conservation of, 21
— density, 140, 141
— equipartition of, 62, 66, 75, 197
— flow of, 145
— free, 42
— internal, 25
— kinetic and potential, 21
— of black body radiation, 188
Enthalpy, 53
Entropy, 40, 42, 53, 82, 200
— constant, 64, 85
EPSTEIN, 245
Equation of state, 23, 25
Equivalence, principles of, 176
— of heat and work, 15, 16, 38, 45
EUCKEN, 52
EUCLID, 179
Euclidean space, 176
EULER, 9
EWING, 135
Exchanges, Prévost's theory of, 12
Exclusion, principle of, 241
Expanding universe, 292

FABRY and PEROT, 115, 116
FARADAY, 12, 13, 137, 145, 246
— dark space, 206
Faraday's laws, 203
FEATHER, 270
FEDDERSEN, 134, 145
Fermat's principle, 10, 250, 251, 253
FERMI, 270
Fermi–Dirac statistics, 209·
Ferromagnetism, 135
Féry's pyrometer, 195
First law of motion, 156, 176
— — — thermodynamics, 26
Fission, 280, 281
FITZGERALD, 152, 154
— — Lorentz contraction, 216
FIZEAU, 87, 93
FLEEMING JENKIN, 127
FLINT, 250
FORBES, 60
Force per unit volume, 137, 138, 139
FOUCAULT, 89
FOURIER, 13, 60, 107, 122
Fourier's theorem, 107, 110
FOWLER, 235
FOX TALBOT, 97
FRAUNHOFER, 11, 104, 112, 282
Free energy, 42, 199
FREEMAN, 122
Free path phenomena, 68
FRESNEL, 11, 92, 93, 96, 109, 114, 148, 154, 182
Fresnel's convection formula, 166
FRIEDMANN, 296
FRIEDRICH and KNIPPING, 236
Full radiation, 186

Galaxies, 290
Galaxy, rotation of, 290, 291
GALILEO, 182
— and the Inquisition, 155
GALLE, 14
Galvanometer, 131
— of Helmholtz, 128
Gamma rays, 87, 264
Gas constant, 23
Gases, kinetic theory of, 62
GAUSS, 12
GAY-LUSSAC, 13, 23, 25, 53, 54, 55

Gay-Lussac's experiment, 26, 29
Geiger counter, 239, 265
GEISSLER, 206
General theory of relativity, 174
Generalization of first law of motion, 176
Geodesic, 176, 177
— equation of, 178
Geometrical optics, 110
Geometry, 9, 156
GLADSTONE and DALE, 100
GLAZEBROOK, 127
GOLDSTEIN, 206
Göttingen, the Seven of, 125
GOUY, 76
GRAHAM, 64, 76
Grating, 112
— concave, 113
— echelon, 113, 114
— formula, 112
Gravitation, characteristics of, 175
— constant of, 11
Gravity, intensity of, 11
GREEN, GEORGE, 12
Green nebulae, 289
GROTTHUS, 203
Group velocity, 108, 254, 255
Groups of transformations, 163, 164
Guard ring, 131
Gullivers, 77
GUTHRIE, 221

H theorem, 80
HAHN and STRASSMANN, 280
Half-value period, 268
HALLWACHS, 223
HAMILTON, Sir W., 9, 10
Hamilton's principle, 249, 250, 252, 253
HAMPSON, 57
Hanover, Ernst August, King of, 125
Harvard observatory, 288, 289
Heat engines, efficiency of, 31, 40
Heat, statistical theory of, 62, 80
Heaviside layer, 147
HEISENBERG, 249
Heisenberg's uncertainty relation, 257
Helium, atom, 234
— discovery of, 104
HELMHOLTZ, 19, 20, 21, 99, 129, 146
— galvanometer, 128
HENRY, 12
— (unit), 132
HERSCHELL, Sir W., 14, 289
HERTZ, 146, 222
HERTZSPRUNG, 283
HESS, 274
Hidden parameters, 301
Hilbert space, 259, 260
HIRN, 19, 21
HITTORF, 203, 206
HOARE, 187
HOFFMANN, 53
HOLBORN, 49
Hooke's law, 11
HUBBLE, 290, 291
HÜCKEL, 205
HUGGINS, 104, 282, 289
HUGHES, 225
HULL, 147
HÜTTEL, 91
HUYGENS, 11, 148
Huygens' principle, 109
Hydrogen atom, 230
Hysteresis, 135

Ideal gas, 25
Indicator diagram, 34
Induced radioactivity, 278, 279
Inductive capacity, 12
Inertial systems, 157, 158
Interference, 109
Interferometers, 114, 148, 149
Internal energy, 25, 27
— — of a gas, 28
Ionic charge, 195, 196
Ionosphere, 147
Isobars, 266
Isothermals, 25
Isotopes, 265, 266

JACOBI, 9, 10, 201, 245

JAMIN, 114, 115
JOHNSTONE STONEY, 211
JOLIOT, IRENE, 269
— M. F., 269
JOLY, 51
JOULE, 16, 17, 18, 19, 21, 25, 26, 28, 29, 37, 38, 45, 51, 53, 54, 55, 123, 129
— (unit), 130

KALUZA, 181, 301
KAMERLINGH-ONNES, 57, 58, 209
KANT, 10
KAROLUS and MITTELSTAEDT, 91
KATER, 11
KAUFMANN, 211, 216
KEESOM, 58
KELVIN, 9, 12, 31, 36, 44, 45, 53, 54, 124, 127, 129, 130, 131, 132
Kelvin's efficiency scale of temperature, 37
— work scale of temperature, 18, 20, 40, 64, 195
Kerr cell, 91
KERST and SERBER, 273
Kinematic relativity, 296, 297
Kinetic theory of gases, 62, 80
KIRCHHOFF, 51, 102, 103, 104, 118, 123, 125, 127, 183
Kirchhoff's law, 185
KLEEMANN, 268
KLEIN, O., 181
KNUDSEN, 78, 79
KOHLHÖRSTER, 274
KOHLRAUSCH, 130, 131, 203
KRAMERS, 246
KRÖNIG, 62, 63
KUNDT, 97
Kundt's law, 97, 98

LAGRANGE, 9
LANGE, 52
LANGEVIN, 135, 136
LANGLEY, 188
Laplace, velocity of sound, 13
LARMOR, 96, 213, 247
— precession, 247
Latent heat, Clapeyron's equation, 35
— — equations of Clausius and Rankine, 43
LAUE, M. VON, 236
LAVOISIER, 51
LAWRENCE, 271
LEES, 60, 61
LEIBNIZ, 252
Leipzig, University of, 125, 126
LEMAÎTRE, 296
Lenz, law of, 135
LE ROUX, 97
Lesser Magellanic Cloud, 290
LEVERRIER, 14
Light, 86, 106
— velocity of, 87, 166
Lilliputians, 77
LINDE, 57, 58
LINDEMANN, 52
Liouville's theorem, 83, 84
LIPPMANN, 117
Liquefaction of gases, 56
Local time, 165
LOCKYER, 104, 235, 282
LODGE, 145, 146, 152, 153
LORENTZ, 101, 148, 152, 166, 212, 213, 216, 246
LORENZ, 101, 127, 128
LOSCHMIDT, 73, 74
Loschmidt's number, 195, 196, 214
LO SURDO, 245
Low temperatures, 58
Luminiferous medium, 154
LUMMER and GEHRCKE, 116
— — PRINGSHEIM, 188, 190
LYMAN, 234
Lyman's series, 233, 237

MCCREA, 296
MACH, 167
MACLAURIN, 9
MCVITTIE, 296
Magnetic permeability and susceptibility, 59, 121, 135
Magnetism, 134
Magneton, 136

Main sequence stars, 285
MAIN SMITH, 240
MARCONI, 147
Mass defect, 277, 278
Mass-energy law, 167, 200, 277, 288
Mass spectrograph, 266
Matrix mechanics, 250, 260, 261, 262
MAUPERTUIS, 9, 10, 249, 250
MAXWELL, 61, 69, 70, 80, 81, 98, 126, 127, 129, 132, 135, 144, 145, 146, 147, 164, 182, 183, 194, 222, 228
Maxwell-Boltzmann exponential, 84, 304
Maxwell's electromagnetic theory, 137
— — — verification of, 145
— law of distribution, 80
MAYER, 15, 16, 25, 28, 38, 51
Mean free path, 69
Mechanical equivalent of heat, 15, 16, 17, 18, 22, 25, 26, 29
MEISSNER and OCHSENFELD, 209, 210
— effect, 210
Melting point of ice, depression of, 36
MENDELÉEFF, 263
Mercury's perihelion, 174, 180
Meson theory, 276
Mesons, 274, 275
Metaphysics and physics, 304
MEYER, V., 53
MICHELSON, 90, 91, 93, 113, 114, 115, 116, 148, 153, 164, 166
— and MORLEY, 93, 151, 154
Michelson's interferometer, 148, 149
Microphysical phenomena, 301
Microscopic vision, 117
MILLIKAN, 215, 216, 274
MILNE, 285, 296, 297
MINKOWSKI, 157, 168, 169, 179
Minkowski's space-time, 176, 178
MOHR, 15
Molecular weight, 52
MOSELEY, 236
Mutations, 303

NATTERER, 49
Nebulae, spiral, 14, 289, 291
— — recession of, 291, 294, 296
— green, 289
— white, 289
Nebulium, 228, 289
Neptunium, 280
NERNST, 52
Nernst rod, 101
NEUMANN, C., 30, 38, 39, 156, 157, 158
— F., 60
— J. VON, 301
Neutrino, 270, 271
Neutron, 239, 269, 270, 279
NEWMAN and SEARLE, 79
NEWTON, 9, 148, 155, 156, 157, 158, 166, 174, 176, 182
— mechanical principles of, 157
— velocity of sound, 13, 87
Newtonian relativity, 155
Newton's *Opticks* (quoted), 97
— rings, 11
NICHOLS, 147
NICHOLSON, 228, 229, 232, 240
Nicol's prism, 91
NORBERT, 112
Nuclear binding energy, 277, 278
Nucleus, atomic, 276

OERSTED, 12
OHM, 12, 122
— (unit), 122, 129
Ohm's law, 203, 205
Orbits of electrons, 238
Oscillograph, 222
Osmosis, 76
Osmotic pressure, 77
OSTWALD, 39

Parallax, 14
Paramagnetic media, 121
Paramagnetism, 135
PARTINGTON, 52, 199
PASCHEN, 190, 234

Paschen–Back effect, 248
PAULI, 240
PELTIER, 12
— heat, 44
Percepts, 304
Perfect gas, 25
Permeability, magnetic, 121
Perpetuum mobile, 20, 32, 45
— — of the second kind, 39
PERRETT and JEFFERY, 292
PERRIN, 75, 76, 210
PETIT, 13, 52
PFEFFER, 76, 77, 78
Phase integral, 231
— rule, 82
— velocity, 107
Photo-electric cell, 91
Photo-electricity, 222
— — Einstein's theory of, 224
Photon, 106, 225
— gas, 85, 192, 194
Physical reality, 302
— laws in living organisms, 303
— world, 304
PICKERING, 235
PICTET, 56
PLANCK, 10, 39, 74, 189, 191, 193, 195, 201, 282, 304
Planck's law of radiation, 191, 193
— oscillator, 193
— quantum hypothesis, 201
PLÜCKER, 206
Plutonium, 280
POINCARÉ, 158
POISSON, 152
Polarization of X-rays, 218
PONCELET, 15
Porous plug experiment, 53
Positron, 259, 275
Potential difference, 124
POUILLET, 12
Poynting's theorem, 145
— vector, 145
Pressure balance, 49
— of radiation, 286, 288
— osmotic, 76, 77, 78
PRÉVOST, 11
Principle of equipartition of energy, 197
— — equivalence, 176
— — exclusion, 241
— — least action, 249, 252
Proper (*eigen*) functions, 259
— time, 177
— (*eigen*) values, 259, 261
Protofluorine, 228
Proton, 230, 263
PROUT, 263
Pyrometers, 195
PYTHAGORAS, 171

Quantum conditions, 200, 201, 231
— — extended, 248
— mechanics, 249
— of action, 193
— — energy, 193
— jump, 231
— jumps in living organisms, 303
— theory, 10, 182

Radar, 222
Radiant heat, 182, 186
Radiation, characteristic, 220
— intensity of, 184
— pressure of, 147, 186
Radioactive decay, 265
— products, 268
Radioactivity and cosmic radiation, 263
Radiobalance, 187
Radiometer pressure, 78
Radium, 264
Radon, 268
RAMSAY, 75, 105, 239
— and SODDY, 239, 265
Range of alpha particles, 268
Rank of tensors, 160
RANKINE, MACQUORN, 19, 20, 43
— A. O., 154
Ratio of charge to mass, 210
— — E. M. to E. S. units, 130, 144
RAYLEIGH, 63, 111, 127, 128, 130, 154, 190, 191

RAYLEIGH and BRACE, 154
Reflecting power, 184
REGNAULT, 25, 49, 51, 53
Relativity, 155
— Einstein's general theory of, 174
— Newtonian, 158, 161
— principle of, 158, 167
— special theory of, 167
Residual rays, 101, 102
Resolving power, spectroscopic, 111, 113
— — of microscope, 227
Reversibility, 30
Reversible adiabatics, 41
— cycle, 31, 38
RICHARDSON, 135, 209, 217, 221, 222
Richardson's law, 221, 225
RIEMANN, 176, 179
Riemannian continuum, 176, 177, 292
— space-time, 178, 181
RITCHIE, 183
RITZ, 230, 240, 244
ROBERTS, 61
ROEMER, 11
RÖNTGEN, 207
ROSSE, 289
ROWLAND, 18, 113, 132, 210, 212
RUBENS, 101
— and ASCHKINASS, 101
— — KURLBAUM, 188, 191
— — NICHOLS, 101
RUMFORD, Count, 12
RUPP, 250
RUSSELL, B., 308
— H. N., 284, 286, 287
RUTHERFORD, 228, 263, 264, 267, 268, 269, 278
— and GEIGER, 239, 265
— — SODDY, 265
RYDBERG, 229, 230
Rydberg's constant, 233, 234, 235

SAHA, 289
Saturation vapour pressure, 47
Scalar quantities, 160
SCHLODOWSKA, MARYA, 264
SCHRÖDINGER, 10, 168, 249
Schrödinger's wave equation, 225, 257, 258
SCHUMANN, 105
SCHUSTER, 299
— and NICHOLSON, 283
SCHWARZSCHILD, 245
Science, its theoretical framework, 298
Screening constant, 237
SEARLE, 132
SECCHI, Father, 282
SEEBECK, 12
Selection rules, 244
Self-inductance, 132
SELLMEYER, 98
Semi-permeable wall, 76
SEN, 296
SHOENBERG, 210
Shower of tracks, 275
SIDGWICK, Mrs., 128
SIMON, 52
Simple group, 108
Slug, 119
SMITH, F. E., 127, 130
SMOLUCHOWSKI, 75
SOMMERFELD, 136, 200, 201, 209
Special relativity, 155
Specific heat of electricity, 45
— — — saturated water vapour, 43
— heats, measurement of, 50
— — of air, 16
— — of gases, 28, 29, 52, 65, 66, 67, 202
— — of solid elements, 52
Spectra, 228
Spectral series, 244
— terms, 230, 244
— types of stars, 288
Spectroscope, 103
Spectrum, 97
— continuous, 103
— of black body radiation, 188
SPENCER JONES, 308
Spin momentum, 240
— — of electron, 240, 241, 277
— — of meson, 277

Spinning electron, 240
Spiral nebulae, 289
Stability, statistical, 303
Stark effect, 245
STARLING, 45, 128, 187
Stationary light waves, 117
— path, 250
— principles, 250
— value, 176
Statistical equilibrium, 81, 84
— mechanics, 62, 82
— theory of heat, 62, 82
Statistics, of Bose and Einstein, 194
— — Fermi and Dirac, 209
Stefan–Boltzmann law, 186
Stefan's constant, 187
Stellar dimensions, 282
— interferometer, 283
— interiors, 285, 286
— temperatures, 282, 283, 286
STOKES, 92
Stokes' law, 215
STONER, 135, 240
Stream velocity, 69
Stroboscope, 90, 91
Superconductivity, 209
Superheated liquid, 72
Supersaturated vapour, 72
Sutherland's constant, 70

Table of electron orbits, 242
TAIT, 15, 60
Tangent galvanometer, 12
TAYLOR, Sir G., 225
Temperature, 63
— of inversion, 54
— scales of, 36
Tensors, 160
Theorem of corresponding states, 73
Thermal conductivity, 6
— — in gases, 70
— neutrons, 279
— phenomena, 15
Thermions, 220
Thermodynamics, first law of, 22, 26
— second law of, 31, 42
Thermoelectric circuit, 44
— phenomena, 12
THOMSON, Sir G., 250
— JAMES, 34, 36
— Sir JOSEPH (J. J.), 132, 205, 210
— WILLIAM (see also KELVIN), 20, 31, 34, 36, 37, 62
Thomson effect, 45
Threshold frequency, 223
Time traveller, 157
Total heat, 53
Tourmaline, 183
TOWNSEND, Sir JOHN, 214
Transformation constant, 267
TRAUBE, 76
TROUTON and NOBLE, 153
TYNDALL, 186

UHLENBECK and GOUDSMIT, 240
Uncertainty relation, 257
Units, arbitrary and derived, 118
— C.G.S., 119
— dimensions of, 121
— electromagnetic system of, 121
— electrostatic system of, 120
— fundamental, 119
— Lorentz–Heaviside, 121
— systematic, 16
University of Leipzig, 125, 126
Uranium X, discovery of, 265
UREY, 266

Values of h and k, 196
VAN DE GRAAFF, 274
VAN DER WAALS, 48, 71, 72, 73
— — — theory of, 70
VAN'T HOFF, 78, 204
Van't Hoff's factor, 204
Vectors, 160
Velocity, group and phase, 91, 92, 93
— of light, 87, 94
Viscosity, of gases, 69
Vital force, 302
VOIGT, 165, 240
Volt (unit), 129
VOLTA, 129

WARBURG, 135
Water, heavy, 267
WATERSTON, 62
WATT, 34
Wave mechanics, 10, 250, 254
Waves, characteristics of, 106
— dilatational and distortional, 87
WEBER, 125, 129, 130, 131, 134, 135, 144
Weber's electrodynamometer, 128
WEISS, 135, 136
WELLS, H. G., 157, 168
Weston cell, 130
WEYL, 181
WHEATSTONE, 89
Wheatstone's bridge, 12
WIECHERT, 211
WIEDEMANN and FRANZ, 207
WIEN, 190, 282
Wien's laws, 188, 189, 190
WIENER, C., 75
— O., 117
WILLARD GIBBS, 62, 64, 82, 194
Willard Gibbs' modules, 64, 85
WILLIAMS, 114, 116
WILSON, C. T. R., 43, 213, 214, 274
— H. A., 215
— W., 181, 222, 238, 247, 248
Wilson–Sommerfeld quantum conditions, 136, 200, 201, 202, 259
WOLLASTON, 11, 282
WOOD, 100
Work, 15, 16
— scale, 37
WORTHINGTON, 47

X-ray spectra and atomic structure, 236
X-rays, 87, 207
— characteristic, 220, 236
— polarization of, 218
— scattering of, 219

YOUNG, THOMAS, 11, 12, 19, 96, 148, 154, 182
YUKAWA, 275, 276, 277

ZEEMAN, 93
— effect, 246