UNIVERSAL
LIBRARY

OU_160163

UNIVERSAL
LIBRARY

43—30-1-71—5,000

OSMANIA UNIVERSITY LIBRARY

Call No. 530/D22I Accession No. 20470

Author Darrow, Karl K

Title Introduction to contemporary physics

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

INTRODUCTION TO
Contemporary Physics

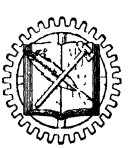
By

KARL K. DARROW, PH.D.

Research Physicist

BELL TELEPHONE LABORATORIES

SECOND EDITION



MACMILLAN AND CO., LIMITED

ST. MARTIN'S STREET, LONDON

1939

COPYRIGHT, 1926, 1939,
BY
D. VAN NOSTRAND COMPANY, INC.

All Rights Reserved

This book, or any parts thereof, may not be reproduced in any form without written permission from the author and the publishers.

First **Published**, *September, 1926*
Second Printing, July, 1927

Second Edition, February, 1939

PATRI MATRIQUE

DILECTIS

EXCERPT FROM THE PREFACE TO THE FIRST EDITION

This book originated from a series of articles published under the title "Contemporary Advances in Physics" in the *Bell System Technical Journal* from 1923 onward. I decided to recast them into a unified account of the phenomena which the nuclear atom-model, the atom-model of Rutherford and Bohr, is designed to interpret: phenomena such as the directly-perceived properties of electrons and ions, the detachment of electrons from atoms, the deflections of flying charged particles by atoms, the transfer of energy from electrons and radiation to atoms, and the regularities of line-spectra and band-spectra.

Anyone desiring to write on any field of physics in which contemporary atomic theories are involved finds himself confronted with difficult decisions about the attitude which he should adopt toward them. It is frequently said, and not always in jest, that the contemporary physics of any day is out of date six months later. This of course cannot be true; if physics were really so mutable, it would not be worth the chronicling. Actually the fact is, that the body of recorded phenomena constituting physics changes comparatively slowly, and nearly always by augmentation; while the current views as to the most expedient and satisfactory models for imitating these phenomena change at times, and especially at these times, with bewildering rapidity. The data of physics are permanent; but at any moment they are liable to be restated in new ways, or re-interpreted with the aid of new analogies. Under these conditions there is no way to write a book exempt from liability to other changes than the inclusion of new data; unless indeed one were to describe the phenomena without the slightest allusion to any theory or atom-model, which would

result in a quite unreadable book. As for the other extreme policy of describing all the theories, no one would dare to recommend it. The only feasible procedure is to separate the data from the models, so far as is reasonably practicable; to choose a few theories for presentation, according to one's fallible estimate of their serviceableness, elegance, or chance of long survival; and to advise the reader to discriminate between the facts of experience and the atom-models devised to copy them, and in dealing with these last to distinguish their essential from their irrelevant features. . . .

PREFACE TO THE SECOND EDITION

It has been a strange experience to re-read the foregoing preface barely a dozen years since it was written. In these years the "body of recorded phenomena" has grown immensely. The first edition antedated the discoveries of the diffraction of electrons, the positive electron, and the interconversion of electron-pairs and light; it antedated all the knowledge of the spins and the magnetic moments of nuclei; and it preceded almost the entire art of transmutation and all its consequences. Theory indeed *was* changing with "bewildering rapidity," now succeeded by a relative stability; for at the time of that edition the correlation of corpuscles and waves was realized only in part; the systematizing of spectra was being attempted without the aid of the concept of the spinning electron; wave-mechanics had not been invented. Introduction of all these things was demanded by the programme of the book, which was and remains—so far as it can be expressed in a single sentence—the description of the elementary particles (and now of their associated waves), followed by that of the major phenomena which are explained by the nuclear atom-model, and by more or less detail as to the manner of the explanation. The climax of the book has moved onward from the section

devoted to spectra into the section devoted to nuclear physics. What was said in that first preface as to the mutability of theory is probably now too drastic as applied to spectra, but is as apt as ever if transferred to nuclear theory, which accordingly I have left almost completely out. Even so, nuclear physics occupies a proportion of the book which may seem excessive to those still loyal to the elder disciplines. I can but guess that the prominence here accorded to it lies somewhere between its prominence in the contemporary journals, and its relative importance *sub specie aeternitatis* (though I do not profess to be able to gauge the latter). Inevitably the book had to expand by a half, and still there was not space for many important subjects (such as, for instance, the whole of statistical physics). I can only hope that the modesty of "Introduction" may atone for the temerity of "Contemporary Physics," and that those who read the book and find only a part of what they sought may feel themselves better prepared for the rest when they encounter it elsewhere.

I am greatly indebted to Messrs. C. D. Anderson, K. T. Bainbridge, C. M. Blackburn, P. M. S. Blackett, Maurice de Broglie, R. M. Bozorth, D. Cooksey, P. I. Dee, J. R. Dunning, J. S. Foster, L. H. Germer, F. Joliot, F. N. D. Kurie, E. O. Lawrence, P. A. Ross, G. P. Thomson, M. A. Tuve and J. R. Zacharias for prints for some of the illustrations.

KARL K. DARROW.

BELL TELEPHONE LABORATORIES,
February, 1939.

ERRATA AND ADDENDA

PAGE

89—In caption of Figure 3.12, for "tellurium" read "tin", and for "eight" read "ten".

92—Add to the table S^{36}, Ca^{46} and Ca^{48}; add also He^3 and Be^8, but with the reservation that the existence of these as of H^3 has been deduced only from experiments of transmutation.

100—Footnote: for masses of proton, deuteron and alpha-particle read 1.00758 ± .00002, 2.01418 ± .00002, and 4.00276 ± .00006 respectively.

101—In line 7, for "on" read "no".

169–170—In fourth line above equation (6.36) insert after "$h\nu$", the words "plus the energy m_0c^2 of the initially-stationary electron"; and then add m_0c^2 to the left-hand member of equation (6.37).

311—In lines 4 and 7, interchange "initial" and "final".

376—In equation (12.73), for $(l - 1)$ in the right-hand member read $(l + 1)$.

399—In line 21, read $(2l + 1)$ for $(2l - 1)$.

401—In line 7 from bottom, for "angular momentum" read "magnetic moment".

551—In line 22, for the proton-mass read 1.0076 instead of 1.0081.

556—In line 19, deduct .0005 from the mass-value 3.0171 to make it equal the *nuclear* masses of H^3 and He^3.

589—All statements about "trans-uranic" elements are now exposed to doubt, owing to the phenomena of "nuclear fission" recognized in January 1939.

CONTENTS

CHAPTER		PAGE
I.	The Fundamental Charge	1
II.	Electrons	34
III.	Ions and Isotopes	62
IV.	The Elements and the Nuclear Atom-Model	101
V.	The Nuclear Atom-Model Attested by Scattering of Alpha-Particles	122
VI.	Correlation of Waves and Corpuscles	139
VII.	Diffraction of Waves by Crystals	190
VIII.	Ionization	230
IX.	Excitation and the Stationary States	277
X.	Analysis of Spectra	297
XI.	Original Atom-Model of Bohr	329
XII.	Introduction to Wave Mechanics	340
XIII.	Magnetism, Electron-Spin and Spectra	378
XIV.	Miscellany as to Spectra	416
XV.	First Period of Transmutation	471
XVI.	Second Period of Transmutation	505
XVII.	Mostly Concerning Neutrons	537
XVIII.	Instability of Nuclei	580
	Subject Index	637
	Author Index	643

Abbreviations Used in References

AJ	Astrophysical Journal.
AP	Annalen der Physik (4th series).
BBS	Bulletin of the Bureau of Standards.
BSTJ	Bell System Technical Journal.
CR	Comptes Rendus de l'Académie des Sciences, Paris.
DPG	Verhandlungen der deutschen Physikalischen Gesellschaft. (References to this periodical given by years to avoid the confusion introduced by the double numbering-system of the Verhandlungen and the Berichte).
JOSA	Journal of the Optical Society of America.
JP	Journal de physique.
JRE	Jahrbuch der Radioaktivität und Elektronik.
NW	Die Naturwissenschaften.
PCPS	Proceedings of the Cambridge Philosophical Society.
PM	Philosophical Magazine (6th series).
PNAS	Proceedings of the National Academy of Sciences.
PR	Physical Review.
PRS	Proceedings of the Royal Society of London. Series A.
PZ	Physikalische Zeitschrift.
RMP	Reviews of Modern Physics.
RSI	Review of Scientific Instruments.
ZP	Zeitschrift für Physik.

CHAPTER I

The Fundamental Charge

In the opening chapters of this book it is the atomic, the corpuscular aspect of Nature which we shall chiefly be regarding. We shall be considering the electrons, those seemingly simplest and most fundamental of all the particles in the world, ultimate if anything which is known be ultimate; and the atoms of the chemical elements, themselves composed in part of electrons; and the ions and the nuclei, which are left behind when from the atoms are removed some or all of the electrons which normally belong to them. We shall examine the methods whereby the masses, the charges, and the magnetic moments of all these particles are measured; for there is no better way, indeed there is often no other adequate way of proving the existence and describing the qualities of a thing, than that of explaining the arts of measurement to which it is subjected. Having been through this discipline, we shall be justified in conceiving the material world as an enormous congregation of particles of these kinds, endowed with mass and charge and with magnetic moments in the fixed and definite amounts which the experiments will have taught us. Later we shall acquaint ourselves with the corpuscles of light which share the universe with the particles aforesaid, and with the quality of angular momentum which is so common among these elementary bodies both luminous and material, that we may justly deem it fundamental as much as charge and mass.

It might seem both simple and wise to acquire the whole of this knowledge before confronting the wavelike aspect of Nature, so much the harder to grasp and comprehend. Yet this is well-nigh impossible to do, and misleading even to attempt; for wave and corpuscle are so implicated with each other throughout the structure of the world, that to separate

them for even the length of a chapter would be a laceration. The fullness of wave theory ought indeed to be postponed until the corpuscles become familiar; but as for the simplest manifestations of waves, the earlier we encounter them the better it will be. I therefore invoke the undulatory aspect of light at the very beginning, in order to prove that matter is atomic.

Light incident on a plane diffraction grating

Think of a beam of monochromatic[1] light coming from a source at a great distance and falling normally on a grating such as the one which Fraunhofer made in 1820: a sequence of parallel wires evenly spaced, the equal gaps between consecutive wires being not too broad. Beyond the grating there will be a beam of light continuing straight onward, not so bright as the oncoming one; and in addition there will be a "first-order diffracted beam" diverging at an angle (we will call it θ) from the direction of the former. (There will be other diverging diffracted beams as well, but the first is all that we need consider). The value of θ will not be influenced in the least by any change in the size or the shape (the cross-section, that is to say) of the individual wires of the grating. It will be controlled uniquely by the "spacing"—by which word is meant, not the breadth of a gap between wires, but the distance (call it d) from any point of a wire or a gap to the corresponding point of the next wire or the next gap. There will be a particular relation, which we shall presently examine, between θ and d. The total number N of the wires (the sidewise extent of the grating) will not affect the value of θ, and yet it will have an influence of its own upon the phenomenon: for the greater this number, the narrower and sharper the diffracted beam will be. So, in these respects, the action of the grating upon the light depends on the totality of all the wires taken together; it depends on the qualities d and N of the grating-

[1] In order not to presume the wave theory of light before introducing it, I here use "monochromatic" in the proper direct sense in which Newton might have used it, had the word existed in his time: as referring to the kind of light which is not spread out nor changed in appearance, when it passes through a prism.

as-a-whole, and not on the qualities of the individual wires thereof.

Here we have reached already the prime objection to the old corpuscular theory of light. We may suppose that the beam is a rain of ordinary particles rushing swiftly on parallel paths to the grating; but let us see where the supposition leads us. Some of the particles will shoot through the gaps, but others will bounce obliquely off the wires, and we may try to interpret the diffracted beam as being made up of those which have bounced. But if this be our idea, then we shall deduce that θ must depend only on the shape of the individual wires and not on their spacing, since the corpuscle which hits a particular wire has dealings only with that one and does not know how far away the others are nor yet how many others there may be.

Imagine however that the beam of light is a succession of pulses or brief pulsations, something like isolated ridges travelling swiftly over a surface of water; for some such a notion was in the mind of Huyghens, when toward the end of the seventeenth century he made the first of the gropings toward what has since become the theory of wave-motion. In Figure 1.1 the four vertical lines on the left are meant to stand for a succession of four such pulses, the dots for the wires of a grating. Imagine the first of the pulses arriving at the grating. Let us suppose that at the instant of its arrival at each wire, there occurs the birth of a spherical pulse or " wavelet " as it shall hereafter be called. After the moment of its birth, each wavelet widens outward from its origin with a speed equal to that of the original pulsation. The curves to the right of the grating in Figure 1.1 may be regarded as arcs of these wavelets, drawn to correspond to four different lapses of time since the instant of birth. (They ought to be drawn as completed circles, but if this were done there would be so much crisscrossing of the curves as to make the drawing almost unintelligible.)

These wavelets are the first of the cardinal features of undulatory motion to appear in this book. Here I can adduce these as facts of experience; for the reader has only to

look at the remarkable photographs (Figures 1.2 and 1.3) borrowed for this book from a treatise on acoustics, to behold pictures of wavelets externalized by experiments on sound. In the topmost views of Figures 1.2 and 1.3, the great circular arc is the pulse due to a sharp and sudden explosive noise—the crack of an electric spark which flashed out behind the black disc, a small fraction of a second before either picture was

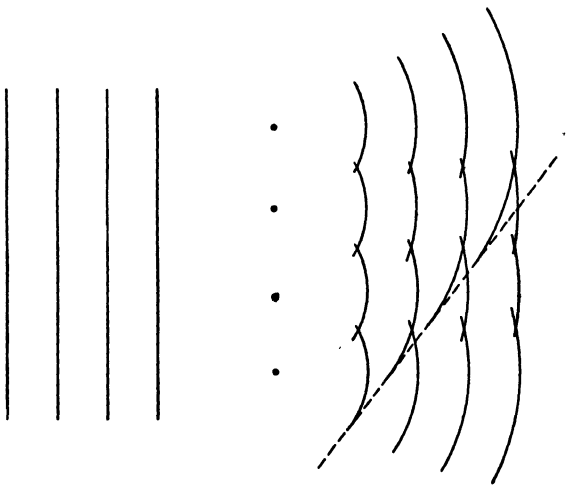


FIGURE 1.1—ILLUSTRATING HOW WAVELETS MERGE TO FORM A DIFFRACTED WAVEFRONT.

taken. The lower pictures of each figure were taken a little later, and in them one sees how wavelets have sprung from each of the steps on the left, from each of the irregularities of the ceiling on the right—wavelets, each of which was born when the pulse arrived at the point of the origin thereof.

Notice next especially how, in the lower pictures of Figure 1.3, the wavelets are in course of merging themselves along their envelope (in the mathematical sense of this word) to form a sort of reincarnation of the original pulse going in the opposite direction. It is the reflected pulse which is thus being formed. The student is now reminded of Huyghens' construc-

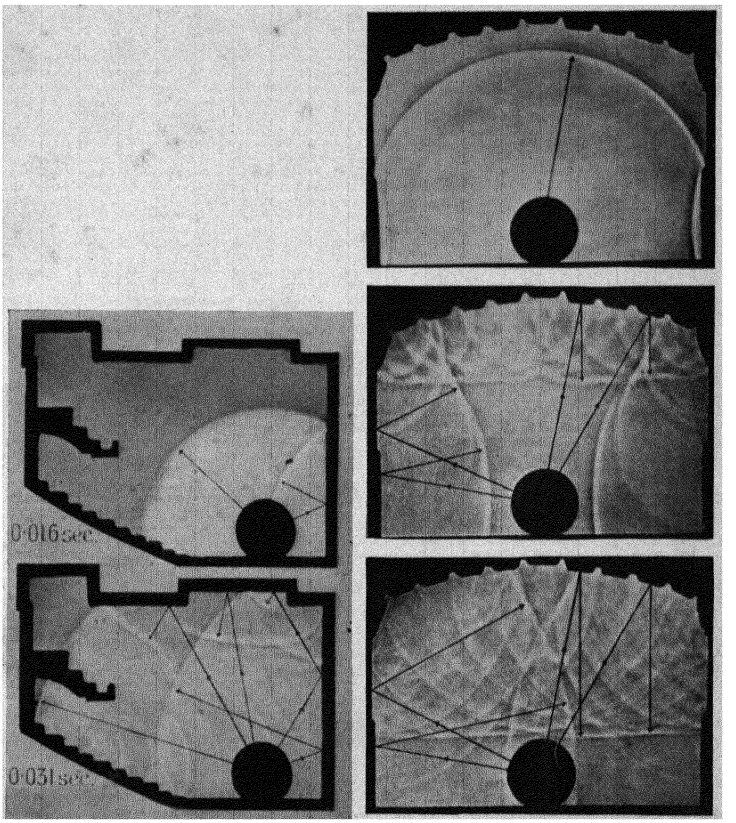


FIGURE 1.2—A SOUND-PULSE, AND WAVELETS WHICH ARE FORMED WHEN THE PULSE ENCOUNTERS SHARP OBSTACLES. (Davis and Kaye, *Acoustics of Buildings;* G. Bell and Sons.)

FIGURE 1.3—A SOUND-PULSE, AND WAVELETS WHICH ARE FORMED WHEN THE PULSE ENCOUNTERS A BROKEN SURFACE, AND THE MERGING OF THESE WAVELETS TO FORM THE PULSE ANEW. (Davis and Kaye, *Acoustics of Buildings;* G. Bell and Sons.)

*

tion, the familiar scheme for showing how waves are reflected and also refracted (we shall use it in Chapter V to interpret refraction). Return to Figure 1.1: the wavelets sketched on the right are likewise in course of recomposing the original pulse (or a pulse somewhat feebler than the original one) travelling onward in the original direction.

Continuing to look at Figure 1.1: think now of the four vertical lines on the left as four consecutive pulsations sketched at a single instant, and of the four sets of wavelets on the right as resulting from four pulsations just like those, which were ahead of those and arrived at the wires a little time before them. Notice the oblique line drawn tangent to four wavelets proceeding from four consecutive wires and due to four consecutive pulses (these four wavelets being drawn more heavily than the others, merely to make them stand out). It is meant to suggest that the wavelets in question may tend to merge along *this* line, composing a slanting pulsation which will travel onward in the direction of the arrow, normal to the line itself. Not the whole of the energy of the wavelets is to go into this obliquely-propagated pulse; some is reserved for the one which goes straight onward, some perhaps for yet others; but a portion of the energy of the oncoming beam, we will suppose, is diverted into a train of pulses such as this one.

By mere inspection we see that the angle θ between the oblique and the straight-onward beams obeys the formula

$$L = d \sin \theta, \tag{1.1}$$

L standing for the distance between consecutive pulses. For trains of pulses with any particular value of L, the angle θ depends on d alone; moreover it depends on d in a special way, $\sin \theta$ varying inversely as d. This is precisely the behavior of the first-order diffracted beam when monochromatic light falls on a grating.

It looks as though we had explained the behavior of a grating toward light, as though we had even derived the well-known grating-formula

$$\lambda = d \sin \theta, \tag{1.2}$$

out of the notion of monochromatic light as a sequence of equally-spaced pulsations like ridges on water. Yet we have done nothing of the sort; and it is important to realize why. The fault will appear if the student will imagine an extra wire interpolated midway in each of the gaps of the grating of Figure 1.1, reducing the spacing to $\frac{1}{2}d$. Adding the extra wavelets to the right-hand side of the sketch, and drawing a new line tangent to four wavelets proceeding from four consecutive wires and due to four consecutive pulses, he will find (of course) that it is oriented much more obliquely than was the original diffracted pulse, its angle θ being now *arc sin* $(L/\frac{1}{2}d)$. This is the first-order diffracted beam corresponding to the new spacing, and experience with light shows that it is there.

But what has happened to the train of pulses diffracted at the angle *arc sin* (L/d), which existed when the extra wires had not yet been added? According to our present conception, nothing should have happened to it! It ought to be still there, enfeebled perhaps to some extent because of the diversion of energy to the new beam. Experience with light, however, shows that it is gone completely: when the spacing of grating is halved, the original first-order diffraction-beam disappears.

To explain this disappearance, we have to assume that midway between the oncoming pulses depicted on the left of Figure 1.1 there are, in effect, negative pulses which bring about the birth of negative wavelets when they impinge upon the wires. Then, in the case of Figure 1.1, there goes off in the oblique direction of the arrow a procession of pulses of which one will be formed uniquely out of positive, the next uniquely out of negative wavelets, the next again uniquely out of positive, and so on—a reproduction (except for intensity) of the oncoming beam with its pulses of alternate signs. Let now the extra wires be added: then the oblique line in Figure 1.1 will be tangent to a succession of wavelets which are alternately positive and negative, and cancel one another so that the resultant pulses will be annulled, or rather, there will be no resultant pulses at all and no beam moving away along the

arrow. So, the beam of monochromatic light cannot be regarded as a sequence of identical pulses, not even when these are supposed to be equally spaced; but it can and must be regarded as an alternation of positive and negative.

This is the great step forward which, in the theory of light, was made by Young and Fresnel about the beginning of the nineteenth century, more than a hundred years after the wavelets of Huyghens. It was not really an undulatory theory which Huyghens had made, even though wavelets be a cardinal feature of wave-propagation. Indeed there was no reason for him to devise a proper undulatory theory; for diffraction by gratings was not discovered (nor was diffraction by holes or slits or edges thoroughly studied) in his time, nor at any time before the era of Young and Fresnel themselves. To say either that the theory was made to fit the phenomena of diffraction, or that the phenomena were studied in the hope of confirming the theory, would be too simple. Both statements are partially true; and the thoroughgoing study of diffraction, which was inchoate when the great step was taken, was made in order to justify it. All that is more than a century past, and there would be little reason for recalling it in a textbook of contemporary physics, but for the fact that in the theory of electricity and matter it is the *same* great step which was made by Louis de Broglie and Schroedinger when they advanced from classical mechanics to the wave-mechanics of today.

To be specific: an endless plane-parallel beam of monochromatic light advancing along the x-direction is represented by the particular succession of positive and negative, the particular sequence of hills and valleys, which is the sine-function:

$$\Psi = A \sin 2\pi(t/T - x/\lambda + \alpha). \qquad (1.3)$$

Here the symbols A, T, λ and α stand for various constants of which the third is the only one that now concerns us, being the only one revealed by experiments on diffraction. It is the crest-to-crest distance, the " wavelength," of the endless train of sine-waves. When we refine the reasoning which led

to equation (1.1), applying it now to sine-waves instead of pulses, we reach the grating-formula:

$$\lambda = d \sin \theta, \qquad (1.4)$$

and using a grating of known spacing d, derive the value of λ from the obliquity θ of the first-order diffraction-beam. The theory fixes the value of the breadth of this beam (in more technical language, the resolving-power of the grating) in terms of the total number of wires N of which the grating is made up. If the beam is broader than this value, it is inferred that the oncoming light comprises sine-wave trains of an appreciable range of wavelengths, each of which is diffracted in the direction given by (1.4) just as though the others were not there. (Some prefer to say that the grating analyzes the oncoming light into these sine-waves, and this is a usage which it is good for us to note in passing, as it suggests that the grating does in reality what we do in imagination when we work out the Fourier expansion of a function which is not a pure sine-wave; but this is anticipation).

Though I have spoken thus far of a grating of wires, everyone knows (as Fraunhofer knew, a hundred and fifteen years ago) that diffraction-beams just as sharp and obeying just the same formula (1.4) are obtained from gratings composed of parallel scratches or grooves incised upon otherwise-smooth surfaces of metal or glass, and that these are generally used. Any sort of regularly-repeated obstacle, hindrance, or impediment athwart the path of a wave-train will cause the formation of a diffracted train (or indeed of several, including those of higher order than the first which are described by prefixing the integers 2, 3, 4 . . . to the left-hand member of 1.4). Resonators—*i.e.* particles capable of entering into sympathetic vibration with the waves—will do it if evenly spaced. Atoms will do it, if they are evenly spaced and the wavelength of the light is smaller than their spacing: this is an instance of a general condition, for a glance at (1.4) will show that if λ is greater than d there is no value of θ that will fit—there is no diffracted beam (a rule that I will later invoke in unexpected

places). Allowance made for this last exception, it may be said in general terms that *any periodicity* athwart the path of a wave-train may evoke diffracted trains. The details of that-which-is-periodically-repeated—the thickness of the wires, the depth or profile of the scratches, the natural frequencies of the resonators, the structure of the atoms—these details are irrelevant to the grating-formula, which is valid whatever they may be. They affect the brightness of the various diffracted beams, sometimes to the extent of quenching one or more of them; but being now concerned with the evaluation of wavelength, the less we think about them the better.

Another thing which it does no good to think about at present is the " real nature " of the quantity symbolized by Ψ in equations (1.3) and other equations like it. The question " just what is it that oscillates? " is re-born whenever an undulatory theory is applied to a new domain of physics. For the time being at least, the best way to answer it is to say: *Ψ is that of which the amplitude-squared is proportional to the intensity of the light.* A fuller answer can indeed be given, in the cases of light and of sound; but even where the fuller answer is available, it is well to hold to the dry and colorless one in the italics, so long as we are thinking of diffraction; for the particular quality of Ψ embodied in those italicized words is the only one which is proved, when the theory of diffraction is tested by experiment. It is, however, quite all right to say that Ψ is a vibration of an aether, if one will be contented with the famous definition of Lord Salisbury: " The aether is the nominative case of the verb *to undulate.*"

Leaving for later chapters the development of the idea of waves, I advance now toward the application for which these first allusions were intended.

The atomic structure of matter

Of the various states in which matter may be found, we will now consider briefly the " crystalline." A properly-developed and self-evident crystal is a piece of solid matter bounded by

wonderfully smooth planes which make sharp angles with each other—angles which are not arbitrary in the least but have specific values characteristic of the sort of crystal which it is. "Cubic" crystals are likely to have faces making right angles with each other, so that a piece of the substance may be a parallelopiped or even a cube; they may also have faces inclined at angles such as 45° or 60° to each other. The implication is, that the crystal is made up of identical "units" arranged in a regular array, like soldiers on parade except that the parade is in three dimensions instead of in two. They are said to form a "lattice," or better, to be "arranged upon a lattice;" for the lattice is strictly defined as an abstraction, like a coordinate-frame, or the network of meridians and circles of latitude which intersect upon a map. It is usually conceived as a network of intersections of three sets of parallel planes, the planes of any set following one another at even intervals of spacing. For convenience of drawing, each plane is sketched (as in Figure 1.4, for example) as a network of two sets of parallel lines. The intersections of three planes are the "lattice-points;" the intersections of two planes are "axes of the lattice." Axes run in three directions, and along each there is a constant spacing from one lattice-point to the next. If the three axes are at right angles to one another, and the spacings along all three are the same, the lattice is "cubic" by definition; a crystal with such a lattice is a crystal of the cubic system (though its form is not necessarily that of a perfect cube). In Figure 1.4 a crystal of the cubic system is suggested, the circles standing for units of the crystalline substance (if the lattice-points were indicated in the drawing, they would be dots at the centres of the circles).

What happens if a beam of monochromatic light, of wavelength smaller than the spacing of the lattice, is directed against a crystal such as this?

A complete answer—or rather, the fullest available answer—to this question requires a book to itself; it is the theory of the diffraction of X-rays by crystals. A partial answer will be given in Chapter VII; an even smaller fragment of the total

theory must suffice in this place. The units of the crystalline substance are impediments to the free onward march of the waves, and therefore they become the centers of trains of wavelets, which spread out, overlap, and by constructive and destructive interference merge with one another into diffracted wavetrains or diffraction-beams. The directions of these beams may be learned by reasoning from the three-dimensional arrangement of the units (we shall do this in Chapter VII), but they may also be learned, and more quickly, by using the mental artifice of *regarding the lattice-planes* (such as those of Figure 1.4) *as semi-transparent mirrors.* Suppose the incident beam of monochromatic light falling obliquely upon the top of the crystal suggested in Figure 1.4: a fraction of it (generally a very small fraction!) will be reflected by each of the horizontal lattice-planes, while the remainder goes onward undeflected, to be partly-reflected and partly-transmitted at the similar planes underneath.

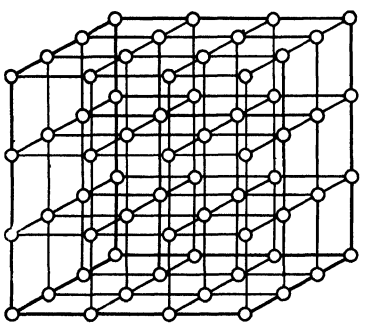


FIGURE 1.4—A CUBICAL SPACE-LATTICE, WITH PARTICLES OF FULL SPHERICAL SYMMETRY INDICATED AT THE LATTICE-POINTS.

Imagine just two of these planes, separated by the spacing D; and a train of waves descending from the upper left, the wave-crests being indicated by the oblique parallel lines. Part of the wave-motion will be reflected from the upper plane, part will go on, part of this part will be reflected from the lower plane. The waves of these two reflected trains will go off in the same direction (toward the upper right); but the wave-crests of the one will coincide with wave-crests of the other and troughs will coincide with troughs, *if and only if* there is a certain relation between λ and D and the angle of incidence θ. The relation is:

$$n\lambda = 2D \cos \theta. \qquad (1.5)$$

(I can safely leave it to the student to derive, with the aid of the diagram).

The relation is deceptively like (1.4), but there is a profound difference, of which the factor 2 which appears in the one and not in the other supplies but a trivial part. Equation (1.4) refers to the special case of *normal incidence* upon a plane grating, and θ to a direction in which a diffraction-beam is always observed whatever the value of λ may be. Equation (1.5) refers to *incidence at the angle* θ upon a crystal grating; if θ conforms to the equation there is constructive interference and there is a reflected beam which is essentially a diffraction-beam, but otherwise there is not.[1] When there is a diffraction-beam, it runs in the direction of specular reflection, at the angle $-\theta$ on the opposite side of the normal to the lattice-planes. The experiment (described on page 222) consists in turning the crystal gradually so as to vary θ and observing always in the direction $-\theta$, and observing the values which θ has when a diffraction-beam appears.

Clearly, *when this experiment is done with light of which the wavelength has been determined with a ruled grating of known spacing, it yields* through equation (1.5) *the value of the spacing of the crystal lattice.*

This sentence has taken much less time to write or to read than the experiment took in the doing! The trouble lay in the fact that if the light is of small enough wavelength to be diffracted by crystals ($\lambda < D$)—if, that is to say, it is of the X-ray spectrum—then it cannot be diffracted by ruled gratings save under a special condition. The condition is, that the angle of incidence must be within a degree (or less) of 90°—the inci-

[1] It can readily be seen that the conclusion is not affected by adding the reflections from a finite or infinite sequence of underlying planes succeeding one another with the spacing D. Moreover it can be seen that the existence of these-and-only-these diffracted beams is contingent upon the wave-train being a succession of positive and negative alternating, not merely a succession of positive pulses; and finally it can be shown that the greater the number of underlying planes, the more sharply these beams are concentrated at the precise values of θ given by (1.5). If there are only a few planes, there is reflection over a widish range of angles adjoining each of those values of θ.

dence, as people say, must be very nearly " grazing." This is partly because at such an incidence the grating appears foreshortened, so to speak—it acts on the light as though its own spacing d were shortened, and this increases the deflection of the diffracted beam, which is very advantageous; but chiefly it is because X-rays are not properly reflected from a surface of metal or glass unless the incidence is nearly grazing, so that only in this case are the wavelets strong. The condition was not fulfilled until 1925, when in the laboratories of A. H. Compton at Chicago there occurred the first of a long series of measurements of X-ray wavelengths with ruled gratings.

Before quoting any of the results, I must turn to another question: what are these " units," by the periodic arrangement of which upon a cubic lattice (or some other kind of lattice) a crystal is built up?

The word " unit " was chosen because of its vagueness. Nothing thus far mentioned in these pages can indicate in the least whether the units are individual atoms (as some of my readers may incautiously have assumed), small groups of atoms, or even big heterogeneous clumps. We can, however, draw some inferences, by invoking the analogue of something which I merely mentioned in describing diffraction by a ruled grating. It was said on page 10 that while the inclinations of the diffracted beams depend only upon the *spacing* of the periodically-repeated elements of the grating—the wires, the slits, the grooves, or whatever these elements may be—yet there is one feature which is controlled by the sizes and shapes of these elements, and this is the intensity-distribution among the diffracted beams of various orders. Similarly, when X-rays are diffracted by a crystal, the intensity-distribution among the various beams occurring for various values of the incidence-angle θ is controlled by the sizes and shapes of the units; and these last may be deduced from it.

To explain the technique of the deduction would be too much of a detour at this point (something will be said about it in Chapter VII): it must suffice to give the general result. In crystals made of a single chemical element, it is invariably found

that the unit consists of two or more identical parts, which usually (at least, usually in cubic crystals) have spherical symmetry. It is these which we identify with the atoms, and justifiably, as will later on appear. The number n of these atoms per unit is often 2, often 4, sometimes higher—never just one. With crystals of definite chemical compounds, the method is capable of showing that there are particles of two or more kinds; thus in the sodium-chloride crystal NaCl there are four particles of one kind and four of another to each unit, and we conclude that they are four atoms of sodium and four of chlorine. This we may restate by saying that there are four molecules NaCl per unit. The number-per-unit of molecules in this sense is one in a few cases (CsCl for instance), but usually two or four or higher.

To come into conformity with usage, I now exchange the word "unit" for "unit-cell," which is defined as follows. With a crystal built upon a cubic lattice of spacing D, the volume of the crystal is divided (in imagination, of course!) into cubes of width D and volume D^3, centered upon the lattice-points, each containing one of the units aforesaid and all together filling the volume of the crystalline substance completely. These are the unit-cells. If V stands for the volume in cubic centimeters of any one of them and N_u for the number of unit-cells, units, or lattice-points per cubic centimeter, then:

$$N_u = 1/V. \tag{1.6}$$

Diffraction of X-rays supplies us with the values of V and N_u, because it supplies us with the value of D on which these things depend. As to any crystal built upon a non-cubic lattice, the unit-cells are volumes having the symmetry of the lattice, but diffraction of X-rays gives us the dimensions of these cells just as well as with the cubic lattice, so that V and N_u are knowable.

Now let ρ denote the density of a crystal, and $\overline{m}$ the average mass of an atom or a molecule, according as the crystal is made of a chemical element or a chemical compound, then:

$$\overline{m} = \rho/nN_u, \tag{1.7}$$

and *thus the diffraction of X-rays by crystals supplies us with the masses of molecules and atoms.*

It would, however, be pointless to give a list of masses thus determined, because we can epitomize the results by referring to a single universal constant N_0.

The universal constant N_0

By studying numberless chemical reactions, the chemists have shown that the elements behave as though they consist of particles having definite weights—one weight-value characteristic of each element—which join with one another in groups to form the ultimate units or " molecules " of the chemical compounds. Thus, in the compound known as common salt, the elements sodium and chlorine are present in the ratio of 23.00 to 35.46 by weight. This was considered to show that sodium and chlorine are made up of atoms having these relative weights. Compounds are not always so simple: thus, in the one which is known as water, the elements hydrogen and oxygen are present in the ratio of 2.016 to 16.000 by weight, but this was considered to show that hydrogen and oxygen are made up of atoms having the relative weights $\frac{1}{2}(2.016)$ and 16.000, there being two atoms of hydrogen and one of oxygen to the molecule of water. There are many reasons for this decision and other decisions like it; some will appear later on in this book, others it is the affair of chemists to expound. Coming straight to the conclusion: there is a list (here reproduced on page 105) of the so-called " chemical atomic weights " of the elements, derived from the analysis of chemical compounds, and expressed in a scale in which the atomic weight of oxygen is taken as 16. It has been assumed for many years that each of these atomic weights is equal to the weight—or let me say, the *mass*—of the individual atom of the corresponding element, multiplied by a factor which must have the same value for all, but a value which no experiment on chemical reactions can reveal.

This factor is the universal constant denoted by the symbol

N_0, known as the " number of molecules per gramme-molecule," and called the " Avogadro number " or the " Loschmidt number." It has been defined from the equation,

$$M = N_0 \bar{m}, \tag{1.8}$$

in which M stands for the atomic (or molecular) weight of an element (or compound), and $\bar{m}$ for the average mass of the atom (or molecule). I did not make the generalization from elements to compounds in the paragraph which precedes, but the student can easily make it for himself.

Now since the chemists have supplied accurate values of M for many crystallizable elements and very many crystallizable compounds, the physicist has only to select the one best suited for diffraction of X-rays, and ascertain N_0 by determining $\bar{m}$. Especially fine crystals are formed, or form themselves, out of the compounds $CaCO_3$ (calcite) and KCl (rocksalt). When crystals like these are employed for diffracting beams of X-rays of which the wavelength is known by measurement with ruled gratings, the value of N_0 which emerges [1] is this:

$$\mathbf{N_0 = 6.02 \cdot 10^{23}}.$$

This value is believed to be correct to the last significant figure which I have quoted.

When I say that a value " is believed to be correct " to a certain significant figure, I mean that the *statistical error*—the uncertainty resulting from the fact that when the routine of the experiment is repeated again and again under seemingly-identical conditions, there are little and insuppressible discrepancies from one performance to the next—is so low that it casts no doubt upon that figure. Usually I shall quote fewer significant figures than many an experimenter sets down: this is because it sometimes happens that the same experiment is performed by two people in different places who arrive at results differing from each other by more than the statistical error which either will admit. The art of assessing the relia-

[1] J. A. Bearden, **PR 37**, 1210 (1931) and **48**, 385 (1935); E. Bäcklin, **ZP 93**, 450 (1935).

bility of data is a very difficult one, requiring as much care as, for example, Birge has lavished upon it in a recent series of papers on the major physical constants. Moreover, there exists at the moment a confusing situation in respect to the values of some of these constants, as will presently be only too obvious; and this also makes it undesirable to accept a precision which may after all be fallacious.[1]

The fundamental constant F (the Faraday)

Think of an aqueous solution of a chemical compound—silver nitrate, to begin with—into which electrodes are dipped which are connected to the poles of a battery. A steady current flows in the circuit, and metallic silver separates itself from the solution at a steady rate and encrusts itself upon the cathode. Let I stand for the strength of the current, T for the time during which it is allowed to flow, IT consequently for the quantity of charge which traverses the solution and W for the weight of silver which is liberated during the passage of that charge. *The ratio IT/W is a constant* (this is one of Faraday's laws of electrolysis). Instead of writing down its value, I will quote the value which IT has when W is equal to M the atomic weight of silver:

$$IT = 2.895 \cdot 10^{14} \text{ when } W = M.$$

(I use the electrostatic unit of charge here, as everywhere else in this book excepting where the contrary is stated.)

This phenomenon is interpreted as follows. In the aqueous solution, the atoms of silver are floating around as charge-bearing particles, "ionized atoms" or "ions" of silver with positive electric charges. When the electrodes are connected to the poles of the battery, the one which is joined to the negative pole acquires a negative charge and becomes the cathode, and the silver ions are attracted to it. Whenever one of these reaches the cathode it neutralizes itself by extracting

[1] R. T. Birge, **RMP** 1, July 1929; *Nature* 137, 187 (1936); *Science* 79, 438 (1934); **PR** 52, 241 (1937) and many earlier notes; DuMond, **PR** 52, 1251 (1937).

therefrom a negative charge equal in magnitude to its own positive charge, whereupon it becomes an ordinary atom of silver and attaches itself to the surface of the cathode.[1] When N_0 atoms (for this is the number of atoms of which the weights total up to M) have acted thus, the total charge (of negative sign) which they have extracted from the cathode amounts to $2.895 \cdot 10^{14}$. This is also the total charge (of positive sign) which was borne by the N_0 ions while they were still ions in the water. We might say that $2.895 \cdot 10^{14}/N_0$ is the average charge of the silver ion. This, however, would be a valueless remark, were it not for the law to which we now attend.

Think of the same experiment being performed with a great multitude of solutions of divers chemical compounds. Many different elements may be liberated at the cathode, some being plated upon it like silver in the first example, others bubbling away in the form of gas. For any one of these, the ratio IT/W is constant, and it is possible to find the value which IT has when W is equal to the atomic weight of the element in question. Now comes Faraday's great law of electrolysis: *when $W = M$ the value of IT is always equal either to the amount just given for silver, or else to a small integer multiple of that amount.*

Assigning the symbol F and the name " Faraday " to that amount (which now takes on the importance of a fundamental constant), we have:

$$IT = nF \text{ when } W = M \begin{cases} F = 2.895.10^{14}. \\ n = 1, 2, 3 \ldots . \end{cases} \qquad (1.9)$$

The total charge (of negative sign) extracted from the cathode by N_0 positive ions of any element, and therefore the total charge of N_0 positive ions of any element, is either one Faraday or some other small integer number of Faradays. Incidentally, the value of n is related to the chemical properties of the element in question, in such a way as to show that chemical

[1] The flow of current onward to the anode is cared for by the negative nitrate ions in the solution.

properties are largely if not altogether of electrical origin: we shall see more of this principle in later chapters.

Therefore the average charge of a positive ion in an electrolytic solution—let it be denoted by $\bar{Q}$—is given thus:

$$\bar{Q} = n(F/N_0) \quad n = 1, 2, 3 \ldots, \qquad (1.10)$$

and this has become a much more important statement than it was when it was made for silver only, being applicable to all elements.

I might now insert the values of F and of N_0 which have just been cited in these pages as derived from experiments so clear-cut and so simple. This I will postpone, however; for not yet has it been proved that all charges of ions are integer multiples of F/N_0! It is only the *average* value of the charge, for ions of any particular kind, for which the proof has yet been given. Notice again equation (1.7): there it is only the *average* value of the mass, for atoms of any particular element, for which the evidence has justified that formula. Now the reader may be disposed to rush on to the assumptions that all the ions of an element are identical in charge and all the atoms of an element are identical in mass. It would, however, be rash to do so, for the latter of these assumptions is known beyond doubt to be false (as will appear in Chapter III), and the former requires further evidence to be marshalled in its favor. To this evidence we now turn.

The charges borne by electrified bodies; the oil-drop experiment

Imagine a very small droplet of some liquid (a non-volatile liquid, for we do not want the drop to be evaporating while we watch it) set adrift in the air. If nothing acts on it but gravity, it will fall: not according to Galileo's law of uniform acceleration which obtains in a vacuum, but according to that other law which has been known for upwards of two thousand years, and which obtains when the falling body is so light that the friction with the air controls. The speed v_0 of the fall is proportional to the weight of the droplet, therefore to the force

which acts upon the droplet; and if another force should act together with that of gravity, the drop would move in the direction of the resultant of these forces, with a speed proportional to the magnitude of that resultant.[1]

Imagine now that a vertical electric field pervades the air: then if the droplet is charged there will be an electric force acting upon it, parallel or anti-parallel to the force of gravity, and this will evoke a speed v proportional to the algebraic sum of the two forces. Now, if the droplet arises from a spray blown into the gas (as from an atomizer) it probably will have been charged from the start; also, if the gas is exposed to X-rays or to the rays from radium or to other well-known " ionizing agents," then the drop will frequently pick up charges of either sign—a process known as " capturing ions "—as it moves through the gas; also if such rays fall directly upon the drop, they may expel electricity from it. It is possible, and not even difficult, to make droplets of such a size that they are clearly visible (shining like brilliant stars when illuminated, and viewed by the observer through a telescope), and that their terminal speeds in air in these conditions are readily measured.

The first of the principles which the experiment proves is most easily shown from formulae, though the formulae will contain more than is required for this first purpose. Let m stand for the mass of the droplet, Q for its charge, g for the acceleration due to gravity, E for the strength of the electric field which may be switched on or off, v and v_0 for the vertical velocities of the droplet when the field is on and when it is off; consider E and v as positive when directed downward; then:

$$v/v_0 = (QE + mg)/mg,$$

$$Q = (mg/Ev_0)(v - v_0). \qquad (1.11)$$

To determine Q from observations on v and v_0 it is necessary to know m, a troublesome requirement. A most important fact

[1] When the force is suddenly changed it presumably takes some time for the droplet to acquire its new terminal speed, but the length of this time proves to be inappreciable.

may, however, be learned without knowledge of m, by holding a single droplet under observation for a long time during which its charge often changes (because it captures ions, or for some other reason). It is found that before and after any alteration of charge, the difference $(v - v_0)$ is always an integer multiple of a fixed common divisor; and this signifies that *all the quantities of charge which appear upon the droplet are integer multiples of a basic amount.*[1] Never, not even at the start, does the drop possess a quantity of charge which is not an integer multiple of that basic amount: the charge (if any) which it brings as it is shot into the gas, the charges which it later acquires in any way whatever, are always proportioned among themselves as though they were assemblages of various numbers of equal units.

The reader now thinks perhaps that it is settled that all electric charges—at least, all that appear upon droplets—are integer multiples of a universal and fundamental constant. Still, it has not yet been proved that the " basic amount " of the sentences just preceding is the same for all droplets. To establish this we have to tackle the problem of measuring the charges of the drops, which as I just said reduces to that of ascertaining their masses.

The mass m of a droplet is estimated by measuring its terminal speed v_0 when the field is switched off so that gravity alone can act, and applying to the datum some formula for the relation between terminal speed and acting force. If we seek for such a formula in classical hydrodynamics, we find one which is based on the assumption that the air in the immediate vicinity of the droplet moves along with the droplet itself, as

[1] "Relationships of exactly this sort have been found to hold absolutely without exception, no matter in what gas the droplets have been suspended nor what sort of droplets were used upon which to catch them. In many cases a given drop has been held under observation for five or six hours at a time and has been seen to catch . . . hundreds of ions. Indeed I have observed, all told, the capture of many thousands of ions in this way, and in no case have I ever found one the charge of which, when tested as above, did not have either exactly the value of the smallest charge or else a very small multiple of that value." (R. A. Millikan, *The Electron*, University of Chicago Press, 2d edition, pp. 71–72.)

though it adhered to the surface of the liquid. This is " Stokes' Law:"

$$v_0 = mg/6\pi\eta a. \qquad (1.12)$$

Here η stands for the viscosity of the air and a for the radius of the drop. (If the drop is exposed to other forces than gravity, one puts v instead of v_0 on the left and the resultant force instead of mg on the right). Since a cannot be observed (the drops are so small that they look like luminous points), we must complicate the equation by invoking the obvious formula,

$$m = (4/3)\pi a^3(\rho - \sigma), \qquad (1.13)$$

giving m in terms of a and the densities ρ and σ of the droplet and of the air (σ must be introduced because the effective weight of the droplet is its true weight minus its buoyancy, as Archimedes knew.) With the aid of these formulae we rewrite (1.11) thus:

$$Q = \left[\frac{162\pi^2\eta^3 v_0}{g(\rho - \sigma)E^2}\right]^{\frac{1}{2}} (v - v_0), \qquad (1.14)$$

where now everything on the right is measurable.

I repeat that when observations are made on a single droplet of which the charge is frequently altered, it is found that the various values of the factor $(v - v_0)$ are found to be integer multiples of a common divisor. Let $(v_1 - v_0)$ stand for this (greatest) common divisor; then the basic amount of charge (I will denote it by e_1) for the droplet in question is this:

$$e_1 = \left[\frac{162\pi^2\eta^3 v_0}{g(\rho - \sigma)E^2}\right]^{\frac{1}{2}} (v_1 - v_0). \qquad (1.15)$$

Now when this quantity is evaluated for different droplets, it is *not* always the same! It varies systematically with v_0 and therefore with a, increasing as they decrease. For a moment we seem obliged to surrender the prospect of proving a basic universal charge; but we can retain it, by criticizing Stokes' law. The assumption that the air sticks to the surface of the liquid and moves with the droplet—this assumption may be

true only in approximation, and may depart progressively further from the truth as one passes from larger to smaller particles. Suppose that there is a relative velocity between the drop and the immediately-adjacent air (technically known as "slip"), which is negligible for large drops but varies inversely as a: then instead of Stokes' law we should write:

$$v_0 = \frac{mg}{6\pi\eta a}\left(1 + \frac{A}{\sigma a}\right). \qquad (1.16)$$

(The density of the air enters in because it is presumed to be the ratio of the diameter of the drop to the average distance between the molecules of air which is the controlling variable). Well, Millikan found that if he replaced Stokes' law by (1.16), and proceeded otherwise without change, arriving thus at a formula for Q more complicated than (1.14) and involving the undetermined constant A as well as the knowable quantities already figuring in (1.14): then, all the data obtained with a great number of droplets of oil falling through air, and whatever the density of the air (within rather wide limits) were compatible with *one* value of A and with *one* value for e_1. Furthermore, when droplets of mercury and particles of shellac were used instead of oil-drops, the data were compatible with values of A identical for the particles of each material among themselves (*i.e.* one value for shellac, another for mercury, a third for oil) and with *one* value for e_1, the same for all materials. Thus in addition to serving their main end, these experiments enabled Millikan to establish the proper correction for Stokes' law and to study the phenomenon of slip: an eminent example of the best technique of science.

The fundamental constant e (the electron-charge), measured by the oil-drop experiment

It having thus been shown that in a very long series of experiments all the charges possessed by droplets of liquid floating in air, and in particular all the charges acquired by these droplets while floating in air, were integer multiples of a common basic amount:

We now generalize from these results, and affirm that the same is true of all electric charges whatsoever. All charges, positive or negative, are believed to be integer multiples of this common basic amount, the fundamental and universal constant denoted by e.

Millikan's celebrated value of e, expressed (as usual) in electrostatic units and amputated of its fourth significant figure, is this:

$$e = 4.77 \cdot 10^{-10}, \quad (\eta = 1823 \cdot 10^{-7} \text{ at } 23^\circ \text{ C.}).$$

To it I have adjoined the value which Millikan used for the viscosity of air, the most difficult to measure of all the quantities which appear on the right-hand side of (1.15); he caused it to be re-determined by various physicists of his school. Within the last three years η has been measured anew in half-a-dozen localities remarkably widely scattered, the new values [1] being prevailingly higher (1829 to 1835). Combining these with Millikan's data, we find:

$$e = 4.80 \text{ to } 4.82 \cdot 10^{-10}$$

Thus the question of the viscosity of air suddenly takes on a cardinal importance in the problem of evaluating e.

The fundamental constant e, inferred from the value of N_0 determined by diffraction

In making the grand general affirmation that all charges are integer multiples of e—so brilliantly supported by these charges on droplets—we are practically [2] assuming in particular that all positive ions of any one kind in an electrolytic solution have equal charges; that equation (1.10) refers to actual, not merely to average, values of these ionic charges; and that (F/N_0), which is the greatest common divisor of these charges, is the same as e. To strengthen these presumptions I may mention that by studying the diffusion of ions in gases—such

[1] See H. R. Robinson, *Nature* 142, 159 (1938).

[2] This is not inevitable, but there would be no advantage in admitting any other possibility.

ions, that is to say, as the droplets pick up while drifting through air—Townsend and Salles and others have shown that the product of their mean charge into N_0 is about equal to F, a welcome confirmation of the likeness between ions in gases and those of electrolytic solutions.[1]

We therefore divide the value of N_0 derived from diffraction of X-rays by ruled gratings and crystals, into the value of F derived from electrolysis, and obtain another and an independent estimate of e:

$$e = 4.81 \cdot 10^{-10} \quad \text{(from diffraction)}.$$

This agrees with Millikan's observations on oil-drops, provided that in interpreting the latter we accept the more recent value for the viscosity of air. If, however, we accept the earlier value for η, there is a discrepancy, large enough to be unmistakable—not attributable, I mean, to statistical errors—and yet small enough to leave us feeling instinctively that it must yield to a deeper study of the methods.

Historical notes

The situation which now prevails because of these three just-quoted estimates of e is so remarkable, that its history deserves to be briefly outlined even though it may have passed away before these words are printed.

The oil-drop experiment was first performed in 1909. One may trace the evolution of many of its features through a train of experiments[2] made during the dozen previous years in England, in the last of which H. A. Wilson observed a cloud of water-drops condensed upon negative ions in air, and measured the downward drift-speed of its upper surface when it was falling first under the influence of gravity alone and then under that of gravity combined with a vertical electric field. More precisely, he observed different clouds, making his measure-

[1] *Electrical Phenomena in Gases*, pp. 236 and 244–245. I say "about equal" because the accuracy of the experiments does not rival that of the measurements of F in electrolysis.

[2] Recounted by Millikan in his book *Electrons Positive and Negative*.

ments on some when the field was applied and on others when it was switched off; this was one of the defects of the method (others being the impreciseness of the edge of the cloud and the volatility of the water) which Millikan undertook to avoid. Repeating the experiment, he found it possible to watch the individual droplets, to make them rise or fall by applying or annulling the field, to observe them in the act of capturing ions. H. Fletcher used droplets of oil in place of water, injecting them with an atomizer into the space between condenser-plates, whereas Wilson had produced the clouds by the cumbrous method of expansion which we shall later find so magnificently suited to experiments of quite another type. Millikan devoted several years to the research, and when in 1917 he turned to other work he left the situation of which I will quote the description written in 1929 by Birge: "Only one really precision method for the evaluation of e was known, and the precision work, by this method, had been carried out by a single individual. Under such circumstances it is very fortunate that the investigations . . . represent a masterpiece of experimental technique, in which every possible cause of error, systematic or accidental, was studied in minute detail, and after many years of work reduced to a final minimum." Only within the last year has the experiment again been done [1] with an accuracy considered to vie with Millikan's.

The diffraction of X-rays by crystals was discovered (as I will relate in Chapter VII) in 1912—thirteen years before these rays were first diffracted by ruled gratings, hence thirteen years before their wavelengths were truly measured and before there was a prospect of attaining to the values of N_0 and of e by the study of diffraction. Physicists were not, however, content to leave the wavelengths of X-rays unguessed. They reversed the trend of the argument which I carried through in pages 10–20 of this book—or rather, the argument of this book is a reversal of theirs: for they derived a value of N_0 by dividing Millikan's e-value into the constant F; and having done this,

1 E. Bäcklin and H. Fleming, *Nature* 137, 66 (1936); see Robinson, *Nature* 142, 159 (1938).

a physicist would select his crystal and his X-ray beam—diffract the one by the other—deduce $\overline{m}$ by (1.8), n from his study of the diffraction-pattern, N_u by (1.7), V the volume of the unit-cell by (1.6), and finally λ the wavelength of the X-rays either by (1.5) if the crystal was built on a cubic lattice, or by the proper corresponding formula if the lattice was other than cubic. Spacings of crystals and wavelengths of X-rays were always evaluated in this way until quite lately, and the so-determined values stand yet in all the tables. Therefore, the discrepancy between Millikan's e-value and the e-value obtained through diffraction made itself manifest as a difference between wavelength-values of X-ray beams measured directly (and, as it now seems, rightly) with ruled gratings, and wavelength-values of identical X-ray beams ascertained indirectly (and, as it now seems, wrongly) by the detour which I have just recounted. In the literature of physics since 1927, it has usually and misleadingly been called: the discrepancy between " grating wavelengths " and " crystal wavelengths."

However, it may be named, that discrepancy has called forth a great deal of good research. I have referred already to re-determinations of the viscosity of air and repetitions of the oil-drop experiment, completed lately or in progress. The theory of diffraction by ruled gratings has been re-studied, and the experiments reshaped so as to counter certain objections.[1] It has been suggested that actual crystals may not be of uniform density, *i.e.* that the excellent regular lattice which the diffraction of X-rays discloses may not extend through the whole of the crystalline mass. There are several forms of this suggestion, none of which can be dismissed offhand: there may be (and sometimes are) flaws in a crystal—there may be (and sometimes is) an occasional or a periodic distortion of the lattice—it may be that the parts of a crystal near its surface have a different lattice-spacing from the parts deep beneath the surface. The last in particular must be taken seriously,

[1] As for instance the objection that in the early measurements the X-ray beam was diffracted by only a small part of the grating. Cf. J. A. Bearden, **PR 48**, 385 (1935), and the literature cited there and by him in **PR 37**, 1210 (1931).

for X-rays are usually diffracted within 0.001 mm of the surface of a crystal. To discuss these possibilities in full would take us far afield (they belong to the difficult specialty of the physics of crystals) and it must suffice to say that according to the consensus of opinion at present, they do not impair the accepted e-value derived from diffraction (based as it is on the use of some of the finest crystals extant.) Of relevant experiments I will mention only one: that of J. W. DuMond and V. L. Bollman of Pasadena,[1] who pulverized calcite to a degree of fineness at which the individual granules were so narrow (.002 mm) that the X-rays traversing a granule were diffracted from every part of it—measured the density of the powder with a pycnometer—observed the diffraction by a method appropriate to powdered crystals (page 211 *infra*)—accepted the value of X-ray wavelength given by Bearden's experiments on ruled gratings and came to an e-value between 4.79 and 4.81. Incidentally, they found for the density of the powdered calcite a value almost indistinguishable from that of the unpowdered crystal.

Why not then discard the e-value 4.77, adopting 4.81 in its place and attributing the discrepancy to some error in the early measurements of the viscosity of air? Partly because, in so fundamental a matter, even more repetition of the oil-drop experiment and even more measurements of η are desirable; but partly also because the situation in respect to the e-value is much more intricate than I have yet suggested. There are several other quantities (such as the fundamental constants h and m, the Rydberg constant, the Larmor precession—we shall make their acquaintance presently) which are linked with each other and with e by interconnections so strong that we dare not be satisfied with a value for any one, but must require a coherent set of values of them all. At present it appears that we depart much further from this ideal if we accept 4.81 for e, than if we keep 4.77—in closing the gap between Millikan's e-value and the e-value from diffraction by invoking the

[1] PR 50, 524 (1936).

new values of the viscosity of air, we seem to have opened a new fissure elsewhere in the system of the physical constants! This situation may be remedied soon after or even before this book comes out; but even if it proves to be so transitory, it will be worthy of remembrance as the singular state of affairs which prevailed in 1938.

Other ways of determining the constant e

In the situation as it stands, nothing is more welcome than an additional way of determining e with a precision of one part in a thousand. There are happily two ways which seem capable of being refined to this extent.

The first [1] is applied to alpha-particles: fast-moving nuclei of helium atoms, of which I will speak in Chapter III and thereafter. It is believed for many powerful reasons that each has a charge $+2e$. Now, an alpha-particle can be detected as an individual: by the flash which occurs when it strikes a fluorescent screen (pages 63, 130), or by the discharge which it sets off in the tube known as a " counter " (pages 518–519), or in other ways which do not matter now. If we could set up, in front of a source of alpha-particles, a fluorescent screen connected to an electrometer, and count the incident alpha-particles until the total of the charges which they had brought to the screen should ascend to a measurable amount—if we could do this, we could evaluate $2e$ merely by dividing the amount by the count, the simplest experiment conceivable! Well, we cannot do quite this, for $2e$ is so small compared with the least amount of charge measurable by an electrometer that we should have to count for many lifetimes; but we can do something practically as good—we can count the particles which fall on a screen subtending a very small solid angle as seen from the source, and measure the charge which accumulates meanwhile on a collector subtending a very large solid angle as seen from the source, and divide the ratio of charge to count by the ratio of these solid angles. (Also, we may count for a short time and collect the charge for a long time, or use a

copious source of alpha-particles for the collecting of the charge and a weak one for the counting.)

This method was first employed in 1908, and several times since: of all but the latest experiment, I need say only that the statistical uncertainty is so great that the results cannot be used to support either 4.77 or 4.81 as e-value against the other, though it is highly important to notice that they disagree with neither.

It should be, but perhaps is not, superfluous to point out that this is a method for measuring the *average* charge of a great multitude of alpha-particles. To argue that the average is also the actual charge of each particle, we must rely on the grand affirmation generalized (page 25) out of the oil-drop experiment, plus the great successes of the theory of the helium atom which involves the assumption that the alpha-particle with charge $+2e$ is the nucleus thereof. There is, however, a worrisome point: some alpha-particles, on their way from source to collector, pick up a negative charge which reduces the net charge of each from $+2e$ to $+e$. This makes the average charge a little less than $+2e$; it is believed that the difference is negligibly small, but perhaps the evidence is insufficient; a further study of this point is promised by Schopper.[1]

The remaining method is applied to electrons—those elementary particles which are already familiar to every student of this book, but thus far have been rigorously excluded from its pages. Electrons, as everyone knows, escape from incandescent metals; if the emergence of any one is entirely unconnected with the emergence of any other, each being in technical language an " independent event," then the numbers emerging in consecutive equal intervals of time should not all be equal—they should be distributed over a wide range of values, in a manner depending upon the average of these numbers. Let N stand for the average number of electrons escaping per second, e for the electron-charge, I for the current, $\bar{Q}$ for the average

[1] E. Schopper, **ZP 93**, 1–21 and **94**, 649 (1935); R. T. Birge and E. McMillan, **PR 47**, 320 (1935); J. M. Shrader, **PR 19**, 422 (1922); E. Regener, **BB** (1909), 948; E. Rutherford and H. Geiger, **PRS A81**, 162 (1908).

amount of charge escaping during a time-interval T very short compared with a second; evidently we have:

$$I = Ne, \quad \bar{Q} = NeT = IT. \tag{1.17}$$

Measurement of I and hence of $\bar{Q}$ is easy, but gives us only the product Ne, and not either factor separately. However, if we could measure the actual values Q_1, Q_2, . . . of the charges emerging during a great number of equal intervals T, we could test the theoretical law of distribution of these values, and (supposing the theory to be verified) we could ascertain from it the value of NT and then get an e-value by dividing this into $\bar{Q}$. The actual experiment is supposed to be equivalent to this, but is far from lucid. In it the incandescent metal is the filament of a thermionic vacuum-tube, the output side of the tube being connected through an amplifying circuit to some device for analyzing variable currents; the irregular emission of electrons from the filament manifests itself as a mixture of innumerable frequencies in the current in the measuring device. The theory is difficult and obscure. Various observers have obtained e-values good enough to recommend the method, but not quite exact enough to favor either of the values 4.77 or 4.81 against the other. The proponents of the method believe it capable of being refined to the necessary degree for making this discrimination.[1]

Name of the fundamental constant e

" Electron " is the name to which the constant e has a better historical right than has anything else. It was invented by G. J. Stoney for the charge borne by the least-charged of the ions in electrolysis—the quantity denoted in these pages by F/N_0. This was in 1891, long before the value of e was more than roughly guessed at. For several years it was strictly reserved for this and then for a similar though broadened mean-

[1] N. H. Williams and H. B. Vincent, **PR 28**, 1250 (1926); W. W. Hull and N. H. Williams, **PR 25**, 147 (1925); J. B. Johnson, **PR 26**, 71 (1925) and earlier literature there cited; T. C. Fry, *J. Franklin Inst.* **199**, 203 (1925). The idea is due to W. Schottky.

ing: it was reserved to signify an amount of charge which we now identify with *e*. Gradually, however, there crept in the practice of applying it to a special kind of corpuscle, with a charge equal to *e* and a mass smaller than that of any other which is known, apart from some of the corpuscles of light. This is regrettable, but the usage is now so well embedded that nothing can be done about it any more. I conform to it in this book, and refer to the constant *e* usually by its symbol, occasionally by the word " electron-charge;" whenever the reader encounters this word, he should remember that its meaning is broader by far than " charge of the electron," for it is the basic amount of charge of which all other amounts (so far as we know) are integer multiples.

CHAPTER II

Electrons

We now are going to conceive the world as populated by small particles of various types, all bearing charges which are integer multiples of the fundamental constant *e*. Integers both positive and negative will be admitted, and so will the integer zero, enabling us to include the special class of the neutral particles in the general class of the charged ones. The types will comprise the electrons, negative and positive; the atoms of the chemical elements; the ions which are formed by detaching negative electrons from or adding them to these atoms; the nuclei of these atoms; and the neutrons. Some of these the reader will instantly recognize as being systems of others: the helium atom, for instance, is a system of two negative electrons and a helium nucleus, and the last-named is itself presumed to be a system of two neutrons and two ordinary nuclei of hydrogen. The electrons, on the other hand, are universally taken to be ultimate particles, and the neutrons are often though not always so considered. All of these types, however, will often be called " particles," from end to end of this book.

(But is this not obviously a wrong use of language, and ought I not to make the opposite choice, reserving the name of " particles " for things like electrons which are still visualized as such, and calling atoms by some such name as " systems" ? Questions of this kind are continually cropping up in physics and in the other sciences; and in general it is found that the obvious answer is not the wise one. One point to be considered is, that we shall never be absolutely sure that even the electron is an ultimate particle; to call it always a " particle-or-system " would be the only strictly-right procedure, but also

it would be excessive—it would be like calling every star in the sky a multiple star, on the ground that better telescopes hereafter may prove it to be such. The essential point is, that in many important experiments—including all of those to be mentioned in this chapter—the atoms or the ions or the nuclei do act as though they were ultimate particles: I therefore treat them and name them as such, till later in the book I come to speak of experiments which prove them to be systems.)

All of these particles have *mass*, and the measurement of their masses is one of the major enterprises of fundamental physics. To say that atoms have mass may sound rather trite; large pieces of matter have it, how then shall the atoms which compose them avoid it? Well, the statement is not so self-evident as it may sound, and the proof of it *was* worth the making; but I will not dwell upon the point. For the electrons, the statement is not commonplace at all. Many people in the past have believed electricity to be an imponderable fluid, and it means a good deal to be able to say at present that electric charge is never disembodied to such an extent that it loses inertia completely: moving charge must always have momentum, and whirling charge have angular momentum.

Measurement of the masses of these atomic and subatomic particles is the leading topic of this chapter. It is a topic much broader than the name suggests. These masses are not measured by weighing—exception being made for the method indicated in Chapter I and culminating in equation (1.7), which, however, is applicable to atoms only, and usually supplies us with a figure which is not the mass of a single kind of atom, but a sort of average among the masses of several. Neither can they be measured by causing one of these tiny particles to interchange momentum with a body heavy enough to be weighable. We have to proceed by computing the paths or trajectories which particles of a constant charge and a constant mass ought to follow in electric and magnetic fields such as we can establish—then observing the paths which electrons and ions and nuclei do actually follow—then proving that the actual paths agree in shape with the theoretical, and finally

showing that by choosing suitable values of mass the agreement can be made complete. *The measurement of masses is the study of trajectories.*

It will transpire that usually it suffices to locate just two points on a trajectory (though there is a method for making the entire path almost continuously visible). It will transpire also that when the path of a particle is through a single field (either electric or magnetic) a study of it will supply the observer not with a value of the mass m, but with a value for a certain combination of the charge-to-mass ratio ne/m and the speed v of the particle. If, however, the path is through both an electric and a magnetic field (superposed or adjacent) the observer may have, in effect, two simultaneous equations for the two unknown quantities ne/m and v, and the experiment can give him the values of both, a welcome enrichment of knowledge. He can never make the one remaining step, and break apart the numerator and the denominator of the charge-to-mass ratio, determining m by itself; having got his value of the ratio, he must presume for its numerator a value taken from an experiment of quite another kind or else from a more-or-less reliable guess, and only thus can he arrive at the m-value. Anticipating one of the results: the masses of the particles named in the third sentence of this chapter are *not* found to be small-integer multiples of a common fundamental unit, as are the charges thereof. There is a fundamental constant denoted often by m (or better by m_0); it is the mass or rather the rest-mass (*vide infra*) of the electron; but it is not believed to be the greatest common divisor of all masses.[1]

We turn now to the theory of trajectories, treating it in Newtonian fashion and disregarding for the time being the modifications due to relativity.

[1] This is hard to phrase impeccably. The mass of the electron is about 1850 times smaller than that of the next lightest material particle (the proton) and it cannot yet be said with certainty that the proton-mass is not an integer multiple of the electron-mass (the same is true *a fortiori* for particles still more massive); but many things combine to indicate that integral ratios are not to be expected.

Trajectories of a charged particle in an electrostatic field

Imagine a particle of charge ne and mass m exposed to an electrostatic field of fieldstrength E. It will be assumed to behave just like a pithball or other charged body of tangible size, responding to the field by an acceleration given as follows:

$$e\mathbf{E} = d(m\mathbf{v})/dt = md^2\mathbf{s}/dt^2. \qquad (2.1)$$

(The position s [distance from an arbitrary origin] and velocity v of the particle, and the strength E of the field, are printed in boldface in the equation to indicate that they are really vectors). This is one of the assumptions which the experiments have verified, over and over again. The experiments, by the way, are usually conducted in gases at low pressure so that no forces due to atoms shall act upon the particles; both the theory and the data correspond to motion in a vacuum.

To take a more specific case, imagine a beam of particles each having the charge-to-mass ratio ne/m, traveling with speed v along the x-axis from the direction of negative x, through field-free space up to the origin ($x = y = z = 0$). Suppose that there is an electrostatic field of fieldstrength E directed parallel to the y-axis and extending over all the region from $x = 0$ to $x = k$: then from the origin onward the particles will describe a parabolic path extending to the point of which the coordinates are

$$x = k, \quad y = \tfrac{1}{2}Ek^2(ne/m)v^{-2}, \quad z = 0. \qquad (2.2)$$

One might locate this point (*e.g.* by placing a photographic plate in the plane $x = k$, and looking for the dot made on it by the impinging particles) and thus ascertain the value of $(ne/m)v^{-2}$, but not that of ne/m by itself nor that of v by itself. The theory itself could be tested by repeating the experiment with many different values of k and of E, and seeing whether the same value of $(ne/m)v^{-2}$ was always obtained: the reader should realize that tests equivalent to this have been made innumerable times.

Trajectories of a charged particle in a magnetic field

In deciding what shall be expected of our particles when moving in magnetic fields, the inference from earlier experience is not nearly so direct. We must begin by inquiring what is the effect of a magnetic field on an electric current, an ordinary current in a wire. Ampere worked this all out (and very much more) in an extraordinary burst of genius, needing only six weeks from the day in 1820 when he heard at the *Académie des Sciences* in Paris of Oersted's observations of the year before in Copenhagen (Oersted had observed that an electric current exerts a force on a magnetic pole).

To begin with the simplest case, that in which the current is at right angles to the magnetic field: there is a force acting on any length L of the wire, its direction being at right angles to both the current and the field, and its magnitude F_L being proportional to L and to the strengths i and H of current and of field. Disregarding for the moment the remarkable direction of the force, we consider the formula for the magnitude. The factor of proportionality occurring in it must depend, of course, on the unit-system chosen for currents and magnetic fields; Ampere chose a unit-system (now called the " electromagnetic ") in which the factor is unity:

$$F_L = iHL \text{ (electromagnetic system).} \qquad (2.3)$$

Now interpret the current as a stream of particles of charge ne, of which there are N per unit length of the wire, all of them advancing with speed v; the amount of charge passing any cross-section of the wire per unit time, which quantity is none other than the current, is given thus:

$$i = Nnev \text{ (electrostatic system).} \qquad (2.4)$$

I have appended the words " electrostatic system " to equation (2.4), because it is coming to be the convention (and a desirable one) in physics to interpret the symbol e as standing for the fundamental charge measured in electrostatic units. In substituting from (2.4) into (2.3), we must therefore insert

the conversion-factor between the two systems, universally known as c. Doing this, we get:

$$F_L = NnevHL/c. \tag{2.5}$$

This is the amount of the force on length L of the wire, containing NL of the charge-bearing particles; dividing it by NL, we get

$$F = Hnev/c, \tag{2.6}$$

and the simplest conceivable inference is, that this is truly the expression for the force on a single charged particle, moving in vacuo in a magnetic field at right angles to the direction of the field. We complete this simplest inference by supposing, that the force in such a case is at right angles to the directions of the velocity and of the field (the question of its sense will be dealt with presently). The inference or assumption, thus completed, is verified by numberless experiments.

Apparatus is so commonly set up with the magnetic field directed at right angles to the beam of charged particles, that we scarcely need to generalize beyond this case, easy as the generalization is (*vide infra*). As soon as the particles enter the field, they are whirled in circular arcs, never leaving the plane perpendicular to the field and never being speeded up or slowed down (one might say that the force is always satisfied with the speed of the particles but never with their directions of motion). Examples of these circular arcs are exhibited shown in figures of this book (2.2 and 6.8). The radius R of a circle may be found by multiplying m into the centripetal acceleration v^2/R (the product is what is misnamed " centrifugal force ") and equating the result to F as given in (2.6):

$$R = mcv/Hne, \tag{2.7}$$

and by measuring the circle the observer may find the value of the combination $(ne/m)v^{-1}$. To be specific, imagine once more the beam of particles travelling along the x-axis from the negative x-direction, until it reaches the origin and enters a field extending from $x = 0$ to $x = k$, which shall now be

a magnetic field of fieldstrength H directed parallel to the y-axis. The particles will now swing around in circular arcs in the xz-plane, the center of the arcs being at ($x = y = 0$, $z = mcv/Hne$); they will extend from the origin to the point of which the coordinates are:

$$x = k, \quad y = 0, \quad z = R - \sqrt{R^2 - k^2} = \tfrac{1}{2}Hk^2(ne/cmv), \quad (2.8)$$

the second of the expressions for z being an approximation which is adequate if the arc is a small fraction of a complete circumference, as is usually true in such experiments. One may locate this point as before by placing a photographic plate in the plane $x = k$, and thus ascertain the value of $(ne/m)v^{-1}$, but not that of ne/m by itself nor that of v by itself. Neither the use of a magnetic field alone, nor the use of an electric field alone, can lead us to the value of the charge-to-mass ratio of particles; but by the use of both, or by either combined with an auxiliary measurement of speed, we can arrive at it.

An important theorem about the circles is the following. The angular velocity ω of the particle in the circle, and hence the frequency $\omega/2\pi$ and the period $2\pi/\omega$ of its revolutions, are independent of its speed! this is easily read from (2.7), considering that ω is v/R:

$$\omega = (ne/mc)H. \quad (2.9)$$

On this theorem reposes the efficacy of the cyclotron, the most powerful of the engines of transmutation (page 508).

The law of the action of a magnetic field on a moving charged particle, in the general case where the angle θ between velocity and field is not restricted to 90°, is easy to formulate. Resolve the velocity into the components $v \cos \theta$ parallel and $v \sin \theta$ perpendicular to the field-direction; the former is permanently unaffected, the latter is affected as it would be if it were the entire velocity of a particle moving at right angles to the field. The motion of the particle is compounded of a steady drift parallel to the field with linear speed $v \cos \theta$, and a steady revolution about a line parallel to the field, with angular velocity $(ne/mc)H$; the path is a so-called "helix" wound

upon the surface of a right cylinder of radius $(mc/Hne)v \sin \theta$. Imagine a line drawn parallel to the field, through any point of the path; the particle must re-traverse the line at regular intervals of time equal to $2\pi mc/neH$, and must never depart from it by more than the diameter of the cylinder.

General equation of motion of a charged particle

All of the special cases foregoing, and innumerably many more, are implicit in the general vector equation for the behavior of a charged particle exposed to an electric field E and a magnetic field H

$$d(m\mathbf{v})/dt = ne\mathbf{E} + \frac{ne}{c}(\mathbf{v} \times \mathbf{H}), \tag{2.10}$$

which for those insufficiently acquainted with vectors I will expand into the triad of scalar equations:

$$\begin{aligned} d(mv_x)/dt &= neE_x + \frac{ne}{c}(v_yH_z - v_zH_y) \\ d(mv_y)/dt &= neE_y + \frac{ne}{c}(v_zH_x - v_xH_z) \\ d(mv_z)/dt &= neE_z + \frac{ne}{c}(v_xH_y - v_yH_x). \end{aligned} \tag{2.11}$$

The sign of the second term in the right-hand member is a source of undying confusion, especially when n is negative as it normally is with electrons. Probably the best thing for the student to do, when confronted with an actual case, is to revive his old memories of the trick of pointing the fingers whereby he was taught to remember the "motor rule" (the thumb and forefinger and middle finger of the left hand are pointed at right angles to each other, and indicate respectively the force, the magnetic field and the current). Never forget, in so doing, that when a negative charge is moving in a positive (negative) direction, it corresponds to a negative (positive) current! However, the only people who are absolutely compelled to remember this rule are those who set up the apparatus

and make the observations, and must find out the signs of the moving particles before they can be sure just what they are observing. Those who read their articles must usually take it on faith that they made no mistake.

Finally, I must dispel the impression, left behind by my words of page 39, that physicists conceive the flow of current in a wire as a uniform motion of charged particles all advancing with exactly the same velocity parallel to the axis of the wire. This image, which in spite of its naïveté has just guided us to the correct expression for the action of a magnetic field on a particle, is superseded at the next stage of sophistication; in which stage we suppose, that the motion of the particles through the current-carrying wire is like that of the molecules of the air in a slow and steady breeze, a very rapid random zig-zag flying ("thermal agitation") on which is superposed the slow and steady drift. The forces which the particles experience from the magnetic field because of their random zig-zags cancel each other out, but there is an unbalanced force resulting from the drift, and its total amount for NL of the particles is given by equation (2.5), if in this equation (and in 2.4) we interpret v as the drift-speed. The inference after equation (2.6) is no longer so immediate, but it is binding.

Measurement of the speeds of particles

As I have stressed already, the paths of charged particles in an electric field give us values of $(ne/m)v^{-2}$, while the paths in a magnetic field give values of $(ne/m)v^{-1}$; thus if we can form a separate estimate of the speed v of the particles in a stream, we can derive the charge-to-mass ratio ne/m from trajectories in either one of these two kinds of field. Now, there is a method of very broad serviceability for making estimates of speed, in cases—like all of those which now concern us—where the particles are so small or so swift (or both) that we cannot watch an individual and time him as he paces off a measured mile. For brevity I will call it " the method of the gates."

For simplicity imagine first a stream of particles of identical speed v, coming along the x-axis from the negative direction; and a pair of walls occupying the planes $x = 0$ and $x = k$, with apertures at the x-axis which can be opened or closed at will by ideally quick-working gates; and a detector beyond the second gate. Suppose that at the zero of time the first of the gates is opened for a time-interval θ, preferably very brief by comparison with the time $T = k/v$ which a particle takes in travelling from the first gate to the second; and that at time t the second gate is opened for the same duration θ. Evidently, *particles—or rather, portions of the stream which we imagine as circumscribed flocks of particles—will be detected beyond the second date, if and only if t is greater than* $(T - \theta)$ *and less than* $(T + \theta)$. If the observer has t completely at his disposal, and a sufficiently good detector beyond the second gate, he thus can determine T with an uncertainty of 2θ, and v with a corresponding accuracy (indeed he might conceivably do even better, for the response of the detector should pass through a maximum at $t = T$, and he might locate the maximum more closely). Moreover, if there is a distribution-in-speed among the particles arriving from the negative x-direction, the observer can map it by varying t, with an exactitude which is greater the smaller the value of θ. Of course, θ cannot be made indefinitely small, else the number of particles coming through would be too small to detect; but this limitation is not so bad as it may seem at first, for one may repeat the experiment over and over again (millions of times in the second, with very swift particles) while using a detector which sums up the fragments of the stream which succeed in arriving at it.

With charged particles, the "gates" are usually slits or grids, across which a transverse field (perpendicular to x) is periodically applied by an alternating-current generator or a vacuum-tube oscillator. The transverse field might be strong enough when at its peak to pull all the particles to the edges of the slit or the wires of the grid, shutting the gate completely. This, however, is not imperative; suppose that the fields at the two gates are sine-functions of time, attaining their zero-

values at instants separated by t; then if t is equal to T the stream, or rather an occasional piece of the stream, will go through the apparatus undeflected and fall completely upon the detector if this is centrally placed upon the axis passing through the gates, while whenever $t \neq T$ the transmitted pieces of the stream will be deflected and will miss the detector partly or completely. This scheme of the transverse fields is the one employed by Van de Graaff, by Tyndall and his school at Bristol, by Loeb and Bradbury and others at Berkeley, for measuring the speeds of ions in gases pervaded by a field; and by Kirchner at Munich and Chaffee at Harvard for measuring the speeds of electrons in a vacuum, as I shall presently relate.

If the observer cannot or will not use a transverse field at the gates, he cannot substitute a pair of actual swinging doors, for they could not be operated fast enough. He can, however, use for his walls a pair of solid discs mounted upon an axle parallel to the x-direction, and for his gates a narrow radial slit or a narrow radial indentation in each of the discs. Then, he must turn the axle with a rapid adjustable angular speed—call it ω—and the time t will be the lapse of time between the passages of the two slits or the two indentations across the axis of x which is the path which the particles are following; he varies ω until the reading of the detector attains its maximum, and the value which it then has gives him the value of T. This scheme has sometimes been used with charged particles of the very slow-moving kinds, but it is especially precious for uncharged particles (such as neutral atoms or molecules, and neutrons) for which there is no other. If at a distance k beyond the first of the discs there is, not a second disc but a stationary plane perpendicular to the x-direction which is able to reflect the particles perfectly—a mirror, that is to say—then the particles will come back along their track and re-pass the disc if and only if the slit or indentation has meanwhile returned to its first position, that is, if the disc makes one complete revolution during the time $2k/v$. This last, with some minor alterations not affecting the principle of the scheme, is the famous "toothed-wheel" method for measuring the speed of light,

dating from 1849; and the other, the "rotating-mirror" method (page 155) of measuring that speed is not fundamentally different.

But is not light a wave-motion? and in performing on light such an experiment as I have just described, is the observer not measuring the rate of travel of waves, rather than the speed of particles? Well, it will later become apparent that all beams, whether of light or of all the various material and electrical types which I listed at the beginning of this chapter, unite the qualities of a stream of particles with those of a train of waves. When the first gate is open for the period θ in one of these experiments, it is letting through something which unites the qualities of a circumscribed flock of particles with those of a segment dissected out of a wave-train. The speed which is measured by means of the two gates is strictly that of the segment or the flock (whichever way we choose to picture it); and it is to be identified with the speed of the particles, but except in special cases it is *not* to be identified with the speed of the wave-crests of the train. These mysteries, however, lie far ahead, and I mention them here only in order to save this chapter from being entirely corpuscular in its point of view.

Negative electrons

It is high time to make this discussion more specific by introducing actual types of particle, and none of them has a better right to being cited first than the most numerous and most ubiquitous of all, the negative electron.

Electrons (henceforth, when I omit to give their sign, I shall be speaking of the negatives only) are released, or extracted, or ejected from so many kinds of matter by so many different techniques as to leave no reasonable doubt that they inhabit every solid and every liquid and the atoms of every gas. It would be wasteful to go into detail about the processes here, since much of this book will be devoted to precisely such detail; but to catalogue them: heat, light, electric dis-

charges, some chemical reactions, the impacts of fast electrons already set free or of charged particles of the other categories—all these are agents capable of detaching electrons from atoms or from bulky pieces of matter. Electrons emerge from certain radioactive substances both natural and artificial, and then are sometimes known as " beta-rays," though nothing but remarkably high speed distinguishes them from electrons of any other provenience.[1] Floods of electrons emerge from the cathodes of discharge-tubes operated at a moderately low pressure and a moderately high voltage, and then are sometimes known as " cathode-rays," though again they do not differ as individuals from any other electrons.

Beta-rays have such enormous speeds at birth, and cathode-rays have such great speeds when they emerge from the region of the discharge, that it is simple to form beams of their particles moving along nearly-parallel lines: all that is needed is a series of collinear slits or otherwise-shaped holes in walls or partitions adjoining the source. This is a rarer quality than we can readily realize, familiar as we have become with this example; Crookes was impressed so much with this quality of the cathode rays that he called them " a fourth state of matter," even while still supposing them to be ordinary atoms. Such beams being procurable, the observer can go ahead with determining the speed and the charge-to-mass ratio by sending them through electric and magnetic fields, or through a field of either kind and through a pair of such gates as I have just been describing.

Electrons boiling out of an incandescent metal ("thermionic electrons") or expelled by ultra-violet light from a metal ("photoelectrons") are by contrast moving so slowly that they cannot be formed into neat beams until they have been accelerated by an electric field. This is not so great a drawback as it might be; for in accelerating them, the observer may dispense himself from one of the stages of the dual process which

[1] Some of these emerge from atom-nuclei, others from the families of orbital electrons surrounding such nuclei (Chapter IV); there is now a tendency to confine the term "beta-rays" to the former.

would otherwise be necessary in measuring the charge-to-mass ratio. Let v_0 stand for the speed of an electron fresh from the source, entering the accelerating field; and v for its speed when it leaves that field and is ready to enter the next stage of the experiment; and V for the potential-difference (in electrostatic units!) between the points of entry into and exit from the accelerating field—then,

$$\tfrac{1}{2}mv^2 - \tfrac{1}{2}mv_0^2 = neV. \qquad (2.12)$$

Now if V is so great that v^2 is thousands of times as great as v_0^2—and this is not at all impracticable—then the observer may ignore the second term on the left of (2.12), writing:

$$ne/m = v^2/2V, \qquad (2.13)$$

and then he may determine v by the method of the gates or $(ne/m)v^{-1}$ with the aid of a magnetic field, and so ascertain the charge-to-mass ratio without having markedly increased the net uncertainty of the experiment by that approximation.

I have continued to write ne/m, as though it were still uncertain which value of n is appropriate to the electron; but it would be absurd to do so longer, in view of the overpowering evidence (a welcome part of which, I recall, is furnished by the fluctuations of thermionic currents) that for the negative electron n is -1. The proper way to write the charge-to-mass ratio of the negative electron is therefore $-e/m$; but the custom is, to write it e/m; and to this I will defer.

Before introducing the value of e/m, I pause to make a forecast. Say we take the value of the fundamental constant N_0 given in Chapter I, derived from diffraction, and divide it into the atomic weight of the lightest element (hydrogen) in order to get the (average) mass of hydrogen atoms which are the lightest of all atoms; and say we divide this mass into e—· we then get a charge-to-mass ratio equal about (there is no need for being extremely precise) to $3 \cdot 10^{14}$. Should the charge-to-mass ratio for the electrons be found to be markedly greater than this, the electron would be proved less massive than any chemical atom, it would be proved a subatomic particle. We

do not even have to invoke diffraction, in order to reach this conclusion. The ratio defined in the second sentence of this paragraph is the ratio of the Faraday constant F to the atomic weight of hydrogen, it is the value for hydrogen of the ratio IT/W defined in Chapter I, as the student can easily see by reviewing that chapter. In brief, it is something *ascertainable from electrolysis*, and it had been ascertained a century ago. There was a sensation in the closing years of the nineteenth century, when J. J. Thomson proved that the charge-to-mass ratio of the electron is thousands of times greater than that long-known value, and hence that the electron (he called it " corpuscle ") is subatomic.

Charge-to-mass ratio of the electron

Beginning with Thomson's pristine measurements of 1897, the measurement of e/m now has a history of forty years—somewhat longer than that of the measurement of e, and decorated by a much larger number of names, for literally scores of physicists have tried to surpass or at least to rival those who preceded them. Even though the methods are in principle all covered in the few pages of this chapter, the conceivable variations in detail are literally limitless. Some of the schemes are remarkably elegant and clever, and quite an interesting book might be written on the art of accurate experiment, based entirely on the attempts to measure this charge-to-mass ratio. Thomson employed the scheme epitomized in equations (2.2) and (2.8), applying first an electric field and then a magnetic field at right angles to a beam of cathode rays, inside the discharge-tube where they originated, and measuring the displacement of the luminous spot which the beam made when it impinged on a fluorescent screen occupying what I earlier called the plane $x = k$ (page 37). He also estimated e/m for slow-moving photoelectrons, availing himself of a principle of which a different example was given on page 40: an electric field is perpendicular and a magnetic field is parallel to the plane of the flat metal plate from which the electrons are emerg-

ing, and owing to the magnetic field these cannot get away by more than a certain distance from the plate, even though the electric field is pulling them outward—the value of the distance, combined with those of the two fieldstrengths, gives e/m (a welcome means of proving the identity of photoelectrons and cathode-rays).

Very early in the game it became apparent that if one writes:

$$e/m = 5.3 \cdot 10^{17},$$

even the second significant figure is practically certain. The electron-mass must therefore be almost two thousand times smaller than the (average) mass of hydrogen atoms, the lightest atoms which there are! it is indeed a subatomic corpuscle with which we are now dealing. This is so small a mass that (as we shall later see) all the electrons in a piece of matter share between them not much more (and usually less) than 1/2000th of the mass and weight of the piece.

To decide the third significant figure of e/m has proved much harder. After thirty years of experimentation it seemed almost certain that the true value must be within .002 of 5.31 (I omit from now on the factor 10^{17}). About that time it became apparent that the situation was somewhat like that which now prevails in respect to e. Certain phenomena of spectroscopy—above all, the Zeeman effect—are interpreted by a theory which enables us to derive a value of e/m from them. The theory will be given in Chapter XIII; for the nonce I will say merely, that it is so simple and forthright that if it is wrong, there is a very serious flaw in the structure of theoretical physics. The value thus obtained for e/m is 5.28.

Accordingly, the question of e/m has revived and returned to the foreground of physics, very much like that of e; and the experiments of the last few years have been closely considered.[1] C. T. Perry and E. L. Chaffee at Harvard accelerated slow electrons, applied the method of the gates to determine their

1 Perry and Chaffee, **PR 36**, 904 (1930); Kirchner, **AP 12**, 503 (1932); Dunnington, **PR 52**, 475 (1937); Shaw, **PR 54**, 193 (1938).

speed and equation (2.13) to calculate e/m; so did F. Kirchner in Munich; they all found 5.28, with a statistical error much smaller than the difference between it and 5.31. F. G. Dunnington of Berkeley achieved ingeniously a scheme (conceived by Ernest Lawrence) for uniting in a single experiment the method of the gates with the deflection by a magnetic field. To revert for a moment to page 40: when electrons in a magnetic field H are travelling in a circle and are describing the complete circumference C of the circle in the time $2\pi mc/eH$, and when the gates are set up at two points of the circle separated from each other by the fraction aC of the circumference, then the time T which a circumscribed flock of electrons takes in getting from gate to gate is $a(2\pi mc/eH)$, and a measurement of T is tantamount to a measurement of e/m. In this experiment, the fields at the gates were parallel instead of transverse to the direction of motion of the particles. The field at the first gate, operating for a short time θ, imparted their speed to the electrons (which initially were slow-moving thermionic electrons) and started them upon the prearranged circular path; the field at the second gate, operating for a short time θ beginning t seconds later than did the first field, was equal and oppositely-directed to that first field, and consequently stopped the electrons—the interval t was that between the momentary opening of the first gate and the momentary shutting of the second, and was equal to T when the current to the detector was at a minimum. Dunnington also arrived at the value 5.28 for e/m. A. E. Shaw, of Dempster's school at Chicago, is applying a peculiar combination of deflecting electric and magnetic fields to accelerating electrons: its special merit is, that it enables a correction to be made for that insidious source of errors, the contact-potential-differences between various metal parts of the apparatus. Shaw likewise reports that the results favor 5.28, falling indeed rather closer to 5.27.

What was formerly known as the "spectroscopic" value of e/m has therefore triumphed over what was formerly known as the "deflection" value. In all probability these contrasted names, lately quite common in the literature, will shortly

vanish; but history should preserve this curious story of experiment being compelled by the pressure of theory to rectify itself. I mention in closing that nearly always the values stated in the journals and the books are values of e/mc.

Estimate of e/m from refraction of X-rays

If a beam of monochromatic light could be sent through a region of space containing a sufficiently large number N of free electrons per unit volume, its phase-speed (Chapter VI) would be altered to an appreciable extent, proportional to N and depending on e/m according to a formula derivable either from classical or from quantal theory. This situation cannot be exactly realized, but is imitated in the laboratory by sending a beam of high-frequency light (X-rays) through a crystal, diamond for instance. The electrons in the crystal, instead of being literally free, are "bound" to the atoms thereof; but the effect of the binding is relatively slight, and the formula can be modified by quantal theory to take account of it. Observations on the refraction of the traversing beam yield the value of the refractive index, hence of the phase-speed; N is obtained by multiplying the number of atoms per unit volume of the crystal (obtained by the method of Chapter I) by the number of electrons per atom (identified with the atomic number of the atom, Chapter IV); e/m is computed. Bearden with this technique [1] obtained the value 5.28.

Estimate of e/m from space-charge phenomena [2]

Imagine a thermionic vacuum-tube, idealized for the present purpose into a pair of infinite parallel metal plates separated by the distance d, one of them (the cathode) incandescent and pouring out great numbers of electrons, the other cold. Let the potential of the cathode be perpetually zero,

[1] PR 54, 698 (1938); he gives the result as (1.7601 ± .0003). Theory: Wheeler and Bearden, PR 46, 755 (1934).

[2] *Electrical Phenomena in Gases*, pp. 321 ff.

that of the anode (V) be gradually raised from zero. At first, owing to the negative charges on the electrons besprinkled throughout the interspace, there will be a potential-minimum between the electrodes. Only the fastest of the electrons emerging from the cathode will be able to pass this potential-minimum and reach the anode; all but the fastest will be returned to the cathode. As V increases, the minimum becomes steadily shallower (also, it steadily approaches the cathode) and a steadily increasing proportion of the emitted electrons succeeds in passing it—*the current between anode and cathode increases steadily with the voltage*, until V becomes so high that the potential-minimum vanishes and all of the emitted electrons go on to the anode (the state of "saturation" of the thermionic current). By making certain approximations, one may derive for the relation between current (i) and voltage the famous "three-halves power formula,"

$$V^{3/2} = (9\pi/\sqrt{2})(e/m)^{-1/2}d^2i, \qquad (2.14)$$

or the corresponding formula for the case in which cathode and anode are coaxial cylinders.

These formulae have been verified over and over again, with a closeness limited only by the errors accepted in the approximations. Closer approximations than these have also been worked out, but even so, this has not yet become a method for evaluating e/m which can rival the others. Nevertheless, it gives a respectably accurate value; and it is interesting to know that by the use of an ordinary thermionic diode, without the assistance of any magnetic field nor the method of the gates, one may demonstrate that electrons have a charge-to-mass ratio far higher than do atoms, and consequently must be subatomic.

Relativistic equations

Without attempting even to enter upon the notions concerning frames of reference, observers, signals, clocks, *et caetera*, on which the special theory of relativity was originally founded,

I quote here two of the changes which must be made when, in any physical argument, relativistic mechanics is submitted for ordinary [1] mechanics. It will presently be seen that nowadays these alterations may be justified by experimental evidence of great directness and power.

First: in the fundamental equation " force is equal to rate of change of quantity-of-motion " with quantity-of-motion expressed as mass-times-velocity mv, we make the alteration by treating m no longer as a constant, but instead as a function —a particular function—of speed v:

$$\mathbf{F} = a(m\mathbf{v})/dt, \quad m = m_0/\sqrt{1 - \beta^2}. \tag{2.15}$$

Here β (the conventional symbol) stands for the ratio of v the speed of the body in question to c the speed of light *in vacuo*. I say " body " instead of " particle " because according to the theory of relativity, this is a principle valid for all material bodies whatever their size, massiveness or complexity. The tests described in this book relate, however, to bodies which are material particles in the sense of this book, *viz.* electrons and atom-nuclei. The constant m_0, called the " rest-mass," varies from one type of particle to another.

Second: in the equation " sum of kinetic energy T and potential energy V is constant " with $\frac{1}{2}mv^2$ put for the kinetic energy, we make an alteration which changes it to this:

$$\begin{aligned} \text{constant} = \text{energy } E &= mc^2 + V \\ &= (m_0/\sqrt{1 - \beta^2})c^2 + V = m_0c^2 + T + V, \end{aligned} \tag{2.16}$$

so that the constant entity—the energy—is now the sum of the potential energy and a term mc^2, in which m stands for that same function of v as we have just encountered. This term may be subdivided into a constant, m_0c^2, and a power-

[1] It is bothersome to decide what to call the pre-relativistic mechanics, now that "classical mechanics" has come to stand for the antithesis to quantum-theory, and therefore to comprise both the relativistic and the pre-relativistic type. Moreover one does not want to choose a word which implies antithesis too strongly, since after all the pre-relativistic is a limiting form of the relativistic mechanics. To speak of "Newtonian mechanics" suggests that there was no progress between Newton and Einstein. I use the term "ordinary mechanics," but with reluctance.

series in even powers of β. The constant is called the " rest-energy "; the power-series is denoted by T and called the " kinetic energy," even though it is not identical with $\frac{1}{2}mv^2$:

$$T = (m - m_0)c^2 = \\ = \tfrac{1}{2}mv^2 + (\text{terms of higher order in } v). \qquad (2.17)$$

The quantity mc^2 has no name of its own, which is unfortunate as it is often called " energy " for short, and this induces people to forget about potential energy.

The tests of the relativistic equations are of four kinds: measurements of the speed of light in circumstances in which the velocity of the observer or that of the source or that of the medium of transmission might be expected to manifest itself—inferences from the theory of spectra—measurements on the trajectories of free electrons—observations on transmutation. The first kind lies outside of the scope of this book (it is usually the foremost in treatises on relativity). The second kind will be mentioned in Chapter XIV, but in the main belongs in books more advanced than is this one. The third and the fourth are appropriate here.

For the measurements on free electrons [1] no addition need be made to the technique, since all the ways of determining e/m involve the learning of v anyhow, as we have lately seen. If the method simplest to apply were the one with the simplest equations, it would consist in producing a homogeneous pencil of fast electrons, measuring v by the method of the gates, bending the pencil into a circular arc by a magnetic field, measuring the radius R of the circle and deducing m from (2.7). Something like this has been done, with a different way of determining v which is useful in many situations and worthy of mention here. The beam is directed along the x-axis, electric and magnetic fields are applied parallel to y and z respectively, and their strengths are adjusted till the beam is found to be entirely undeflected, whereupon:

$$E/H = v/c. \qquad (2.18)$$

[1] An excellent critical review of the work up to 1926 is to be found in *Handbuch der Physik* 22 (article by W. Gerlach).

(The student can easily prove this with ordinary mechanics from (2.3) and (2.8), and it remains valid with relativistic mechanics). If the beam comprises electrons of various speeds, the field-pervaded region serves as a " filter," only those going through undeflected which obey (2.18); by varying E/H one may transmit, independently and successively, electrons of whatever speed one wishes to expose to the deflecting field beyond. Such was the scheme of A. Bestelmeyer, A. H. Bucherer and others prominent in the early history of the problem: they used beta-rays, their sources providing them with wide ranges of speeds from which they selected one value after another of v by filtering. Hupka, on the other hand, accelerated slow electrons by using a high voltage (up to 90,000), and subsequently deflected them by a magnetic field; he interpreted the data by using (2.12) with the relativistic expression (2.17) for T inserted in place of $\frac{1}{2}mv^2$, and (2.7) with the relativistic expression (2.15) for m. The discovery that m varies with v was made (by W. Kaufmann, about 1905) by what in Chapter III we shall study [1] under the name of " parabola-method "; and similar schemes have been used by C. E. Guye and others.

The outcome of these experiments was: *apparent agreement with the relativistic equations.* Especially good checks were made from $\beta = 0.35$ to $\beta = 0.63$ by Hupka, and from 0.39 to 0.85 by Neumann. All the experimenters used their data to test not only the relativistic but another theory, which the data seemed to invalidate. Meanwhile the prestige of the theory of relativity grew to be so great that for ten years nobody undertook researches of this type; but in 1937 Zahn and Spees [2] began to re-examine the methods, and they have criticized severely all the prior work, restoring the problem to the agenda of contemporary physics. Incidentally, appeal

[1] Pp. 74–80 *infra.* As will there be explained, the deflecting fields and the photographic plate are so disposed that if the beam consists of particles with a continuous range of speeds but all possessing the same e/m-ratio, they imprint a parabola upon the plate. In the Kaufmann experiments with a beam of electrons having a broad range of speeds, the curve upon the plate was distinctly not a parabola.

[2] **PR 53**, 511 (1938) and earlier papers there cited.

can now be made to measurements of electron wavelength (page 225).

The values of e/m for the negative electron, given in foregoing pages and in other books and tables generally, are thus to be regarded as estimates of the *limiting* value approached by e/m as v approaches zero. Strictly, they should be denoted by e/m_0, with which usage m_0 means what it means in (2.15). The obvious question " how large may v become before the increase in e/m is perceptible, or must be taken account of? " cannot properly be answered without a full description of whatever apparatus the inquirer has in mind. Usually the inquiry is phrased in terms of kinetic energy, in which case a rough but perhaps-not-too-misleading answer is " 20,000 electron-volts "—*i.e.* so long as $\frac{1}{2}mv^2$ is less than 20,000 ev, a perceptible error will rarely if ever be made by presuming the ordinary instead of the relativistic equations. For alpha-particles and other atom-nuclei, the answer in terms of kinetic energy is an enormously larger figure, since the larger m_0, the smaller the v corresponding to a given T. (This incidentally is the reason why the cyclotron is effective with nuclei though not with electrons: equation (2.9) departs steadily further from the truth as v increases, but with the more massive particles the departure does not become serious until T is already tremendous, beyond ten *Mev*). The test of relativity by the trajectories of free nuclei is therefore almost or quite beyond our present powers; but it is precisely for these particles that the fourth of the mentioned tests becomes available.

The fourth test will be described at length in the first chapter on transmutation: at this point a brief mention will suffice. Briefly, then, there are countless examples of a general type of process, in which two nuclei A and B encounter each other, and from the scene of the encounter there spring two different nuclei C and D: as though A and B had merged, and the body resulting from the merger had exploded into two fragments different from the original two components. Often all four of the nuclei are recognizably the atom-nuclei of known elements, of which the rest-masses have been ascertained by

methods which have been mentioned already and are to be fully described in Chapter III. Often also, one of the initial two (A and B) and one or both of the final two (C and D) are possessed of measurable kinetic energy amounting to millions of electron-volts. Let us write an equation of conservation of mass:

$$m(\mathrm{A}) + m(\mathrm{B}) = m(\mathrm{C}) + m(\mathrm{D}). \qquad (2.19)$$

The accuracy of measurement is in many cases sufficient to enable us to say, that *this equation* is not verified if for the masses m we put the rest-masses of all the nuclei, but *is verified if we put mass-values which are larger than the rest-masses in the proportions indicated by relativity in view of the speeds of these nuclei.*

If we multiply all the terms in (2.19) by c^2, and refer back to (2.16), we see that (2.19) has become an equation of conservation of $(E - V)$. This is tantamount to recognizing (2.19) as an equation of conservation of mass and energy both, provided that we may consider potential energy V as constant during the process. Now insofar as V is due to external electric and magnetic fields, it *is* constant during the process. Putting it equal to zero in this case and in any other case in which it does not change, we have,

$$m = E/c^2, \qquad (2.20)$$

the often-quoted mass-energy-relation of Einstein. In this chapter we have been concerned with rest-mass m_0 and with kinetic energy T. To the former corresponds the rest-energy m_0c^2, to the latter corresponds the extra mass T/c^2. That there is no essential difference between rest-energy and kinetic energy, rest-mass and extra mass, is shown by the processes described by (2.19); for in these processes, what was rest-mass before the event exists in part as extra mass after the event, or reversely. This principle becomes even more striking when it is confirmed by the interconversion of electrons and light, which we shall study in Chapter VI.

The positive electron [1]

The first of positive electrons to be recognized as such streaked its way across an expansion-chamber in Pasadena on the second day of August, 1932. Figure 2.1 is a copy of the photograph which revealed to C. D. Anderson this the most famous of all individual particles. To interpret it I must refer to various passages elsewhere in this book: Chapter III, for the description of the expansion-chamber and of " tracks "

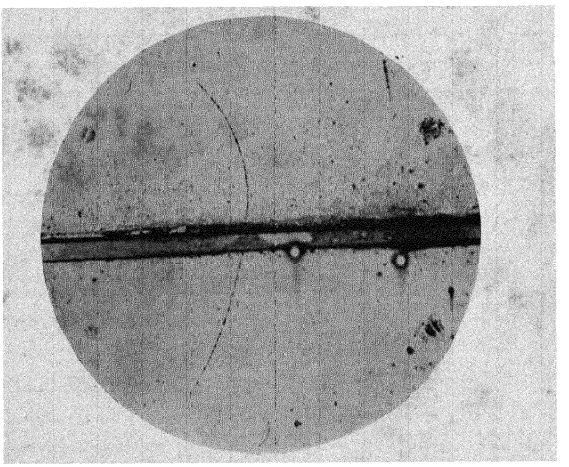


FIGURE 2.1—TRACK OF A POSITIVE ELECTRON POSSESSING INITIALLY 63 MEV OF KINETIC ENERGY, AND TRAVERSING A LEAD PLATE 6 MM THICK. (C. D. Anderson.)

such as this one—Chapter VIII, for the way in which this track was identified from its aspect as that of an electron—pages 39–42 above, for the way in which the sign and the speed of this electron are derived from the curvature imprinted upon its path by an applied magnetic field perpendicular to the plane of

[1] Anderson, *Science* **76**, 238 (1932); **PR 41**, 491 (1933); Blackett and Occhialini, **PRS 139**, 699 (1933); Chadwick, **PRS 142**, 23 (1933) and (with Blackett and Occhialini) **144**, 235 (1934); Curie and Joliot, **CR 196**, 1105, 1581, 1885 (1933); Meitner and Philipp, **NW 21**, 286 (1934); Thibaud, **PR 45**, 781 (1934).

the picture. The fact that the path is concave to the left told Anderson (when combined with his knowledge of the sense of the magnetic field) that it was made either by a negative particle going downward or by a positive particle going upward. Had the metal plate not been there, the ambiguity would have subsisted, and anyone can guess what Anderson would have judged. However, the plate *was* there, and the curvatures of the two parts of the path above it and below gave testimony that in traversing these 6 mm of lead the particle had either gained or lost a great deal of energy (some 40 *Mev*), according as it was going downward or upward. That an electron should lose so much energy in traversing so much lead was entirely plausible; but that it should *gain* so much proved even more incredible than the possibility of its being an electron of sign opposite to that of all the myriads which thus far had been observed. Nevertheless it is no secret that Anderson hesitated long before he came to the decision which so much later work has ratified.

This particle was one of the phenomena of cosmic rays. (It was in order to study these that Anderson had arranged the experiment, inserted the partition into the chamber, and set up this last with its axis horizontal instead of vertical as had always previously been the rule). So also were the next-to-be-observed of the positive electrons: Blackett and Occhialini used the expansion-chamber for photographing "showers," phenomena in which great numbers of electrons spring from a single point usually located somewhere in the midst of a block of heavy matter; they, too, had a magnetic field, and in single showers they observed that some of the electron-tracks were curved oppositely in sense to others. Instead of reproducing one of their photographs, I choose for the fewness of the tracks a picture from the Pasadena school (Figure 2.2): this exhibits three electrons of one sign and two of the other, the negatives being in the majority.

Almost at once Blackett and Occhialini found one of the two ways of producing positive electrons which are controllable by man: irradiation of dense matter by light of extremely high

frequency.[1] We shall consider in Chapter VI the process, which is nothing less sensational than the conversion of corpuscle; of light into pairs of electrons of opposite sign. Early in 1934 Joliot and Curie discovered the other way, or rather the first of many ways belonging to the other type: they created, by the art of transmutation, unstable or radioactive nuclei which after a certain length of time disintegrated by emitting positive electrons.[2] More than fifty of these are known by

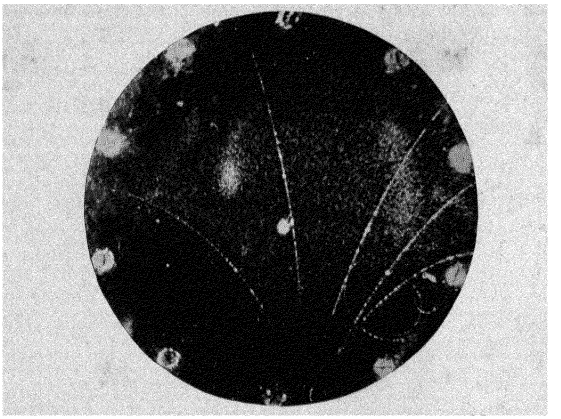


FIGURE 2.2—SMALL COSMIC-RAY SHOWER COMPRISING THREE NEGATIVE AND TWO POSITIVE ELECTRONS. (Anderson *et al.*)

now, and many of them can be produced by more than one technique: we shall make the acquaintance of some in Chapter XVIII.

So complete is the confidence of physicists that the positive electron must be identical with the negative one in mass, that

[1] At first they, as well as Joliot, Curie, Meitner and Philipp, produced the positive electrons by irradiating matter with neutrons and photons (from the reaction $Be^9 + He^4 \rightarrow C^{12} + n^1$, Chapter XVII) simultaneously; they later showed the photons to be the agents. Joliot and Curie had earlier observed the positive electrons so produced, but rather than accept such an interpretation of the curvature of the tracks, they had assumed them to be those of negative electrons going toward the source of the irradiating rays.

[2] They observed the emission of positive electrons before detecting that an appreciable length of time supervenes between it and the impact of the particles (α-particles at first) which provoke the transmutation.

no one as yet (to my knowledge) has troubled to determine its e/m ratio with anything approaching the precision achieved with negative electrons.[1] Employing an ingenious system of deflecting fields especially suited to very weak streams of charged particles, Thibaud in 1933 arrived at the conclusion that the masses of the positive and the negative electron cannot differ by as much as fifteen per cent.

We shall see in Chapter VI that the positive electron is evanescent. Everything indicates that the particle of Figure 2.1 converted itself into light by uniting itself with a negative electron at or very near to the end of its visible trail, having probably come into being out of light at a point not far from where the trail becomes visible.

[1] The aspect of the trails proves that the particles are much nearer in mass to the electron than they are to any of the positively charged particles (atom-nuclei) known heretofore, but does not fix the mass very closely.

CHAPTER III

Ions and Isotopes

The previous chapter was devoted to that lightest of all charged particles, the electron, and largely to the experiments whereby the laws of its motion in fields are proved and the value of its charge-to-mass ratio is determined. This chapter has the same purpose in respect of the other charged particles which fall within the purview of this book. It will transpire that these are atoms of chemical elements or molecules of chemical compounds, lacking one or more from their normal quotas of electrons. The best generic names for them are "ionized atoms and molecules" and "positive ions." In measuring their charge-to-mass ratios we are, in effect, ascertaining their masses; now, the mass of any one of them is thousands of times as great as that of the electron or electrons which if added to it would make it into a completed atom or molecule; therefore we are ascertaining the masses of molecules and atoms, with a minute error (and even this error is not really one, for we can correct it whenever we need to). The student, recalling that a method of calculating atom-masses was given in Chapter I, may think that all this is merely an excessive confirmation; but it turns out to be infinitely more, as the story of the discovery of isotopes will show.

In saying that all ions have charge-to-mass ratios I am, of course, implying that they all conform to the laws of motion described in Chapter II as valid for electrons. Only their masses, and to a much lesser extent their charges, distinguish them from the electrons (in the range of the phenomena which are the concern of this chapter). This, however, is not a trivial distinction. In one respect the high masses of the ions are a simplifying factor: positive ions have very much lower

speeds than do electrons of identical kinetic energy, and consequently do not transgress the limits beyond which the relativistic equations must be used, unless their energy-values are greater by far than are encountered in nearly all experiments. The difference in mass is manifest also in various ways, depending on the mode of observation—this affords a good reason for dwelling, in this chapter, on diverse ways of observing electrons as well as ions, and making the acquaintance of some of the most interesting apparatus of modern physics.

First we consider the type of ion which transcends all the rest in historic importance, for it provided in quick succession the first visible sign of the individual atom, the first evidence for the structure of the atom, and the first attainment of transmutation: the particle well named by accident, for it is in truth the alpha (but not the omega) of modern physics.

Alpha-particles

Almost as soon as radioactivity was discovered, the rays emanating from radioactive bodies were classified into gamma-rays (later to be identified as light of very high frequency), beta-rays (soon to be identified as electrons), and alpha-rays. All three are capable of causing a screen of fluorescent stuff, zinc sulphide, for instance, to shine; but the effect is most brilliant by far, when produced by the last-named; and when Crookes examined with a simple microscope such a screen under bombardment by alpha-rays, he saw that the shining consisted of tiny bright stars—the " scintillations "—blazing out and vanishing instantaneously all over the screen. When Crookes beheld this (in 1902) he became the first man to perceive an individual act of an individual atomic, nay even subatomic particle; for these flashes were signs of the impacts of individual bodies upon the screen, which bodies were in fact the alpha-rays or " alpha-particles," and were themselves charged atoms of helium.

The evidence that alpha-particles are ionized atoms of helium was amassed in a very few years, chiefly by Rutherford.

For instance, he (with Royds) showed by experiment that when they were allowed to penetrate through the walls of a sealed tube filled with a gas very carefully freed of helium, there was a sure indication that helium was gradually appearing in that gas; for the tube was equipped with electrodes for producing a luminous discharge, and at first the spectrum comprised only the lines of the original gas, but during the bombardment the lines of helium appeared. Also he observed that when a piece of lead was bombarded with alpha-particles and subsequently melted, it gave out helium, though other pieces of the same lead not bombarded gave out none. Here and there in the next few pages, and elsewhere in the book, I will mention other items of evidence which altogether make this one of the best-proved theorems of physics. Though unneeded as part of the proof, I mention here for its own sake the fact that minerals containing radioactive substances also contain helium due to disintegration of those radioactive atoms which have already perished.

Charge and mass of the alpha-particle

We have already seen (page 30) how the *charge* of the alpha-particle was evaluated as $+2e$. The *charge-to-mass ratio* was evaluated by Rutherford and Robinson, from deflections in electric and magnetic fields, as $1.446 \cdot 10^{14}$; rounding it off, I write

$$2e/m_\alpha = 1.45 \cdot 10^{14}. \qquad (3.1)$$

Multiply numerator and denominator on the left by N_0: the new numerator is twice the Faraday constant F; if the alpha-particle is truly a helium atom lacking two electrons, then except for the masses of $2N_0$ electrons (a trivial correction, as I said) the new denominator is the atomic weight 4.00 of helium. Now, the ratio of $2F$ to the atomic weight of helium is $.447 \cdot 10^{14}$. The agreement is amply good enough to clinch the identification of the alpha-particle. To get the mass of the particle we must divide this ratio into $+2e$, thus involving ourselves again in the question of the e-value. Choosing 4.80

more for its roundness than for any other reason, I set down for the characteristics of the alpha-particle: charge $2e = 9.6 \cdot 10^{-10}$; mass $m_\alpha = 6.6 \cdot 10^{-24}$; charge-to-mass ratio $1.45 \cdot 10^{14}$.

The alpha-particle a bare atom-nucleus

Though I am outrunning the development of my subject by saying so already, it is important for the student to know that the alpha-particle differs in one fundamental way from almost (not quite) all of the other ions to be mentioned in this chapter. Though all of them are atoms lacking one or more of the electrons of their normal quotas, nearly all of them are also atoms retaining some of the electrons of their normal quotas, and roughly as large as they would be if the quotas were complete. Alpha-particles however and a few other kinds of positive ions (above all, those of atomic hydrogen), are atoms lacking all of their electrons, and reduced to their "nuclei," relatively so small that we may still consider them as particles even when we shall be thinking of atoms as systems. Later on there will be much evidence for this. Meantime I mention only that particles which cleave through glass walls thick enough to hold together and to keep an ordinary gas enclosed, without harming those walls (as in the experiment of Rutherford and Royds) —and which traverse without deflection eight centimeters of ordinary air, a thickness in which they must meet with hundreds of thousands of atoms—such particles must be much smaller than the atom.

Expansion-method of C.T.R. Wilson

The wonderful method of C.T.R. Wilson for making visible the paths of alpha-particles, and those of fast electrons and of fast-moving ions generally, is based upon two fundamental facts. First, all of these particles *ionize* a gas which they traverse; by which is meant, that they detach electrons from its molecules or atoms as they pass. Sometimes these electrons are hurled far off, sometimes they are so little removed that they recombine practically instantly with the atoms whence

they were torn; but often they remain for an appreciable time, unrecombined and close to the path of the ionizing particle, and so also do the positively-charged atoms from which they were torn. The path is then marked by these " negative ions " (the freed electrons, by the way, often adhere to neutral atoms or molecules) and " positive ions "—or rather, it is not marked as yet, for the ions are invisible. The path can, however, be made visible; and here enters the second fact on which Wilson's invention is based.

Suppose that the gas is mixed with water-vapor, and that just before or just after the passage of the ionizing particle it is suddenly cooled to such an extent that the vapor is super-

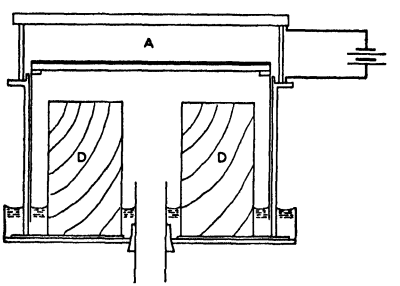


FIGURE 3.1—EXPANSION-CHAMBER. (After C. T. R. Wilson.)

saturated. The cooling is achieved by letting the gas expand, this being the only way which is rapid enough; and now let us pause to consider Wilson's apparatus, the " expansion-chamber " or " cloud-chamber," sketched in Figure 3.1. It is a short fat cylinder, usually a few inches wide (but rectangular chambers have lately been made a couple of feet across.) It is filled with moist clean air; is closed on top by a slab of glass, through which the observer may see or photograph what happens inside; and is closed on the bottom by a piston-head which may be caused to fall suddenly and sharply through a preassigned distance, being halted by some buffers (*D*) inserted at an adjustable level. (The sudden fall of the pistonhead is traditionally produced by means of a valve which opens a

connection, through the pipe indicated at the bottom of the sketch, between the space under the piston-head and a large evacuated jar not shown; when the valve is opened the air rushes suddenly out. Other ways have lately been introduced, but on the whole the Wilson chamber remains a fine example of an invention of which the earliest form was almost the best.)

When the pistonhead drops, the moist air above *B* expands and cools; and now comes the second point. The supersaturated water-vapor has a tendency to condense, but it may not yield to this tendency unless it can find what I will call a "foothold." When fogs occur in the atmosphere, the foothold is supplied by solid bodies, granules of dust or smoke, on each of which a droplet of water is formed. The air in the Wilson chamber is carefully freed of dust in advance, and this foothold is lacking. If the percentage of expansion and therefore the cooling are very considerable, the vapor will make the best of a bad situation, and condense into droplets anyhow. If the percentage of expansion and the cooling are moderate, it will prefer to remain in the form of vapor. *There is*, however, *a particular range of expansion-percentages*—it runs around 35% in ordinary cases—*for which the water-vapor will accept a charged atom as a foothold, though in the absence of charged atoms it will remain as vapor.*[1] Therefore if a fast-flying ionizing-particle traverses the chamber just before or during or just after the expansion, its path will be made visible as a trail of droplets, or—if these are so close together that the eye does not resolve them—as a continuous pencil or line of mist.

Such trails were first observed by Wilson in 1911. He had invented the method years before, but had only seen the chamber filled with an evenly-spread mist, because he was observing the total effect of too long an exposure to ionizing particles,—many trails had spread out and merged. Later he used a strong electric field to clear out the "old" ions in

[1] There is even a portion of this range for which the vapor will condense upon negative but not on positive ions, a most interesting fact (H. A. Wilson availed himself of it in making the water-droplets which he used in one of the experiments which led up to the oil-drop technique. Cf. page 26, *supra*).



FIGURE 3.2—DROPLETS FORMED IN THE EXPANSION-CHAMBER ON INDIVIDUAL IONS LEFT BEHIND BY A TRAVERSING FAST ELECTRON. (C. T. R. Wilson; *Proceedings of the Royal Society*.)

advance of an exposure, and observed only those produced in a very brief period of time close to the moment of expansion. This idea showed him the trails, and made him the first man to see individual acts of individual subatomic particles, exception being made for those who had seen the scintillations of alpha-particles which Crookes had discovered; it made him also the inventor of one of the most profitable methods of research, and also the illustrator-in-chief of modern physics, as this book like so many others will exemplify.

Figure 3.2 is a Wilson-chamber photograph of the trail of a fast electron, showing the individual droplets remarkably well; many of them appear in obvious pairs, one member of each pair having been formed on a positive ion, the other upon the negative ion consisting of the electron torn away from that positive ion and (probably) of a molecule to which the electron has adhered. (The two members of each pair have been drawn away from one another by an electric field which acted during the appreciable time-interval between the passage of the ionizing particle and the expansion which produced the droplets). Other figures will be described in other connections, as

FIGURE 3.3—EXPANSION-CHAMBER PERVADED BY X-RAYS AND SHOWING TRAILS OF ELECTRONS RELEASED BY THESE; THE COMPTON PROCESS (Chapter VI) IS RESPONSIBLE FOR THE SHORT COMMA-LIKE TRACKS AND THE PHOTO-ELECTRIC PROCESS (Chapter VIII) FOR THE LONG SINUOUS ONES. (C. T. R. Wilson.)

we reach them; but I call the student's attention in advance to Figure 3.3 (p. 69) showing the sinuous tracks of moderately fast electrons with their " thin " appearance due to relatively wide spacing of the droplets—to Figures 2.1 and 2.2 (pp. 58, 60) showing the tracks of electrons moving so rapidly that their paths are no longer sinuous though the trails remain thin—to Figures 3.5 and 15.3 (pp. 72, 479) showing the straight tracks of alpha-particles with their " thick " or continuous appearance due to very close spacing of the droplets (tens of thousands per cm!)—and once more to Figure 2.2, exhibiting the splendid circular arcs described by charged particles moving in magnetic fields as explained in Chapter II.

Collisions between alpha-particles and atoms [1]

The beautiful straightness of alpha-particle trails is one of their distinctive features: it is due to the high momentum which, owing to their large mass (large, that is to say, compared with that of the electron!) is correlated with their high kinetic energy. Yet very occasionally, a trail is noticed which has a sharp corner; and sometimes a second trail originates at the corner, diverging from the first one. Sometimes this means that transmutation has occurred; but for the present we are interested in the cases where the particle has simply made an elastic impact with another object, having a mass of the same order of magnitude. The equations of elastic impact supply us with a great deal of information about the struck object, or let me say, the struck particle.

Let m stand for the mass of the striking, M for that of the struck particle; u for the initial and v for the final speed of the striking, and V for the final speed of the struck particle (we assume that the latter was practically stationary before the impact); θ, ϕ for the angles made by the final velocities of striking and of struck particle, respectively, with the initial

[1] P. M. S. Blackett, **PRS 104**, 349 (1925); Blackett and D. S. Lees, *ibid.* 136, 338 (1932); P. Auger and F. Perrin, **CR 175**, 340 (1922); cf. also M. Akiyama, *Jap. Jour. of Physics*, 2, 279 (1923).

velocity of the striking particle. The equations (which I write in the non-relativistic approximation) are that of conservation of kinetic energy of translatory motion:

$$\tfrac{1}{2}mu^2 = \tfrac{1}{2}mv^2 + \tfrac{1}{2}MV^2, \qquad (3.2)$$

and those of conservation of momentum,

$$mu = mv\cos\theta + MV\cos\phi, \qquad (3.3)$$

$$0 = mv\sin\theta + MV\sin\phi,$$

and these are three equations involving the five quantities v/u, V/u, M/m, θ and ϕ, from which we can determine any one of the five in terms of any two others.

Photographs of alpha-particle collisions, occurring in hydrogen and in helium (substituted for air in the Wilson



FIGURE 3.4—TRAILS OF AN ALPHA-PARTICLE AND A PARTICLE OF ONE-QUARTER ITS MASS (PRESUMABLY A HYDROGEN NUCLEUS) WITH WHICH IT COLLIDED. (Blackett; *Proceedings of the Royal Society.*)

chamber) are shown as Figures 3.4 and 3.5. Evidently θ and ϕ can be ascertained from such a photograph, if the observer can allow for the foreshortening of the tracks: this is done by making two simultaneous photographs of the process from different directions, and double photographs such as these two figures are now very common. Evaluating M/m from these pictures, Blackett got 0.253 from the former and 0.981 from the latter; these values agree with the ratios (0.252 and 1.000) of the masses of hydrogen atom and helium atom to mass of

alpha-particle, as we have already determined them! In an extensive study of impacts in hydrogen, Blackett and Lees made calculations from 49 collisions, 16 of which were outstandingly clear, while 33 others were ranked somewhat lower (in the main they involved slower alpha-particles) as material for computation: from the first sixteen they derived $M/m = 0.2531$ with a statistical uncertainty of .0011, from the rest 0.2529 with a statistical uncertainty of .0020.

Notable as these experiments are, this is hardly likely ever to compete with the other methods of determining atomic

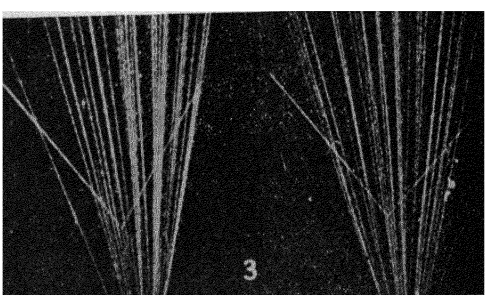


FIGURE 3.5—TRAILS OF AN ALPHA-PARTICLE AND A PARTICLE OF THE SAME MASS (PRESUMABLY A HELIUM NUCLEUS) WITH WHICH IT COLLIDED. (Blackett; *Proceedings of the Royal Society.*)

masses! (Literally tens of thousands of alpha-particle trails must be photographed before one with a sharp corner and long diverging trail is found; and the trail of the struck atom would be too short to measure if M/m were greater than 10 or thereabouts). Nevertheless it is worth while to know and remember how by a chain of evidence consisting only of three links—measurement of the charge of alpha-particles, measurement of the charge-to-mass ratio of these particles, and measurement of the collisions which they make in gases—it has been proved that gases contain particles having masses of the order of 10^{-24} gramme, these masses being proportional to the chemical atomic weight of the gas.

The struck particles, or more precisely the particles which

recoil from the impacts, are not complete atoms, but nuclei which have lost certainly some and almost certainly all of their attendant electrons. The impinging alpha-particle knocks a nucleus right out of its family of electrons.

Positive ions

Thus far we have been studying the peculiar type of positive ion which is emitted by various radioactive elements; it is obtainable also from processes of transmutation, and so are other ions such as the proton and the deuteron (not to speak of positive and negative electrons), but these will be more fittingly placed in other chapters. Now we take up the positive ions which are produced by the more ordinary, or the less extraordinary, processes of the laboratory.

Ionized atoms, like free electrons, may be obtained in a variety of ways. They emerge from incandescent metals, as electrons do; only, in such a case they are usually atoms not of the metal itself, but of sodium or potassium or some similar element of which a small amount is mixed with the metal, and the emission puts an end to itself by exhausting these impurities. Ions of the alkali metals and the alkaline-earth metals, and some others, may be obtained by heating properly-chosen salts or mixtures of salts of these elements, or making these salts or mixtures into targets which are then bombarded by electrons or used as the anodes in discharge-tubes. If a metal is immersed in a gas, and certain conditions are fulfilled,[1] every atom of the gas which impinges on the metal will cede an electron to it and fly off as a positive ion. (It is possible to construct a gas-filled two-electrode tube in which a steady current flows because this process is continually occurring at the anode; if the voltage is not too high the current is limited by space-charge, and conforms to equation (2.14) or to whatever similar equation is required by the shape of the electrodes;

[1] The metal must be kept hot so that it will not cover itself with a film of atoms of the gas, and metal and gas must be so chosen that the energy required to remove an electron from the former must be greater than that required to remove an electron from an atom of the latter (Chapter VIII).

and the charge-to-mass ratio of the ions can be obtained accurately enough to identify them). Ionized atoms or molecules may be produced, under very controllable and analyzable conditions, by projecting electrons or light against a gas; we shall be examining this process at length in Chapter VIII. The oldest, the classical, and to this day the most prolific method of procuring positive ions, is to take them from a gas which is the theater of an electric discharge.

In many discharges the voltage is low, and it is necessary to apply special fields to extract the ions and to accelerate them after they have been extracted. In some, however—notably

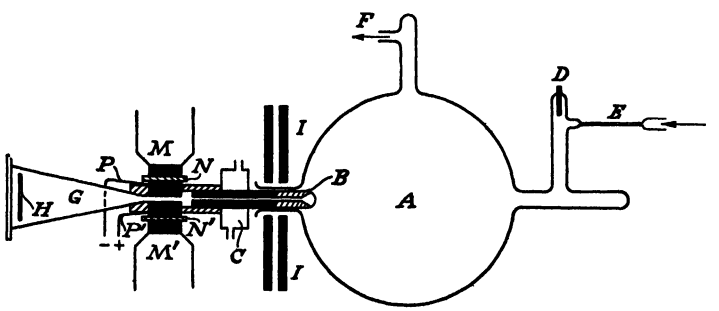


FIGURE 3.6—J. J. THOMSON'S APPARATUS FOR POSITIVE-ION ANALYSIS.

the self-sustaining glow-discharge in a gas at a pressure of the order of 0.1 or 0.001 millimeters of mercury—it may be raised to thousands or tens of thousands of volts; and in such a case the electric field in the discharge-tube spontaneously disposes itself in such a way that some of the ions in the discharge are sharply accelerated with great force in the region in front of the cathode, and fall upon the cathode with great energy, moving in directions perpendicularly to its surface. If a narrow hole or channel is drilled right through the cathode, the ions which fall toward it will shoot clear through it and emerge at the far end as a beam or pencil of so-called " canal rays " (*canal* being here a childish mistranslation of the German word for *channel*).

Thomson's form of apparatus for analyzing such a beam is sketched in Figure 3.6. The discharge occurs in A: the cathode

B is perforated by the long thin channel, from which the beam emerges to pass between the metal blocks at N, N' which serve both as magnet pole-pieces and as condenser plates; having been deflected by the fields between these blocks it travels onward through the field-free region G till it impinges on the plate (a photographic plate) normal to the plane of the paper at H. (The region to the left of the channel may be kept at a pressure very much lower than that in the discharge-tube A, by a powerful pump connected directly to it).

Suppose the x-axis laid along the axis of the channel, the y-axis vertical, the z-axis at right angles to the plane of the paper, the origin at the point where the beam enters the interspace between the blocks N, N'. A beam of ions approaches with speed U along the x-axis; at the origin it enters a region extending from $x = 0$ to $x = k$, in which a homogeneous electric field E and a homogeneous magnetic field H are directed parallel to the y-axis; it is deflected by these fields, and cuts the plane $x = k$ at a point of which the coordinates (compare equations 2.2 and 2.8) are:

$$x = k, \quad y = \tfrac{1}{2}\frac{ne}{MU^2}Ek^2, \quad z = \tfrac{1}{2}\frac{ne}{MUc}Hk^2, \qquad (3.4)$$

and continues thence along a direction of which the inclinations to the coordinate-axes are given by the derivatives:

$$\left(\frac{dy}{dx}\right)_{x=k} = \frac{neEk}{MU^2}, \quad \left(\frac{dz}{dx}\right)_{x=k} = \frac{neHk}{MUc}, \qquad (3.5)$$

so that if a photographic plate is placed in the plane $x = l$, the beam strikes it and makes an imprint at the point of which the coordinates are:

$$x = l, \quad y = \frac{ne}{MU^2}Ek\left(l - \frac{k}{2}\right), \quad z = \frac{ne}{MUc}Hk\left(l - \frac{k}{2}\right). \qquad (3.6)$$

It may be pointed out that this result is the same as though the entire deflection occurred at a single point midway in the region of the field, at $\left(x = \frac{k}{2}, y = 0, z = 0\right)$.

Thus wherever upon the photographic plate an imprint is found, the values of ne/M and of U for the ions making that imprint can be computed. In practice, it usually happens that the ions in a beam coming out of a discharge-tube are moving not all with the same speed, but with speeds distributed from some low value up to a definite maximum $U_{\text{max.}}$; in this case the imprint of the ions upon the plate is a parabolic arc conforming to the equation

$$z = \sqrt{\frac{ne}{M}\frac{H^2k(l-\frac{1}{2}k)}{Ec^2}}\,y, \tag{3.7}$$

and extending from large values of z inward to the point defined by

$$\begin{aligned} y_{\text{min.}} &= \frac{ne}{MU^2_{\text{max.}}}Ek\left(l-\frac{k}{2}\right), \\ z_{\text{min.}} &= \frac{ne}{cMU_{\text{max.}}}Hk\left(l-\frac{k}{2}\right), \end{aligned} \tag{3.8}$$

where it suddenly breaks off. In practice, it also usually happens that several kinds of ions are mixed together in a beam coming out of a discharge-tube, and the ions of each kind have speeds distributed over a certain range ending at a maximum speed $U_{\text{max.}}$ which is that acquired by an ion which falls unimpeded from rest through the entire potential-drop between the terminals of the discharge-tube. This being the case, the value of $U_{\text{max.}}$ and the value of $\frac{ne}{M}$ for each kind of ion are related to the potential-drop V by the equation

$$\tfrac{1}{2}MU^2 = neV, \tag{3.9}$$

and the student can easily see that each kind of ion will produce its own parabolic arc and that all of the arcs will terminate along one and the same line parallel to the z-axis.

In actual apparatus the electric and magnetic fields are, of course, not perfectly constant over a sharply bounded range and exactly zero elsewhere; but it remains true that ions with a common value of charge-to-mass ratio and different values of

U delineate a parabolic arc upon the plate and parabolic arcs due to different kinds of ions lie apart but end upon a line perpendicular to the direction of the fields. These features are beautifully clear in Figures 3.7 and 3.8, which are reproductions of plates made in this very way; they are so oriented that the direction parallel to the fields, the y-direction of the present convention, is horizontal. The only change from the simplified case is, that the coefficients of ne/MU^2 and of ne/MU in the equations (3.4) for y and z are replaced by more complicated functions of the fieldstrengths and the dimensions of the apparatus.[1]

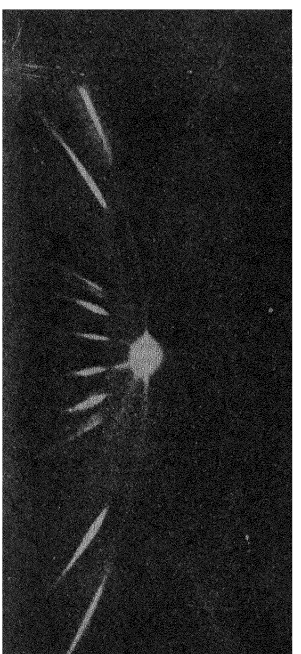


FIGURE 3.7—PARABOLIC ARCS PRODUCED BY POSITIVE IONS OF VARIOUS TYPES IN THOMSON'S APPARATUS. (J. J. Thomson, *Rays of Positive Electricity*; Longmans Green and Co.)

Thomson has lately recalled [2] that when he first attempted the experiment, the only marks on the plates were made by particles having the charge-to-mass ratio of hydrogen ions, even though he put several different gases into the tube. Apparently this was because the gas between channel and plate was still too dense, and stopped all of the ions excepting those formed from the hydrogen ineradicably present in the main tube (hydrogen ions, like alpha-particles, are bare nuclei and therefore very penetrative); for when Thomson improved what would now be called his vacuum technique, the expected ions appeared, and, in fact, a good many more. (One gathers

[1] J. J. Thomson, *Rays of Positive Electricity*, pp. 16–21.

[2] *Recollections and Reflections*, p. 353.

that before this success, he had begun to infer that positive charge might be carried by hydrogen atoms only.)

Suppose, to begin with, that the tube is filled with a monatomic gas, mercury or helium for instance. The experimenter can predict just where, upon the plate, should lie the parabola made by ions of charge e; for let M stand for the (average) mass of an atom, W for the atomic weight of the element in question—then, as I have indicated at several places already, the charge-to-mass ratio e/M should be equal to F/W, and

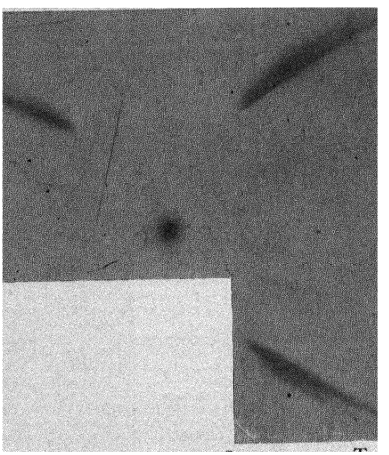


FIGURE 3.8—PARABOLAE DUE TO LITHIUM, SHOWING THE TWO ISOTOPES OF THIS ELEMENT. (From the original plate by G. P. Thomson, also published in *Philosophical Magazine*.)

this locates the parabola to be expected. Actually there will be a parabola at or near the predicted location, but also there will usually be many more. Some are accounted for by assuming values 2, 3, 4 . . . for n in the general expression ne/M for charge-to-mass ratio—assuming, in words, that they are produced by " multiply-charged ions," atoms lacking two or more electrons. Some are accounted for by doubling M—assuming, in words, that they are produced by diatomic molecules lacking one or more electrons. (This seems to be in discord with what I said about the gas being monatomic, but it

appears that ions are formed in the discharge which correspond to molecules too rare or too unstable, or both, to be detected by the ordinary methods of chemical analysis). If the tube is filled with a diatomic gas, hydrogen for instance, parabolae are seen which are explained by assuming ionized atoms, ionized diatomic molecules and even ionized triatomic molecules. (It was thus that " triatomic hydrogen " H_3 was discovered; Thomson having observed ions with charge-to-mass ratio one-third as great as the value for the hydrogen atom, there commenced a search for a gas composed of molecules H_3, and this was soon successful [1]). If the tube is filled with a gas which is a chemical compound, there are not only ionized molecules having the formula of the compound, but ions corresponding—according to the testimony of their values of ne/M—to many other groupings of the atoms of the elements involved in the compound; thus R. Conrad, putting the gas CH_4 (methane) into his tube, got parabolae corresponding to more than a dozen different groupings of C and H atoms (including even C_4H_4). If the tube contains a mixture of gases, ions emerge which are built out of contributions from two or more of them; and mementoes of prior fillings are likely to persist a long time.

We see thus that a good photograph obtained by the parabola method (as Thomson calls it) is, in the hackneyed phrase, a mine of information. Even the complexities, such as those due to the liability of the ions to changes of charge as they travel, may be turned to good account. Many analyses have been made of the ions issuing from discharge-tubes; but I pass them over, for the phenomena of a self-sustaining discharge belong to a specialty far too extensive and deep to be treated in this book, and we shall be more interested (in Chapter VIII) in the production of ions by controllable electron-impacts. There remains, however, to be cited one of the greatest discoveries of modern times, a gift of the parabola-method.

[1] J. J. Thomson, **PRS 89**, 1 (1913); A. J. Dempster, **PR 8**, 651 (1916); G. L. Wendt and A. S. Landauer, *J. Am. Chem. Soc.* **42**, 930 (1920).

Isotopes

In 1912 Thomson applied the parabola-method to various samples of gases derived from residues of liquid air. Neon, one of the elements found in these residues, has a value 20.2 for its atomic weight, corresponding to a presumptive mass $(20.2)/N_0$ for the atom of neon, and a presumptive charge-to-mass ratio $eN_0/(20.2)$ for a neon atom lacking a single electron. A parabola was found agreeing roughly with this value; but it was always companioned by another much fainter parabola, corresponding to a value roughly 10% lower, therefore to a mass some 10% higher; and there were two others with values twice as great as these, suggesting doubly-ionized particles of the same masses. It was suspected at first that only the less massive ions were ionized atoms of neon Ne^+, the others being ionized molecules such as NeH_2^+ or CO_2^{++}. (Here I have introduced the usage of writing as many plus signs above and to the right of the symbol for the atom or molecule, as there are supposed to be electrons lacking from this last). One of these suspicions was dismissed because the proportion of CO_2 in the gas filling the tube could be varied without affecting the relative intensities of the parabolae; the other was finally discarded because no evidence for combinations between hydrogen and neon could be found.[1] Eventually it was proved, by refining the technique of measurement, that *neither mass agreed with* $(20.2)/N_0$. One was $(20.0)/N_0$, the other was $(22.0)/N_0$; and both belonged to neon.

Neon therefore is an element with more than one kind or "species" of atom (at least three kinds are now assured!) the different kinds differing in mass. This is true, as I shall presently relate, of almost every element; an element whereof all the atoms are alike is very much of an exception.

But does not this sap the base of the inference, from the

[1] Had as much been known then as now about the chemical potentialities of hydrogen and of the rare gases under the conditions of the discharge, this idea might not have been so soon and happily rejected. It was contended also that molecules do not have double charges ($n = 2$), but this has turned out to be a rule with so many exceptions that it cannot be invoked as a crucial argument.

laws of chemical combination (page 16), that matter consists of atoms? Not at all! For, to return to my previous illustration of sodium chloride: there is no essential contradiction between the statement that the chemical atomic weight of chlorine is 35.46, and the statement (presently to be confirmed) that chlorine consists partly of atoms having mass 35.00 and partly of atoms having mass 37.00. The first of these is a short way of embracing many facts about chemical compounds, one of which is, that any sample of sodium chloride NaCl large enough to be analyzed by the standard processes of chemistry contains sodium and chlorine in the proportions 23.00 : 35.46 by weight. This is compatible with the idea that sodium and chlorine each consist of atoms all exactly alike, each sodium atom weighing 23.00/35.46 times as much as each atom of chlorine. But it is also compatible with the idea that chlorine consists of two kinds of atoms and sodium of one, each Na atom weighing 23.00/35.00 as much as a Cl atom of the lighter kind and 23.00/37.00 as much as a Cl atom of the heavier kind; provided that every sample of NaCl large enough to be analyzed by the standard processes of chemistry contains chlorine atoms of the two kinds perfectly mixed in the proportion 76 : 24 by number. Moreover, it is compatible with any, even the wildest conceivable diversity of masses among the atoms of either element or of both; provided always that in every sample of NaCl large enough to be analyzed, the various kinds of atoms of each element be perfectly mixed in such proportions as to yield an *average* mass of 23.00 for those of sodium against an *average* mass of 35.46 for those of chlorine.

Among these infinitely many possible assumptions about the masses of the atoms of an element, which then should be adopted? During the nineteenth century, when all of the data of chemistry and physics were compatible with all of them, it was natural to adopt the simplest one of all, and suppose that all of the atoms of each individual element are exactly alike in mass. Indeed any other choice would have been contrary to one of the first principles of theorizing, which is, not to make the theory more complex than the available facts require. One

fact (it will soon be stated) did indeed suggest that it might be advantageous to choose some more complicated distribution of masses among the atoms; but the suggestion was not by any means insistent, for it was not heeded. The simplest choice encountered no definitely adverse facts, till certain observations were made on radioactive elements in 1905; and the observations of Thomson in 1912 on the positive ions of neon were the first which, quite discrediting that simplest choice, likewise showed how to discover that one among all the others which must be made in its place. The kinds of atoms belonging to an element are recognized by obtaining the atoms as fast-moving ions travelling in a beam, and deflecting them in fields which separate them according to their charge-to-mass ratios and therefore according to their masses. Charge-to-mass ratios, and accordingly masses, are computed from the deflections of the separated beams; relative "abundances" of the several kinds of atoms are obtained by measuring the relative strengths of the several separated beams; the chemical atomic weight is N_0 times the "weighted average" of the masses of all of these atoms, weighted according to their relative abundances.

The different kinds of atom belonging to an element are called the "isotopes" thereof, the word being formed from Greek roots to imply that they occupy the same place (in the Table of the Elements, that is to say). The name was coined by F. Soddy, in allusion to a fact of radio-activity—the fact which, following him, we express by saying that different isotopes of an element may be differently radioactive. This occurred before Thomson's discovery; I have defied the chronology, because radioactivity is best to treat by itself in another part of this book. There is a slight lack of unanimity in the use of the word even now; if an element has but one kind of atom, some people say that it has one isotope, some say that it has none; I follow the former usage.

Search for isotopes

The discovery made upon neon was not followed up for several years—because of the outbreak of the war, I suppose; but ever since the first of the post-war years, the analysis of the elements into their isotopes—the "search for isotopes," to put it shortly—has been one of the most exciting of the romances of physics. Aston of Cambridge did the most of it, having analyzed single-handed nearly all of the elements—an achievement of a kind reserved for very few in the evolution of any science. He derived his ions from discharge-tubes. Some of the elements, being gaseous or at least volatile, could be used directly as filling for the tubes; some have gaseous compounds which are just as serviceable, their ions comprising individual atoms or simple groupings of atoms; some were incorporated into the anode of the tube, the ions springing from its substance rather than from the gas. For some of the elements it took a long time to discover suitable compounds, and four of them (palladium, gold, iridium and platinum) never surrendered at all to Aston, being finally analyzed by Dempster of Chicago who derived their ions from spark-discharges between electrodes of the metals in question. These very ions, and others, have subsequently been produced for analysis by Bleakney of Princeton, by the device of bombarding the vapors of the metals with strong electron-streams; and Dempster in earlier years achieved the analysis of magnesium and other elements, by using the ions emitted from heated salts.

Every element known has by now been analyzed, and the epic period of the hunt, one may think, is now over; but a very great deal remains to be done. In the first place, any such statement as "The element E is found to have x isotopes" must always be regarded as an abbreviation for "The element E has x isotopes which have been found, and if there are any others, they must constitute less than one part in k of all the atoms of the element." It is then the business of each observer to state his value of k, and in general the business of all experimenters to improve their methods so as to increase k. In the

second place, there is still a good deal to be done in measuring relative abundances among the isotopes of a single element, so as to get the chemical atomic weight thereof. In the third place there is unlimited scope for improving the accuracy of the measurement of masses—not only (nor even largely) because these are wanted for computing the chemical atomic weight, but chiefly because the masses and their differences are fundamental in nuclear physics, and uncertainties in their values which might be considered trifling by any other standard are important when judged by the standards prevailing in that particular branch of physics.

Of the field-dispositions used in the study of isotopes, the simplest (from the point of view of the expositor) is that employed initially by Dempster and at present by several physicists, including W. Bleakney, A. O. Nier, and others of the school of J. T. Tate. In this method the ions are moving slowly when formed, are accelerated by a measured voltage, are then admitted to a well-evacuated chamber pervaded by a magnetic field transverse to their velocity, and in this field they describe complete half-circles. The principle of the scheme has already been stated (Chapter II, pages 39–41); the student can easily confirm that if V stands for the accelerating-potential (in electrostatic units), H for the magnetic fieldstrength and R for the radius of the half-circles, then (if the original kinetic energies of the ions are truly negligibly small compared with eV):

$$ne/M = 2c^2V/H^2R^2. \qquad (3.10)$$

Dempster's early realization of the scheme is sketched in Figure 3.9: his ions, emerging from the heated ribbon or filament sketched above G, were accelerated by the potential-difference prevailing between it and the diaphragm at S_1, described their half-circles in the chamber B and arrived at the plane of the slit S_2. There might have been a photographic plate in the plane S_2, in which case ions of different ne/M-values would have made their imprints simultaneously at different places on the plate; we shall meet in another chapter with cases where the like is done. In these experiments,

however, it is customary to vary either V or H, and plot against this variable the reading of the electrometer at E; whenever there is a peak on the curve, ions of a specific ne/M-value are describing the particular half-circle which brings them to the slit S_2, and its radius (known once for all) is substituted into (3.10), together with the appropriate values of H and V. When Dempster made the experiment with a salt of magnesium

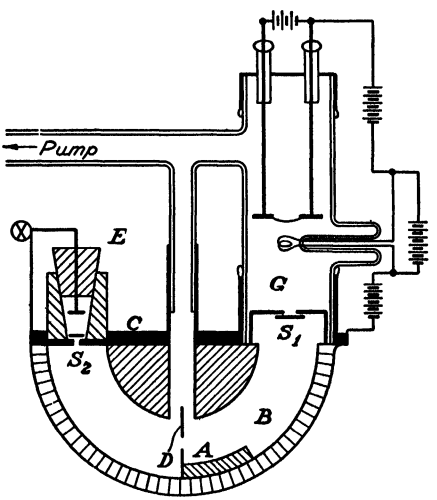


FIGURE 3.9—ANALYSIS OF IONS BY MEANS OF ELECTRIC ACCELERATION FOLLOWED BY MAGNETIC DEFLECTION THROUGH 180°: DEMPSTER'S EARLY APPARATUS. (After A. J. Dempster.)

on the incandescent ribbon, the curve exhibited three peaks corresponding to the masses $(24.0)/N_0$, $(25.0)/N_0$ and $(26.0)/N_0$, instead of the single peak corresponding to $(24.3)/N_0$ which would be required by the chemical atomic weight 24.3 of magnesium, if all magnesium atoms were alike.

In recent variants of this design, there is an elaborate system of diaphragms and fields in the chamber corresponding to G of Figure 3.9, which serves to project a nearly-uniform beam of ions upon the slit S_1; the radius R is large, having been raised to 22 cm. by Sampson and Bleakney; the pump-

ing-system maintains a much better vacuum in the deflection-chamber B than in the acceleration-chamber G. Owing to these and perhaps to other improvements, the peaks in the recently-published curves are marvellously sharp (Figure 3.10); yet even so, the use of this method is restricted to determining relative abundances of isotopes and for detecting isotopes of exceeding rarity, other types of apparatus being better suited for evaluating masses with extreme precision. Ions are some-

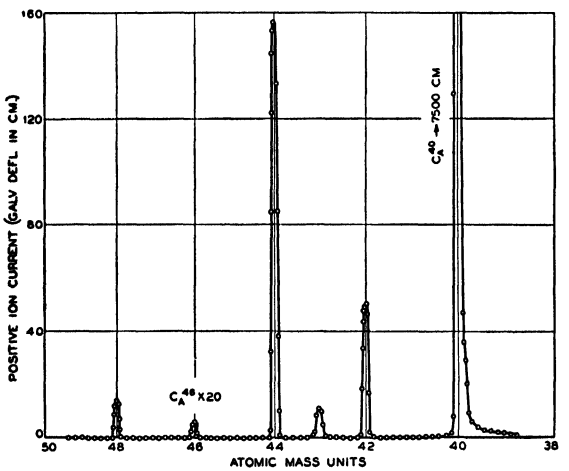


FIGURE 3.10—THE ISOTOPES OF CALCIUM, EXHIBITED BY APPARATUS OF THE SAME PRINCIPLE AS THAT OF FIGURE 3.9. (Nier; *Physical Review.*)

times obtained straight from heated salts, sometimes by bombarding with a controlled electron-stream a vapor contained in the compartment G.

As for the other types of apparatus now in use (notably Aston's with which the most analyses have been made, that of K. T. Bainbridge of Harvard which is shown in Figure 3.11, and the contemporary scheme of Dempster), in all of them the ions are deflected first in an electric and then in a magnetic field. This sounds like the parabola-method, but the theory in all these cases is far more complicated, largely because the

system is designed with regard to something not yet mentioned here. In the account of the parabola-method I wrote as though the ion-beam entering the deflecting-fields were infinitely narrow and were made up of particles all moving in exactly parallel lines. This is of course an unattainable ideal; one might approach it by adding slit after slit to the system, but in approaching it one would cut off so many of the ions that long before the ideal was attained, the ions which were

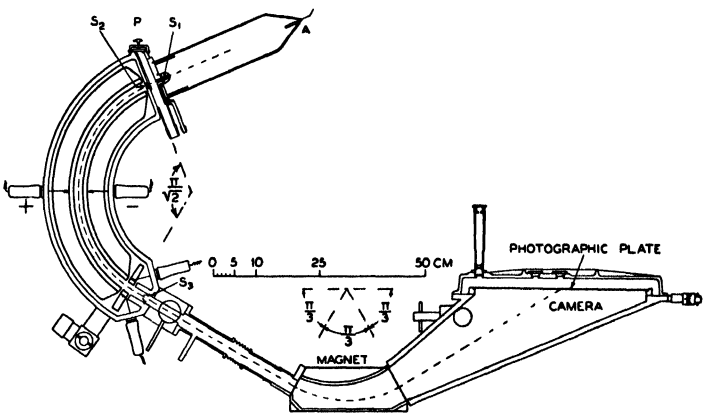


FIGURE 3.11—MASS-SPECTROGRAPH OF K. T. BAINBRIDGE: A DISCHARGE-TUBE WHICH IS THE ION-SOURCE; ELECTRIC FIELD DEFLECTING IONS THROUGH $\pi/\sqrt{2}$ RADIANS, MAGNETIC FIELD DEFLECTING THEM THROUGH 60°; MASS-SPECTRUM FORMED ON PHOTOGRAPHIC PLATE. (*Physical Review.*)

left would be too few to be detected. It is therefore highly desirable to design the apparatus in such a way that a moderately-divergent beam of ions may enter the deflecting field, and reconverge at such a rate as to form a neat focus in the plane where it leaves the field. The scheme of Figure 3.9 itself exhibits this feature, for a moderately-divergent beam entering through a small aperture at S_1 will reconverge so as to form a focus in the plane of S_2 (this is one reason why such apparatus is normally designed with a full half-circle arc in the deflecting chamber). This has long been known; more recently A. L.

Hughes and V. Rojansky have proved a similar theorem for a peculiar type of electric field. Suppose that the curving walls of the deflection-chamber in Figure 3.9 (which are concentric circles) are electrodes charged to a potential-difference such that the electric fieldstrength in the interspace varies inversely as the distance from the center of curvature; then, refocussing of a divergent beam entering at S_1 will occur, not, however, in the plane of S_2 which is at the distance of π radians measured around the arc, but in the plane which is at the distance of $\pi/\sqrt{2}$ radians. The curving electrodes shown on the left of Bainbridge's device (Figure 3.11) are thus proportioned and thus charged. It is also possible to achieve a certain degree of "velocity-focussing" whereby ions of equal charge-to-mass ratio and slightly different speeds are brought to the same point, as though a part of each parabolic arc of Figure 3.11 were shrunk into a point. Recent apparatus is designed so as to give somewhat of both types of focussing, and there is even a design theoretically capable of giving perfect focussing, now being developed by Bleakney; but it would transcend the scope of this book to go into further detail about this problem.

In all these types of apparatus in which the two deflecting fields are used, a photographic plate is set athwart the ion-streams after they emerge from the second field, and on it appear the imprints of several kinds of ions side by side—for instance, all of the isotopes of a single element, or those of several elements close together in atomic weight, or perhaps ions of charge e and a particular mass M_1 together with other ions of charge $2e$ and a mass M_2 which is nearly equal to twice M_1. The incoming beam being usually defined by long narrow slits, the imprints have themselves the aspect of lines in an optical spectrum (as in Figure 3.12), whence the names "mass-spectrum" for such a display and "mass-spectrograph" for the instrument. Often the microphotometer curves of such plates are published instead of the plates themselves.

Continual progress is being made in sharpening these lines, reducing the ratio of line-breadth to spacing of successive lines progress, therefore, both in separating ions for which the

charge-to-mass ratios are nearly the same, and in measuring the ratio itself. With the very earliest mass-spectrograph, the isotopes of the heavier elements could scarcely be distinguished, but now the so-called "resolving-power" [1] may surpass 10,000, and charge-to-mass ratios are currently being measured with a statistical uncertainty of less than one part in 10^5. This last statement, however, needs a severe qualification, which it is high time to give.



FIGURE 3.12—MASS-SPECTRA OF TELLURIUM (EIGHT KNOWN ISOTOPES, SEVEN DISCERNIBLE IN THIS PICTURE) AND IODINE (ONE KNOWN ISOTOPE). (K. T. Bainbridge).

Though my language thus far has implied that charge-to-mass ratios of positive ions have been measured "absolutely"—*i.e.* in C.G.S. units, like e/m for the negative electron—this is not so. No one, so far as I know, has even attempted to push such a measurement to the utmost accuracy attainable today (the best, I suppose, is that of Rutherford and Robinson upon the alpha-particle, cited on page 64). Probably it has been regarded as a futile experiment, on the ground that for an element with a single isotope the charge-to-mass ratio of an atom lacking a single electron must be equal to the ratio of the Faraday F to the chemical atomic weight W (page 19); yet the situation in regard to the values of the major constants (page 333) suggests that it may be worth the doing. However, the universal practice is to evaluate the ratios for the various kinds of ion relatively to the ratio for the principal isotope of oxygen, the mass which is then arbitrarily expressed as 16; or else relatively to the ratio for some other kind of ion serving as an intermediate standard, which itself is referred to that of

[1] Defined as $R/\Delta R$, where R and $R + \Delta R$ stand for the charge-to-mass ratios of two ions of which the imprints can be distinguished clearly—but there is enough vagueness in the words "can be distinguished clearly" to make it dangerous to quote numerical values without a thorough analysis of the experimenter's criteria.

oxygen as the ultimate standard. A part of the art of isotope-measurement consists in getting, close together upon the same plate, the lines of the desired ions and the lines of oxygen or of the intermediate standards. (We shall later see—in Chapter XV—that temporary errors existed in the table of atomic masses, because some of them were referred to helium as an intermediate standard, at a time when the ratio of charge-to-mass ratios between helium and oxygen was erroneously given).

Thus: what is measured, is the ratio of the charge-to-mass ratios of the desired ion and of some other ion taken as intermediate or ultimate standard: what is aimed at, is the ratio of the charge-to-mass ratios of the desired ion and of the commonest kind of oxygen ion; and what is printed in the tables is, the mass of the desired atom in a unit-system in which the mass of the commonest kind of oxygen atom is 16. The student, remembering how the chemical atomic weight of an element is defined (page 16), will guess the reason for this choice of unit-system. If oxygen and some other element E had each a single isotope only, then the mass of the atom of E in this unit-system would be equal to the chemical atomic weight of E. If oxygen had a single isotope only and E had two or more, then the weighted mean (weighted according to the abundances) of the atomic masses of these would be equal to the chemical atomic weight of E. Unluckily oxygen has (at least) three isotopes, a circumstance which was unknown when this unit-system was adopted; therefore in reasoning from atomic masses given in this unit-system to chemical atomic weights, a correction has to be made; but it amounts only to 2.7 parts in 10^4 (in the downward sense) and may usually be disregarded.

Atomic masses from hydrogen through argon have been tabulated after critical study by Livingston and Bethe in *Reviews of Modern Physics* **9**, 373 (1937). I quote the values 1.00813 and 2.01473 (both ± .00002) for the two chief isotopes of hydrogen, and Bainbridge's later value 4.00386 ± .00006 for helium.

Some general statements about isotopes

First, as to the *notation*. An isotope is symbolized by the chemical symbol of the element to which it belongs, adjoined above and to the right (or, sometimes, above and to the left) by its "mass-number" and below and to the left by the "atomic number" of the element: thus, $_2He^4$, $_{11}Na^{23}$. The mass-number of the isotope is the integer to which the value of its atomic mass, given in the unit-system just defined, lies closest. The atomic number of the element will be defined in the next chapter. With a few important exceptions, no isotope has a particular name or chemical symbol of its own. The exceptions are: (*a*) the isotopes which are radioactive and are found in Nature (instead of being made by transmutation), and (*b*) the second-to-be-discovered of the isotopes of hydrogen, which has the names "deuterium" and "heavy hydrogen," the chemical symbol D, and the complete symbol $_1D^2$ or $_1H^2$. (The commonest isotope of hydrogen is sometimes called "protium" when it is being explicitly contrasted with deuterium).

Second, as regards *alternative ways of detecting isotopes*. We shall see in later chapters that the phenomena of radioactivity, and also certain features of optical spectra, show that a single element may have more than one kind of atom. Radioactive substances alike in chemical properties but different in activity were the first-known examples of isotopes. Isotopes discovered with the mass-spectrograph or the parabola-method have been confirmed from spectra; some have themselves been discovered by interpretation of spectra, notably deuterium and the scantiest two of the three known isotopes of oxygen. Sometimes the masses, or the abundances, or both, of the isotopes of an element can be computed with worth-while accuracy from spectra.

Third, as regards *number of isotopes*: About two hundred and seventy stable kinds of atom are known, in addition to a steadily-increasing number of which the nuclei, and therefore the atoms themselves, are unstable and transitory. These are

TABLE 3.1

STABLE ISOTOPES OF THE ELEMENTS.[1]

Element	Z	A	Element	Z	A	Element	Z	A	Element	Z	A
H	1	* 1	K	19	40	Ge	32	70	Mo	42	98
		2			41			72			100
		3	Ca	20	*40			73	Ma	43	—
He	2	4			42			74	Ru	44	96
Li	3	6			43			76			98
		* 7			44	As	33	75			99
Be	4	9	Sc	21	45	Se	34	74			100
B	5	10	Ti	22	46			76			101
		*11			47			77			102
C	6	*12			*48			78			104
		13			49			80	Rh	45	101
N	7	*14			50			82			*103
		15	V	23	51	Br	35	*79	Pd	46	102
O	8	*16	Cr	24	50			81			104
		17			*52	Kr	36	78			105
		18			53			80			106
F	9	19			54			82			108
Ne	10	*20	Mn	25	55			83			110
		21	Fe	26	54			*84	Ag	47	*107
		22			*56			86			109
Na	11	23			57	Rb	37	*85	Cd	48	106
Mg	12	*24			58			87			108
		25	Co	27	57	Sr	38	84			110
		26			*59			86			111
Al	13	27	Ni	28	*58			87			112
Si	14	*28			60			*88			113
		29			61	Y	39	89			114
		30			62	Zr	40	90			116
P	15	31			64			91	In	49	113
S	16	*32	Cu	29	*63			92			*115
		33			65			94	Sn	50	112
		34	Zn	30	*64			96			114
Cl	17	*35			66	Cb	41	93			115
		37			67	Mo	42	92			116
A	18	36			68			94			117
		38			70			95			118
		*40	Ga	31	*69			96			119
K	19	*39			71			97			120

[1] After Livingston and Bethe, with additions from subsequent work of Dempster and of Nier. Where two or more isotopes of an element are known, and one constitutes more than 50% of the atoms of the element, this one is starred.

TABLE 3.1—*Continued*

STABLE ISOTOPES OF THE ELEMENTS [1]

Element	Z	A	Element	Z	A	Element	Z	A	Element	Z	A
Sn	50	122	Ce	58	*140	Dy	66	164	Os	76	188
		124			142	Ho	67	165			189
Sb	51	*121	Pr	59	141	Er	68	162			190
		123	Nd	60	142			164			192
Te	52	120			143			166	Ir	77	191
		122			144			167			*193
		123			145			168	Pt	78	192
		124			146			170			194
		125			148	Tm	69	169			195
		126			150	Yb	70	168			196
		128	—	61	—			170			198
		130	Sm	62	144			171	Au	79	197
I	53	127			147			172	Hg	80	196
Xe	54	124			148			173			198
		126			149			174			199
		128			150			176			200
		129			152	Lu	71	175			201
		130			154	Hf	72	176			202
		131	Eu	63	*151			177			204
		132			153			178	Tl	81	203
		134	Gd	64	152			179			*205
		136			154			180	Pb	82	204
Cs	55	133			155	Ta	73	181			206
Ba	56	130			156	W	74	180			207
		132			157			182			*208
		134			158			183	Bi	83	209
		135			160			184	Th	90	232
		136	Tb	65	159			186	U	92	234
		137	Dy	66	158	Re	75	185			235
		138			160			*187			*238
La	57	139			161	Os	76	184			
Ce	58	136			162			186			
		138			163			187			

[1] After Livingston and Bethe, with additions from subsequent work of Dempster and of Nier. Where two or more isotopes of an element are known, and one constitutes more than 50% of the atoms of the element, this one is starred.

distributed among about ninety chemical elements, in a very uneven way: tin has ten isotopes at least, xenon nine, several have eight or seven, and only a few may still be ideally simple in the sense of having one kind of atom only. All such statements are limited in the way described on page 83 *supra*, which I will now exemplify by quotations. Nier, observing cadmium, and failing to find any isotopes having the mass-numbers 107, 109, 115 or 118, says in effect that if such atoms exist they cannot constitute more than 1 part in 2350—1 in 2350—1 in 800—1 in 14700 respectively, of the total number of atoms in a sample of cadmium. According to Sampson and Bleakney, there is no second isotope of manganese comprising as many as 1 in 15,000 manganese atoms, and no second isotope of sodium comprising as many as 1 in 50,000 sodium atoms; according to Nier, no second isotope of beryllium which is even 10^{-5} times as abundant as Be^9. Anyone wishing to select an element which may really have only one isotope can probably make no better choice. When the first edition of this book appeared in 1926, it could still be affirmed that seven out of the ten lightest elements apparently had only one kind of atom apiece; now there is only one among them (fluorine) of which this may still be said (and even for fluorine it cannot be said if unstable isotopes are included). All statements about numbers of isotopes may thus be understatements; yet some of them may be over-statements too, for some ions which have been supposed to be ionized atoms may actually be ionized molecules (thus, an ion once interpreted as an ionized atom of a cadmium isotope having mass-number 115 is now interpreted as an ionized molecule of cadmium hydride CaH containing a cadmium atom of mass-number 114).[1]

Fourth, as regards *abundances and chemical atomic weights*. Most of the weighted-means of atomic masses of isotopes have agreed from the start with the chemical atomic weights of the

[1] Hydrides have been prolific sources of error, as volatile organic compounds containing hydrogen were used in some of the early investigations. Bainbridge circumvented this danger by employing discharges in neon, in which the metals which were being examined were evaporated and ionized (A in Figure 3.11 is the discharge-tube).

elements in question. In some cases there were bad disagreements, but—incredible to relate, for anyone brought up in the tradition that chemical atomic weights represent the *ne plus ultra* of accuracy—they were shortly reconciled by revisions of the *chemical* data. Among the elements for which this has happened are krypton, xenon, selenium, caesium, niobium and tantalum. To take one example: the chemical atomic weight of xenon, given as 130.2 up to the time when Aston obtained the weighted-mean of 131.27 ± 0.04 from mass-spectra, was revised to 131.26 ± 0.01 in a subsequent redetermination.[1] So great an impression has been made by these changes, that wherever a discrepancy exists at present it is expected (at least by physicists) to yield to a re-determination of the chemical atomic weight: iridium offers an outstanding instance at the moment. In respect of statistical uncertainty, the best of the evaluations of atomic masses of individual isotopes already surpass the best of the evaluations of chemical atomic weights of elements, and the advantage which they hold seems likely to increase.

Fifth, as regards *chemical atomic weights and integers*. The chemical atomic weights of the various elements (*cf.* the table on page 105), considered as a whole, exhibit a singular feature. I preface it by recalling that as long ago as 1815, W. Prout suspected from the few crude data then available that the atomic weights of the elements are integer multiples of that of hydrogen, and from this premise he inferred that all atoms are groupings of hydrogen atoms. Exacter measurements soon disposed of his premise, and yet something of it survived; for in the table of the atomic weights, many of the values do lie amazingly close to integers—too many to let us disregard the fact as a mere accident, too few to let us accept it as a universal rule.

[1] Whytlaw-Gray *et al.*, **PRS 134**, 7 (1931). It was contended that the "chemical" determination of the atomic weight of a noble gas like xenon, being a density-measurement interpreted by Avogadro's law, is really a "physical" experiment. To this Aston aptly replied that such a determination owes its value to a proper purification of the gas preceding the experiment, and this is an essentially chemical process.

The fact stands out most clearly if we review the lightest elements (for which, incidentally, the values are very reliable). Take the lightest twelve, for instance; set aside oxygen, as being the standard; of the other eleven, the values for *seven* lie between $(I - 0.02)$ and $(I + 0.02)$ where I stands in turn for each of the seven integers to which they severally are nearest, but for the *four* others the values (to wit, 6.94, 10.82, 20.18, 24.32) lie so far from their nearest integers as to vitiate all hope that the departures may be traced to experimental errors. This is a disconcerting situation. Had *all* the atomic weights been integer multiples of the least among them, it would have been received as a beautiful and satisfying law of Nature, vindicating Prout's idea. Had they been distributed completely at random, with no greater a percentage of close approximations to integer values than could be plausibly ascribed to chance, it would have been regarded as too commonplace to be surprising. The actual compromise is most curious; the probability of such a distribution occurring by chance has been estimated, and is fantastically small.

After a few elements had been resolved into isotopes, there was for a short time a prospect that this famous mystery might suddenly and totally be solved. Among these twelve lightest of the elements, for example, the four with the atomic weights sharply different from integers were early shown to be mixtures of isotopes; while the other eight (including oxygen) appeared at first to consist of only one kind of atom apiece, though additional isotopes present in scanty proportions were later to be discovered for all but two of the eight. A new rule was foreshadowed: wherever it is actually a single kind of atom to which the measurement refers, will not the chemical atomic weight be integral? and wherever the chemical atomic weight is definitely not an integer, will it not mean merely that the element in question is a mixture of different isotopes, for any one of which by itself (could it be separated) the value would be integral?

Well, it has turned out that while there would be something of the truth in such a rule, this formulation would be too

extreme. In the table of the chemical atomic weights, the grosser departures from integers are in fact due to mixing of isotopes, but some of the smaller ones are inherent in the atoms: the masses of atoms are not precisely integer multiples of the least among them, nor of any common divisor of the order of magnitude of the least among them. For example, the atomic masses of the lightest isotopes of hydrogen, of helium and of oxygen stand to one another in the ratios 1.008 : 4.003 : 16.000. This seems at first to be a painful frustration of Prout's hypothesis, becoming obvious just at the moment when the discovery of isotopes seemed to be removing the obstacle to that hypothesis. And yet, it is just what is needed to make Prout's notion tenable!

The point may be paraphrased as follows. If one assumes that a body is made up of particles, one is perforce assuming also that the particles stick together. " Stick together " is of course a layman's term; a physicist will say, that energy must be supplied to tear the particles apart. But energy has mass, and accordingly the mass of the system of particles must increase when they are torn apart. Reverse the process: when the particles come together and form the body, its mass must be less than the aggregate of the masses which they had when they were free. If a stable atom-nucleus is composed of x particles of one kind having mass m when free, its mass must be less than xm. If it is composed of x_1 particles of one kind and x_2 particles of another kind having masses m_1 and m_2 when free, its mass must be less than $(x_1m_1 + x_2m_2)$. The amount by which its mass falls below xm or $\Sigma x_i m_i$ will be the best measure of its stability.

The recently-introduced unit of mass—one-sixteenth of the mass of an O^{16} atom—cannot be the mass of a fundamental particle of which nuclei are built up: this is proved by the facts that the lightest nucleus of all has a mass distinctly (about 0.8%) *above* the unit, while many others have masses which are slightly above its integer multiples. As the book approaches its end, it will be stressed at various points that all nuclei but one are now considered as aggregates of two kinds of funda-

mental particles each having (when free!) a mass 0.8% to 0.9% above the unit. (The one exception, the H^1 nucleus or proton, is considered to be a free particle of one of these types; of the other, the " neutron," we shall make the acquaintance in Chapter XVII). We might tabulate the differences between the actual mass-values of the isotopes and the values of $(x_1m_1 + x_2m_2)$, putting for m_1 and m_2 the masses of proton and neutron when free, and for x_1 and x_2 the integers which make the differences least. However, m_2 is not exceedingly well

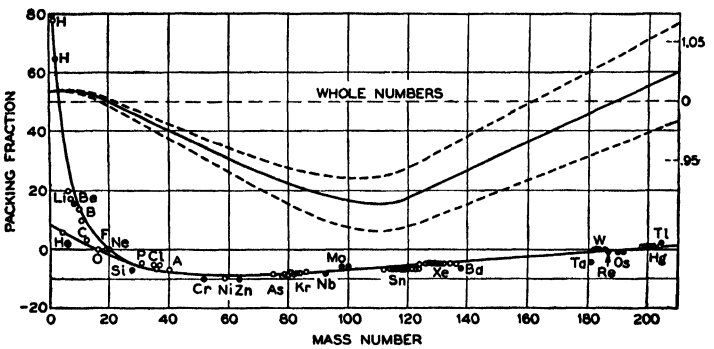


FIGURE 3.13—PACKING-FRACTION CURVE TAKEN FROM A REPORT OF A ROYAL SOCIETY DISCUSSION OF 1929, WITH SUBSEQUENT VALUES FILLED IN FROM DATA OF ASTON AND BAINBRIDGE.

measured as yet, and anyhow the model is somewhat speculative: it seems wisest accordingly to keep to the old practice, and tabulate or plot the differences between the mass-values and the corresponding mass-numbers. A technical term must be learned here: " packing-fraction," which signifies the excess of mass-value (in terms of the unit lately defined) over mass-number (the nearest integer below the mass-value) divided by mass-number and multiplied by 10^4. One might call it the excess of mass-value over mass-number, expressed in per cent of per cent of the latter.

Figure 3.13 exhibits what is known as a " packing-fraction curve." Its most notable characteristic—the downward con-

vexity, with the crossings of the horizontal axis near the mass-numbers 20 and 175—was early established by Aston, the bifurcation at the left-hand end somewhat later. The plotting of additional points to fix the trend of the curve more closely is now one of the major occupations of physicists who study isotopes. Hitherto the values in the upper parts of the range of masses have been inferior in number and accuracy to those in the lowest part, but with the lapse of time this is steadily ceasing to be the case as isotope after isotope is being redetermined with the best available technique. Many values have lately been added by Dempster [1] (Figure 3.14).

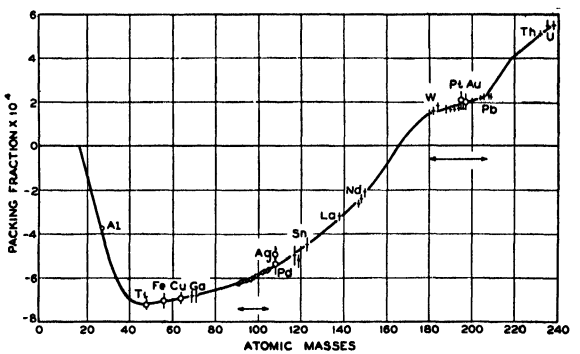


FIGURE 3.14—PACKING-FRACTION CURVE FITTED TO DATA BY DEMPSTER. (A. J. Dempster: *Physical Review.*)

The words " nucleus " and " nuclear " have crept into this section, as they did into the section on alpha-particles, even though their formal introduction was scheduled for Chapter IV. It would have been foolish to attempt, in a discussion of the masses of atoms, to conceal the fact that these relate to a certain portion of the atom which is called the " nucleus " or " kernel "; the rest of the atom consisting of negative electrons of which the total mass is less than 1/3600 (except for

[1] PR 53, 869 (1938); for work before 1933, see Aston, *Mass-spectra and Isotopes* (Arnold).

H^1, for which it is about 1/1850) of the nuclear mass.[1] It is on this account that in the last paragraph but one, the mass-values were interpreted on the basis of a theory of the structure, not of the atom but of the nucleus. Here we have made our first entry into nuclear physics, not to be followed by another until at the end of Chapter XIV we shall add the properties of angular momentum and magnetic moment as two more qualities of kernels. I end this chapter by reflecting once more on the history of Prout's conception. Probably it happens seldom in the history of science that a great idea is based at first on measurements too crude to be significant—then meets an obstacle when the measurements are improved—then finds the obstacle suddenly vanishing—then meets again the seeming obstacle in an attenuated form, but only when it can be recognized as no impediment at all, but as a help and indeed a necessary help to the idea.

[1] Thus the masses of the nuclei—"proton," "deuteron" and "alpha-particle" respectively—of the three atoms named in the last sentence of page 90 are 1.00810 ± .00002, 2.01470 ± .00002 and 4.00380 ± .00006 respectively. Slight as they are, the differences are not always negligible in nuclear physics.

CHAPTER IV

The Elements and the Nuclear Atom-model

The periodic table and the procession of the elements

The Table of the Elements is an instance of a wonderful achievement resulting from the independent researches of many, coordinated eventually by the wide vision of a few; for the intuitions of those who arranged it would have had on material on which to play, had it not been for the enormous labors first of those—chiefly chemists—who identified and separated the elements, and then of those—chemists and physicists alike—who measured their physical and chemical properties one by one.

It is scarcely possible to imagine the amount of work expended in fitting each of the elements into its proper place in the table. The very notion and concept of an " element," not by any means self-evident, had to be defined out of wide experience. Multitudes of substances had to be analyzed to ascertain whether they were elements, or compounds, or mixtures. Multitudes of compounds had to be analyzed to ascertain what are the combining-weights and what the valences of the elements participating in them. These features of an element, and certain others, such as its behavior towards high-frequency radiation and fast-flying alpha-particles, can be studied even though it be never observed uncompounded with others; but many of its properties cannot be ascertained unless a certain amount of it is separated from all other elements. Thus it was clearly important to undertake the task of isolating each element, and having isolated it, to measure its density, its elastic properties, resistivity in the solid and in the liquid phase, temperatures of change of phase, crystalline

habit, optical spectrum, magnetic susceptibility in each phase, and a host of other properties besides. Measurements upon all of these assist in the allocation of an element to its due position in the table, and the table is based upon them all.

Some of these properties of the elements fall within the scope of this book, but most of them lie beyond. Yet they all vary from element to element in ways which betoken an interconnection and an interrelation among them all. This is an extremely important fact. It proves, for one thing, that the elements are affiliated with each other. It proves, for another thing, that in selecting those properties of the elements which the nuclear atom-model is at present competent to imitate, we are not choosing properties which are eccentric nor untypical nor irrelevant. On the contrary: in the manner of their variation from element to element, these particular selected properties resemble the others so much, that any atom-model competent to simulate them must contain some feature which embodies that mode of variation common to them all, and to that extent—partial and feeble though it may be—is a competent model for them all.

The periodic properties and the periodic table of the elements

When the elements are arranged in the order of their chemical atomic weights (with three exceptions which will presently be named), and the variations of their several properties from one to the next all along this " procession of the elements " are examined, it is found that many of these properties vary periodically and with the same periodicities. This is illustrated by the Figures 4.1 and 4.2, which are graphs representing the variations of the atomic volume (the volume of one gramme-atom of the solidified element at some particular temperature), the compressibility, the coefficient of thermal expansion, the reciprocal of the melting-point and the magnetic susceptibility, along the line of elements. Such graphs could be drawn for a number of other properties. This is the fact which the arrangement of the elements in the Periodic Table is meant to make conspicuous.

Consider the Periodic Table as it is displayed, for instance, on page 105. (This is only one among several ways of displaying it, another of which will be introduced as Figure 4.3.) The elements are arranged almost exactly in the order of increasing atomic weight, the line doubling back on itself at stated intervals and proceeding from left to right in each row. The similarities between the elements sharing a column are so

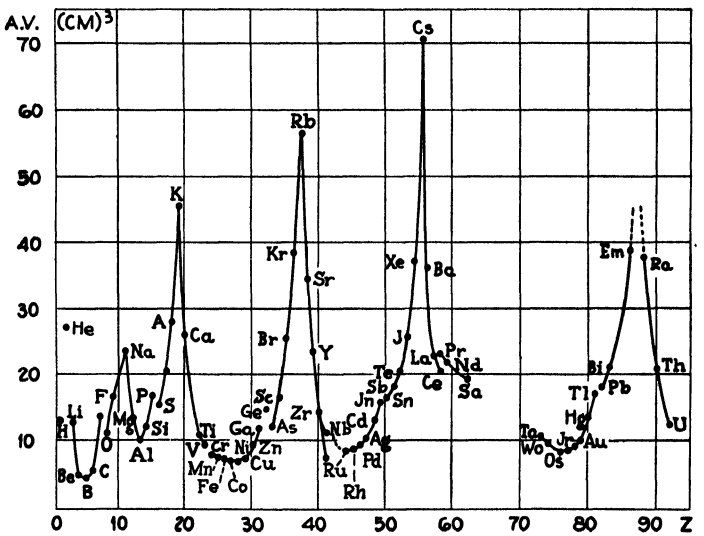


FIGURE 4.1—INSTANCE OF A PERIODIC FUNCTION OF ATOMIC NUMBER: THE ATOMIC VOLUME. (After Sommerfeld.)

pronounced, that anyone having even a modest acquaintance with chemistry or physics feels it quite natural to see them thus assembled. The elements in the left-hand side of the first column are known as the alkali metals, those in the left-hand side of the second column are the alkaline-earth metals, those on the right-hand side of the seventh column are the halogens, those in the ninth column are the noble gases. These particular names do not matter for our present purpose; the essential point is, that each of these sets of elements forms a natural

group distinctive enough to bear a name. The elements in the first column are all univalent (though some of them are in addition bivalent or trivalent in some compounds), those in the second column predominantly bivalent, those in the third column trivalent; and in general an element in the nth column is found to be n valent in some at least among its compounds, except that the elements of the ninth column rarely if ever

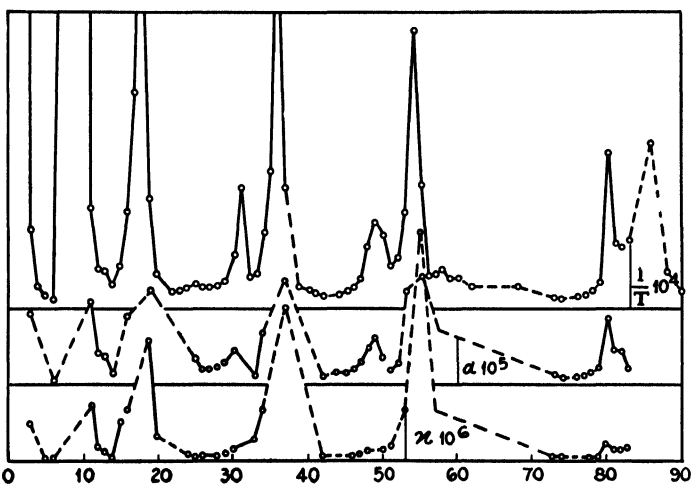


FIGURE 4.2—INSTANCES OF PERIODIC FUNCTIONS OF ATOMIC NUMBER: ABOVE, RECIPROCAL OF MELTING-POINT; IN THE MIDDLE, COEFFICIENT OF THERMAL EXPANSION; BELOW, COMPRESSIBILITY. (After Sommerfeld.)

enter into chemical combination. Many elements enter into various combinations with various valences, and these complexities likewise are recurrent in the Periodic Table.

Passing over to the physical properties, one finds likenesses quite as striking among the elements belonging to a single column. The elements of the first column have doublet spectra, those of the second have singlet and triplet spectra, and to each of the other columns there corresponds a distinctive set of spectrum-types. These types will be described in Chapter X; here, the only essential point to realize is that

PERIODIC TABLE OF THE ELEMENTS

(Values of atomic weights taken from the seventh Report of the Committee on Atomic Weights of the International Union of Chemistry; G. P. Baxter, *et al*, *J. Am. Chem. Soc*, 59, p. 2215.)

I	II	III	IV	V	VI	VII	VIII			O
1 H 1.0078										2 He 4.002
3 Li 6.940	4 Be 9.02	5 B 10.82	6 C 12.01	7 N 14.008	8 O 16.000	9 F 19.00				10 Ne 20.183
11 Na 22.997	12 Mg 24.32	13 Al 26.97	14 Si 28.06	15 P 31.02	16 S 32.06	17 Cl 35.457				18 A 39.94
19 K 39.096	20 Ca 40.08	21 Sc 45.10	22 Ti 47.90	23 V 50.95	24 Cr 52.01	25 Mn 54.93	26 Fe 55.84	27 Co 58.94	28 Ni 58.69	
29 Cu 63.57	30 Zn 65.38	31 Ga 69.72	32 Ge 72.60	33 As 74.9	34 Se 78.96	35 Br 79.916				36 Kr 83.7
37 Rb 85.48	38 Sr 87.63	39 Yt 88.92	40 Zr 91.22	41 Nb 92.91	42 Mo 96.0	43	44 Ru 101.7	45 Rh 102.91	46 Pd 106.7	
47 Ag 107.880	48 Cd 112.41	49 In 114.76	50 Sn 118.70	51 Sb 121.76	52 Te 127.61	53 I 126.92				54 Xe 131.3
55 Cs 132.91	56 Ba 137.36	RARE EARTHS	72 Hf 178.6	73 Ta 180.88	74 W 184.0	75 Re 186.31	76 Os 191.5	77 Ir 193.1	78 Pt 195.23	
79 Au 197.2	80 Hg 200.61	81 Tl 204.39	82 Pb 207.21	83 Bi 209.00	84 Po	85—				86 Rn 222
87—	88 Ra 226.05	89 Ac	90 Th 232.12	91 Pa 231	92 U 238.07					

RARE EARTHS

57 La 138.92	58 Ce 140.13	59 Pr 140.92	60 Nd 144.27	61	62 Sa 150.43	63 Eu 152.0	64 Gd 156.9
65 Tb 159.2	66 Dy 162.46	67 Ho 163.5	68 Er 167.64	69 Tm 169.4	70 Yb 173.04	71 Lu 175.0	

the character of the spectrum of an element depends upon its column in the table. The space-lattice formed by the unit-cells when an element crystallizes likewise depends, though not exclusively, upon the column in which the element belongs. The alkali metals are set apart from all other elements as distinctly by physical as by chemical properties; they have greater conductivities for electricity and heat, higher threshold wavelengths for the photoelectric effect, lower thermionic work-functions than any of the rest. For such properties as these, which are measurable by numbers, the periodicity is most vividly shown by assigning consecutive integers to the elements, and plotting the numerical values of the properties against these integers. This is what was done in Figures 4.1 and 4.2 for four properties and will be done elsewhere in this book for others.

The curves show that the variations, while distinctly periodic, are by no means smooth. The most striking deviation from perfect evenness lies in the period itself. The deviation may be expressed by saying that the first row of the table is two elements long, the next two rows contain eight elements apiece, the next two rows comprise eighteen elements each and the next following row thirty-two. (The last row ceases after six elements, but not in such a way as to imply that a new row is about to begin.) Another way of describing the unevenness is to say that, instead of every eighth element being a noble gas (as one could expect from the manner in which the Periodic Table begins) the elements which are noble gases are the 2d, 10th, 18th, 36th, 54th, and 86th. There is a curious relation among these numbers: the first is $2 \cdot 1^2$, the others are obtained from it by successively adding $2 \cdot 2^2$ once, then again, then $2 \cdot 3^2$ once, then again, then $2 \cdot 4^2$ once. This shows that the periods of the Periodic Table, while not all of the same length, yet are regular in a no less striking way.

Another scheme for representing the Periodic Table (Fig. 4.3) shows a sort of compromise between the evenness with which the periods of the Periodic Table commence and the unevenness in which they end. The periods of eighteen ele-

ments apiece are represented as periods of eight, like their predecessors, but with intrusions of blocks of ten elements apiece; while the period of thirty-two is represented as a period of eighteen, into the midst of which an additional block of fourteen extra elements has shouldered its way. This particular set of elements, the rare-earth elements, is distinguished

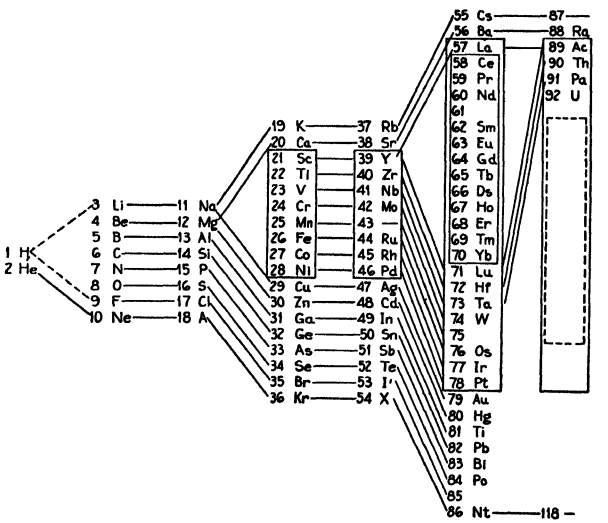


FIGURE 4.3—SCHEME FOR EMPHASIZING CERTAIN ASPECTS OF THE PERIODIC TABLE. (After Bohr.)

by so great a uniformity and similarity among all its members that it is very difficult to separate any two of them which happen to have been originally mixed; the difficulty is surpassed only by that of separating isotopes. It is because of this great interresemblance that the table on page 105 is printed with all of them crowded together into one place.

The non-periodic properties and the procession of the elements

Next it is natural to enquire whether, in addition to all these periodic properties, there exist any properties of the elements which increase or decrease steadily from one end of the table to the other, displaying no periodic variations. Before 1905 only one property was known of which this statement is even approximately true. The chemical atomic weight does almost always rise in passing from one element of the table to the next higher one. The " almost " of the last sentence is an important qualification, although it operates only in three places of the table; argon is placed before potassium, cobalt before nickel, tellurium before iodine, although in each of these pairs the element first named has the greater atomic weight. It must have taken courage as well as acumen to reverse these pairs when the Periodic Table was first estab lished, since at that time the atomic weight seemed to be the really fundamental quality of each element, though now it is believed to be inconsequential and in many cases, as we have seen in the foregoing chapter, is actually known to be deceptive. Even in the much more numerous cases in which a step upward along the table involves a step upward in combining-weight, the steps in atomic weight are by no means all equal. Nevertheless the atomic weight was precious as the solitary instance of a property varying not periodically but steadily with the location of the element in the table, and it was invaluable as a guide in arranging the table in consecutive rows and columns, although in three cases it was a misleading guide.

Not until the discovery of alpha-particles and of X-rays were other properties of the elements discerned which vary not periodically but steadily along the Table. During the first decade of the twentieth century this feature was demonstrated, by rather imprecise measurements, for the penetration of alpha-particles into layers of the elements; for the absorption and scattering of X-rays by the elements; and for the so-called " hardness " of the characteristic X-rays generated by impacts of high-speed electrons against the elements. This rather

vague concept of " hardness " was translated into a quantity very definite and measurable indeed, the frequency of the characteristic X-rays, when by virtue of the discovery that crystals serve as diffraction gratings for X-rays it became possible to measure the wavelengths and to compute the frequencies of the rays. Making the first measurements of these frequencies, and making them for a long series of elements, H. G. J. Moseley discovered that they increase smoothly along the Periodic Table from each element to the next, showing (for the lines of highest frequency) not a trace of periodic variation. Moseley in fact discovered that the variation is according to a simple law; for he found that the square root of the frequency of a particular X-ray line (for the characteristic X-rays of an element comprise a number of distinct recognizable lines each with its own frequency) is a linear function of the integer numbers attached consecutively to the consecutive elements. In passing from one element to the next, the square root of the frequency of an X-ray line increases always by the same amount. This is Moseley's Law. It is illustrated by Figure 4.4, in which the square root of the frequency of the particular X-ray line known as the $K\alpha$-line

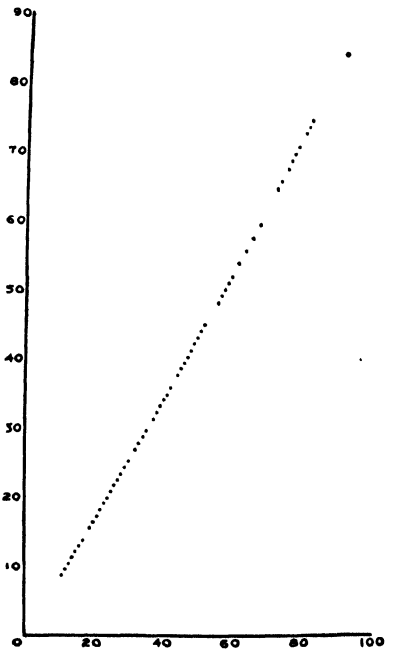


FIGURE 4.4—INSTANCE OF A NON-PERIODIC FUNCTION OF ATOMIC NUMBER: THE K-SQUARE ROOT OF THE K-FREQUENCY OF THE X-RAY EMISSION SPECTRUM. (Data from Siegbahn, Stevenson, Cork and Rechou.)

is laid down for element after element along the procession at equal spacings; and by Figure 4.5, in which the frequencies

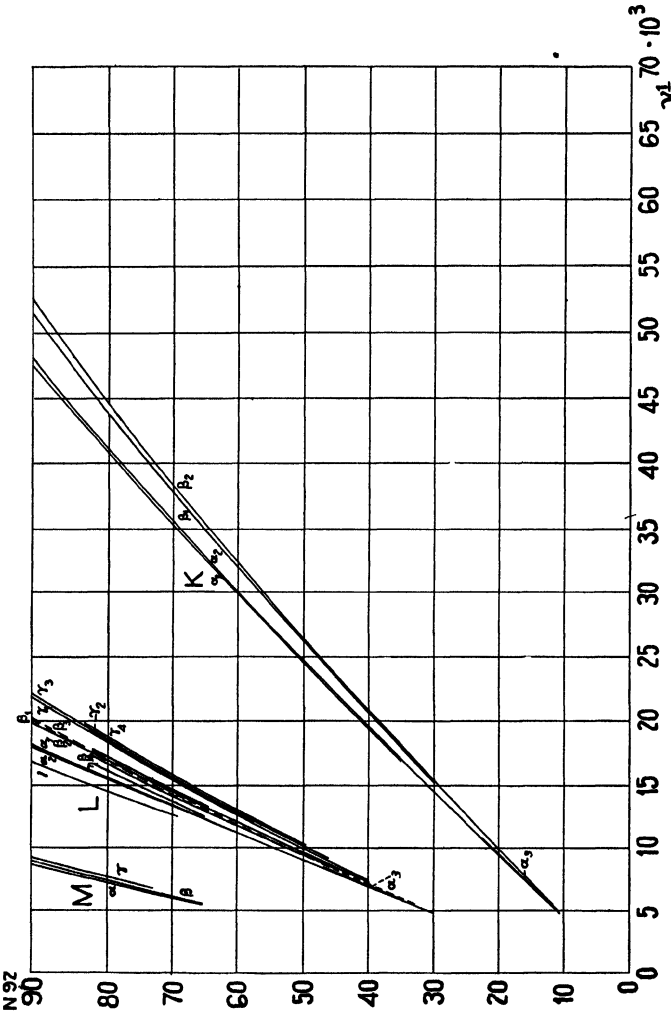


FIGURE 4.5—INSTANCES OF NON-PERIODIC FUNCTIONS OF ATOMIC NUMBER: THE SQUARE ROOTS OF THE FREQUENCIES OF THE MOST CONSPICUOUS LINES IN X-RAY EMISSION-SPECTRA. (After Ledoux-Lebard and Dauvillier.)

of a multitude of other individual and recognizable lines are represented by the smoothly drawn curves.

Moseley's Law may be so interpreted as to become one of the foundation-stones of the nuclear atom-model, and in that sense I shall presently interpret it; but its greatest service is one which needs no model. The mere discovery of a property of the elements varying linearly over long stretches of the Periodic Table reinforced the credit of that table, and vastly enhanced the confidence of physicists and of chemists in the essential rightness of its arrangement. Moreover, the discovery settled some debatable questions still outstanding in regard to the table, and caused it to be improved in certain ways.

The first question to settle is naturally the question about the " reversed pairs " of elements previously cited—argon and potassium, tellurium and iodine, cobalt and nickel—in each of which the Periodic Table demands one order of succession and the " atomic weights " demand the opposite order. The X-ray frequencies of these elements require for each pair the order demanded by the Periodic Table, and exclude the order suggested by the " atomic weight." This result is in itself sufficient to dethrone the " atomic weight " from its rank as the fundamental property of the elements.

The next question involves the gaps in the Periodic Table. Those who established the table fifty years ago had found that when they wrote the symbols of the elements in their order of increasing " atomic weight " (except for the reversed pairs) it was necessary in some places to put consecutive elements into non-consecutive columns, leaving gaps which seemed to be awaiting elements yet undiscovered. Mendelejeff pointed out three such gaps, which were subsequently filled by the afterward-discovered elements gallium, germanium, and scandium. These lacunae were detected only through a study of the valences and the other chemical and physical properties of the elements adjoining them. Of course, when a property of the elements was discovered varying usually by equal steps from one to the next, the lacunae became much easier to discern and immeasurably more striking to demonstrate. Where there is really a gap between two consecutive known elements, the square root of each X-ray frequency should jump

twice as far as where there is none. At the time of Moseley's discovery, the chemists admitted the existence of six gaps, at the places numbered 43, 61, 72, 75, 85 and 87 in the table. Across each of these the square root of an X-ray frequency executes a leap twice as great as the normal leap; the presence of the lacuna is confirmed. Afterwards, the elements belonging in two of these places (72 and 75) were discovered. In each case, the discovery was tested by ascertaining that X-ray frequencies of the substances presumed to be the new elements lay in the proper places upon the curves drawn originally by Moseley. Measurements of X-ray frequencies have likewise settled questions as to the number and ordering of the known rare-earth elements, the answers to which previously depended upon chemical evidence excessively difficult to gather and to interpret.

The foregoing passage suggests that Moseley's discovery simply confirmed what chemists already suspected; this is nearly true, so long as we ignore the interpretations which can be placed upon Moseley's data. The fact that it is nearly true is wonderful praise to the chemists, who were able to systematize a vast mass of miscellaneous data to so great an extent that they detected the reversed pairs and the gaps in the procession of the elements, and recognized that the atomic weight is not altogether a reliable guide in ordering that procession. It is worth while to point out, however, that from such data no one could ever have detected the absence of an entire column of missing elements, had there been such a one. This might have happened. Until in the early nineties Rayleigh became curious about the disparity in atomic weight between atmospheric nitrogen and nitrogen from ammonia, and so was led to discover a gas (argon) unknown till then which was mixed with the one and not with the other, no one suspected the existence of a group of elements destined to form an entire ninth column of the Periodic Table; but there *is* such a group, and four of its other members were swiftly discovered after argon gave the clue. Had this discovery been postponed until after Moseley's, the gaps in the X-ray-

frequency curves would have revealed that it was still waiting to be made. Conversely, the fact that there are no more gaps in the series of X-ray frequencies, apart from those already mentioned, is powerfully convincing evidence that save for those vacancies (and possibly for elements lying beyond uranium) the Table of the Elements stands complete. No theoretical model is required to reinforce this conclusion; one has but to look at Figure 4.4, and reflect that if any new element with a new X-ray frequency were discovered, and did not fill one of the gaps in that line of dots, it would disrupt the harmonious progression of the line.[1]

Localization of periodic and non-periodic properties in the atom-model

There is an important Rule concerning periodic and non-periodic properties of the elements, which influences the design of the nuclear atom-model, as it must that of any other.

The Rule may be stated in this manner: *The properties of an element which are not changed by change of phase nor by change of state of chemical combination are non-periodic*; *while the periodic properties of an element depend upon its phase and its state of chemical combination.*

The first part of the Rule may be illustrated from the behavior of the elements towards high-frequency radiation and fast-flying charged particles. The large-angle scattering of alpha-particles (as we shall see in Chapter V) follows the same laws whether the scattering-substance be a gas such as argon, or a solid such as gold-foil. The absorption of X-rays of a given type by a stratum containing a given amount of sulphur commences (as we shall see in Chapter VIII) at nearly the same frequency, whether the stratum be a lamina of crystalline sulphur, a layer of pure "amorphous" sulphur, a film of some sulphide or a sheet of some sulphate. In the graphs of various X-ray frequencies of the elements presented in

[1] This reflection has already served to discourage attempts such as were frequently made only a few years ago, to postulate new and light elements to account for lines of unknown origin in spectra of nebulae or the aurora.

Figure 4.5, the measurements for some of the elements were made upon elementary gases, for some upon the pure solid elements and for some upon compounds; yet they all lie at even intervals along the same smooth curves.

On this account, the non-periodic properties of the elements are attributed to inner regions and internal features of the atom, nearly or quite sheltered from external circumstances such as the influences of neighboring atoms.

On the other hand, the periodic properties of the elements depend upon phase. Some of them, such as crystalline habit and rigidity, apply to the solid phase only; others, such as viscosity and compressibility, have extremely different values in the different phases; others, such as the optical line-spectra, are found by experiment to belong to the gaseous phase alone. Furthermore, when two or more elements enter into chemical combination, the resulting compound has an optical spectrum, a viscosity and a compressibility, a magnetism and a space-lattice of its own, which need not be and frequently are not identical nor even similar to those of any of the elements which make it up.

On this account, the periodic properties of the elements are attributed, directly or indirectly, to frontier regions and superficial features of the atom which are liable to alteration by the influence of other atoms which come near.

We have had examples of properties of the elements which vary with the periodicity of the Periodic Table, and examples of other properties which progress at a more or less uniform rate from one end of the table to the other; now we will consider some properties which fall in neither class.

Radioactivity

It seems to be the fate of radioactivity, until the very last chapter of this book, to be mentioned always as an incident to something else and never for its own sake. I have already had occasion to say that radioactive substances emit alpha-particles (of charge $+2e$) and negative electrons (of charge $-e$);

they also emit corpuscles of light, but this fact is not important for the present. The radioactivity of any substance is fully described by specifying the kinds and the speeds of the electrified particles and the energies of the corpuscles of light which it emits, and the rate at which it emits them. The chemistry of these bodies has been studied so thoroughly as to form a large department of the science of chemistry (" radiochemistry ") all to itself. One of its prime results is, that radioactive substances are chemical elements rather than chemical compounds. Another is, that the radioactivity of any one of them is quite unaffected by putting it into chemical combination with any other element; and this is supplemented by a great deal of evidence that radioactivity cannot be altered by anything, by temperature or change of phase or exposure to light or X-rays or anything else—for during the early years of the present century, much of the experimentation in this field consisted of vain attempts to find some way of altering it.

Radioactivity is, therefore, an extreme example of a property which must be localized in a part of the atom (or, of the atom-model) which is sheltered and immune from the outer world and from the proximity of other atoms. Thus the student will not be surprised to learn that it is not a periodic property of the elements.

Hyperfine structure

A considerable part of this book will be devoted to optical spectra, and we shall find that in the main they form a periodic property of the elements: those of the elements in the first column of the Periodic Table, for instance, are strikingly like one another and strikingly unlike those of the elements in the second and in the ninth and indeed in all the other columns, so that in this respect lithium near the beginning of the table is far more like caesium beyond the middle thereof than it is like its neighbors helium and beryllium. Yet there is an important detail known as " hyperfine structure " which is like radioactivity, in that it is neither a periodic nor a simply

progressive quality of the elements. Moreover it is like radioactivity, in that it differs for different isotopes of an element. It is true that nobody has yet unmixed sufficiently the isotopes of any element, to be able to see separately the hyperfine structures belonging to each; but it is possible to analyze the hyperfine structure of the mixture of isotopes which constitutes an element such as cadmium or mercury, in such a way as to show convincingly which lines belong to which isotopes; whereas in respect of nearly all the other details of spectra, the number of the isotopes of the element in question obviously makes no difference, and it is necessary to suppose that in these respects all the kinds of atom belonging to the element behave exactly alike.

We have met by now with three properties—mass, and radioactivity, and hyperfine structure—which for the individual element vary from one isotope to another, and for the procession of the elements are neither periodic nor simply progressive. (The number of isotopes per element is also neither periodic nor simply progressive, as I indicated earlier.) We now return briefly to radioactivity.

The displacement-law of Russell, Fajans and Soddy

Doubtless it is already known to every student of this book that radioactivity is a spontaneous transmutation: the radioactive element—or rather, the radioactive isotope of an element—is changing into an isotope of a different element. The change is described by the "displacement-law" which bears the names of Russell, Fajans, and Soddy. When an alpha-particle is emitted, the atom whence it came is transformed into an atom of the element lying two steps nearer the beginning of the table; when an electron is emitted, the atom whence it came is transformed into an atom of the element lying one step onward toward the end of the table. (In practice one must be careful not to confuse these electrons with others ejected by secondary processes!)

The mass which departs with the electron is a negligible

fraction of that of the atom; that which departs with the alpha-particle is not so small, and yet we know of plenty of cases in which two isotopes of a single element differ in mass by more than that amount. It is, therefore, not the departure of the mass which has changed the element, but something different. Now we begin to see that it is *electric charge* which makes the difference between one element and another. Yet we cannot suppose that the mere detachment of an electron is always enough to transmute an element! Electric discharges in gases are populous with atoms which have temporarily lost electrons (such are the ones which are observed in the mass-spectrograph), and so are electrolytic solutions, but when these ions are re-neutralized they are atoms of the same element as before. When a radioactive atom loses an electron in the way which I have just mentioned, it, too, is soon re-neutralized by capture of another electron; but evidently the latter electron cannot occupy the same place as the former vacated, for else we should have the original atom over again instead of an atom of a different element. It begins to be obvious that the atom-model must have compartments, one to contain the alpha-particles and the electrons which are emitted in radio-activity, the other to contain the electrons which are detached when the atom is ionized and replaced when it is re-neutralized.

Positive ions again

We return for a moment to the positive ions which were so prominent in Chapter III, in order to learn a principle which is one of the foundations of the nuclear atom-model. It is possible to identify, in the optical spectrum of a discharge in a gas, the lines due to the various kinds of positive ion inhabiting the discharge. It turns out that the optical spectrum of an ion is entirely different from that of the neutral atom from which it is formed by ionization; moreover, that if successive more-and-more highly charged ions are made by detaching electrons one after another from an atom, each has a characteristic spectrum of its own entirely different from all the rest.

The character and structure of the optical spectrum of an atomic system change, whenever the number of electrons in the system is changed. Now to turn for an instant to electrochemistry: it is well known that one may dissolve in water two salts of chromium (to take but one element as an example) and obtain two different kinds of ion Cr^{++} and Cr^{+++}, wandering around in their sustaining " atmosphere " of water, which in mobility and refractivity and magnetic susceptibility are altogether different. Many elements may thus be caused to supply two or more kinds of ion in aqueous solution, others only one; but in any case, the qualities which these particles display are qualities which are definitely of the ion, and not of the neutral atom whence it is derived.

The fundamental principle is this: *when an ion is formed by detaching* n *electrons from an atom of some element, it resembles in many of its properties the neutral atom of the element* n *steps nearer to the beginning of the procession.* The resemblance is most striking in respect of optical spectrum (of ions in a gas), magnetic susceptibility (of ions in a liquid), and scattering of X-rays (by ions in a crystal lattice); but it is not by any means confined to these few qualities.

The principle sounds suspiciously like a turning-upside-down of the second part of the displacement-rule of Fajans and Soddy; but we shall presently see that even this makes sense instead of nonsense.

Magnetism in the atom

Neutral atoms isolated in space have magnetic moments which are measurable; the like has just been implied of the ions in electrolytic solutions, and the proof may also be made for atoms and molecules in gases, liquids and solids. Details will be given in later chapters; for the present I simply state that magnetic moment is a property inherent in the atom and the ions, and must be explained by the atom-model. Now we know that on the scale of objects large enough to be seen and handled, a magnet may be simulated by a circulating electric

current. If, therefore, our atom-model is built of electrified particles, we may provide it with magnetic moment by assuming that these particles are in incessant circulating motion.

The nuclear atom-model

From all the foregoing data we are impelled to make an atom-model in which there are detachable electrons, and a positive charge to balance them, and some at least of the charges are in motion so as to make the system inherently magnetic. The model must be at least tripartite, including an outer region varying periodically in its character from element to element along the procession, varying also with circumstances such as proximity of other atoms—an inner region almost uninfluenced by circumstances, varying steadily in its character from one element to the next along the pro cession—an inmost region immune from circumstances, to which we shall ascribe such capriciously varying qualities as the mass and the radioactivity of the atom. The outer regions will be occupied by the detachable and replaceable electrons—scores of them, in certain kinds of atom, and yet always so few that the total of their masses is less than .0002 of the whole of the mass of the atom, all the rest of which is left to be put into the inmost compartment. This inmost compartment must also be capable of ejecting alpha-particles and negative electrons with great energy. This amounts almost to saying that electrons and alpha-particles must be imagined existing in the compartment; yet not quite, for one might conceive that charge and mass are distributed continuously through it and that they come out as identical particles only because of some peculiarity of the agencies which carve them out; even as the raindrops falling out of a cloud are fairly uniform in size although their material exists in the cloud quite otherwise than as rain.[1] The atom-model must also account for the

[1] I give myself some credit for having said this in the 1926 edition, years before it was seriously propounded that the electrons and alpha-particles emerging from nuclei are formed just as or shortly before they are emitted; yet in a few more years this idea may be discarded (or extended to all the electrons of the atom?).

displacement-law of Russell, Fajans and Soddy, and for the principle concerning the properties of ions.

These demands are fulfilled by the nuclear atom-model. It consists of a nucleus of charge $+Ze$ surrounded by a family of negative electrons, the number Z being an integer varying from one element to another, the very integer which marks the place of the element in the procession—the integer beside the symbol of the element in the tables on page 105 and pages 92–93. I give it henceforth its name of "atomic number." When the family is complete with Z electrons, the model corresponds to the neutral atom. As we pass along the procession from beginning to end, we distinguish each atom-model from its predecessor by adding a new electron to the family and a new positive charge $+e$ to the nucleus (we may or may not add mass to the nucleus along with positive charge, but if we do it has no influence upon the number of electrons). Before we have got very far along the procession we can begin to distinguish between outer and inner electrons of the family, the former relatively exposed to the influence of other nearby atoms, the latter relatively sheltered. We may suppose that the electron-family is arranged in consecutive layers or shells. This enables us to impose a steady progression of properties upon the inner electrons, by imagining that as Z increases the inner shells retain their general conformation but are more and more strongly drawn inward by the increasing nuclear charge. It also enables us to impose a periodicity upon the outer electrons, by imagining that as Z increases a new outer shell is completed at periodic intervals. Detachment of n electrons from the model for the Zth element converts it into a model possessing $(Z - n)$ electrons, therefore like the model for the neutral atom of the $(Z - n)$th element, yet not identical with this last because its nuclear charge is still Z instead of $(Z - n)$. Most features of optical spectra depend primarily on the number of electrons and secondarily on the nuclear charge, and therefore are identical for the isotopes of a single element, and similar though not identical for atoms and ions for which $(Z - n)$ is the same while Z has different values. Mass and

radioactivity are nuclear properties, and hyperfine structure in spectra is ascribed to a peculiar interaction between nucleus and electrons; all three may therefore vary from one isotope to another of a single element. Emission of an alpha-particle reduces the nuclear charge by 2, emission of a negative electron increases the nuclear charge by unity, and the size of the electron-family must diminish by two or increase by one in order to re-neutralize the atom; hence the displacement-law. Chemical properties are mainly attributable to the outer electrons, hence their periodicity. X-ray spectra are mainly attributable to the inner electrons, hence their smooth progression with Z. Magnetism is tentatively explained by requiring the electrons to revolve around the nucleus; this turns out to be necessary but not sufficient; we shall in addition be requiring the electrons and the nuclei to rotate inalterably upon themselves.

Such is the general design of the atom-model, to the perfecting of which a great part of the thought of physicists has been devoted for nearly thirty years, resulting in universal assent. The assent is commanded not by statements so vague as these which preceded, but by close numerical agreements with experience which appear when the atom-model is made precise by sharply-defined numerical assumptions. The earliest and to this day the most striking of the proofs depends, however, on nothing more precise than I have already stated. It is supplied by a study of the deflections which alpha-particles, the nuclei of helium atoms, suffer when they traverse atoms of other kinds. We now take up this proof.

CHAPTER V

The Nuclear Atom-Model Attested by Scattering of Alpha-Particles

This chapter is devoted to the *statistical* study of the deflections of alpha-particles by atoms. We have already beheld in Chapter III, with the spectacular aid of Wilson-chamber photographs, individual deflections of this type. Each one of them by itself was capable of showing (by equations 3.2 and 3.3) that the deflection was due to a force exerted between the alpha-particle and another body, which had sensibly the entire mass of an atom of the gas. As to the nature of this force and the way in which it varies with the distance between the colliding bodies, the individual deflection can teach us nothing. The statistical study of many deflections can, however, instruct us; and this will now be exemplified.

We will assume, tentatively at first, what the experiments will confirm: to wit, that the "other body" in every such collision is the positively-charged nucleus of the atom in question, and that the force between it and the alpha-particle (the latter being itself, we remember, the nucleus of a helium atom!) is solely the electrostatic force due to their positive charges. We will even go temporarily beyond what the experiments will confirm, and assume that the nuclei are infinitely small, mere geometrical points invested with charge and mass. I use ordinary mechanics, for the quantal treatment (which anyhow lies beyond the scope of this book) was not achieved for nearly a score of years after the classical method had guided Rutherford to the great discovery. Quantal theory arrives at the same result, except in one case which will be mentioned at the end of this chapter. This is a happy coincidence between the old mechanics and the new; for if they had not coincided, the data of 1911 might have agreed with the as-yet-nonexistent

quantal theory and disagreed with the only theory then extant or possible, and in that case the nuclear atom-model might have been stillborn.

Orbit of an alpha-particle passing near to a massive nucleus

Imagine a moving alpha-particle of charge $+2e$ and mass M, and a nucleus of charge $+Ze$ and a mass so much greater than M that we may regard it as infinite. Strictly, of course, we should always assign to the latter the mass which we know it possesses, from observations on positive ions, but most of the experiments relevant to the problem were performed with elements having atoms several times more massive than the alpha-particles, and the data are not accurate enough to discriminate between the consequences of assigning the proper mass to the nucleus and the consequences of supposing it to be infinitely massive and therefore stationary. Therefore we proceed to determine the orbit of a particle of charge $2e$ and mass M in the field of a stationary particle of charge Ze.

This orbit is a branch of a hyperbola (say SBR in Fig. 5.1); the nucleus is at that focus F_2 which is not enclosed by the branch in question. The initial and final velocities of the alpha-particle are directed along the two asymptotes ROV and TOS, so that the angle VOS which is designated by ϕ is the measure of the deflection suffered by the particle. Had the particle continued undeflected along its original path, its least distance from the nucleus would have been F_2H. This distance is designated by p and called the *impact parameter*. It is obvious that if the field of the nucleus is the same in all directions, the deflection ϕ can depend only upon p and upon the initial speed U of the alpha-particle. The theory culminates in a relation between ϕ, p and U. We proceed to prove the foregoing statements and to determine the relation.

The most familiar form of the equation of the hyperbola is

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \qquad (5.1)$$

being referred to a coordinate-system of which the origin is at

the center of the hyperbola, while the x and y axes are oriented as is shown in Fig. 5.1. The symbol a stands for the distance OA, the symbol b is defined by the equation

$$\frac{b}{a} = \cot\left(\tfrac{1}{2}\phi\right), \tag{5.2}$$

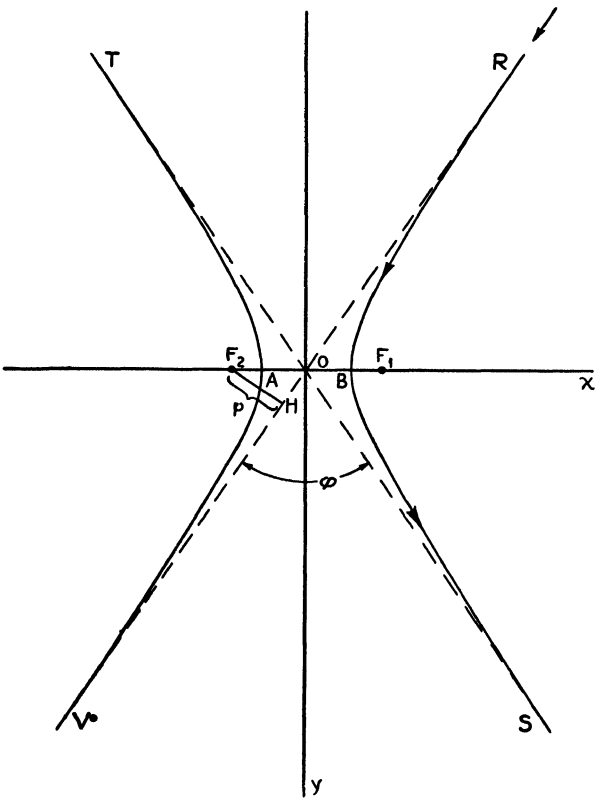


FIGURE 5.1—HYPERBOLIC ORBIT.

in terms of the angle ϕ between the the asymptotes. An additional symbol ϵ is defined by the equation

$$\cot^2\left(\tfrac{1}{2}\phi\right) = \epsilon^2 - 1. \tag{5.3}$$

The " foci " F_1 and F_2 of the hyperbola are located upon the x-axis at the points distant $\pm a\epsilon$ from the center. It is easy to show that:

$$\overline{F_2H} = p = b = a \cot (\tfrac{1}{2}\phi) \tag{5.4}$$

so that the problem becomes that of determining the relation between a, ϕ and U.

Transforming into polar coordinates (r, θ) with the origin at the focus F_2 and the direction $\theta = 0$ pointing along the x-axis, by means of the substitutions

$$x + a\epsilon = r \cos \theta; \quad y = r \sin \theta, \tag{5.5}$$

we arrive at the equation for the branch SBR of the hyperbola in the form in which we shall use it

$$r = \frac{a(\epsilon^2 - 1)}{\epsilon \cos \theta - 1}, \tag{5.6}$$

and at the derivative thereof

$$\left(\frac{dr}{d\theta}\right)^2 = \frac{r^4}{a^2(\epsilon^2-1)^2}\epsilon^2 \sin^2 \theta = \frac{r^4}{a^2(\epsilon^2-1)} - \frac{2r^3}{a(\epsilon^2-1)} - r^2 \tag{5.7}$$

All this is geometry. We must now prove that the path actually described by a particle of charge $2e$ and mass M in the electrostatic field surrounding a charge Ze stationary at F_2 is precisely the curve SBR characterized by (5.6) and (5.7). Now the path actually described by the moving particle is such that the energy and the angular momentum of the system remain constant throughout.

Constancy of the angular momentum $Mr^2 d\theta/dt$ is expressed by writing

$$Mr^2 \frac{d\theta}{dt} = P = \text{constant}. \tag{5.8}$$

Constancy of the energy of the system is expressed [1] by

[1] Or by adding K and V, differentiating the sum of them with respect to time, equating the derivative to zero and introducing (5.8), one obtains the well-known equation *force = mass × acceleration*, which to some readers may seem a better starting-point.

equating the sum of the kinetic energy K and the potential energy V of the system to a constant C. For the kinetic energy we have

$$K = \tfrac{1}{2}M\left[\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\theta}{dt}\right)^2\right]$$

$$= \tfrac{1}{2}Mr^2\left(\frac{d\theta}{dt}\right)^2\left[1 + r^{-2}\left(\frac{dr}{d\theta}\right)^2\right]. \qquad (5.9)$$

For the potential energy we have

$$V = 2e \cdot \frac{Ze}{r} = \frac{2Ze^2}{r}, \qquad (5.10)$$

an arbitrary additive constant being fixed by putting $V = 0$ when the moving particle and the nucleus are infinitely far apart. Combining (5.8) with (5.9) and forming the expression for the total energy, we have

$$K + V = \frac{P^2}{2Mr^2}\left[1 + r^{-2}\left(\frac{dr}{d\theta}\right)^2\right] + \frac{2Ze^2}{r} = C. \qquad (5.11)$$

Solving this equation for $\left(\frac{dr}{d\theta}\right)^2$ we obtain

$$\left(\frac{dr}{d\theta}\right)^2 = \frac{2MC}{P^2}r^4 - \frac{4Ze^2M}{P^2}r^3 - r^2. \qquad (5.12)$$

Comparing this expression with (5.7) we perceive that they are identical in form. Thus it is proved that the orbit of the moving particle is a branch of a hyperbola, the nucleus being in that focus which the branch in question does not enclose. By adjusting the coefficients in (5.12) to the corresponding coefficients in (5.7) we learn which hyperbola it is. We have

$$a^2(\epsilon^2 - 1) = \frac{P^2}{2MC}, \quad a(\epsilon^2 - 1) = \frac{P^2}{2Ze^2M}. \qquad (5.13)$$

The two constants a and ϵ of the hyperbola are expressed in terms of the angular momentum and the energy of the system; and the determination of the orbit is complete.

Now when the moving particle is infinitely far away from

the nucleus, and approaching along the line ROV with the speed U, the angular momentum and the energy of the system have respectively the values

$$MUp = P, \quad \tfrac{1}{2}MU^2 = C, \tag{5.14}$$

which they retain throughout the orbit and which are to be substituted into (5.13). Remembering (5.4), we find from these

$$\cot(\tfrac{1}{2}\phi) = \left(\frac{MU^2}{2Ze^2}\right)p, \tag{5.15}$$

equivalent to

$$\tan\phi = 2\frac{\dfrac{MU^2p}{2Ze^2}}{\left(\dfrac{MU^2p}{2Ze^2}\right)^2 - 1}, \tag{5.16}$$

which is the relation between deflection ϕ, impact parameter p and speed U assigned as the goal of our theory. Here also we may introduce the expression for the closest approach of the particle to the nucleus, the measure of which is the distance F_2B in Figure 5.1, which is equal to

$$F_2O + OB = a\epsilon + a = \frac{MU^2}{2Ze^2}\left(1 + \sqrt{1 + \frac{pMU^2}{2Ze^2}}\right)$$
$$= p^2\cot\tfrac{1}{2}\phi\{1 + \sqrt{1 + \cot\tfrac{1}{2}\phi}\}. \tag{5.17}$$

The reader may develop for himself the formulae for other cases actually or possibly of physical interest; for instance, the cases in which M_0 the mass of the nucleus should not be treated as though infinite, in dealing with which cases one must refer the motions of alpha-particle and nucleus to a coordinate-system moving with the center of mass of the two, and will arrive at the formula

$$\tan\phi = \frac{\left(\dfrac{M_0}{M}\right)\sin 2k}{1 + \left(\dfrac{M_0}{M}\right)\cos 2k}, \left(k = \arctan\frac{(M_0+M)2Ze^2}{M_0MU^2p}\right). \tag{5.18}$$

In dealing with deflections of electrons, the potential energy of the system must be placed equal to $-Ze^2/r$, and it is necessary to replace (5.6) by the expression for the branch of the hyperbola enclosing the focus, which is done by replacing $(\epsilon \cos\theta - 1)$ by $(\epsilon \cos\theta + 1)$ in the expression for r.

Equations such as (5.17) and (5.18) differ from those earlier (Chapter III) laid down for encounters between alpha-particles and atoms, in that they involve a quantity which is not accessible to measurement. Equation (5.17), for instance, is proposed as a relation between the observable deflection ϕ of the alpha-particle, the observable initial speed U of the alpha-particle, and the impact parameter p which cannot be observed at all. Such a situation occurs in most attempts to imitate phenomena by atom-models. If the quantity p could be observed, the postulated atom would not be a model, it would be a fact of experience. For any observed deflection one can calculate the corresponding value of p by (5.17); but this is not a test of the theory. The individual deflection teaches nothing; but through a statistical study of many deflections it is possible to test the atom-model.

The principle underlying the statistical study may be grasped by imagining a lamina of metal or a layer of gas so thin, that an alpha-particle projected into it would be far more likely to suffer either no deflection at all or else a deflection by a single nucleus, than to suffer two or more deflections by different nuclei, before it leaves the metal or the gas. (This statement is rather vague, we shall reconsider it at another place; for the time being it will suffice.) Consider such a lamina, containing Q atoms per unit area and suppose that N alpha-particles are projected vertically against it. Imagine drawn around each nucleus a sphere of radius p_0; the traces of these spheres upon the plane of the lamina are circles, each enclosing the area πp_0^2. The particles deflected through angles greater than the angle ϕ_0, calculated by substituting p_0 for p in (5.15), are those which are initially moving in lines intersecting the plane of the lamina within one or another of the circles. The number of these particles deflected through more than ϕ_0

will stand to N in the same ratio as the total area of the Q circles stands to unit area; that is to say

Number deflected through ϕ_0 or more

$$= NQ\pi p_0^2 = NQ\pi \left(\frac{2Ze^2}{MU^2}\right)^2 \cot^2 (\tfrac{1}{2}\phi_0). \qquad (5.19)$$

Thus, by counting the total number of alpha-particles directed against a lamina, and those out of this number which are deflected through more than a preassigned angle, it should be possible to determine Z. The actual method is based upon this principle. Differentiating (5.19) we find, for the number of particles deflected through angles greater than ϕ and less than $\phi + d\phi$,

$$dN = NQ \cdot 2\pi p \cdot dp = NQ \cdot 2\pi p \left(\frac{dp}{d\phi}\right) d\phi = N \cdot F(\phi) d\phi, \qquad (5.20)$$

in which by definition

$$F(\phi) = Q \cdot 2\pi p \left(\frac{dp}{d\phi}\right) = Q\pi \left(\frac{dp^2}{d\phi}\right). \qquad (5.21)$$

The function $F(\phi)$ is the *distribution-in-angle* of the deflected or " scattered " alpha-particles. Any atom-model which entails a functional relation between p and ϕ entails also a corresponding form for the function $F(\phi)$; and this function is the observable thing whereby the atom-model can be tested. The model which we have been considering, entailing as it does the relation (5.16) between p and ϕ, entails for $F(\phi)$ the following form:

$$F(\phi) = Q\pi \left(\frac{2Ze^2}{MU^2}\right)^2 \cot (\tfrac{1}{2}\phi) \operatorname{cosec}^2 (\tfrac{1}{2}\phi). \qquad (5.22)$$

Determination of the distribution-in-angle of deflected alpha-particles [1]

The distribution-in-angle of deflected or " scattered " alpha-particles is studied either by projecting a stream of

[1] Bibliography for this section: E. Rutherford & J. Chadwick, **PM 50**, pp. 889–913 (1925) and references there cited; P. Blackett, **PRS 102A**, pp. 1–17 (1922); P Auger & F. Perrin, **CR 175**, pp. 340–343 (1922).

particles normally against a piece of thin metal foil, and counting the scintillations which appear upon a small fluorescent screen which is located successively at various places along a circular arc centered at the scattering foil; or by photographing (from two directions simultaneously) a very great number of alpha-particle tracks rendered visible in a gas by the expansion-method, and measuring the angles of all the deflections which occur.

To test whether the distribution-function $F(\phi)$ actually has the form given in (5.22), it is sufficient to make counts of particles scattered within each of several restricted ranges of angle, due allowance being made for whatever change occurs in the amount of the primary stream of alpha-particles during the somewhat long-drawn-out measurements. E. Rutherford in 1911 was the first to do this, and his work was continued in 1913 by H. Geiger and E. Marsden in a series of experiments well meriting the name of "classical." For instance, they determined $F(\phi)$ for alpha-particles of a particular speed scattered by gold foil, locating the fluorescent screen successively at closely adjacent positions over the angular range from $\phi = 5°$ to $\phi = 150°$; the arrangement is diagrammed in Figure 5.2. The number of alpha-particles falling upon the screen when the line from the scattering-foil to its center is inclined at ϕ to the primary stream is proportional to $F(\phi)/\sin\phi$; consequently to $\mathrm{cosec}^4(\frac{1}{2}\phi)$ provided that $F(\phi)$ truly conforms to (5.22). As a matter of fact, Geiger and Marsden found that the quotient of the number of scintillations by the function $\mathrm{cosec}^4(\frac{1}{2}\phi)$ fluctuated within a range not

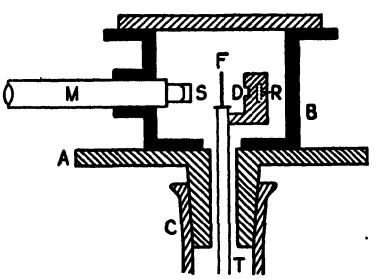


FIGURE 5.2—APPARATUS OF GEIGER AND MARSDEN. SOURCE OF ALPHA-PARTICLES FIXED AT *R*, SCATTERING FOIL FIXED AT *F*; SCREEN *S* AT END OF TELESCOPE *M* CAN BE REVOLVED AROUND AXIS *T* CARRYING *F*. (After their drawing).

exceeding 30 per cent of the extreme value as ϕ was increased from 5° to 150°, while both the numerator and the denominator of this quotient were steadily decreasing to four millionths of their initial values. This is apparently the most thorough investigation of $F(\phi)$ that has yet been published. The distribution-in-angle for argon and for air was subsequently shown to agree with (5.22) by P. M. S. Blackett, who used the expansion method. In another laboratory P. Auger and F. Perrin arrived for argon at the same result, but with the qualification " the scanty measurements made do not permit precise verification."

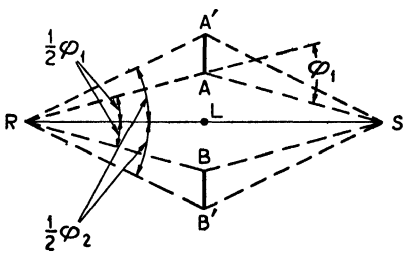


FIGURE 5.3—ILLUSTRATING CHADWICK'S METHOD OF DETERMINING NUCLEAR CHARGES.

To measure the nuclear charge Ze it is necessary to be convinced that $F(\phi)$ has the form expressed in (5.22), and to know the values of N and Q.

The number Q of atoms in unit area of the scattering foil of metal or stratum of gas is determined from the weight of the metal or gas and from the mass of the individual atom, hence eventually from the value of N_0 (Chapter I).[1]

The symbol N stands, when the scintillation-method is used, for the number of alpha-particles falling per unit time upon the

[1] Thus the uncertainty of N_0 limits the accuracy attainable for Z. J. Chadwick, however, states that the actual limit of accuracy is set by the "probability" fluctuations in the rate of emission of alpha-particles from the source, which cause an uncertainty in Z exceeding 1 per cent even when the observer has counted 4000 scattered particles and 8000 in the primary stream; this uncertainty could be lowered only by still more protracted countings.

foil. This must be determined indirectly,[1] or else by bringing the fluorescent screen into the path of the primary stream behind a revolving disc opaque except for a narrow slit, so that only a known fraction of the stream small enough to be countable arrives at the screen. The uncertainties of the indirect methods prevented Geiger and Marsden from determining Z for gold with accuracy better than 20 per cent. With the aforesaid direct method J. Chadwick[2] made measurements upon platinum, silver and copper; he did not, however, count particles scattered through several ranges of angle so as to determine the form of $F(\phi)$, but assumed that for these metals as for gold it would conform to (5.22), and with this assumption calculated Z from counts of the particles scattered within a certain definite range of angles. These values of Z rank as the best-established of any, and are placed first in Table 5.1. Subsequently Rutherford with Chadwick counted the particles scattered by gold within the ranges of angle from streams of two distinct speeds, and fixed Z for gold within 2 per cent of 79.

When the expansion-method is used, N signifies simply the total number of tracks photographed and examined for deflections. By this method the values cited for argon and "air" were obtained. The value for air is due to C. T. R. Wilson, who used very fast electrons instead of alpha-particles, and determined the proportion scattered through angles exceeding 90°, assuming without test that the distribution-function conforms to (5.22) except for the replacement of the alpha-particle

[1] By independent knowledge of the rate at which the source is emitting alpha-particles, or by waiting until this rate has diminished (through decay of the source) by a known large fraction and then counting the particles in the much-weakened primary stream.

[2] Chadwick's arrangement is diagrammed in Fig. 5.3. The alpha-particles striking the screen at S are those which leave the source at R in directions leading toward a wide flat ring of metal foil normal to the plane of the paper and cutting it at A and B, and which are deflected in the foil through angles equal to twice that between their original direction and the line RS. It is obvious that the angle of deflection ϕ suffered by an alpha-particle reaching S depends on the distance of the point where it struck the ring from the inner edge thereof, so that the entire range of values $\phi_1 < \phi < \phi_2$ is represented among the counted particles.

charge $2e$ by the electron charge e. Naturally it should be considered as a mean between those for nitrogen and for oxygen, strongly weighted in favor of the former. The value for argon was obtained by Auger and Perrin in the course of their work already cited.

The assertions made at the last three entries of the table are based on the assumption that $Z = 79$ for gold; for the procedure in making these estimates amounted in principle to counting the particles scattered through angles within a given range by gold-foil, using this datum to determine N, and using this value of N to interpret counts of particles scattered under precisely the same circumstances by foils of the other metals.

TABLE 5.1

Element (Atomic Number)	Nuclear Charge (Authority)
Platinum (78)	77.4 ± 1 } Chadwick
Silver (47)	46.3 ± 0.7 } Chadwick
Copper (29)	29.3 ± 0.5 } Chadwick
Gold (79)	79 ± 2 (Rutherford & Chadwick)
Argon (18)	19 (Auger & Perrin)
Air (7, 8)	6.5 (C. T. R. Wilson)
Pt; Ag; Cu; Al (13); Sn (50)	Within 5% of atomic number if $Z = 79$ for gold (Geiger and Marsden)
Aluminium (13)	13.0 ± 0.5 (if $Z = 79$ for gold) (E. S. Bieler)
Magnesium (12)	12.6 ± 1.0 (if $Z = 79$ for gold) (E. S. Bieler)

All of the data, it is seen, just fall short of being precise enough to prove that the atomic number is the only integer compatible with the measured value of Z. Of course they fall far short of proving that Z is integral or nearly so. Additional data would be welcome; but the tedium of these experiments is evidently so great, and the results are so generally regarded as being known in advance, that a request for more would probably be thought unreasonable.

Next it is necessary to consider the influence of the speed of the alpha-particles upon their scattering. A glance at (5.22) shows that the theoretical distribution-in-angle function there written is a function of this speed U as well as of ϕ; it is, in fact, a function of the type

$$F(\phi,U) = f(\phi)\cdot g(U), \qquad (5.23)$$

in which $g(U)$ is specifically of the form

$$g(U) = \text{const} \cdot U^{-4}. \qquad (5.24)$$

As C. G. Darwin has shown, the distribution-function $F(\phi,U)$ would still be of the type (5.23) if we were to assume that the field around the nucleus of the atom-model varies as an arbitrary power $-n$ of the distance; but in this general case we should have not (5.24), but

$$g(U) = (U^{-2})^{\frac{2}{(n-1)}}. \qquad (5.25)$$

By measuring the number of alpha-particles deflected through angles lying within a given fixed range, for instance the number of scintillations appearing upon a screen which is kept in a fixed position, while the speed of the primary particles is varied through a sufficient range, one can determine the form of $g(U)$ without needing to know anything about $F(\phi)$. (However, the procedure is not justified unless one has reason to believe that $F(\phi, U)$ is truly separable into factors depending each upon only one of the variables.) Wide variations of the value of U are impracticable, as Nature furnishes no alpha-particles exceeding 2.10^9 cm./sec. in speed,[1] and when these are slowed down to about 10^8 cm./sec. they cease to function as alpha-particles. Nevertheless, Geiger and Marsden found with gold that when U was varied over such a range that U^4 was diminished in the ratio 1 : 9.22, the product of U^4 by the number of scintillations upon the fixed screen fluctuated only within 15 per cent of its mean value; and Chadwick found with platinum that when U^4 was diminished in the ratio 2.3 : 10, the data when compared with (5.25) were incompatible with any values of n but those lying between 1.97 and 2.03.

Similar results were obtained by Blackett and by Auger and Perrin with argon. In such experiments as these, in which the expansion-method is used, the speed of the particle dimin-

[1] The occasional particles ejected from certain radioactive elements with speeds higher than this (cf. L. F. Bates and J. S. Rogers, **PRS A105**; 1924) are too few to be used in these experiments; but the cyclotron (Chapter XVI) is capable of enlarging the range of available U-values.

ishes continually as it approaches the sharply-definite end of its visible track. The law of dependence of the speed U upon the distance l from the end of the track is known by separate measurements. Imagine each of the tracks photographed in the course of the experiments to be divided into small intervals of length dl; consider all of the intervals which are at a common distance l from the ends of their tracks; count all of the deflections which occur in these intervals, and which lie between preassigned limiting values ϕ_1 and ϕ_2; designate the number thus counted by $N(\phi_1, \phi_2, l)dl$, and determine $N(\phi_1, \phi_2, l)$ for as many values of l as possible. This function of l, when translated into a function of U by using the empirical relation between U and l, is proportional to $g(U)$. Auger and Perrin apparently found by this procedure that $g(U)$ is of the form $const. \cdot U^{-4}$. Blackett reversed the procedure, and determined the relation between U and l by assuming that $g(U) = const. \times U^{-4}$.

Continuing to speak of the atom as though it were the nuclear atom-model, we may now say that the evidence shows that the field surrounding the nucleus is an inverse-square field. This statement is subject to limitations; to understand these, we must consider the distance of closest approach of the alpha-particle to the nucleus, which corresponds to " perihelion " in astronomy, and is called the " apsidal distance." This distance is given by (5.17) as function of Z, U, and p, hence as function of Z, U, and ϕ; hence we can calculate it for any observed deflection provided we know U and accept a value for Z and are convinced that the distribution-in-angle conforms to (5.22). Now with gold it has been shown that $F(\phi, U)$ conforms to (5.22), over a certain range of speeds for a particular angular range, and over a certain range of angles—viz., from 5° to 150°—for a particular speed. For this speed, and accepting $Z = 79$, it is calculated from (5.17) that the apsidal distance is 3.6X (that is, $3.6 \cdot 10^{-11}$ cm) for the particles deflected through 5° and 0.31X for the particles deflected through 150°. We infer that the field of the nucleus is an inverse-square field at least at distances from 0.3X to 3.6X outward from the center. The outer boundary of this region

of "Coulomb field" must in fact be farther out than 3.6X, since the curvature of the path of a particle is notable out as far as several times the apsidal distance. The estimate just given for the outward extension of the Coulomb-field region is therefore not excessive. Blackett's study of argon led similarly to 0.74X and 54X as conservative estimates of the inner and outer radii of the inverse-square-field region. Somewhere beyond must lie the principal habitat of the orbital electrons, but as to these the alpha-particle can tell us nothing further.

The purpose of this chapter—to recite the earliest if not the greatest single piece of evidence for the nuclear atom-model—is now fulfilled. We must not, however, presume the nucleus to be a mathematical point! Somewhere there must begin a region, within which the situation is very unlike that of empty space pervaded by a Coulomb field. One might hope to intrude upon it with very fast alpha-particles; the particles which would traverse it (if any did) would be the fastest and the most-deflected; with particles of predetermined speed, the best hope of success would probably be offered by the nuclei of smallest charge. Rutherford, in fact, observed that the formula (5.22) for $F(\phi, U)$ is not an adequate description of the data when alpha-particles are scattered by protons (here it was the distribution of the recoiling protons which was studied); and so did Bieler for his data of the scattering by aluminium and magnesium. Bieler interpreted his data as signifying an inverse-fourth-power attraction superposed upon the inverse-square repulsion, balancing it at 0.0344X from the nucleus and fading into relative insignificance only two or three times farther out. I mention this only because it shows how far inward one must (in imagination) go, in order to find the inward limit of the Coulomb-field region; for nowadays it is the custom to interpret such data by invoking the quantal theory of scattering, and also to supplement them by invoking phenomena of transmutation.

The quantal theory of scattering lies beyond the scope of this book, but I will quote one of its consequences, which is

not only remarkable but finely verified.[1] Though in general the classical and the quantal theory lead for an inverse-square field to precisely the same result—the distribution-function $F(\phi)$ of (5.22)—there is one case, or set of cases rather, where they do not: the cases in which the scattered and the scattering particles are identical. One may perhaps understand this a

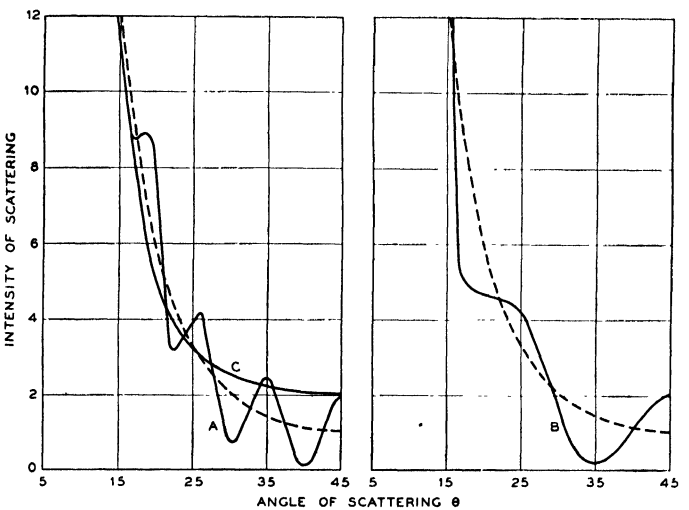


FIGURE 5.4—THE FUNCTION $F(\phi)$ BY CLASSICAL THEORY (DASHED CURVE) AND QUANTAL THEORY (CURVES *a*, *b*, *c* FOR $U = 1, 2, 8$ TIMES 10^8 CM/SEC RESPECTIVELY. (Blackett and Champion, *Proceedings of the Royal Society*.)

little by realizing that in quantal theory, one imagines particles as being guided by waves, and applies the mathematical treatment to the waves; and that when the scattering and the scattered particles are identical their waves are capable of mutual interference, whereas otherwise they are not. When the ratio

[1] Theory: Mott, **PRS 126**, 259 (1930). Experiments: on scattering of alpha-particles by alpha-particles, Chadwick, **PRS 128**, 114 (1930); Blackett and Champion, **PRS 130**, 380 (1931). Scattering of protons by protons: Gerthsen **AP 9**, 769 (1931); in subsequent work (Tuve, Heydenburg and Hafstad, **PR 50**, 806 (1936), **51** 1023 (1937)) the penetration of the protons into the neighborhood of the nucleus is such that $F(\phi)$ is greatly altered by the departure of the field from inverse-square character.

of the quantal $F(\phi)$ to the classical $F(\phi)$ is plotted against ϕ, it is seen to vary up and down, passing through maxima greater than unity and minima less than unity. The number and closeness of the maxima and minima increase as the speed U of the to-be-scattered particles diminishes, there being a good deal of variation even over the relatively small U-range of available alpha-particles.[1] The quantal function $F(\phi)$ depends on the " spin " of the particle, a quality to be described in Chapter XIII. Mott has worked out the formula for two cases: that of " spinless " nuclei, of which alpha-particles afford an instance; and that of particles " of spin $\frac{1}{2}$ " of which electrons and protons are examples. A detail is, that at 45° the ratio of " quantal " to " classical " scattering is 2 : 1 in the former and 1 : 2 in the latter case, whatever the speed.

The quantal theory is tested by projecting alpha-particles into helium. It is desirable to use the slowest alpha-particles of which scattering can be observed, partly because the difference between the quantal and the classical formulae is then at its greatest and partly because there is then the least danger of $F(\phi)$ being distorted because of intrusion of the particles into the regions close to the nuclei where the inverse-square law is not valid. Blackett and Champion (after a previous test by Chadwick) accordingly used the expansion-chamber in the manner just mentioned, so that their data included U-values ranging from 8.25 down to 1.15 times 10^8 cm/second. With these they came to a notable confirmation of the quantal formula over the classical.

Having now been forced to take account of undulations even in regarding an effect so conspicuously corpuscular as that of scattering, we may no longer delay in admitting waves and particles to parity in our cosmos.

[1] As U decreases the maxima and minima of the ratio become maxima and minima of $F(\phi)$ itself, and finally the distribution-in-angle becomes a familiar diffraction-pattern, that of waves of wavelength λ diffracted by small obstacles (or apertures) of diameter D; here λ stands for the wavelength associated with the particles by the Rules of Correlation of Chapter VI, and D for one-half the distance of closest approach of a "scattered" and a "scattering" particle when one flies straight toward the other.

CHAPTER VI

Correlation of Waves and Corpuscles

In turning from the corpuscular to the wavelike aspect of Nature, we are turning away from the easier part of our task—easier, at any rate, for the heirs of the Anglo-American tradition (J. G. Crowther has wittily suggested that this is because the *ball* has been for ages the favorite plaything of the Anglo-Saxon races). In attacking the harder part, it seems well to begin by analyzing the long-established undulatory theories such as those of water-waves, sound, and light.

The very name "wave-theory" calls up a picture of a procession of water-waves over the surface of a pool, crest following on trough and trough on crest in an indefinite sequence. Of all the features of such a procession, the most striking is this regular alternation of valleys and hills; and the next most striking is the constant onward gliding of the sequence. These constitute "wave motion" to the eye. A little observation shows that the water does not glide onward with the waves—a floating leaf will rise and fall as the waves pass under it, and have no tendency to drift with them. The onward motion of the waves is motion of form, not of substance. The substance (the water) has a different motion of its own: hydrodynamical theory describes how it moves, and is confirmed by trial. This is not the simplest of cases, and it is better to think of transverse waves on a very long tense cord. These travel along the cord, a succession of valleys and hills like ripples on water; but every short length or "element" of the string is moving at sensibly a right angle to that onward march of waves.

Let the x-axis lie along the line with which the string coincides when not vibrating, and let the vibrations occur in the

xz-plane. For the displacement ζ of any element of the string, we have:

$$\begin{aligned} \zeta &= A \cos (nt - mx) \\ &= A \cos 2\pi(\nu t - \mu x) \\ &= A \cos 2\pi(t/T - x/\lambda) \\ &= A \cos (2\pi/\lambda)(ut - x), \end{aligned} \tag{6.1}$$

wherein the names and meanings of the constants are as follows:

λ, the wave-length or crest-to-crest distance;
μ, the wave-number;
ν, the frequency;
T, the period;
u, the speed with which the waves glide along, hereinafter to be called the wave-speed or phase-speed;
A, the amplitude.

Strictly there should be an additive constant inside each of the brackets ($\zeta = A \cos (nt - mx + c)$, etc.) but we may as well imagine the zero of time so chosen as to annul it.[1]

Here we meet the *sine-function* (or the cosine-function, or the exponential function with imaginary exponent—it matters little which we call it). The ubiquity of this function is one of the major features of physics, and not alone of those branches of physics which we are inclined to call contemporary or new, but of the whole of the science from the middle of the eighteenth century onwards. This will become apparent, I think, as the book develops; and it will also become apparent that wherever that function occurs it betokens a differential equation dominant in the background, a differential equation belonging to that protean family known as the " wave-equations." For the moment let me simply state with emphasis, that whenever in physics we speak *with definiteness* of frequency, of period, or

[1] The word "phase" is sometimes applied to this additive constant and sometimes to the argument of the cosine-function, an ambiguity which renders it a dangerous word to use.

of wave-speed (or of how many other things later to be mentioned!) we are tacitly or explicitly assuming a sinusoidal wavetrain.

Instead of a string running along the x-axis, imagine an elastic solid pervading the whole of space, or at least a region of which the linear dimensions are very large compared with λ; and suppose that the displacement of every particle thereof is parallel to the axis of z, and conforms to (6.1). This is the old-fashioned wave-model of a beam of plane-parallel plane-polarized monochromatic light proceeding in the x-direction. To obtain the new-fashioned wave-model for the same thing, we write E_z for ζ, and say that it stands for electric field-strength; the equation otherwise remains the same. (There is a similar and simultaneous equation for H_y, the magnetic fieldstrength directed along the y-axis; but for the time being we need not think of it.)

It is rather strange to realize how little of this elaborate wave-model does actually serve to represent observable features of light. In a beam of light, we never see anything in vibration, and therefore we neither perceive nor measure the frequency ν. We do not see any waves, hills, valleys, crests nor troughs, and therefore we do not perceive nor measure (in any strict sense of the verb " to measure ") the wavelength. We do not actually perceive nor measure the wave speed or phase speed, though we do measure something which happens to be equal to it *in vacuo* but not in material media. Apart from polarization (a very important quality, of which we shall have too little space to speak), and momentum (to be discussed further on) there are only two qualities of light which we can measure. One of these is interpreted as being proportional to A^2, while the other is *a speed which may or may not be* (!) *the same as* u. The former is the intensity of the light, which in a truly plane-parallel monochromatic beam would be the same everywhere and always. Thus, of all the quantities appearing in (6.1) which is the description of the wave-model of a plane-parallel monochromatic beam of light, only the coefficient A has an immediate meaning in terms of what is observed, and *it* is

constant both in time and space. It would not be worth while to invent a wave-model for such a case as this.

The model begins to show its worth when we superpose a pair of similar beams not coinciding in direction. Say that we have two such wave-trains as (6.1) describes, travelling, however, not along the x-direction but along two directions in the xy-plane inclined respectively at angles θ and $-\theta$ to the x-direction, and thus intersecting one another at angle 2θ. Adding the displacements, we get:

$$\begin{aligned}\zeta_1 + \zeta_2 &= A \cos (nt - mx \cos \theta - my \sin \theta) \\ &\quad + A \cos (nt - mx \cos \theta + my \sin \theta) \\ &= 2A \cos (my \sin \theta) \cos (nt - mx \cos \theta)\end{aligned} \tag{6.2}$$

Here is a wave-motion of which the amplitude varies from place to place, being in fact a sinusoidal function of y, the square of its amplitude being given by the equation,

$$(\text{Amplitude})^2 = 4A^2 \cos^2 [(2\pi \sin \theta/\lambda)y]. \tag{6.3}$$

If upon sending two identical beams of plane-parallel light across one another's paths at inclination 2θ we should observe that the intensity of the light varies with place according to (6.3), the wave-model would be justified.

The aforesaid is a typical experiment on the "interference of light." The intensity of the light is found to vary as $\cos^2 (ky)$; here k stands for a quantity which is ascertained by measuring (for instance) the distance between two successive planes of zero light-intensity (the so-called "width of a fringe") and which is indeed found to be proportional to $\sin \theta$ for beams of a given color, and to depend on the color of the beam for a given value of θ. When k is determined and θ is measured, the wavelength λ can be computed: that is to say, we can assign the value which the crest-to-crest distances in the wave-model must have, in order to produce an interference-pattern of the scale which is observed.

Such is the standard and ultimate way of ascertaining wavelengths, by the Michelson interferometer for instance. It

may seem quite unlike the method of the ruled grating mentioned in Chapter I, and quite unlike the method of the crystal lattice mentioned there and to be more fully expounded in Chapter VII; but fundamentally they are all alike—all of them consist in causing beams of light to overlap, producing thus a distribution-in-intensity or " pattern " which fits some particular value of λ, and ascertaining that value. To interpret the pattern produced by grating or lattice, we have to think of wavelets continually expanding from the lines of the former or the unit-cells of the latter; but in interpreting the pattern produced when plane-parallel beams intersect, we can circumvent the wavelets and arrive at the result by the simple algebra of (6.2), though a full analysis will show that the wavelets are still there.[1] Always the so-called " wavelength " is evaluated by making observations on a distribution-of-intensity produced by overlapping beams, and called an interference-pattern or a diffraction-pattern.

Next we examine one of the subtlest features of the wave-model: the meaning of the velocity, or rather the meanings of the velocities, which are inherent in it.

Speeds in undulatory theory

The constant u of (6.1), called the " phase-speed " or " wave-speed," is the speed at which the wave-crests of the undulatory model glide onward. But we never perceive the wave-crests, and so cannot measure their speed. The model described by (6.1) corresponds to an unlimited beam of uniform monochromatic light. In such a beam, there is nothing which visibly moves. To get at something of the nature of speed which is measurable, one must cut out a segment of such a beam, and measure how long it takes to go from one place to another. This is done in all the schemes for measuring the " velocity of light," the toothed-wheel method (page 44) being the most lucid.

However, the speed with which a segment of a wave-train gets from place to place is *not necessarily* the speed of the

[1] Cf. any treatise on optics.

wave-crests. Now it certainly seems foolish to affirm that the rate of progress of a chopped-off segment of a wave-train can be anything else than the speed of the waves! Careful thinking is demanded, to pierce to the origin of the distinction; but the distinction itself is (happily) a fact of experience—when a limited group of ripples is traversing the surface of a pond, one may actually see the individual crests running faster than the group, each in turn advancing to the front of the slowly moving undulatory area of water and dying out as it gets there, while others arise at the rear.

To find the cause of this behavior, one must realize that a train of waves of finite length is never a perfect sine-wave such as (6.1) describes. Even though every single wave of the limited train should be a perfect arc of a sine-curve, even though the train should be broken off with absolute sharpness at each end—even in this apparently ideal case, the train could not be called perfectly sinusoidal. To be unlimited in both directions is of the essence of a perfect sinusoidal wave. A segment of a wave-train *must* be regarded as a resultant of superposed sine-waves of the type (6.1), having various amplitudes and various frequencies, which there is a method for determining when the exact shape of the train is known—Fourier's method.

The onward gliding of a segment, therefore, is the outcome of the onward gliding of the individual infinite wave-trains of different wavelengths, into which it must be resolved. Now if all of these have the same phase-speed—if u is independent of λ—the segment will travel undistorted and u will be its speed. But if u depends on λ, the segment will travel at a rate which cannot agree with all, and may even differ greatly from all, of the values of u represented among its component trains. This is not the worst: the segment changes its form as it proceeds, and in the course of time it may spread out until it mingles with other preceding or following segments, or disperses into the indistinguishable, like a puff of smoke in a breeze or a drop of ink in water. In such a case we should not speak of its speed at all, unless we hold in mind that we are giving a name to something which is essentially somewhat

indefinite, and grows more and more indefinite the longer time goes on. This indefiniteness prevails in *every* case, save the exceptional one of a wave-model in which phase-speed is independent of wavelength; and we shall presently find it exalted as part of a principle.

It is perhaps unfortunate that the study of light and of sound gives us too good a chance to evade these subtle questions: for light *in vacuo* and sound in air (the latter with certain limitations) exemplify the simple special case in which u is independent of λ and is the speed of segments as well as of wave-crests.[1] No such simplification occurs in the cases of light in material media, waves on the surface of water, radio waves in ionized air, electrical waves along cables, and—first of all in interest, among those within the purview of this book —the case of the waves associated with electrons and with atoms.

Group-speed

It seems pretty hopeless to expect to work out any general rules for the speed of a segment, considering in the first place that it is essentially indefinite (unless u is independent of λ) and in the second place that it depends on the distance between the ends of the segment and on the abruptness or lack-of-abruptness of these ends. There is, however, a peculiar quantity, entirely determinate so soon as u is given as a function of λ, which is not precisely a segment-speed but rather a limiting value which the speed of a segment approaches as its length increases: it is called "group-speed."

Say we have a pair of infinite sinusoidal wave trains, equal in amplitude but of slightly different wave-lengths. Super-

[1] For the case of light *in vacuo*, the chief evidence for this all-important statement (that u does not vary with λ) is astronomical. If light of one color traveled faster than light of another, a luminous star emerging from behind a dark one would be seen first in the faster-travelling hue; in fact there would be a sequence of colors, the same for every emergence of every such star, and spread out over a time-interval proportional to the distance of the stars. Nothing of the sort has ever been observed, although there are plenty of luminous stars revolving around dark ones which regularly occult them.

posing them, we obtain the familiar pattern of an endless sequence of beats (Figure 6.1). The maximum or center of a beat occurs where a crest of one sine-wave coincides with a crest of the other—the minimum between beats, where crest falls together with trough. Denote the wave-lengths by λ and $\lambda + \Delta\lambda$. Supposing $\Delta\lambda$ to be small (vanishingly small, to be strict) compared with λ, one may readily see by mere inspection that a wave-length is the same fraction of the distance D between consecutive beat-maxima, as the discrepancy $\Delta\lambda$ is of the wave-length:

$$\lambda/D = \Delta\lambda/\lambda. \tag{6.4}$$

Now if the two component wave trains advance with equal speed, the beats are simply carried along with a speed equal to theirs. But if the velocities of the component waves are not the same, the velocity of the beats is not the same as either, nor their mean, but something totally different.

To see this, imagine that you are moving along with one of the sine-waves; for definiteness, that you are riding on the crest B of the train with the shorter waves (Figure 6.1). At a certain moment, say $t = 0$, it coincides with a crest A of the other sine-wave, and you are at the top of the beat. Meanwhile the other train is moving relatively to the first; for definniteness suppose that the longer waves move faster, so that relatively to the shorter they are gliding upward. After a certain time they have gained on the shorter waves by a distance equal to $\Delta\lambda$. When this time has elapsed the top of the beat is no longer where you are, but where the crest A' of the long-wave train coincides with the crest B' of the short wave. It has dropped back through the distance λ while the crest A was getting ahead by the distance $\Delta\lambda$. Therefore when the longer waves travel faster than the shorter, the beats travel more slowly than either.[1a]

[1a] This is the case with the wave models which figure in quantum mechanics, and usually also with light waves. Sometimes the opposite is the case with light waves—*i.e.*, in a region of anomalous dispersion, shorter waves travel faster than longer; but in such cases the waves are rapidly absorbed, and this alters the situation. Cf. A. Sommerfeld and L. Brillouin, Ann. d. Physik 44, 177–202, 203–240 (1914).

We now deduce the formula for the actual value of the speed of the beats. Denote by u and $u + \Delta u$ the phase speeds of the two wave trains λ and $\lambda + \Delta\lambda$; by g the speed of the beats. These are speeds conceived as relative to the medium. Relative to a crest of the former wave train, the speed of the

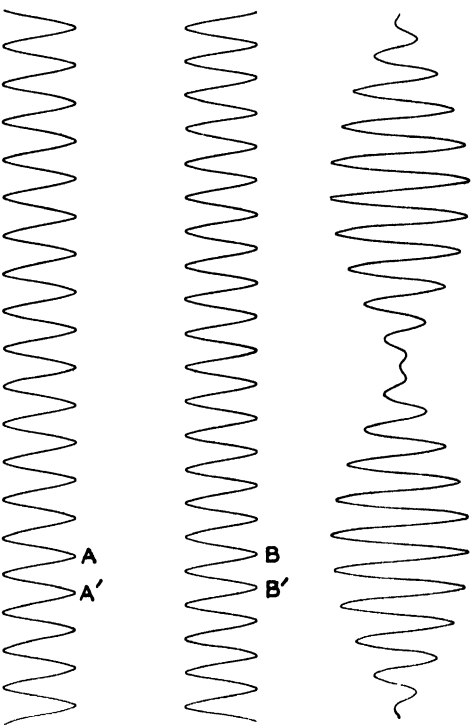


FIGURE 6.1—ILLUSTRATING BEATS AND GROUP-SPEED.

latter train is Δu, that of the beats is $(g - u)$. Relative to the former wave train, the latter moves a distance $\Delta\lambda$ while the beats are moving a distance λ in the opposite sense, therefore with a minus sign. Hence:

$$(g - u)/\Delta u = -\lambda/\Delta\lambda. \qquad (6.5)$$

Solve for g; go over to the differential notation, to signify that the conclusion is strictly valid in the limit of vanishing difference of wave-length;

$$g = u - \lambda(du/d\lambda). \tag{6.6}$$

A more elegant form of the equation is obtained by remembering that phase speed u is the product of wave-length λ and frequency ν, and acting accordingly:

$$g = -\lambda^2(d\nu/d\lambda) = d\nu/d(1/\lambda) = d\nu/d\mu. \tag{6.7}$$

This quantity g is known by the name of " group speed." We shall later find it playing an important rôle.

Beginnings of the fusion of corpuscular and undulatory theories

I have recalled that in the classical and familiar undulatory theory of light, the fundamental features—the waves themselves, their wave-length, their frequency, their phase speed—are *not observable* in any ordinary sense of this word. The values of λ and ν and u are not directly measured: they are deduced from certain data of experiment, with varying degrees of indirectness. Wave-length is the nearest of the three to actual data: though it is not observed, its value is deduced from distribution-in-intensity in overlapping beams. No experiment ever conceived can detect the phase speed, though there happens to be, in the special case of light *in vacuo*, a coincidence by virtue of which we may affirm that it is equal to the speed of a broken-off piece of a beam, which is observable. Frequency is the most remote of the three, for in practice ν is just a symbol for the quotient of phase speed by wave-length.[2] Not one of these things has ever become observable, as the atoms of electricity and matter have become observable of recent years; and it is not to be foreseen that any of them ever will.

[2] This statement, like certain others in the chapter, requires modification if sound and radio waves are taken into account. The continuity between radio waves and what is here called "light" imparts a quality to the wave theory of light which as yet finds no parallel in the undulatory models of electricity and matter.

The student perhaps is wondering why I have spent so much space in recalling what has long been known about light. The reason is this: there is now a wave theory of electricity and of matter having certain features which seem at first sight strange, which come as a surprise and a stumbling-block to many physicists, and yet they are essentially the same as the features of the theory of light which I have just been quoting, which have been known and readily accepted for half a century or more.

Consider, for example, a beam or stream of electrons all going in the same direction with the same speed—such a one as we so often envisaged in Chapter II. Now it will be suggested that upon this corpuscular picture of the beam, we superpose the image of a wave train. But this is not because some new method of observation has made us able to perceive vibrations or advancing wave crests in the stream—nothing of the sort has happened, any more than with light. What has been newly observed is a distribution-of-intensity in overlapping beams of negative electricity, resembling that which occurs in overlapping beams of light, and interpretable like the latter by a wave model, provided that we identify square-of-amplitude in the model with intensity in the actual situation. Square-of-amplitude is measured, and wave-length is deduced; but neither wave-length, nor phase speed, nor frequency is observed. The rate of progress of a chopped-off segment of the beam is measurable, but the phase speed is not to be identified with it; the coincidence which occurs in the case of light *in vacuo* does not occur in this case, but we should not be more surprised than at its failure to occur in the case of light in material media or radio waves in ionized air.

There is, I repeat, no suggestion of throwing away the corpuscular image of the beam of electricity, and installing the wave model in its place. We must adopt the wave model in order to account for the diffraction pattern of the overlapping beams; but we must keep the corpuscles because the expansion-chamber has shown their tracks to us. It is now our problem to make them exist together in our minds.

Intensity, as I have said, is represented in the wave model by square-of-amplitude. But it also has something to represent it in the particle theory: it is the total charge borne by the streaming particles, per unit time, across unit area set normally to their course. This quantity is the number of electrons per unit volume, multiplied by their speed and by the charge of each. The two last factors are constant throughout the diffraction pattern of a beam of electrons of uniform speed.[3] In dealing with light, one converts it into heat and so measures its energy, and in the particle-model of light the intensity is the number of corpuscles per-unit-volume multiplied by their speed and by the energy of each. The square of the amplitude of the waves thus corresponds to the number-per-unit-volume, the " concentration " of the particles. If we say that the concentration of the particles is everywhere proportional to the square of the amplitude of the waves, we make the first step towards union of the theories.

This step has been suggested by the diffraction-pattern which appears when the original electric beam is broken into many by a grating or a crystal. Say that instead of meeting such an obstacle, the beam continues indefinitely onward through empty field-free space. It then is represented by two images: (*a*) a train of plane-parallel sine-waves with uniform amplitude, and (*b*) a cloud of electrons dispersed at random throughout its extent and all advancing in the same direction with the same velocity. Both images together represent it; neither is complete without the other, though the former seems superfluous so long as one considers none but unimpeded beams.

[3] This passage is open to criticism for evading an important issue. If one uses the simple corpuscular picture for a pair of overlapping beams, one must imagine some corpuscles going in one direction and some in another; and in the extreme case of beams inclined to one another at 180°, the *net* flow across any surface is nil, and yet we do not say that the intensity vanishes. This belongs to the complex of difficult and delicate problems with which the "principle of indefiniteness," hereinafter mentioned, attempts to cope.

The two theories mutually adjusted, to enable both to cope with deflection and refraction

There are classical ways of deflecting a beam of electricity, without impeding it by matter. For instance, it will be bent into a parabolic arc if it traverses an electric field inclined to its original direction. This is familiar, and familiar also is the explanation by the corpuscular theory: each electron suffers a force and answers by an acceleration. If the wave model is to subsist, the wave train must be bent along with the paths of the corpuscles; we cannot have the electrons going one way and the waves another. Can we devise such an interrelation between electrons and waves, or more generally, between waves and particles? To achieve this, we shall find that we are obliged to assume a certain relation between wave-length of waves and momentum of corpuscles. This has been verified by experiment, and is one of the most important laws of nature.

To derive it, I will take a somewhat artificial case, the limit of an actual case; it has the advantages of simplicity and of a close analogy with a famous phenomenon of light. Suppose that we have two regions of empty space at different electrostatic potentials. They shall be separated by a plane, say the xy-plane; and the potential gradient from one to the other shall be so steep, that we may deem it concentrated at the plane of separation, so that there is a potential discontinuity between the regions $z > 0$ and $z < 0$. For definiteness, suppose that the latter region is at the higher potential—higher by V electrostatic units than the former—so that electrons coming from above are speeded up as they pass through the plane $z = 0$.

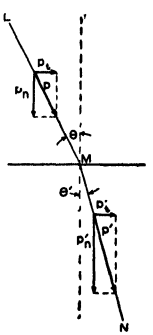


FIGURE 6.2—ILLUSTRATING REFRACTION INTERPRETED BY THE CORPUSCULAR IMAGE.

Now imagine a stream of electrons of momentum p, approaching the plane along a direction inclined at θ to the normal—LM in Figure 6.2. As an electron goes through the

plane, the magnitude of its momentum is changed to a new value p', but the tangential component p_t of the momentum remains unaltered, since there is no component of field strength parallel to the plane. Figure 6.2 is almost self-explanatory; as one readily sees,

$$\sin\theta/\sin\theta' = p'/p. \tag{6.8}$$

Now if we vary θ while leaving p the same, the right-hand member of this equation remains unchanged; for the momentum of an electron is fixed when its kinetic energy is fixed, and the kinetic energy of these electrons increases by the constant amount eV when they go through the plane, whatever their direction of approach from above. We have therefore:

$$\sin\theta/\sin\theta' = p'/p = \text{quantity independent of } \theta. \tag{6.9}$$

But this is the law of optical refraction, which, as every physicist remembers, is compatible with the wave theory of light. We are finding what has been known since the days of Newton and Huyghens, that the law of optical refraction is compatible both with the wave theory and with the corpuscular theory. We shall, therefore, be able so to contrive our wave model, that the waves will bend through the same angle as the electrons when they pass the plane, and follow the same direction as the electrons thereafter. However, we shall not be able to make them go at the same speed as the electrons, as will now appear.

Figure 6.3 represents the classical Huyghens construction for the passage of a wave front through a refracting surface, and is almost self-explanatory. The wave front is perpendicular to the plane of the paper; AA' is its trace at a certain instant, say $t = 0$; BB' is the trace, at a later instant $t = T$, of the part which has not yet reached the surface; CB is the trace, at $t = T$, of the part which has traversed the surface. Denote by u and u' the speeds of the wave front above and below the surface. Evidently the perpendicular distance from AA' to BB' is uT; we need in addition to know the distance from A to CB, and this is supplied by Huyghens' argument, which

is as follows: the refracted wave front is the envelope of spherical wavelets, which start from the various points of the surface at the instants when the incident wave front reaches these points, and spread outward with speed u'. This argument gives $u'T$ for the length of the line AC, and we readily derive:

$$\sin \theta / \sin \theta' = u/u' = \text{quantity independent of } \theta. \quad (6.10)$$

We are trying to represent a beam of electricity both as a stream of corpuscles and as a train of waves, which must travel

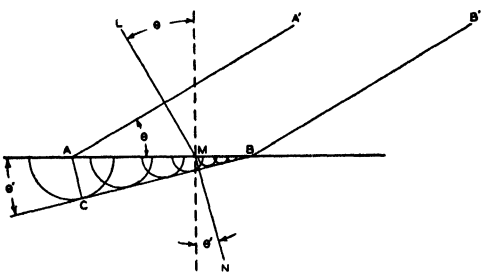


FIGURE 6.3—ILLUSTRATING REFRACTION INTERPRETED BY THE UNDULATORY IMAGE.

the same course, and must therefore bend alike as they pass across a potential discontinuity. By comparing equations (6.8) and (6.10) we see what we must assume in order to achieve this: to wit, the relation:

$$u/u' = p'/p. \quad (6.11)$$

At the crossing of the discontinuity, phase speed must go down (or up) in the same proportion as electron-momentum goes up (or down)! *Phase speed and particle momentum must vary inversely*:

$$u = \text{const.}/p. \quad (6.12)$$

With electron speeds not approaching too closely to that of light, this is almost the same as saying that wave speed and particle speed are inversely proportional to one another—a hard but inevitable saying. Equation (6.12) disposes of all chance of assuming that wave speed and electron speed are

identical; were they supposed to be equal at any one point of the trajectory, they would depart from each other elsewhere.

The interrelation (6.12) is necessary and sufficient to hold corpuscle-stream and wave-train upon the same path, not merely when they intersect an idealized surface of potential-discontinuity where the deflection is sharp and sudden like a corner, but also while they traverse an ordinary field of force in which their common path is gradually curved. This is readily accepted, because such a gradually-bending trajectory may be regarded as the limiting case of a succession of infinitesimal sharp bends occurring at infinitesimal potential-discontinuities; this is the method of the differential calculus for treating a continuously-turning path.

In dealing with light, we say that a sharp corner in the path of a beam occurs at a sharp boundary between two distinct homogeneous media each having a constant refractive index, and that gradually-bending paths occur when the light-beam is traversing a non-homogeneous medium with a continuously varying refractive index. This suggests that "refractive index" and "potential" are terms which we apply in a given situation, according as we are using the wave-model or the corpuscle-model—that when we pass from corpuscles to waves or *vice versa*, we may replace the potential by some function of the refractive index and *vice versa*; and such is indeed the case.

The "crucial" experiment of Foucault and Fizeau

For most of the last eighty years it has been the custom to say that the wave-theory and the corpuscle-theory of light contended with each other, until in 1850 an experiment by Foucault gave the victory to the waves. We will now drop back into the past, and examine what actually was proved by Foucault and Fizeau,[1] and after them by Michelson.

[1] Foucault usually gets all the credit, but Fizeau and Bréguet were working at the same time with the same incentive, and using the same method with differences of detail; and they announced their result only six weeks later. Indeed, at the meeting of the Académie des Sciences (May 6, 1850) at which Foucault reported his success, Fizeau said that if the sun had shown that day or the day before, they too would have had data to present.

All of these experimenters used the rotating-mirror method. The source of light shines continuously, and the continuous beam falls on a rotating mirror and is reflected in a continuously varying direction; a second, stationary mirror receives the reflected beam during a very small fraction of each complete revolution and sends it back, so that the twice-reflected beam is a series of segments cut from the primary beam. When these segments reach the first mirror on their return they find that it has revolved through a small angle θ since they left it, and accordingly it reflects them onward not quite along the path to the source of light, but along another path inclined to that one at the angle 2θ. The experimenter measures 2θ and the rate of rotation of the mirror, and deduces how long the segments of the light-beam require to travel from the rotating to the stationary mirror and back again.

The incentive to Foucault and Fizeau was supplied by Arago, who had asked in effect: is it the speed of the wave-crests in the wave-theory, or the speed of the corpuscles in the corpuscular theory, which agrees with the measured speed of a segment of a beam of light? Arago had written: "These experiments . . . will permit no further hesitation as between the rival theories. They will settle *mathematically* (I employ this word on purpose), they will settle mathematically one of the greatest and most disputed questions of natural philosophy." He had proposed a question to Nature, and had allowed her the choice between two answers only. Everyone then thought that she must choose the former or the latter. Foucault and Fizeau found that the segments of light traveled more slowly in water than in air. They reported that she had given the former answer; but they had not heard distinctly; for her response was: *Neither.*

Michelson's experiment, performed by the same method but with much greater refinement, made this abundantly clear. He found that the segments took (1.76 ± 0.02) times as long to go a given distance in carbon bisulphide as in air. But the refractive index of carbon bisulphide, in the range of the spectrum where Michelson's source of light was brightest, is about

1.63; so that the primitive wave-front-theory gives 1.63 for the ratio of the speeds in air and CS_2, and the corpuscle-theory gives $(1.63)^{-1}$.

If this had been discovered by Foucault and Fizeau, it would have been a poser! but in the interim, the notion of group-speed had been developed. The refractive index of carbon bisulphide relative to vacuum varies rapidly with wave-length; since u in vacuum is constant, u in the liquid depends on λ. Ascertaining the function $u(\lambda)$ from the refractive index, and then forming its derivative $du/d\lambda$, one finds that in the region of the visible spectrum, the term $\lambda(du/d\lambda)$ amounts to about 7% of the term u, on the right-hand side of equation (6.6). Group-speed *in vacuo* being the same as phase-speed, and both being almost identical with the corresponding speeds in air, we conclude that Michelson's value for the ratio of segment-speeds falls close to the ratio of group-speeds. (In water, the difference between phase-speed and group-speed is small enough to permit us to believe that Foucault and Fizeau, with their earlier and cruder apparatus, missed a similar difference between the phase-speed and the segment speed).[1]

The lessons of this story? First, that not even an Arago should ask a question of Nature, and prescribe all the answers in advance. Second, that the speed of a segment of a light-beam—the only " velocity of light " which it is possible at all to measure—differs from the speed of the waves in the wave-model, and yet the difference does not disqualify the wave-

[1] The numerical values in this section are taken from a review of Michelson's work by J. Willard Gibbs (*Am. Jour. Sci.* 31, 62; 1886) which so far as I know is the latest critical discussion of the data. The problem is more complex than I have intimated, not only because Michelson observed light covering a very wide range of wave-lengths so that u and $du/d\lambda$ both extend over wide ranges of values, but also because different parts of a wave-front are reflected from different parts of the mirror at different moments, and therefore from *differently inclined* parts. Quite a controversy went on during the eighteen-eighties in the pages of *Nature* as to what it was that Foucault had really measured. Rayleigh at first (*Nature* 24, p. 382; 1882) thought it was g; then changed his mind (25, p. 52; 1882) and decided it was u^2/g; then was convinced by Schuster (33, pp. 439–440; 1886) that it was really $u^2/2(u-g)$. J. W. Gibbs then took a hand (33, p. 582; 1886) and contended that after all it was really g. The controversy seems to have rested there. It may be added that Michelson's data eliminate u^2/g, but do not quite discriminate between g and Schuster's expression.

model. Third, that even as this difference did not disqualify the wave-theory, so the fact that segments of beams move more slowly in water (and carbon bisulphide) than in vacuum cannot disqualify the corpuscular theory: for the corpuscles of light, as we shall be finding out, are undeniable. Fourth, that one must always remember that the *right-hand member of* (6.8) —the equation of refraction derived from the corpuscular model—*is a ratio not of velocities but of momenta.* With slow-moving material particles momentum is proportional to speed, and the ratio is the same as the ratio of velocities; but one must be prepared to make some different assumption, when dealing (for instance) with light.

The principle known as "least time" for the waves, as "least action" for the corpuscles

Select a couple of points, L and N for instance, on the path of the beam of Figure 6.2, or, to be more precise, on the path of one of the corpuscles of the beam. Go over to the wave-model and to Figure 6.3; the path in question (LMN) now has a new significance—it is a curve to which a wave-front remains perpendicular, in going from L to N. One can draw any number of paths from L to N, each consisting of a straight line drawn from L to some point of the refracting surface other than M, and another straight line drawn from this point onward to N. Each of these is a conceivable path for a particle travelling from L to N; likewise, a wave-front may be conceived as going from L to N in such a way as always to be at right angles to any one of these. None of these, however, is the actual path. The actual path is outstanding among all the rest by the feature, that the time spent by the wave-front in going from L to N is less by this route than it would be by any of the others (always supposing that by any route it would be obliged to travel with speed u in the upper medium and u' in the lower).

This is Fermat's "principle of least time," which is easily proved by simple geometry for the situation of Figures 6.2

and 6.3 (a single plane refracting surface). It can be generalized to the case of a beam which in passing from L to N traverses a continuous field of force, though then it may be strictly true only when the comparison is made between the actual path and others which differ infinitely little from the actual path. Cases can be contrived in which the actual path is described not less, but more, rapidly than some of the adjacent paths; while if the refracting surface is replaced by a lens whereby light starting out from L in various directions is refocussed at N, the time spent by the wave-front along each of these routes is the same. The proper title for the principle is therefore " principle of stationary time "; the proper wording (in technical language) is, that the actual path is the one for which the first variation of the time is zero; and the proper formulation is:

$$\delta \int u^{-1} ds = 0, \tag{6.20}$$

the integral being a line-integral along the actual path, between any two points thereof, of the reciprocal of u the wave-front-speed or phase-speed.

This principle is evidently *not* true for the corpuscles! the actual path—LMN of Figure 6.2—is not the one by which they could go in the shortest time from L to N. This difference enhances the prestige of the waves. It supplied, indeed, the earliest argument for what we now know as the wave-theory of light—a striking historical fact, not so well known as it should be. Fermat had faith in " economy of time " as a fundamental principle of Nature, and postulated—this was as long ago as the middle of the seventeenth century, before Huyghens and before Newton—that light must spend the shortest possible time in going from the start to the finish of its path. Reasoning thence in the opposite direction from that which we have just been taking, he came to (6.10); so that before the notion of pulses or the notion of waves existed, he found a reason for supposing that the speed of light is what Huyghens was later to identify with the speed of pulses, and what we now identify with the phase-speed of waves.

The corpuscles have a principle of their own, which we discover by substitution from (6.12) into (6.20), thus:

$$\delta \int p ds = 0. \tag{6.21}$$

The line integral of the momentum of a particle between any two points of its path is stationary with respect to all other paths joining these points and differing infinitely little from the actual one.

This is the *principle of least action*: the name of the integral $\int p ds$ being "action," while the adjective "least" is not so suitable as would be "stationary", for the reasons stated above. Maupertuis offered it as a substitute for Fermat's principle of least time, at an epoch (the mid-eighteenth century) when it seemed that in spite of Fermat's argument the corpuscular theory of light had vanquished the undulatory theory. We of the twentieth century are not obliged to discard the one in order to employ the other: conceiving of a beam of light or of electrons or of matter as a stream of corpuscles associated with a train of waves, we may make use of both.

The function W, and its level-surfaces considered as wave-fronts

The action is a function of the two points on the path of the corpuscle, between which the integral is taken. I denote it henceforward by A. If one of the points (call it P_0) be regarded as fixed, A is a function of the coordinates x, y, z of the other point (call it P).

The element pds of the integrand of A is the scalar product of the two (parallel) vectors **p** the momentum of the corpuscle and **ds** the line-element of the path, and therefore may be written:

$$pds = p_x dx + p_y dy + p_z dz. \tag{6.22}$$

Thus A is invested, by virtue of its definition, with the properties:

$$\partial A/\partial x = p_x, \quad \partial A/\partial y = p_y, \quad \partial A/\partial z = p_z. \tag{6.23}$$

The components of the momentum of the particle at any point of its path are equal to the values there prevailing of the derivatives of A with respect to the coordinates.

These properties are shared by the function W defined as follows,

$$W = A - Et, \tag{6.24}$$

in which E stands for the energy of the particle (the *total* energy, not merely the kinetic!) which remains constant all along its path. W has a fourth property, which I now write with the other three:

$$\partial W/\partial x = p_x, \partial W/\partial y = p_y, \partial W/\partial z = p_z, \partial W/\partial t = -E. \tag{6.25}$$

At any particular instant, W has a system of " level surfaces," that is to say, surfaces over each of which its value is constant.[1] These surfaces are perpendicular to the path of the particle at the points where they intersect it. Now let t vary. The level surfaces wander through space, each area-element of each moving parallel to its own normal. *Their speed is given by the equation:*

$$\text{Speed of level surfaces} = E/p. \tag{6.26}$$

This important equation is easily proved. Imagine a vehicle (I use this word deliberately, to avoid confusion with the particle having the momentum p) travelling with the speed u through the point P in the direction normal to the level-surface and tangent to the path of the particle. Say that it passes through P at the instant t, the function W having then and there the value W_1. At the instant $t + dt$ it is at another point, and the function W has then and there the value:

$$W_1 + dW = W_1 + (\partial W/\partial x)dx + (\partial W/\partial y)dy + (\partial W/\partial z)dz + (\partial W/\partial t)dt. \tag{6.27}$$

[1] The functions W and A are liable to Bridgman's caustic comments on functions which are defined all over the place but have physical significance only along one curve (*The Nature of Physical Theory*, p. 68).

By (6.25) the last term on the right is $-Edt$, the three previous terms add up to pds; by our assumption, ds is udt; hence:

$$W_1 + dW = W_1 + (pu - E)dt. \qquad (6.28)$$

If the vehicle is moving with the speed E/p, the coefficient of dt vanishes, the vehicle as it travels keeps pace with the level-surface for which $W = W_1$, and that particular speed is characteristic of both; which is what was to be shown.

Equation (6.26) is equation (6.12), with a specific value assigned to the constant which in (6.12) was left indeterminate. The level-surfaces of W fulfil the condition which is required, in order that waves and corpuscles shall keep to a common path; and in addition they present a special value for the constant which, so far as that condition is concerned, might still be arbitrary.

Thus far the argument has served to enhance the prestige of the waves, by associating them with the fundamental function of the corpuscles which bears the name of " action "; also, to aid the prestige of this function by linking it with the wave-fronts. Still there is something missing: in the foregoing argument there is no allusion whatever to vibration or periodicity, wave-length or frequency. The primary feature of what we nowadays call " wave motion " is lacking; only a secondary feature, the wandering of wave-fronts through space, is present.

Looking back again into history, we find that the notion of regular vibration did not enter into the theory of light for many decades after Fermat and Huyghens. It is certain that Huyghens did not have it; when he first drew the diagram which reappears in this book as Figure 6.3, he was picturing to himself the passage of a single pulse across the boundary between the two media. Indeed, there seems to be nothing in Huyghens' construction which implies a regular sequence of pulses or wave-fronts. In his book of 1690, he shows a sketch of spherical wave-fronts proceeding from various points of a candle flame; those diverging from a single point follow each other at equal distances; but he is careful to explain that this

is an accident of the drawing, that in reality the pulses emerge from the radiant points at random intervals!

Nevertheless it can be argued that Huyghens was assuming a regular sequence without realizing it. To translate his "pulse" into a modern picture, we must imagine that the displacement at any point of the region traversed by the oncoming light remains zero up to a certain instant, then rises suddenly to an appreciable value, falls off suddenly to zero again, and continues at zero until the arrival of another pulse. Each pulse would be a sort of thin film of finite displacement, travelling onwards through space towards the refracting surface.

Now such a pulse must be regarded as a superposition of sinusoidal wave trains of various wave-lengths and amplitudes. If all of these component wave trains have the same phase speed, all is well; the pulse marches onward undistorted. This is the case so long as light moves through vacuum. But when light enters a material medium, the phase speeds of the different wave trains draw apart. *The pulse becomes distorted, and the distortion must commence at the refracting surface itself.* Huyghens may have been justified in drawing a straight line athwart the path of the light to represent a single pulse in the upper medium of Figure 6.3, provided that he meant the upper medium to be a vacuum. But he was not justified in drawing another straight line to represent the pulse in the lower medium, glass or water or whatever it may have been; for in crossing the boundary the thin film of displacement would have changed into a hazy volume spreading out in various directions.

Indeed there is no way to design a pulse that shall go through a refracting surface with a sharply-defined deflection. The only wave which surely obeys Huyghens' rule is the member of an endless sequence of sine-waves. The only wave-front to which Huyghens' construction can securely be applied is the wave-front which belongs to a sinusoidal wave train of definite wave-length and frequency. It is paradoxical but not entirely extravagant to say that Huyghens was postulating a full undulatory theory without ever realizing it.

A periodic function is introduced which for its wave-fronts takes the level surfaces of the action

Now the undulatory theory of electricity and matter enters definitely upon the scene. The analogy with light may be traced in the following steps:

At a refracting surface, the only kind of beam of light which is refracted at a definite angle is the special kind which by definition is called " monochromatic." Also the only light which will form an interference pattern corresponding to a definite value of wave-length is precisely this monochromatic kind.[1] Therefore monochromatic light is light of a single wave-length, and the only kind of light which is refracted at a definite angle in passing through a refracting surface is a sinusoidal wave train. Now at a surface of potential-discontinuity, the only kind of beam of free electricity (electrons) which is refracted at a definite angle is a beam consisting of electrons of a single kinetic energy. We have identified a function W associated with this beam, of which the level surfaces are refracted at the same definite angle. This establishes a partial analogy with light but not a complete one. To complete the analogy we assume that the level surfaces of W are also those of another function which is itself a sinusoidal wave train: an oscillatory function, of which W itself is the " argument " or " phase."

The introduction of this wave train into the picture, the assertion that it possesses real and physical importance, are distinctive features of the new mechanics. They are due primarily to Louis de Broglie and to Schroedinger.

These are, of course, the very waves of which I earlier spoke, those of which the phase speed u is inversely proportional to the corpuscle momentum p according to equation (6.12). By insisting that their wave-fronts coincide and travel with the level surfaces of W, we fix the hitherto undetermined constant appearing in that equation. Writing E for the total energy—not merely the kinetic energy!—of a corpuscle, ν

[1] See footnote 1, page 2.

and λ for the frequency and wave-length of the waves, we get:

$$u = \nu\lambda = E/p. \tag{6.30}$$

To have fixed the phase speed by fixing the constant is not, however, of the first importance, u being unobservable. What *is* of first importance is the conclusion with respect to the wave-length. We derive:

$$\lambda = (E/\nu)/p = \text{const.}/p. \tag{6.31}$$

The ratio of E to ν is considered a constant, because, first, the total energy of a corpuscle remains the same throughout its entire course; and, second, anything properly named "frequency" should be the same throughout the entire system in question. The latter statement is due to our instinctive feeling that when two material bodies are in contact with one another, they cannot vibrate and also remain in continual contact unless their frequencies are the same.

We will now conclude the list of our assumptions by supposing that the ratio of E to ν is not merely constant throughout the entire course of a single beam, but has always the same value for every beam in every locality. It shall be a *universal* constant. I denote it by h, thus revealing the climax already foreseen, no doubt, by nearly all the students of this book; that it is to be identified with Planck's constant first introduced in the theory of black-body radiation ($h = 6.55 \times 10^{-27}$ $\text{g}\cdot\text{cm}^2\cdot\text{sec.}^{-1}$).

Rules of correlation of corpuscles and waves

Thus we have come to the "Rules of Correlation" lying at the base of the new theoretical physics, and governing the interrelation between the wave-aspect and the particle-aspect of Nature. I give them in the reverse of the order of their discovery:

$$p = h/\lambda, \tag{6.32}$$

$$E = h\nu = hu/\lambda. \tag{6.33}$$

Evidence both direct and indirect for each of these two Rules is scattered through the following chapters of this book; at this point it is suitable to speak of some of the most direct.

To test the first of the Rules for beams composed of any type of particle, one must be able on the one hand to measure the momentum of the particles, and on the other hand to form diffraction-patterns from the beams. With *charged particles* (electrons and ions) the momentum is measurable, and has been so for forty years, as we have seen in Chapters II and III; and diffraction-patterns are obtained by projecting the beams against or into crystal lattices, as we shall see in Chapter VII. The earliest diffraction-patterns of this type were obtained with electrons in 1927, first by C. Davisson and L. H. Germer of the Bell Telephone Laboratories, and shortly afterward (and independently) by G. P. Thomson of Cambridge. With *uncharged material particles* (neutral atoms) the speed may be determined by the method of the gates and the momentum computed by multiplying it with the atomic mass, and diffraction-patterns may be formed with the aid of crystal lattices. With *light*, the technique of determining wavelength is ancient (*i.e.*, ancient in the sense of modern physics—more than a century old!) but the momentum of the individual particle cannot be separately measured: there are phenomena in which both of the Rules are interwoven (as we shall immediately see) in such a way that both are verified conjointly, but it is not possible to measure p by itself in one experiment and compare the result with λ determined by itself in another.

To test the second of the Rules of Correlation for beams composed of any type of particle, one must be able to measure the energy of the particles, to form diffraction-patterns from the beams, and to measure the phase-speed of the waves. The third requirement puts an entirely new face upon the situation, and restricts the prospects of success to beams of light *in vacuo*, for which alone (owing to the coincidence of phase-speed and group-speed, as we have lately seen) the phase-speed u is measurable. With light, however, there is again a disability like that which inhibits the testing of the first of the Rules; so long

as the particles exist, one can measure neither their momentum nor their energy! This would be thoroughly discouraging, were it not that phenomena occur in which these particles vanish, and in vanishing transfer their energy and their momentum to electrons; and once they are transferred, these quantities are measurable! Leaving to Chapter VII the direct test of the first of the Rules, I now introduce some of these phenomena in which the second Rule by itself or the two of them conjointly are verified for light.

Test of the second Rule of Correlation by the photoelectric effect

The " external photoelectric effect of metals," usually abbreviated to " photoelectric effect," is the outflow of electrons from metals exposed to light. The study of this effect was one of the most exciting episodes of the seven or eight years following 1906, for the phenomenon was the first-to-be-discovered among those in which the corpuscular aspect of light is strikingly manifest and undeniable. In particular, the intensity I of the light controls the number N (per unit time and per unit area of illuminated surface) of the emerging electrons, but not the kinetic energy K of the individual electron; N is proportional to I but K is constant. This is definitely what is to be expected if the energy of the light is borne in corpuscles, each of which is entirely absorbed by a single electron inside the metal. It is definitely not what is to be expected from the " pure " undulatory theory of light prevailing throughout the nineteenth century. Many attempts were made to escape from this last conclusion; it was (for example) suggested that the electrons within the metal are attached in such a way to the atoms, that they are able to suck up energy continually and gradually from the light-waves until they have gathered up the amount K (or somewhat more), whereupon they detach themselves and rush out. This idea was spoiled by the fact (among others) that when a piece of metal is suddenly exposed to light, the outflow of electrons begins so soon (instantaneously, so far as can be told!) that if one were to apply the idea to the earliest

among them, it would be like imagining that a rowboat moored at the entrance to New York Harbor may in a moment absorb all the energy of all the waves rolling against the entire Atlantic seaboard. Many vivid similes and metaphors were used, for twenty years and more, to illustrate the situation; it would be pleasant to revive a few of them, if they did not already seem *passé*.

While it is assumed that the entire energy E of a " photon " (I use this name hereafter for " corpuscle of light ") is transferred to an electron, it is not assumed that E is the same as K the kinetic energy possessed by the electron just after it leaves the metal. What *is* assumed can best be explained by introducing at once the law relating K for the *fastest* emerging electrons to λ the wavelength of the incident light:

$$K = A_1/\lambda + A_2. \qquad (6.34)$$

Here A_1 and A_2 stand for constants, of which the values are given by the experiments. Now, the theory of the photoelectric effect of metals is based on the idea which follows. A metal contains a tremendous multitude of so-called " free " electrons, having kinetic energies ranging up to a maximum value eW_i, and normally unable to escape because if one of them were to leave the metal, it would have to traverse a " surface potential-barrier " which would take from it an amount eW_a of its kinetic energy, and eW_a is larger than eW_i. The escaping electrons are presumed to be those which have added, to their initial kinetic energies, the entire energy E of a photon and have thus built up a sum which is superior to eW_a. Specifically it is assumed that the term A_1/λ on the right of (6.34) is the photon-energy E, and the term A_2 is the difference $(eW_i - eW_a)$. The fastest emerging electrons are those which while still in the metal had the maximum kinetic energy eW_i; the slower electrons had a lower initial stock of energy, or perhaps lost more of their total stock than the minimum possible loss eW_a while on their way out. This theory is sustained by many facts, which lie beyond the scope of this

book; I mention as one, that A_1 is the same for all metals while A_2 varies from one to another.

Accepting this interpretation, and writing c for the phase-speed of light and ν for c/λ, we find:

$$E = (A_1/c)\nu, \quad h = A_1/c, \tag{6.35}$$

as the epitome of all this theory and these results.

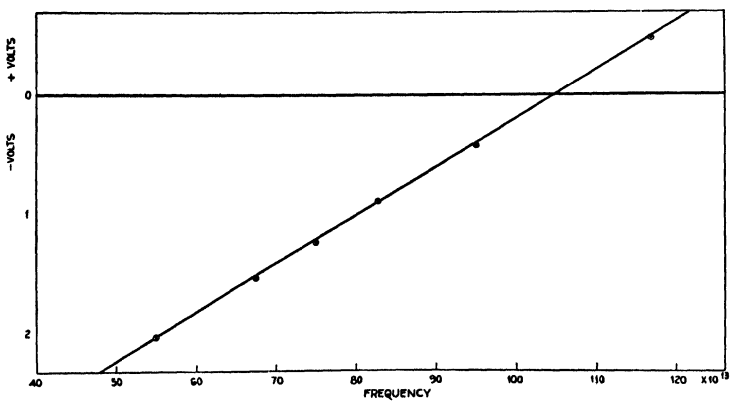


FIGURE 6.4—LINEAR RELATION BETWEEN K THE MAXIMUM KINETIC ENERGY OF PHOTOELECTRONS EXPELLED FROM A METAL AND c/λ THE FREQUENCY OF THE LIGHT EXPELLING THEM. (R. A. Millikan; *Physical Review.*)

The numerical values of h resulting from these and other experiments will be treated later. I now point out that in all of this section, not a word has been said about momentum or about the conservation thereof, in these encounters between photons and electrons in which we are supposing the entire energy of one to be transferred to the other. It will presently transpire that we have incurred a fallacy, in thinking simultaneously of the energy-transfer as being complete and of the electrons as being initially free. We shall be obliged to suppose that in the photoelectric effect there is a linkage between the electrons and the metal, such that the momentum belonging originally to the photons is passed straight onward to the latter while the electrons keep most of the energy. Now we

examine a phenomenon, in which there is no metal present to serve as a receptacle for the momentum, and consequently the process is quite novel in detail.

The Compton effect

The equations which now appear are the counterpart of equations (3.2) and (3.3), little as they resemble those. Those equations described elastic impacts between material particles; these, elastic impacts between a photon and a material particle, the latter being an electron. The lack of resemblance is due partly to this distinction, and partly to the fact that in dealing with the material particle, I use the relativistic expressions for momentum and for energy. (As many a student must have found out for himself, the attempt to make the equations look simpler and be more easily solvable by using ordinary mechanics has precisely the opposite result—this theory is an example of the wisest policy proving to be the easiest). As in (3.2) and (3.3), we shall assume that the "struck" particle—the electron—is stationary before the impact. Another similarity: in those equations we assumed that the striking particle survived the impact, though with diminished energy and momentum, and this assuredly astonished no one—here we shall have to make the same assumption, and this time it may seem stranger. The symbols ν, λ shall designate frequency and wavelength of the photon before the impact; ν', λ' the frequency and wavelength of the surviving or "scattered" photon after the impact; v the final speed of the electron; ϕ and θ for the angles made with the direction of the incident photon by those of the recoiling electron and the scattered photon respectively.

Conservation of energy requires that the energy of the incident photon, $h\nu$, be equal to the sum of the energy of the scattered photon, $h\nu'$, and that of the recoiling electron, which latter in relativistic mechanics (Chapter II) is given by the expression,

$$E = mc^2 = [m_0/(1 - \beta^2)^{\frac{1}{2}}]c^2, \quad \beta = v/c, \qquad (6.36)$$

where m_0 stands for the rest mass of the electron. Thus:

$$h\nu = h\nu' + [m_0/(1 - \beta^2)^{\frac{1}{2}}]c^2. \qquad (6.37)$$

Conservation of momentum requires the following two equations, referring respectively to the momentum-components parallel and perpendicular to the direction of the incident photon, in the plane common to the three directions of incident photon, scattered photon, recoiling electron:

$$h\nu/c = (h\nu'/c)\cos\theta + [m_0 v/(1 - \beta^2)^{\frac{1}{2}}]\cos\phi, \qquad (6.38)$$

$$0 = (h\nu'/c)\sin\theta + [m_0 v/(1 - \beta^2)^{\frac{1}{2}}]\sin\phi. \qquad (6.39)$$

Eliminating ϕ and v between these three equations, we arrive at a relation between ν', ν and θ, which may immediately be translated into the following relation of simpler aspect:

$$\lambda' - \lambda = (h/m_0 c)(1 - \cos\theta). \qquad (6.40)$$

This is a prediction of the wavelength which should characterize the photons rebounding, or " scattered " as the term is, in the direction θ. Eliminating θ and ν' between (6.37), (6.38) and (6.39), we get a relation between v, ν, and ϕ, which is in effect a prediction of the speed and kinetic energy of electrons recoiling in the direction ϕ. Eliminating v and ν' between those three equations, we get a relation between ν, θ and ϕ which is in effect a prediction that when photons are falling along parallel lines into a flock of stationary or slow-moving electrons, then whenever a photon rebounds in a direction θ there is also an electron simultaneously recoiling in a corresponding direction ϕ which is to be computed from that relation.

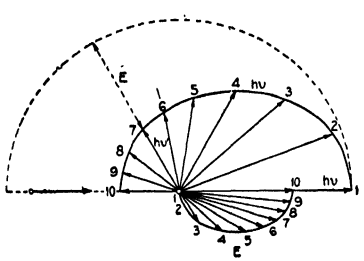


FIGURE 6.5—ILLUSTRATING THE ENERGY-RELATIONS AND DIRECTION-RELATIONS OF THE PHOTON AND THE ELECTRON DEPARTING FROM AN ELASTIC IMPACT: INITIAL ENERGY OF PHOTON BEING E, INITIAL KINETIC ENERGY OF ELECTRON BEING ZERO. (After Debye.)

What I have just been describing as though it were an imaginary process is in fact the " Compton effect," which occurs when a beam of X-rays impinges on a piece of matter. Scattered X-rays spring forth from the matter; what was first of all observed was a " softening," as people used to say, which distinguished these from the incident or primary X-rays—the scattered rays are more absorbable than are the primary. No accurate test of this or any theory could, however, have been made, until the spectrum of the scattered rays should be discovered by diffraction and compared with that of the inci-

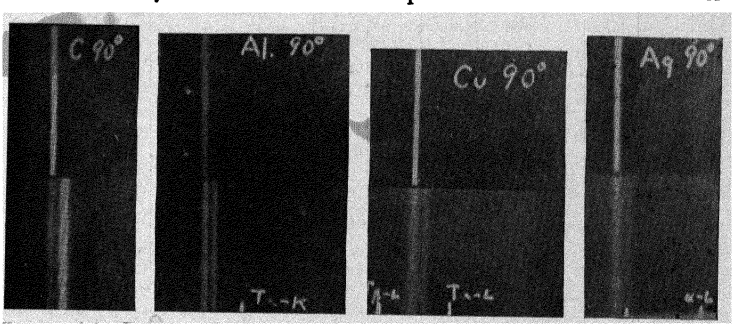


FIGURE 6.6—COMPTON EFFECT: ABOVE, SPECTRUM OF THE PRIMARY X-RAYS (A SINGLE LINE, $\lambda = 710X$); BELOW, SPECTRA OF THE X-RAYS SCATTERED BY FOUR ELEMENTS $Z = 6, 13, 29, 47$ (TWO LINES, OF VARYING INTENSITY-RATIO). (Photograph by P. A. Ross.)

dent rays. This was first achieved by A. H. Compton in 1922, and together with the data he published the theory embodied in these equations, for which reason the effect bears his name. The equations were also published by P. Debye early in the following year.

If the primary X-rays consist of a single frequency or spectrum-line, the scattered rays exhibit this line at the original wavelength, accompanied by another—the " modified," " shifted " or " Compton " line—with a wavelength somewhat greater (Figures 6.6, 6.7). If the primary rays comprise several lines, each reappears in the " scattered spectrum " attended by a companion shifted by the identical amount toward greater

wavelengths. The unmodified lines are composed, in the corpuscular picture which we are now employing, of photons rebounding with energy and momentum sensibly unchanged. These must have rebounded from impacts with bodies of mass so great, that if it is substituted for m in (6.40) the "shift in wavelength" $(\lambda' - \lambda)$ given by that equation is imperceptibly small. The mass of an atom is large enough to fulfil this condition: accordingly it is assumed that the unmodified photons have rebounded from impacts against electrons which remained firmly attached to atoms throughout the encounters. It is however the modified lines which concern us now: for their shifts in wavelength agree with (6.40). The shifts in question are independent of the wavelength and frequency of the primary rays; and proportional to $(1 - \cos\theta)$ for rays of any wavelength, the spectrometer being revolved around the scatterer so as to pick up rays scattered at different angles when this relation is being tested. The ratio of the observed shift to $(1 - \cos\theta)$ agrees with h/mc; by which is meant, that the value obtained for h/mc by identifying it with that ratio is consistent with the values of h and m and c obtained by other experiments—this will be discussed elsewhere.

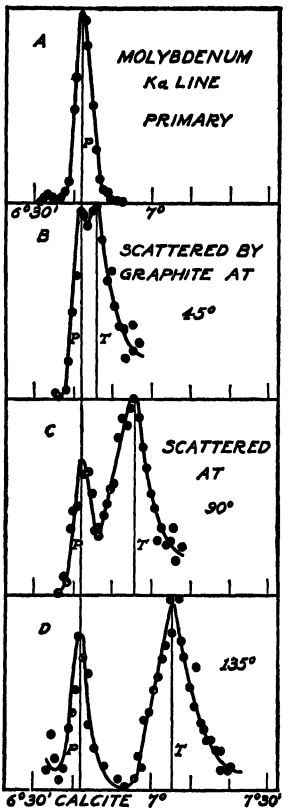


FIGURE 6.7—COMPTON EFFECT: SHIFTED AND UNSHIFTED LINES IN THE SCATTERED RADIATION, EXHIBITING THE DEPENDENCE OF SHIFT ON θ; THE VERTICAL LINES T INDICATE THE POSITIONS COMPUTED FOR THE SHIFTED LINE BY THE COMPTON–DEBYE EQUATIONS. (A. H. Compton; *Physical Review*.)

As the theory reposes wholly upon the assumption that electrons recoil from the impacts of photons, the search for recoiling electrons becomes an essential part of the test. It has long been known that fast electrons spring from matter during exposure to X-rays, a fact which is made splendidly apparent by expansion-chamber photographs such as Figure 3.3 (page 69) taken while X-rays are traversing the chamber: one sees the tracks of electrons starting from whatever solid piece of matter is in the way of the X-rays, and also tracks beginning in the irradiated gas itself. Not all of these, however, may be interpreted as recoiling or " Compton " electrons. In addition to the Compton process, there is another whereby X-rays liberate electrons from atoms, a process in which an entire photon is absorbed by an atom. I will speak of this alternative process in Chapter VIII; for the present it is enough to say that the electrons liberated by it have energies usually so markedly different from the values derived from the Compton-Debye equations, that there is no difficulty in distinguishing them. They being set aside, the others agree roughly with the theory. I am obliged to say " agree roughly," for the difficulty in measuring the kinetic energy of these electrons under the conditions of the experiment is evidently so great, that discrepancies of ten and even of twenty per cent are considered tolerable.[1] Since the quantitative test of the equations is so well made by study of the photons, there seems to be no forcible reason for requiring more than a qualitative test from the electrons.

The Compton effect is a much bigger subject than I have indicated yet. The electrons involved in it are not initially stationary, and in many cases they are not entirely free, being bound (though loosely) to atoms: there is a broadening of the shifted line, *i.e.*, a distribution-in-energy of the scattered photons, which is due to these circumstances, and has been studied by DuMond and his school at Pasadena and by the school of

[1] Compton and Allison, *X-rays in Theory and Experiment*, pp. 215–219. The chief, though not the only, method consists in observing the "ranges" (Chapter XV) of the electrons.

Scherrer at Zürich. The probability of the effect, in its dependence on such things as θ, ν, and the nature of the scattering substance, is interesting and extensively studied: Here I mention that it increases with ν, so that at the upper end of the spectrum (high-frequency X-rays and gamma-rays) the shifted lines are much stronger than the unshifted, and at the lower end (the optical spectrum) they are not even perceptible. Furthermore, it is greater for elements of low atomic number than elements of high, as Figure 6.6 makes vivid. Not all of these details are understood, but the value of the Compton effect as a proof of the Rules of Correlation stands apart from them all.

Interconversion of electrons and photons

According to the second Rule of Correlation, light of frequency ν consists of corpuscles of energy $h\nu$. According to Einstein's relation (2.20), these corpuscles have mass $h\nu/c^2$. According to experiment, electrons have mass m which varies with speed in the manner indicated by relativistic theory. Accepting relativistic theory completely, we assign to any electron the energy mc^2, which we may divide into a pair of terms as we did in Chapter II:

$$mc^2 = m_0c^2 + K, \tag{6.41}$$

m_0 standing for the rest-mass and m_0c^2 for the rest-energy of the electron, and K for a function of the speed v of the electron which we may well call " kinetic energy," though it is a power-series in v^2 of which only the first term ($\frac{1}{2}m_0v^2$) agrees with the ordinary expression for kinetic energy.

Imagine a photon disappearing, and two electrons of opposite sign coming into being in its place. This imagined process can occur without altering the sum-total of electric charge in the world. Moreover it can occur without violating the principles of the conservation of energy and mass, if it obeys the following law,

$$\Sigma K = h\nu - 2m_0c^2, \tag{6.42}$$

ΣK standing here for the sum of the kinetic energies of the positive and negative electrons. It cannot occur by itself without violating the principle of the conservation of momentum; but perhaps it might evade this requirement just as another process, the photoelectric effect of metals, manages to evade it—that is, by the transfer of most of the momentum (and very little of the energy) of the photon to a nearby solid body or even a nearby atom-nucleus.

But *can* such a process occur at all? The idea would have sounded fantastic, not many years ago. It might possibly have arisen without the aid of quantum mechanics, after electrons of both signs were known; but of this no one can ever be sure, for similar though not identical ideas had been derived from quantal speculations by Dirac even before the discovery of the electron of positive sign. Dirac at first imagined a negative electron paired with a proton, and later conceived of a process in which two photons unite and become an electron-pair (the converse of this process, as we shall presently see, has now been confirmed). The process implied by (6.42) was conjectured by Blackett when he observed cosmic-ray showers comprising particles which, as he recognized from the ionization-densities and the curvatures of their tracks, were electrons some of one sign and some of the other. Later he produced free electrons both positive and negative by playing gamma-rays on pieces of metal, a technique which has since been followed by many.

By far the most vivid sign of this process of "pair-production" is a cloud-chamber photograph showing two electron-tracks springing from the same point and bending away from each other in the magnetic field. The common point of origin may be in the gas of the chamber, or on the surface of an inserted piece of metal. Pairs born in the gas are preferred over pairs born in a solid, for which last there is danger that their members may lose a good deal of kinetic energy before emerging into the gas, and there is danger also of fortuitous pseudo-pairs of which the members spring from different processes occurring at nearby points. (With both

types of pairs there is danger of mistaking a sharp deflection of an electron-track for the origin of a pair, a possibility stressed by Leprince-Ringuet.) But to get even a few of these preferred pairs, with gamma-ray sources no stronger than have been available until lately, one must make a great many expansions. Curie and Joliot had the good fortune to find, on a plate made in 1933 with argon in the chamber, a lovely example since borrowed for many a book (here it is Figure 6.8); but Simons and Zuber four years later had to take 1380 pictures with an argon-filled chamber in order to observe 52 pairs, eighteen only

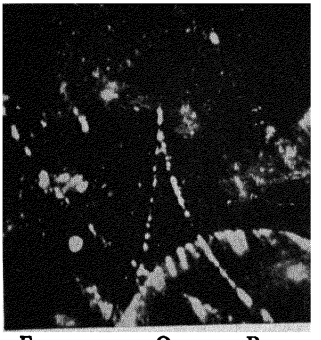


FIGURE 6.8—CREATION OF AN ELECTRON-PAIR OUT OF A PHOTON IN AN EXPANSION-CHAMBER FILLED WITH ARGON; PRESUMABLY IT TOOK PLACE VERY CLOSE TO AN ARGON NUCLEUS. (F. Joliot and I. Curie.)

of which showed both tracks neatly enough to be measurable. These and other experiments [1] for the test of (6.42) were done with the gamma-rays from thorium C″ which consist mainly of photons of energy 2.62 Mev. The prediction accordingly is, that the sum ΣK of the kinetic energies of the members of a pair should be 1.6 Mev. One is disappointed to find that the most that can be said of the data is, that for nearly all of the observed pairs ΣK appears to lie between 1.4 and 1.8 Mev—so difficult it seems to be to assess the electron-energies! The remaining pairs are taken to be fortuitous, or mere simulations

[1] Curie and Joliot, **JP 4**, 494 (1933); Immelman, **NW 24**, 61 (1936); Klarmann and Bothe, **ZP 101**, 489 (1937); Simons and Zuber, **PRS 159**, 383 (1937).

due to electron-deflections, or else are ascribed to photons of exceptional energies. With pairs springing from metals, the only prediction that can be made is that the sum of the observed kinetic energies should never exceed 1.6 Mev: out of 22 examples arising from lead, Anderson and Neddermeyer found only one which did, the majority of the rest lying between 1.0 and 1.6 Mev.[1]

Metal plates exposed to gamma-rays emit not only pairs, but a greater number of electrons not thus companioned. The presumption is, that *all* of the positives and *some* of the negatives are really members of pairs, their mates having failed to escape from the metal or having for other causes eluded detection; while the rest of the negatives come from the orbital electron-families of the atoms, having been expelled either by the Compton process or by another yet to be studied (page 248) under the name of " photoelectric effect." The testable consequence is, that the distributions-in-energy of the positive and the negative electrons should extend to different upper limits; the former we already know, the latter is that of the " photoelectrons " and is nearly one Mev higher (in the notation of page 250 it is the energy of the incident photons minus the K-extraction energy of the metal atoms). Chadwick, Blackett and Occhialini irradiated lead with thorium-C″ gamma-rays, and observed the tracks of more than 4000 negative and 400 positive electrons; the distribution-in-energy curves dropped to the axis at points 0.99 Mev apart, sustaining the theory.[2] Support for the theory is also derived from experiments in which metals are exposed to gamma-rays of photon-energy slightly under one Mev; positive electrons should not be found, and are not.

[1] Anderson and Neddermeyer, **PR 43**, 1034 (1933); see also Anderson, *Nature* **133**, 313 (1934), where a similar conclusion is drawn from rather more data.

[2] Chadwick, Blackett and Occhialini, **PRS 144**, 235 (1934). Though the difference between the upper limits agreed with the theory, the limits did not; the discrepancy was taken to imply an error in the routine for deducing the true upper limit from the trend of the curve in its neighborhood, a very ticklish question. Similar results were reported by Crane and Lauritsen (Internat. Conference on Physics, London, 1934; pp. 135 ff. Also **PR 46** 531; 1934).

Gamma-rays from artificial radioactive bodies, and from nuclei newly born of transmutation, are sometimes detected by the pairs which they produce in cloud-chambers; not only are they detected, but their photon-energies are deduced by adding one Mev to the ΣK-values of the pairs; but this is material for Chapter XVII.

The conversion of electron-pairs into light is attested by 0.5-Mev photons emanating from radioactive bodies engaged in generating positive electrons, and also from metals into which these are absorbed. The process cannot be the reverse of (6.42), for if it were it would yield one-Mev photons; it is presumed to consist of the coming-together of two slow-moving electrons of opposite sign and the leaping-forth of two nearly identical photons in two nearly-opposite directions, a process in which conservation of momentum and of energy can be assured without the good offices of a neighboring nucleus. The sharpest evidence is, of course, the latest: J. R. Richardson placed expansion-chambers containing targets beside the radioactive bodies emitting positive electrons, and observed the Compton-recoil electrons springing from the targets; they manifested the presence of 0.5-Mev photons [1] (Figure 18.8, page 630). Richardson's sources were samples of N^{13}, V^{48} and Cu^{64}, embedded in matter thick enough to absorb the whole of the positive electrons. (There are signs of a continuous distribution of photon-energies upward from 0.5 Mev, and this may mean that now and then a positive electron unites with a negative while its speed is still appreciable.) McMillan earlier had detected the photons from N^{13} by an ionization-chamber, and had measured their absorption-coefficient in lead; this, as we shall see in Chapter VIII, is an index of their energy; here it indicated 0.5 Mev. Earlier yet, Joliot and Thibaud had used counters and photographic plates respectively for detecting rays emitted by metal screens onto which they were focussing beams of positive electrons; they, too, measured the absorption-coefficient, and it led them to the same conclusion.

[1] References given and experiment described on page 629. Prior work: McMillan, **PR 46**, 868 (1934); Thibaud, **PR 45**, 781 (1934); Joliot, **JP 5**, 299 (1933).

Thus it appears that the career of a positive electron is a quick brief passage from the place of its birth to the place where, having been much slowed down in the interim, it encounters a negative electron and is transmuted into light.

It is easy to imagine an alternation of processes, commencing when a high-energy photon enters an absorbing stratum of matter: it becomes an electron-pair, then the positive member of the pair is slowed down and merges finally with a negative electron in the absorber to become a photon-pair. This brings us back to the experiments of Chao, Gray and Tarrant, who were probably the first to detect any immediate consequence of the interconvertibility of electrons and light; they observed photons of about the right absorption-coefficient springing from pieces of metal which were being bombarded by photons of energy superior to one Mev.[1] Even earlier it had been noticed that photons of energy so high suffer a greater absorption in matter than could be accounted for by the processes then known: we revert to this story in Chapter VIII.

Phase speed, group speed and corpuscle speed of electricity and matter

Returning to equation (6.7) and substituting from equations (6.32) and (6.33), we find:

$$g = -\lambda^2(d\nu/d\lambda) = dE/dp. \tag{6.45}$$

By the classical mechanics of material particles, we have:

$$p = mv, \quad E = \tfrac{1}{2}mv^2 + U, \tag{6.46}$$

from which we derive at once the result

$$g = v. \tag{6.47}$$

The speed of the corpuscles in a beam of matter or electricity—atoms or electrons—is equal to the group speed of the waves.

[1] Gray and Tarrant, **PRS 143**, 681 (1934) and literature there cited; Chao, **PR 36**, 1519 (1930). Gray and Tarrant also observed a less absorbable radiation which may have consisted of the one Mev photons derived from the process reverse to (6.42). The intensities of these radiations relative to each other and to the primary beam present various unsolved problems.

The same theorem is obtained from relativistic mechanics, the required formulae being these:

$$E = m_0c^2/(1 - \beta^2)^{\frac{1}{2}} + U, \quad p = m_0v/(1 - \beta^2)^{\frac{1}{2}}, \quad (6.48)$$

$$E = c(m_0^2c^2 + p^2)^{\frac{1}{2}} + U, \qquad (\beta = v/c). \qquad (6.49)$$

As I have already said a number of times, frequency and phase speed are not observable. There are special circumstances which enable us to estimate them in the case of light, but these are lacking in the cases of electricity and matter. They are therefore of minor importance, or perhaps of none at all; but people generally seem to want to be told their values, so values have to be assigned. From equation (6.33) we get E/h for the magnitude of the frequency. Here E stands not for the kinetic energy alone (an error easy to fall into) but for the *total* energy, and indeed for the total energy in the sense of relativity. The relativistic expression for E would then be $(mc^2 + U)$; m standing for the mass of the corpuscles (depending on their speed) and U for their potential energy. The frequency is the quotient of this expression by h. For the phase speed we then have the quotient of $(mc^2 + U)$ by p, which is $(c^2/v + U/p)$. Since U involves an arbitrary additive constant, both phase speed and frequency are indeterminate to the corresponding extent; but since both are unobservable, this indeterminacy makes no difference.[1]

Pressure exerted by a beam of particles

Imagine a homogeneous beam of N particles per unit volume, impinging upon a target which swallows it up completely. The pressure P exerted by the beam upon the target is the sum of the momenta of all the particles which traverse unit-cross-section of the beam in unit time—therefore Nvp. The energy conveyed into the target per unit area per unit time (call it I) is the sum of the energies of all those particles—

[1] One often sees the phase speed equated to c^2/v, an expression obtained by setting $U = 0$. It is dangerous to use or remember this expression, as it looks like an approximation to the right-hand member of Eq. (6.12), which is not the case at all.

therefore NvE. For the ratio of I to P we have accordingly:

$$I/P = E/p = u. \tag{6.50}$$

Here is a way of determining the phase-speed of the waves associated with the particles, if I and P can be measured. P is obviously measurable, if the beam is strong enough. I poses a problem of a harder kind. The normal way of ascertaining the energy acquired by a target is, to ascertain the amount of heat developed in the target. If the entire energy of the particles entering the target is converted into heat, that procedure is a valid way of measuring NvE; otherwise it is not.

With *light* it is always assumed that the entire energy of the photons is convertible into heat, and is so converted in the target under the conditions of the experiment. If this is correct I can be measured; and so can P (the pressure of light was first observed in 1900 by P. N. Lebedew, and soon after by E. F. Nichols and G. F. Hull). *The ratio I/P*, while not determinable with anything like the accuracy attainable in measuring the speed of light, *agrees with c the phase-speed of light in vacuo.*

With *material particles*, both the potential energy U and the energy m_0c^2 associated with the rest-mass may evade conversion into heat; what we can measure is not the ratio I/p, so again the phase-speed eludes our attempt to fix its value by experiment.[1]

Principle of indefiniteness

Consider again the beam of corpuscles of uniform momentum, and the accompanying sinusoidal wave train, with which the argument started. Remember that the corpuscles are *not* concentrated at the wave crests, nor in any other particular part of the waves; the phase of the vibration at any particular

[1] If we make the mistake of putting the kinetic energy for E in the relation $u = E/p$, and in addition put $\frac{1}{2}mv^2$ for the kinetic energy and mv for the momentum of the particles, we reach the fallacious expression $\frac{1}{2}v$ for the phase-speed. In the early years of this century, various physicists supposed that the corpuscular theory of light required that I/P should be one-half the measured speed of light, and gave this as an argument against the corpuscular model.

point has nothing whatever to do with them; it is only the *amplitude* of the waves which is related to the concentration of the particles, the two being proportional to each other. Thus a plane-parallel sinusoidal wave train, of definite wavelength, uniform amplitude, and endless extent, corresponds to a flock of corpuscles which all have the same definite momentum but are uniformly scattered throughout space.

The content of this last sentence may be restated in a variety of ways. Here is another: an endless sinusoidal wave train describes a flock of corpuscles, with an absolutely definite value of momentum but an absolutely indefinite value of position.

What sort of a wave train, then, could describe a crowd of corpuscles confined to a single very limited region of space? The natural answer is: a wave motion of which the amplitude, instead of being uniform all over space, is everywhere zero except within that limited region.

This answer, however, leads straight to a paradox. Such a wave motion may be resolved (by a well-known, highly-developed, and much-employed mathematical technique) into an infinity—or, in certain cases, a finite number—of sinusoidal wave trains of suitably chosen wavelengths and amplitudes. These, however, have different phase speeds (in the cases of electricity and matter). Therefore arises the result of which I spoke in discussing the fate of a segment of a beam of light passing through a dispersive medium. The distribution-of-amplitude changes as time goes on. We may have assumed that at $t = 0$ the amplitude differs from zero only in a region of length L; but this cannot continue true; at other instants of time, the region in which the amplitude is other than zero must be broader than L (perhaps, in special cases, narrower).

Our wave picture is, therefore, unable to represent a flock of corpuscles which stays together forever within a strictly-limited volume (whether this volume be supposed to stand still or to move onward with uniform speed). Analyzing it further, we perceive the difficulty. The sinusoidal wave trains of different wavelengths, into which our initial distribution-in-

amplitude was resolved, correspond to different values of corpuscle momentum. They, therefore, correspond to different values of corpuscle speed. Our wave picture thus describes a situation, in which a crowd of corpuscles is confined at $t = 0$ to a limited volume of space, but the various corpuscles have various speeds, so that they run away from one another and the crowd spreads out in the course of time.

Moreover, the wave picture appears to be quite unable to describe a situation in which a crowd of corpuscles is confined to a limited volume of space, and also to a limited range of momenta and speeds. This is the paradox aforesaid. It might easily be taken to mean that the wave model is fundamentally incompetent. We can easily conceive a cloud of corpuscles all having the same speed and all located close together, and yet the wave picture refuses to represent it. This seems a fatal flaw of the wave theory. Heisenberg, however, took a startling way out of the dilemma affirming in substance that what the wave theory cannot describe, cannot exist. The conception of a cloud of particles all having nearly the same position and nearly the same momentum cannot be transcribed into waves; therefore—according to Heisenberg—it does not correspond to anything in reality.

This is an illustration of the so-called " principle of indefiniteness," or " principle of uncertainty."[1] By carrying the reasoning further, it can be shown that the more narrowly limited is the region to which the corpuscles are confined, the broader is the range of momenta which must be postulated in order to save the wave picture. It can further be shown that the product of the range in position (the breadth of the region) by the range in momenta is roughly constant. In more customary phrasing: the product of the indefiniteness in position by the indefiniteness in momentum is a constant. A rough sort of proof is the following:

[1] More commonly designated as "principle of uncertainty" or "principle of indeterminacy"; introduced into theoretical physics by Heisenberg, who called it *Unbestimmtheitsprinzip*. It is more extensive than the two illustrations here given will suggest.

Turn back to Figure 6.1 and consider a single "beat" between two consecutive minima of amplitude, imagining that all the other beats are blotted out. This is a distribution in amplitude describing a flock of corpuscles confined within the region of length D between the extremities of the beat. Repeating equation (6.4):

$$D = \lambda^2/\Delta\lambda, \tag{6.51}$$

where λ stands for either of the very nearly equal wave-lengths of the two sinusoidal wave trains whereof the overlapping produces the beat, and $\Delta\lambda$ for the difference between these wavelengths. D is the "indefiniteness in position," which denote by Δx; and the corresponding "indefiniteness in momentum" Δp is the difference between the values of momentum corresponding to the two wave-lengths. We thus obtain:

$$D \cdot \Delta p = \Delta x \Delta p = (\lambda^2/\Delta\lambda)\Delta(h/\lambda) = (\lambda^2/\Delta\lambda)(h\Delta\lambda/\lambda^2) = h, \tag{6.52}$$

so that the product of the two indefinitenesses is not only a constant, but is equal to the famous constant h. Because of the approximations and specializations made in this particular proof, we cannot accept this result too literally; the general conclusion is, that the product of the indefinitenesses is of the order of magnitude of h.

One more example of the principle must suffice in this place. Imagine a beam moving parallel to the x-direction, and of endless extent in all directions: that is to say, both the concentration of the particles and the amplitude of the waves are independent of x, y, and z. Now confine this state of affairs to the region $x < 0$, by supposing the plane $x = 0$ to be occupied by an opaque screen interrupted only by a narrow slit of width D, the central line of which runs along the y-axis. What exists in the region $x > 0$?

According to the simple corpuscular picture, the particles for which $\frac{1}{2}D < z < -\frac{1}{2}D$ will traverse the slit and continue onwards undeflected, while all the others are stopped by the screen. Therefore in the region $x > 0$ there will be a broad

flat beam of thickness D, limited by the planes $z = \frac{1}{2}D$ and $z = -\frac{1}{2}D$. But according to the wave picture, the beam will broaden steadily after it emerges from the slit. This is one of the phenomena of diffraction, observable in the cases of water waves and sound waves, and derivable from the mathematics of the wave theory.

The wave picture thus describes a broadening beam. This, in the corpuscular theory, would be a stream of particles possessing not only momentum along the x-direction, but also a distribution of transverse momenta along the z-direction. The undulatory theory seemingly cannot be fitted to a beam of finite breadth consisting of particles all moving parallel to one another. Again it will be affirmed (following Heisenberg) that what the waves cannot describe does not exist. A limitation on the breadth of a beam of electricity, matter, or light, entails the arising of transverse momenta among the corpuscles of the beam.

The distribution of intensity in the broadened beam is usually studied (in the case of light) by observing the intensity on a screen placed parallel to the screen containing the slit, at a distance from the latter many times as great as the wavelength. There are fringes parallel to the z-direction, with a central maximum of intensity at $z = 0$, and symmetrically placed maxima of first, second and higher orders, with minima between. It is well known that the narrower the slit, the more this pattern of maxima and minima widens out. Now all the particles lie within the slit as they pass the plane $x = 0$, so that the breadth D of the slit may be regarded as the initial "indefiniteness" Δz of the z-coordinate of the corpuscles. Similarly the breadth of the diffraction pattern on the distant screen may be regarded as a measure of the "indefiniteness" Δp_z which must be assigned to the transverse momenta p_z of the corpuscles, in order to preserve the wave picture. One of these two indefinitenesses varies inversely as the other, a fact which recalls equation (6.52). By making certain approximations, a relation resembling equation (6.52) can be derived, as follows:

Let us evaluate the transverse momentum which a particle must have, if from the slit it goes toward either of the minima adjoining the central maximum on the distant screen. (Actually there would be no particles going in just these directions, since the intensity at the minima is zero; but there will be corpuscles moving in directions differing from these by infinitely little.) Denote by θ the angle between the lines drawn from the slit to the central maximum and to either minimum. The former of these is the x-axis. As for the latter, denote by p_x, p_z the x and z components of the momentum of a particle travelling along it; as one readily sees:

$$p_z/p_x = \tan \theta. \tag{6.53}$$

The angle θ conforms to the equation:

$$\sin \theta = \lambda/D, \tag{6.54}$$

which is proved in all textbooks of optics.[1]

If now we suppose that the ratio of wave-length to slit-width is so small that $\sin \theta$ and $\tan \theta$ may be replaced by θ, and if we assume the first of the Rules of Correlation, we get:

$$p_z/p_x = \lambda/D = h/Dp_x. \tag{6.55}$$

The difference between the values of p_z corresponding to the two first-order minima which bound the central fringe on either side is thus equal to $2h/D$. If we agree to call this the " indefiniteness in transverse momentum " Δp_z, we obtain:

$$D\Delta p_z = \Delta z \Delta p_z = 2h. \tag{6.56}$$

This seems to differ from equation (6.52) by the presence of the factor 2. But, as before, we must not take this factor seriously. Had the argument been carried through with respect to the second or third-order minima instead of the first-order, we should have obtained 4 or 6 in place of 2. The sharpest

[1] If we divide the wave front in the slit into strips, these strips may be paired off in such a way that the wavelets proceeding along the direction θ from either strip of any such pair are completely annulled by those proceeding from the other of the pair, before they reach the distant screen.

permissible statement is, that the product of the indefinitenesses of z-coordinate and z-momentum is of the order of magnitude of h. The " principle of indefiniteness " is itself indefinite.

Principle of indefiniteness and breadth of spectrum lines

Return to equation (6.52); on the left, multiply the first factor and divide the second factor by c, obtaining:

$$(D/c)(c\Delta p) = (\Delta x/c)(c\Delta p) = \Delta t \Delta E = h. \qquad (6.57)$$

(Remember that the last equals-sign implies not an exact equality, but rather an order-of-magnitude agreement, or something even rougher.) We interpret c as phase-speed, so restricting the argument which follows to the case of *light*.

Again we are dealing with a beam composed not of one sinusoidal wave train but of two having different wave-lengths, and again we realize that this involves, on the one hand a variation in the wave-amplitude from a maximum down to zero over a distance $\frac{1}{2}D$, on the other hand a distribution-in-momentum of the associated corpuscles extending over a range Δp. We recall the inference: the members of a limited flock of corpuscles circumscribed within a length Δx may not all have the same momentum, but must exhibit a distribution-of-momentum extending roughly over the range $h/\Delta x$. Now we are denoting D/c by Δt, because it is the time which an entire " beat " of the wave-system, corresponding to a limited flock of corpuscles, takes to pass a fixed point of the beam; and we are denoting $c\Delta p$ by ΔE because the Rules of Correlation make the energy E of a corpuscle equal to pc. The new inference, or the old inference in the new form, is that the members of a limited flock of corpuscles passing a given point within a length of time Δt may not all have the same energy, but must exhibit a distribution-of-energy extending roughly over the range $h/\Delta t$.

Locate the " given point " at the very atom whence the light is proceeding. It seems as though we were forcing ourselves to conclude that if an atom is predestined to emit a great number of corpuscles within a finite time Δt, their energies

must be spread over a range $h/\Delta t$. But, as we shall see in later chapters, an atom is predestined to emit only one corpuscle of light of any particular kind. (It may, so to speak, be wound up afresh at a later time and restored to its original state, after which it may emit another photon; but this prospect is not taken into account by the waves.) The beam of waves emanating from the atom is correlated, not with a great multitude of corpuscles sprinkled throughout its extent according to its amplitude, but with one corpuscle only.

We deal with this situation by imagining a very great number of atoms simultaneously existing in the same state and each capable of emitting a photon; and we conclude that if they are all predestined to launch their photons within a time Δt, then these must exhibit a distribution-in-energy extending over the range $h/\Delta t$. Then, for any one atom among the multitude, the energy of the photon which it is predestined to launch may be anywhere in the range $h/\Delta t$. Both the time at which the emission is to occur, and the energy of the emitted particle, are indefinite; and the product of the two indefinitenesses is of the order of magnitude of h.

It is not, however, assumed that the waves emanating from an atom form a sharply limited group; it is assumed that they constitute a continuously damped wave train, with an amplitude which decreases exponentially as time goes on, so that the displacement at the "given point" is described not by $A \cos 2\pi(\nu_0 t - a)$ or $A\ exp\ 2\pi i(\nu_0 t - a)$ as in equation (6.1), but by $A\ exp(-t/T) \cos 2\pi(\nu_0 t - a)$ or $A\ exp[-(t/T) + 2\pi i(\nu_0 t - a)]$. Such a wave-motion is a sum of ideal sinusoidal wave trains, their frequencies extending strictly over an infinite range. In the corresponding corpuscular image we have two infinities to deal with, that of the time over which the emission of the photon from any atom may occur, and that of the energy-range within which the energy of the photon may lie. Yet of all the sinusoidal trains the one with the greatest amplitude is the one with the frequency ν_0, and the others are of progressively smaller amplitudes the further their frequencies depart from ν_0. This means that of all the energy-values the

one which the photon is likeliest to have is $h\nu_0$, and the others are progressively less probable the more they depart from $h\nu_0$. Moreover the greater the "relaxation-time" T, the more unimportant become the component wave trains of which the frequencies differ greatly from ν_0, and the more improbable become the energy-values which depart considerably from $h\nu_0$. Thus, though any damping however small entails an infinite range of photon-energies, yet practically the probable values—or at least, the not grossly improbable values—of E become more and more concentrated around $h\nu_0$, the smaller the damping is made and the longer the emission continues. We may consider the relaxation-time as a rough measure of the indefiniteness in time of emission of photons, and the energy-range comprising (say) the fraction $(1 - 1/e)$ of the photons as a rough measure of the indefiniteness in their energy; then of these two indefinitenesses T and ΔE, the one increases as the other decreases. By analyzing the damped wave train by the Fourier method into its components, it may be shown that the product of these indefinitenesses is constant and of the order of magnitude of h:

$$T\Delta E = h. \tag{6.58}$$

This equation is reached by way of another characterizing the wave-image separately, with which I close this chapter. The photons constitute a spectrum-line which is not ideally sharp and narrow, but has a breadth which is often perceptible. If we define the "half-breadth" B of the line as the difference between the two frequencies, at each of which the intensity of the light is one-half what it is at the center of the line, we find from the Fourier analysis,

$$TB = 1, \tag{6.59}$$

from which we proceed to (6.58) by associating the corpuscles with the waves according to the second Rule of Correlation.

We turn now to the study of the phenomena of diffraction, by means of which the first of the Rules is tested for electrons and atoms.

CHAPTER VII

Diffraction of Waves by Crystals

To diffraction by crystals men owe the first clear proof that beams of X-rays are undulatory; the first and to this day the only direct proof that beams of electrons are undulatory; and the major portion of the vast contemporary knowledge of the arrangement of atoms in crystals, especially the finer details thereof. The last of these contributions is by far the most voluminous, forming already a very large part of the science of crystallography. To enter into this mighty subject would be to distort the programme of this book, in which the main importance of diffraction lies in the evidence which it furnishes for the wavelike aspect of electricity and of light, and the values of wave-lengths with which it supplies us. Yet it would be almost intolerable to omit all reference to the service of diffraction in elucidating crystal structure; and accordingly this chapter will include (and largely consist of) a description of the four great techniques for adjusting beam and crystal so as to produce intelligible diffraction, and the elements of the notation for the principal planes and directions in a lattice. We shall confine ourselves to the cubic lattice, thus simplifying the task but avoiding a large number of the most interesting types of crystal structure. At the end we shall observe how the measurements upon electron-beams and atom-beams confirm the first of the Rules of Correlation of Chapter VI.

This episode in the history of physics began on a day in 1912 when von Laue and Ewald at the University of Munich were discussing what might happen to *long* waves traversing a lattice capable of scattering them, and it occurred to the former to inquire what should be expected if the waves were

shorter than the spacing of the lattice. Von Laue predicted such a diffraction-pattern as we shall presently be studying, and proposed the test to his experimental colleagues Friedrich and Knipping, who undertook it with X-rays. Many crystals disclose their lattices because of the angles between the facets and the cleavage-planes which are natural to them. It was therefore easy for Friedrich and Knipping to select a lattice known in advance to be cubic. As for the spacing, they could not know it in advance, for the facets and the cleavage-planes do not reveal how many atoms there are to the unit-cell. It turned out however that the spacings of their crystals and the wave-lengths of their X-rays were so proportioned to each other, that not only did diffraction occur but the scale of the diffraction-pattern was superbly convenient to observe; and this is true of crystals generally and of all X-rays excepting those produced by apparatus of immensely high voltage. As for electron-beams, the first to observe diffraction with these were Davisson and Germer very early in 1927 at the Bell Telephone Laboratories,[1] following on whose discovery came shortly the independent researches of G. P. Thomson at Cambridge. Since at this time the first of the Rules of Correlation had already been proposed by L. de Broglie, while the spacings of crystal lattices had been established closely by the work with X-rays, the predictions to be tested were quite definite. Ever since, the measurements of electron-beam wave-lengths have been made for the sake of testing the rule, while X-ray wave-lengths have been measured for the purposes of X-ray spectroscopy. As for material particles more massive than electrons, the scanty evidence concerning diffraction of these (all of it less than ten years old) will be mentioned at the end of the chapter.

I need not re-define "lattice" or "unit-cell" (for the latter I shall often be using "atom-group" as a synonym), having spoken of these concepts in Chapter I. The atom-groups, denoted in Figure 1.4 (page 12) by spheres, are represented

[1] Their research upon diffraction by single crystals of metal was an outgrowth of the work on the scattering of electrons by polycrystalline masses of metal, described in Chapter V of the first edition of this book.

in later figures (Figures 7.1 and 7.2) by perfectly unsymmetrical outlines similar to large commas, after the usage of Bragg. This is meant to suggest their customary lack of perfect spherical symmetry. In many cases the asymmetry is such that the commas should be unsymmetrical in three dimensions; the student might conceive of them as smooth on one side, rough on the other.

Now to visualize the diffraction-pattern, let us conceive the crystal set at the centre of a transparent bulb painted inwardly with fluorescent stuff, so that wherever X-rays or electrons fall upon its inner surface it will shine. The crystal should be very small in comparison with the bulb, and the incident beam of waves extremely narrow. We suppose that this beam comes in at a window, and follows a diameter of the bulb, and the unscattered part after flowing through the crystal goes out through another window. If the rays are plane-polarized X-rays, the electric vector will remain constantly parallel to some direction at right angles to the beam. If they are unpolarized X-rays, the electric vector will run or swing rapidly around in the plane at right angles to the beam. Thus far it is customary to use in crystal analysis X-rays which either are unpolarized, or have the slight degree of polarization usually imprinted on such rays at their excitation in a discharge-tube or a Coolidge tube.[1]

This "imaginary bulb" is very nearly realized in practice, although instead of a fluorescent screen it is customary to use a photographic film on which the imprints of the diffraction beams are permanent spots. The film is usually flat instead of spherical, so that the rings presently to be mentioned are liable to be distorted. In many experiments an ionization-chamber (for X-rays) or a Faraday chamber (for electrons) is swung around in arcs over a spherical surface centered at the crystal, and the current which it reports is plotted as a function of its position; the curves then display peaks wherever the chamber is so placed as to capture a beam. In what fol-

[1] Because the electrons which produce the rays in falling onto the target of the tube are all moving along parallel lines when they impinge upon it.

lows I shall use the terms "diffraction beam," "diffraction peak" and "diffraction spot" almost as synonyms.

On the inner coating of the bulb, then, we shall see the diffraction-pattern of the crystal. It will be an assemblage of luminous spots arranged in a symmetrical array recalling the symmetry of the crystal lattice. In making this statement I am anticipating what is presently to be proved, or rather to be deduced from the fact that the atom-groups of the crystal are marshalled on a lattice. Indeed, if the incident waves are monochromatic or nearly so, we shall not see even spots unless by happy accident or by a careful choice of wave-length. Most waves are not diffracted by a three-dimensional crystal lattice. If we could reduce our crystal to a single plane of atom-groups forming a two-dimensional network, there would be a pattern of spots for any wavelength whatever. If we could isolate a single row of atom-groups, there would be rings of light on the coating of the bulb, instead of only spots. And if we could remove all but one of the groups, and then in some magical way magnify the luminescence which the waves that this survivor scatters produce at the wall of the bulb, then we should see the wall shining all over with a continuous brightness, sinking to zero perhaps nowhere, perhaps at occasional points.

Reverse the process, starting from the solitary atom-group. The intensity of the scattered waves of which it is the source varies continuously with direction, and the brightness of the wall of the bulb varies correspondingly from point to point. If this brightness is proportional to that intensity, its distribution over the spherical wall is the scattering-pattern or diffraction-pattern of the atom or group of atoms. However, it is immeasurably too faint to see.

Add now to the solitary group a great number of similar and similarly oriented groups of atoms, forming altogether a long and evenly spaced row. The luminescence now fades out everywhere except over certain rings upon the wall, but along these rings it is enhanced. The newly added atom-groups have amplified the waves scattered by the original one along

the directions leading to these rings; in compensation they have effaced those scattered in other directions. These are effects of interference.

Add next a great number of such rows of similar and similarly oriented groups, forming altogether a lattice in a plane. The rings now all fade out except for occasional spots, which however are much brighter than they were before. Interference between the waves scattered from the various atom-groups has exalted the brightness of certain points on the wall of the bulb, to the detriment of all the rest; it has amplified the diffraction-pattern of the single group in certain directions, and destroyed it in the others. The spots due to the atoms of a single plane have been observed with electron-waves, but never with X-rays.

Add finally a great number of such planes, thus building the three-dimensional crystal. Now except for certain wave-lengths the spots vanish altogether. For those exceptional waves, however, they are intensified into visibility. One group of atoms by itself would have produced a complete diffraction-pattern—at least we suppose that it would—but never one intense enough to be perceived. But when it is joined with an enormous number of its peers in a lattice, they all conspire to enhance not indeed the entire pattern, but the intensities at certain of its points. The physicist, if he varies the direction in which the rays fall upon the crystal or the wave-length of the rays or both, can observe a great number of these intensities which the crystal has so obligingly amplified for him; and out of them he can reconstruct the entire diffraction-pattern of the individual atom, which but for this amplification would have been forever out of his reach. [1]

I must not leave the impression that the diffraction-pattern of the atom-group is amplified equally in all the directions in which the lattice amplifies it at all. Amplification depends on

[1] If the atoms or atom-groups of a substance would orient themselves all alike without at the same time spacing themselves at regular intervals, and without being too much crowded together, the entire pattern would be amplified instead of certain spots only.

direction. If the observer sees two spots produced by diffraction-beams inclined say at 45° and at 60° to the primary beam, he is not to infer that the ratio of their brightnesses is the ratio of the intensities of the waves scattered at 45° and at 60° by the individual group. A correction factor must be employed to translate one ratio into the other. Not forgetting it but not taking it into account, we will calculate the directions in which amplification occurs.

We start as before from the solitary atom-group. In Figures 7.1 and 7.2 this is depicted as a two-dimensional figure, irregular and utterly without symmetry. It ought to be conceived as a three-dimensional, perfectly unsymmetrical mass. This depiction is meant to imply that if a stream of waves strikes the atom-group on any side, the intensity of the waves scattered in any direction—what above I called the diffraction-pattern of the group—is or may be a thoroughly unsymmetrical function of direction. Also it may depend on the side of the group on which the primary waves impinge, or, in better words, on the direction whence they come. Of course, in any particular case, it might turn out to be a symmetrical function of direction, or it might be the same function whichever the side of the atom which the primary waves encounter; but we should not rely in advance on either of these simplicities. I must qualify this statement somewhat: the crystallographers have their ways of learning more or less (and sometimes a good deal) about the symmetry of the atom-group, and thus foretelling certain aspects of its diffraction-pattern. Moreover in many cases it is possible to make in advance a good estimate of the absolute intensity of the scattered waves. It will do no harm to remember these encouraging facts; nevertheless it will be best for us to keep our ideas fluid by supposing an atom-group of absolute asymmetry, the scattering-pattern of which is one of the ultimate objects of the quest.

I must state with all emphasis that whatever may be foretold or measured in the scattering-pattern for electromagnetic waves (X-rays) need not always be valid in the scattering-pattern for electron-waves.

Consider now the amplification which ensues when other atom-groups are added to the first one, to form with it an evenly spaced row.

Denote by a the distance between corresponding points in adjacent groups (Figure 7.1). Suppose the primary waves to emanate from a point P, while the observation of the scattered intensity is made at a point Q (on the wall of the bulb, to continue the picture). In practice the distance a is so submicroscopically small, that P and Q are practically infinitely far away; yet it is easiest to begin by thinking of them as nearby

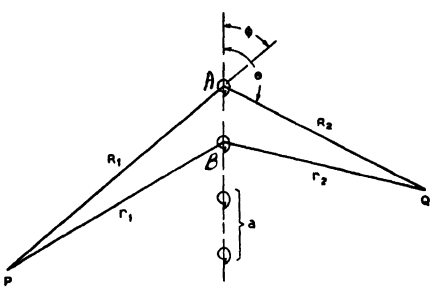


FIGURE 7.1—ILLUSTRATING DIFFRACTION BY A SINGLE FILE OF ATOM-GROUPS OR UNIT CELLS.

and passing to the limit. Designate then by R_1, R_2 the distances from any atom-group A to P and Q, by r_1 and r_2 the distances to these points from the adjacent group B. We shall not assume that R_1, r_1 are coplanar with R_2, r_2, though owing to the flatness of the page, the drawing must make them seem so. Denote by φ, θ the angles between the direction of the row of atom-groups and the directions in which the primary and the scattered waves advance, respectively—these latter being the directions from A *away from* P and *towards* Q, respectively.

Under what condition will the waves scattered by the atom-groups A and B reinforce one another best at Q? Best reinforcement will occur, greatest enhancement of the effect of either atom by the presence of the other, when the waves from

both arrive at Q in identical phase. This will occur when Q is so located that the path from P to Q *via* A either is equal to the path from P to Q *via* B, or differs from it by an integer number of wave-lengths.[1]

$$(R_1 + R_2) - (r_1 + r_2) = (R_1 - r_1) + (R_2 - r_2) = n\lambda;$$
$$n = 0, \pm 1, \pm 2, \ldots \qquad (7.1)$$

Let P and Q recede to infinity; in the limit:

$$R_1 - r_1 = a\cos\varphi, \qquad R_2 - r_2 = -a\cos\theta \qquad (7.2)$$

and the condition for optimum reinforcement at Q is this:

$$a(\cos\theta - \cos\varphi) = n\lambda. \qquad (7.3)$$

In all directions for which θ satisfies this equation, there will be maximum amplification of the diffraction-pattern of A (or B) by the presence of B (or A). For every such value of θ, there will be a ring on the wall of the bulb. If the two atom-groups by themselves could make the wall fluoresce brightly enough, we should see annular fringes. They would be broad and hazy, for though the cooperation between the scattered waves is best at the definite angles determined by equation (7.3), it is also good over quite a range of nearby angles. Equation (7.3) would give the locations of the central rings of the broad fuzzy bright fringes. Thus, when visible light is sent through a pair of parallel similar slits in a screen, one sees hazy fringes superposed on the diffraction-pattern which either slit by itself can produce; and the formula analogous to equation (7.3) locates the central lines of these.

This allusion to parallel slits in a screen will simplify the next steps. It is well known that when to a pair of parallel slits, new ones just like them are added at equal intervals one after the other, the fringes are not displaced. The bright fringes shrink, the dark ones widen, but their central lines remain unshifted. As more and more slits are added, as more

[1] This statement implies the tacit assumption that there is a constant phase-difference (whether it is zero or not is of no importance) between the primary waves striking an atom-group and the scattered waves leaving it—the very important assumption of *coherence*.

and more lines are ruled on a metal surface to constitute a grating, the dark fringes encroach steadily on the bright ones, and it becomes easier to locate the central lines of these latter with precision. As they grow narrower, they brighten, the energy which was lavished over a wide angular range being gathered into a small one as it is progressively increased by the addition of new slits or rulings. So there is a double gain. In the limit, nothing remains but the central lines, and these are brilliant. And in the limit, these lines are still located where the formula derived for only two slits predicted that the maxima of brightness should be found.

Now in the same way, when to a pair of like and likewise-oriented atom-groups additional such groups are added so as to form an evenly spaced row, the annular fringes on the wall of the ensphering bulb are not changed in location but in distinctness. The bright fringes contract into brighter rings, the dark ones broaden into (relatively) dark bands. *The multitude of the atoms sharpens the diffraction-pattern.* In the limit, the bright rings are very sharp and brilliant, and they are still located at exactly the angles predicted by equation (7.3) derived for two atom-groups only. Remember however that a ring may not be equally bright all around its circuit. It is only an enhancement of the scattering-pattern of the individual atom-group; and this in general will vary from one point to another.

So much for the single row or file of groups of atoms! We must now pass to two dimensions, and predict the diffraction by a plane in which groups are arranged in a network. In the plane, rows of atoms lie side by side at equal spacings. We might start with one of the rows, and estimate how its diffraction-pattern is amplified by the cooperation of a second and then a third and a fourth and eventually an infinite number of added rows laid parallel with it at equal intervals. Owing to this equality of intervals and the new periodicity which it entails, the diffraction-rings of the pattern of the individual rows will be amplified not uniformly, but only at certain points; precisely as owing to the equal intervals between the atom-

groups of the single row these amplified the pattern of the individual group not uniformly, but only over certain rings. However we can reach this result in another way, by considering three atom-groups forming a triangle, as formerly we considered two forming a pair.

Start then as heretofore from a solitary atom-group A, and add to it two others B and C (Figure 7.2). They must not all three be collinear; as a rule it is best that B and C should be the two groups nearest A.[1] As before P stands for the source of the primary waves, Q for the point (on the wall of our imaginary bulb) where the scattered waves are to be measured. The question is: under what condition do the scattered waves from A and B and C all three reinforce one another best at Q? Under what condition do the waves from all three groups arrive at Q in identical phase?

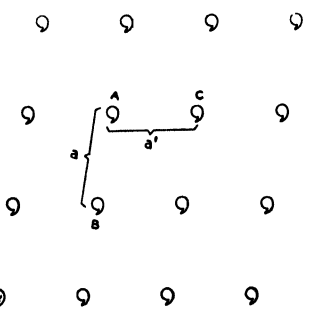


FIGURE 7.2—ILLUSTRATING DIFFRACTION BY A PLANE OF ATOM-GROUPS OR UNIT CELLS IN REGULAR ARRAY.

Evidently we have only to restate the condition that the waves from A and B arrive in identical phase, and supplement it with one exactly like it for the waves from A and C. We have only to repeat equation (7.3), and beside it write another like it in which a is replaced by a', the distance from A to C; φ by φ', the angle between the direction of the primary waves PA and that from A to C; and θ by θ', the angle between the direction of the scattered waves AQ and that from A to C. When the waves from all three atom-groups reinforce one another, both equations prevail:

$$a(\cos\theta - \cos\varphi) = n\lambda, \qquad (7.3)$$

$$a'(\cos\theta' - \cos\varphi') = n'\lambda \qquad (7.4)$$

[1] The choice of groups to serve as B and C is purely a question of expediency. The same results are reached whichever two we choose, so long as they are not collinear with A; but the results when reached are in a form which depends upon the choice, and is most convenient when AB and AC are parallel to the crystallographic axes in the plane.

In these equations n and n' stand for integers but they need not stand simultaneously for the *same* integer. In all directions for which θ and θ' satisfy equations (7.3) and (7.4), there will be maximum amplification of the diffraction-pattern of any atom-group by its pair of neighbors.

Now equation (7.3), with various integer values of 0,1,2, . . . , substituted for n one after the other, described a system of rings on the wall of the bulb—parallel rings like latitude-circles, with the poles at the points where the bulb is intersected by the line drawn through its center parallel to AB. Likewise equation (7.4), with various integer values for n', describes a system of rings oblique to the first, having *its* poles at the points where the diameter drawn parallel to AC through the center of the bulb reaches the wall. There are intersections between the rings of the two systems, and these intersections are the point—the discrete, the finitely numerous points—where the diffraction-pattern of the atom-group is most greatly amplified. So long as there are but three of the groups, these points are merely the centers of broad, hazy, bright (of course, in practice, utterly invisibly dim) blotches. But when to the three groups we add enormously many others to form the extensive two-dimensional network of which a section is depicted in Figure 7.2, interference eats away the edges of these patches and enhances their centers, and in the limit nothing remains of the diffraction-pattern except brilliant dots at the intersections of the rings.

The next and last step follows immediately. The crystal is built up of planes of atom-groups laid parallel to one another at equal intervals. Suppose the space above and below the plane of Figure 7.2 to be thus stratified; select, from the stratum just above or just below that figured on the page, an atom-group close to A, preferably the closest. Call it D; let a'' stand for the distance from A to D, φ'' for the angle between the direction in which the primary waves advance and the direction $\vec{AD}$, θ'' for the angle between the direction from A towards Q and that from A to D. When the scattered waves

from all *four* groups ABCD reinforce one another best at Q, all these three equations are valid simultaneously:

$$a(\cos\theta - \cos\varphi) = n\lambda, \tag{7.3}$$

$$a'(\cos\theta' - \cos\varphi') = n'\lambda, \tag{7.4}$$

$$a''(\cos\theta'' - \cos\varphi'') = n''\lambda, \tag{7.5}$$

n'' standing for a third integer, which may or may not be the same as either of the other two. In all directions for which θ, θ', θ'' conform with equations (7.3), (7.4), and (7.5), there will be maximum amplification of the scattering-pattern of any atom-group by its triad of neighbors.

Now in thus adding a third equation to the previous one and two, we have made the conditions so severe that save in exceptional cases they are quite unfulfillable. Equations (7.3) and (7.4) confined the amplification-effects for which we seek to the points of intersection of a few rings belonging to two families, oblique to one another upon the wall of the bulb. Equation (7.5) supplies a third family of rings oblique to both, having for its poles the points where the diameter parallel to AD reaches the wall of the bulb. Agreement of phase between the waves scattered from A and B and C and D, optimum amplification, can occur only if and where a ring of the third family cuts rings of the first and the second just where these happen to cut one another. And when the set of four atom-groups ABCD is repeated over and over again in an extensive crystal lattice, these points of optimum amplification are the only points where the amplification is great enough to enhance the diffraction-pattern into visibility. Visible luminous spots will appear on the coating of the wall of the bulb, only if three rings, one from each family, happen to intersect at the same point—*only by coincidence*, in the popular sense of the word.

To bring about such a coincidence, there are four practicable ways.

I. If the incident beam is monochromatic, or comprises a narrowly limited range of wave-lengths, the observer can rotate

the crystal—presenting it to the beam under varying aspects, and so in effect varying the direction of incidence (the angles φ of the equations). Now and then the desired coincidence occurs, in a direction which depends on φ. To catch the diffracted beam one must accordingly use an ionization-chamber or collector which can be freely moved along an arc of the surface of the imaginary bulb, or else a photographic film occupying a part of that surface. If the incident rays are not monochromatic the experimenter thus obtains their spectrum (Figure 14.4, page 433.) This method is known by Bragg's name.

II. If the incident beam is monochromatic or nearly so, and its wave-length can be varied continuously, the observer can keep the crystal motionless and the direction of incidence constant, and yet count on the coincidence turning up occasionally (Figures 7.6, 7.7).

This is the method used with electrons by Davisson and Germer, the method whereby it was first shown that free negative electricity has a wavelike quality. To waves of this kind it is especially adapted, as their wave-lengths can easily be varied by varying the voltage-rise which endows the electrons with their speed. A Faraday chamber is used as collector; as in Bragg's method, it must be capable of swinging in an arc over the surface of the imaginary bulb, so as to catch the diffracted beam, of which the direction will vary with λ. Figure 7.3 represents an apparatus of Davisson and Germer; C is the Faraday chamber, the arc along which it is moved is obvious; the orientation of the crystal T with respect to the electron-beam proceeding from the electron-gun at the left, is adjustable. Figures 7.6 and 7.7 show curves obtained with this apparatus.

III. If the incident beam is a mixture of wave-lengths covering a very wide range, we can keep the crystal motionless and the direction of incidence constant, and yet count on the coincidence turning up for some of the wave-lengths (Figures 7.4, 7.5).

This is Laue's method, the first to be applied to the analysis of crystal structure, and the one whereby it was first proved that X-rays are waves—as people said at the time, which was

1912. Nowadays we say that it was proved that X-rays are an example of the correlation of corpuscles and waves.

To make the spots appear, a fluorescent coat or a photographic film is spread on a flat screen instead of the spherical bulb. The locations on the screen can be deduced from those on the imaginary bulb by simple projection. Different spots are likely to be due to components of different wave-length in the primary beam, which will probably not be equally intense

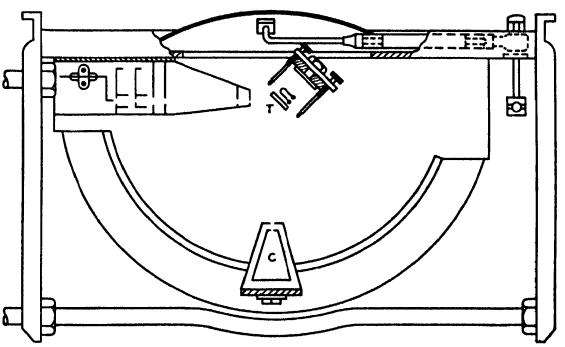


FIGURE 7.3—ILLUSTRATING HOW A CRYSTAL (T) IS MOUNTED SO THAT IT MAY BE ORIENTED AT WILL TO THE BEAM OF ELECTRONS EMERGING FROM THE ELECTRON-GUN ON THE LEFT, WHILE THE COLLECTOR C MAY BE SHIFTED ALONG THE ARC TO FIND THE DIFFRACTION-BEAMS. (Davisson and Germer.)

—a thing to be remembered when deducing the diffraction-pattern of the atom-group.

IV. If the incident beam is monochromatic or nearly so we can present to it not a single stationary crystal but a confused mass of tiny crystals oriented in every way whatever. The desired coincidence will certainly occur for some among the crystals.

This is the "powder method" invented and applied to X-rays by A. W. Hull and by P. Debye and P. Scherrer and applied to electron-beams by G. P. Thomson. The term implies that the chaotic mass of little crystals is obtained by pulverizing a large one; but small pieces or thin films of ordinary metals are likely to present quite as complete a chaos.

The diffraction-pattern when formed on the wall of the bulb or on a flat screen set normal to the direction of the primary waves, consists not of separate spots but of continuous rings (Figure 7.8).

Such in outline are the four great methods for the analysis of waves by crystals and of crystals by waves. Each has its own field; each is beautifully adapted to certain problems as nature or art present them, not so useful or altogether useless for others. So, for instance, the first is enormously the best for the spectroscopy of X-rays; the second is the outstanding method for long electron-waves (of the order of an Angstrom) and thin films of matter; the third is very useful in the analysis of reasonably large crystals formed by intricate chemical compounds, and the fourth is invaluable for substances like the metals of which the crystals are often very small and tangled up together. But the fourth does not carry the investigator so far into the delicate details of crystal structure, as do the third and the first, when large crystals are available; the third and the first are impotent, when large crystals are not to be had; and the second would be exceedingly toilsome with X-rays. The complete crystal-structure laboratory now contains apparatus for the first method, the third, and the fourth. Perhaps it will not be long before the second also is demanded.

Before entering into details I wish to comment on four assumptions which have crept unsignalized into these deductions.

(i) *The assumption of the unlimited perfect crystal.* As I have already said sufficiently, a crystal composed of only a few atom-groups would produce broad hazy spots instead of sharp ones, as a grating with only half-a-dozen rulings would cast a spectrum of indistinct wide fringes instead of fine sharp "lines." It takes a multitude of atom-groups or rulings to produce the efficient destructive interference which etches out the borders of these blotches and leaves the centers standing up in high relief—which in technical language makes the resolving-power high. In actual crystals, are there atom-groups enough?

In actual crystals, there are usually plenty. Infinite resolv-

ing-power or infinite sharpness of diffraction-pattern would of course imply infinitely many groups arrayed at perfectly regular intervals; but infinity and perfection are not workable ideas in physics. The practical question is, whether the actual defects of the diffraction-pattern from infinite sharpness arise because the crystal lattices are not prolonged enough, or from other causes. Usually they are due to other causes, which are numerous; for instance, appreciable breadth of the primary beam and appreciable size of the crystal. In the rare cases

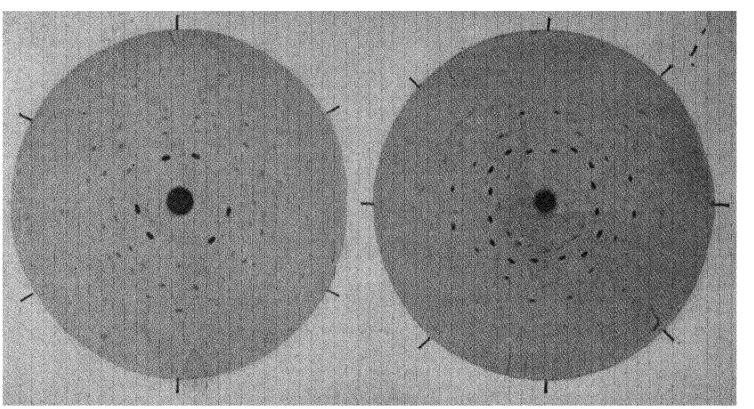


FIGURE 7.4—THE HISTORIC FIRST LAUE PATTERNS OBTAINED BY FRIEDRICH AND KNIPPING WITH A CRYSTAL OF ZINCBLENDE ORIENTED IN TWO WAYS RESPECTIVELY TO A BEAM OF X-RAYS COMPRISING A BROAD RANGE OF WAVELENGTHS.

where the fuzziness of the diffraction-spots betokens that the crystal lattice is limited, the fact is often of scientific importance. The size of exceedingly small—submicroscopic—crystals can be determined in this way; the dimensions of colloid particles, and of the crystals in metals, are estimated thus.

(ii) *The assumption of stationary atoms.* This of course is faulty, for the atoms are in thermal agitation. Their oscillations make the spots of the diffraction-pattern somewhat hazy, and alter furthermore their relative intensities. Out of this circumstance the physicists have drawn some profit; they have

estimated the amount of the thermal agitation and determined how it alters as the temperature goes down. It decreases, of course; not however in such a way as to imply entire standstill at absolute zero, but quite the contrary.

(iii) *The neglect of absorption.* Since the energy of the scattered waves is drawn from that of the primary beam, this cannot retain its amplitude unaltered as it progresses through a crystal lattice; and calculations based on the assumption that all the atom-groups of a crystal scatter equally cannot be cor-

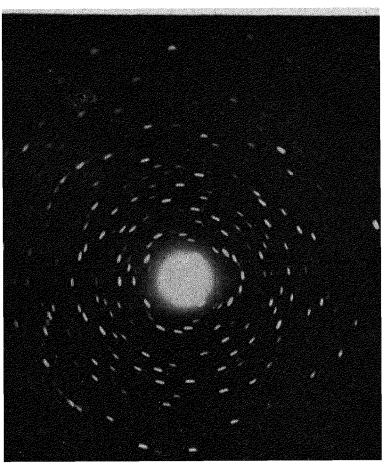


FIGURE 7.5—LAUE PATTERN OF X-RAYS OF A MICA CRYSTAL. (Bozorth.)

rect, for they do not all have incident waves of the same amplitude to scatter. This might well be serious, in a large crystal. It turns out however that large crystals are seldom if ever perfect; instead, they are likely to consist of smaller crystals tilted with respect to one another. The tilts are very small, but they are sufficient to suspend the consequences which should strictly follow from the assumption that absorption may be neglected.

(iv) *The neglect of refraction.* It has been tacitly assumed that there is neither bending of path nor change of wave-length when primary waves enter a crystal lattice or scattered waves emerge. Refraction however entails both of these phenomena;

and refraction will in general occur. However with X-rays it is so slight (the refractive index is so nearly unity) that it need seldom be allowed for in crystal analysis, and must indeed be looked for with care and skill if it is to be detected. The like is true for short electron-waves. With the long electron-waves first studied, however, the refraction is considerable; and during the interpretation of the earliest data, there was serious confusion.

We will now work out the details of the diffraction-pattern of a *cubic* lattice. This kind of lattice is much the easiest to treat, and Nature has graciously bestowed it on many of the commonest crystals. It must however be emphasized that *never* is the unit-cell a single atom—in elements it comprises two or four atoms, in compounds it comprises two at least and often many more. The arrangement of the atoms in the unit-cell affects the relative intensities of the spots or rings of the diffraction-pattern; I mention this effect again (page 224).

In the cubic lattice, the nearest neighbors of any atom-group lie at equal distances from it along three directions at right angles to one another. Moreover any atom-group in the entire (perfect) lattice can be reached from the one first chosen by a sequence of three displacements, one along each of the three directions and each an integer multiple of the minimum distance, or "spacing," or "grating-constant," or "edge of the unit cube." In other words, any atom-group can be brought into coincidence with any other by a sequence of three such shifts. This way of putting it brings out the point that all the groups in the lattice are oriented alike.

Denote by a the magnitude of the spacing. Install a system of rectangular coordinates with its origin at some one atom-group, say A of our previous picture, and its axes along the three directions stated. Among the six neighbors of A we pick out three to serve as B, C, D of our previous picture; say the three groups shifted from A through the interval a in the *positive* senses of the three axes, so that the coordinates of the four shall be:

$$A(0, 0, 0); B(a, 0, 0); C(0, a, 0); D(0, 0, a).$$

The directions of the diffraction-spots are to be deduced from the general equations (7.3), (7.4), and (7.5), with all the simplifications from which we benefit, thanks to the lattice being cubic. The three spacings are now all equal; but the greatest advantage is, that the various cosines which figure in the equations are now direction-cosines, and we can avail ourselves of the theorems to which direction-cosines conform. There are two of these which we shall use: the theorems that the sum of the squares of the direction-cosines of any line is unity, and that the cosine of the angle between any two lines

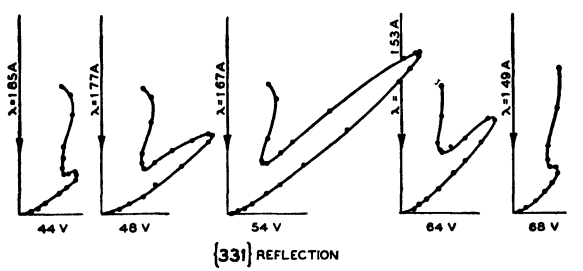


FIGURE 7.6—GRADUAL APPEARANCE AND DISAPPEARANCE OF A DIFFRACTION-BEAM OF ELECTRON-WAVES AS THE WAVELENGTH PASSES THROUGH ONE OF THE VALUES SATISFYING EQUATIONS (7.6). (Davisson and Germer).

is the sum of the three products formed by multiplying together corresponding direction-cosines of the two. Denote by α_1, α_2, α_3 the direction-cosines of the primary, by β_1, β_2, β_3 those of the diffracted beam. The first three are the quantities cos φ, cos φ', cos φ'' of equations (7.3), (7.4) and (7.5); the second three are cos θ, cos θ', cos θ''. To bring the notation into harmony with usage I further write h_1, h_2, h_3 for the integers n, n', n''.

Then by translating equations (7.3), (7.4), and (7.5) into the new notation and adding two more supplied by the first of the foregoing theorems, we form a family of five equations:

$$a(\beta_1 - \alpha_1) = h_1\lambda, \qquad (7.6a)$$

$$a(\beta_2 - \alpha_2) = h_2\lambda, \qquad (7.6b)$$

$$a(\beta_3 - \alpha_3) = h_3\lambda, \tag{7.6c}$$

$$\alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1, \tag{7.6d}$$

$$\beta_1^2 + \beta_2^2 + \beta_3^2 = 1. \tag{7.6e}$$

They involve seven variables: the wave-length, the direction-cosines of the primary or oncoming beam, the direction-cosines of the scattered or outgoing beam.

Now that we have this family of equations, the features of the four great methods of crystal analysis can be restated in few words. In the methods of Laue and of Davisson (II and III) the crystal and the primary beam are fixed in space, which amounts to prescribing values for α_1, α_2, α_3; then only four equations are left (7.6*d* has fallen out) and they contain four variables (λ, β_1, β_2, β_3) so that diffraction-spots can appear only for special sets of values of these four, to be obtained by solving the equations. In the original method of Laue, a wide range of values of λ is constantly provided, and the screen extends over a wide range of values of β_1, β_2, β_3; consequently the chance of observing spots is good. In the method of Davisson and Germer the different values of λ are realized in succession and a movable Faraday chamber searches out the peaks—a process much more long-drawn-out, but whenever one finds a peak one knows the wave-length to which it is due. With the other two methods the value of λ is prescribed, and consequently one at least of the direction-cosines α must be variable; therefore the crystal with its implanted coordinate-frame must be revolved, or else a multitude of crystals oriented every way must be placed in the path of the incident beam.

In the foregoing passage it seems as if I had taken for granted that the integers h_1, h_2, h_3 have fixed unchangeable values. As a matter of fact they may be any three integers at all. Strictly speaking, there is a different quintet of equations for every conceivable triad of integral values of the "indices" h. One might infer that in Laue's experiment the screen would be found completely covered with spots due to all the different triplets. However it turns out that only the spots for which

all the integers are small stand out strongly enough to be seen. Meanings for these integers must now be found; but before finding them I will deduce two more equations out of the quintet.

Squaring and adding the left-hand members of equations (7.6*a*, 7.6*b*, 7.6*c*), doing the same with the right-hand members, equating the sums and substituting from (7.6*d*, 7.6*e*), we obtain:

$$2 - 2(\alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3) = \frac{\lambda^2}{a^2}(h_1^2 + h_2^2 + h_3^2). \qquad (7.7)$$

Now by the second of the theorems concerning direction-cosines, the quantity in parentheses on the left is none other than the cosine of the angle between the direction of advance

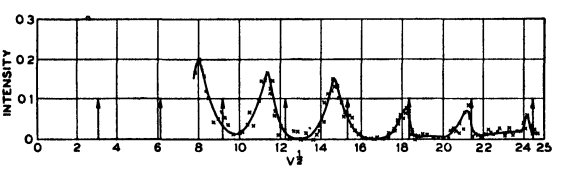


FIGURE 7.7—DIFFRACTION-PEAKS OBTAINED WITH FIXED CRYSTAL AND FIXED COLLECTOR (Φ CONSTANT) BY VARYING THE WAVELENGTH. (Davisson and Germer.)

of the primary beam, and the direction of the scattered waves which go to form the spot or diffraction-maximum of the indices h_1, h_2, h_3. If we conceive the diffraction-beam as the path of a portion of the energy which came with the primary stream and was deflected out of it, then this is the angle of deflection. Call it Φ. We have:

$$1 - \cos\Phi = \frac{\lambda^2}{2a^2}(h_1^2 + h_2^2 + h_3^2), \qquad (7.8)$$

$$\sin\frac{1}{2}\Phi = \frac{\lambda}{2a}\sqrt{h_1^2 + h_2^2 + h_3^2}. \qquad (7.9)$$

As the reader will observe, there is no allusion here, explicit or implicit, to the orientation of the crystal. This is therefore the appropriate equation for the "powder method," in which crystals turned every way are presented all together to the

primary stream, and no one knows the orientation of any particular one—indeed the individuals are often too small to be seen.

Equation (7.9) describes a cone, having for its origin the mass of assembled crystals, for its axis the direction of the primary beam, for its apical semi-angle the angle Φ. Such a cone intersects any sphere centered at the crystals (our imaginary bulb), or any plane at right angles to the primary wave-stream,

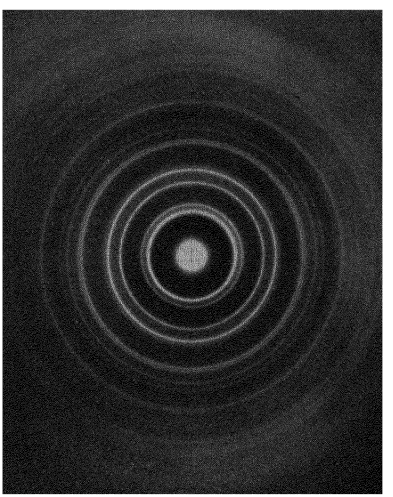


FIGURE 7.8—DIFFRACTION-RINGS OF ELECTRONS WITH A VERY THIN FILM OF POLYCRYSTALLINE SILVER. (Germer.)

in a *ring*. There are in principle as many of these rings as there are triads of integers (h_1, h_2, h_3) except that when two different triads have the same value of the sum-of-squares ($h_1^2 + h_2^2 + h_3^2$), their rings coincide. They appear as dark circles on a photographic film so placed as to coincide with such a sphere or such a plane, exposed and subsequently developed. Each of these circles consists of the diffraction-spots with the appropriate indices cast by the various crystals. If the crystals are few, one sees the individual spots (the ring looks ragged and spotty, like a star-cluster); if they are few and small the

spots are hazy; if instead of being turned at random they favor certain orientations, the circles are not evenly dark all the way round. But these are matters for later study.

If we measure the radii of the first few rings and calculate from them the values of $(1 - \cos \Phi)$ for the corresponding cones (a simple matter of geometry) we should find that these stand to one another as $1 : 2 : 3 : 4 \ldots$—provided, that is, that the lattice is cubic and the incident waves are nearly monochromatic. This is verified by experience for X-rays and electron-waves. The first ring consists of spots having the indices (1, 0, 0) or (0, 1, 0) or (0, 0, 1); the indices for the second ring are (1, 1, 0) or (0, 1, 1) or (1, 0, 1), while those for the third are (1, 1, 1). The reader can easily guess the indices which yield other small integer values for the sum (h_1, h_2, h_3)—but not in every case. There is no triad of integers of which the squares add up to seven, and there is none for which they add to fifteen. If the radii of the rings are plotted as a function of their ordinary number, there are breaks in the smooth curve beyond the sixth and the thirteenth, as if two were absent from the regular sequence. To express it more graphically than accurately, the seventh and the fifteenth rings are missing. Other rings also may be wanting, for the atoms in the atom-groups may be so disposed that in certain directions the individual groups scatter no waves whatever; and even if the lattice were so proportioned that waves in such a direction would be tremendously amplified, there would be nothing to amplify and no diffraction-spot. But this also is a subject for later study.

If we measure the radius of any ring and calculate sin $\frac{1}{2}\Phi$, and then change the wave-length and repeat the process, the values so obtained for the sine should stand to one another in the ratio of the wave-lengths. If the waves in question are electron-waves, then since their wave-length is inversely as the speed of the electrons, the radius of each ring should vary inversely as the speed of the electrons falling upon the crystals.[1]

[1] More precisely, the radii should vary inversely as the momentum of the electrons. Thus at high speeds the relativistic dependence of mass on speed enters in (see the closing pages of this chapter).

This was verified by G. P. Thomson. Knowing the values of the spacing a of the metal crystals which he had used—these values having been deduced in earlier days from the diffraction-circles produced by X-rays of known wave-length, scattered by the crystals of such metals—Thomson also determined by equation (7.8) the actual wave-lengths of the electron-waves. With X-rays the powder-method is seldom used to evaluate wave-lengths, Bragg's being the better when available; it serves chiefly for the study of lattices. But with X-rays and electrons both, the splendid array of the diffraction circles which spring forth when a beam of either kind is sent against a mass of tiny crystals is the most easily adducible, the simplest and perhaps the most vivid and striking evidence that there is something wave-like in the nature of either.

So much for the fundamental equation of the powder method! We will now derive, from equation (7.6), the fundamental equation of the method invented by Bragg.

We have seen that the diffracted beam forming the spot with indices (h_1, h_2, h_3) is inclined to the primary beam at a certain angle Φ for which we have found the formulae (7.8) and (7.9). We may conceive that this *deflection* is due to a *reflection* of part of the incident wave-motion from a mirror or mirrors traversing the crystal, so tilted that their plane bisects the angle Φ. The picture would be legitimate even if there were nothing physical corresponding to these "mirrors"; but we shall presently see that they are not imaginary. The normal to their plane makes supplementary angles θ_0 and $\theta = \pi - \theta_0$ with the primary and the diffracted beams respectively; and $(\theta_0 - \theta)$ is the angle of deflection Φ.[1] Combining these statements, and choosing the positive sense of the normal so that it shall make an acute angle with the diffracted beam, we find:

$$\theta = \tfrac{1}{2}\pi - \tfrac{1}{2}\Phi; \quad \theta_0 = \tfrac{1}{2}\pi + \tfrac{1}{2}\Phi. \qquad (7.10)$$

[1] It is the custom to say that the normal to a mirror makes *equal* angles with the incident and the reflected beams, but this manner of statement implies that the positive senses along the directions of the beams are defined oppositely—one *with*, the other *against* the sense in which the waves are advancing. The convention adopted here is the more logical, and leads to more symmetrical equations.

We wish to deduce the direction-cosines of the normal—denote them for the moment by the symbols γ_1, γ_2, γ_3—from the conditions that it makes the angles θ_0 and θ with the rays having the direction-cosines α_1, α_2, α_3 and β_1, β_2, β_3 respectively. These conditions are thus expressed by the aid of equation (7.9):

$$\alpha_1\gamma_1 + \alpha_2\gamma_2 + \alpha_3\gamma_3 = \cos\theta_0 = -\sin\frac{1}{2}\Phi = -\frac{\lambda}{2a}\sqrt{h_1^2 + h_2^2 + h_3^2}, \quad (7.11)$$

$$\beta_1\gamma_1 + \beta_2\gamma_2 + \beta_3\gamma_3 = \cos\theta = +\sin\frac{1}{2}\Phi = +\frac{\lambda}{2a}\sqrt{h_1^2 + h_2^2 + h_3^2}, \quad (7.12)$$

and the reader can easily show by means of equations (7.6*a*, *b*, *c*) that they are satisfied when:

$$\gamma_1 = \frac{h_1}{\sqrt{h_1^2 + h_2^2 + h_3^2}}; \quad \gamma_2 = \frac{h_2}{\sqrt{h_1^2 + h_2^2 + h_3^2}}; \quad \gamma_3 = \frac{h_3}{\sqrt{h_1^2 + h_2^2 + h_3^2}}. \quad (7.13)$$

If therefore a diffraction-spot (h_1, h_2, h_3) can with any reason be attributed to a reflection of part of the incident energy from mirrors traversing the crystal, then these mirrors must be so tilted that their normal is pointed in the direction defined by equation (7.13).

This equation being attained, we are prepared to discern the physical meanings of the integers h.

One of their meanings is evident. They state the "order" of the diffraction-spot, in the sense in which that word is used in describing the spectra cast by optical gratings. An ordinary ruled grating supplied with light of a single wave-length forms, it may be, six or seven or even more different "lines" on the focal plane of the lens installed in front of it. These correspond to diffraction-spots on the surface of our imaginary bulb. Take any one of these lines, say the nth. The paths of the light from the source *via* consecutive rulings of the grating to this nth line differ by precisely n wave-lengths; the contribu-

tions of any two adjacent rulings to the wave-motion at this line differ in phase by n complete cycles. It is named the line, or the image, or the diffraction-maximum *of the nth* order. There may be lines of positive, of negative and of zero order; the line of zero order is commonly called the "direct image" of the source.

Now in the same sense a diffraction-spot with the indices h_1, h_2, h_3, cast by a crystal, is of three orders simultaneously; these indices are its orders with respect to the three principal directions of the crystal lattice. Referring to our quartet of atom-groups ABCD: the paths of the waves from the source *via* A and B to the diffraction-spot differ by h_1 wave-lengths, those *via* A and C differ by h_2 and those *via* A and D by h_3 wave-lengths. The (ooo) spot is in the prolongation of the incident stream, and is called the "direct image."

This is one important meaning of the indices h. Happily there is another which is much more picturesque—happily indeed, for otherwise it is likely that the art of crystal analysis would have developed more slowly than it has, while the art of X-ray spectroscopy might not even have begun for years. It does not always happen, perhaps indeed it rarely happens, that the earliest formulation of a theory is the one best adapted to make it widely understood and useful. Someone other than the founder meditates upon the idea, and tries to re-express it to himself, and hits upon a novel way of stating it—a new aspect to it, possibly, or perhaps nothing more than a new distribution of the emphasis, laying the greatest stress upon some feature which to his forerunner seemed minor. Then suddenly the theory appeals to the world of physicists in general. Experimenters see what can be done with it, how it can easily be tested and how it can be applied after the tests are passed; and progress for a time is furiously fast. Something much like this befell the theory of the diffraction of waves by crystals. The form in which I have thus far developed it (except while describing the powder method) is the one in which it was clothed by the brilliant inspiration of Laue. However it was W. L. Bragg who made the theory well known

and widely used all the world over, by singling out and featuring the fact that *each diffraction-beam is due to its own special set of atom-strata in the crystal, which reflect it as light is reflected by a pile of parallel mirrors.*

The integers h are then no other than the indices, by which the crystallographers denote these strata. For, to the student of crystallography, the strata are very real—as much so as the rows of atom-groups, by referring to which I developed the theory of the diffraction-spots in Laue's way. We started with the individual group and went on to the row, and then constructed the plane by laying rows down side by side; but we might have started with planes, and defined the rows of atom-groups as the lines or edges where two planes intersect, and located individual atom-groups at corners where three planes meet. This is the historical way; for the planes are the prominent feature of any well-developed crystal. The smooth flat facets which are the boundaries of every well-formed crystal are parallel to important strata, they are themselves examples of important strata. In studying a crystal otherwise than by diffraction, the first step is to measure the directions (relative, of course) of all the available facets. Before the invention of analysis by diffraction, this was often the last step also; but if the crystals available have grown up really well, it is a very long step. Having taken it, the crystallographer proceeds to visualize the crystal as a region of space which is intersected and partitioned by flocks of planes, long sequences of evenly spaced planes parallel to the facets; and he locates the atom-groups at their intersections. How then shall we harmonize these inferences of his with the implications of the diffraction pattern?

A good way to unify the two procedures is to explain the notation by which the crystal planes are named. In doing this, I shall often speak of planes "containing atom-groups," meaning in the strict sense planes containing lattice-points around which atom-groups are placed.

Return to the cubic lattice and to the basic set of four atom-groups ABCD, so chosen that the lines AB, AC, AD are

three edges of the fundamental or "unit" cube. Complete the cube by adding four more atom-groups EFGH. We will pick out the planes which contain three or four of these eight groups. They will be the most populous with atom-groups of all the planes traversing the cube or "cell"; hence the most populous of all the planes traversing the crystal. I have spoken above of the "important" strata of the crystal, meaning those which are likely to be parallel to facets. The word "important" is vague, and it is better that it should remain somewhat vague; but, in a rough way, the more populous a stratum is the more likely it is to be "important" in that sense, and also in the sense that it produces strong diffraction-spots. We will therefore consider chiefly the planes, which cut through the unit cube in such a way as to traverse three or four of the atom-groups. Put the origin of coordinates at A, the x, y, z axes through C, B, D respectively.

Commence with the plane containing BCD, normal to the direction AE which is one of the principal diagonals of the cube. We wish a symbol for this plane that shall describe its orientation, which is its important feature, a symbol that shall denote not only the plane BCD, but equally all those which are parallel to it, such as FGH and the parallel planes through A and E and other unit cells. We might use the three direction-cosines of the normal to this family of planes; or we might use the three intercepts of BCD on the three coordinate-axes, multiplied by an arbitrary factor to convert them into convenient integers and show that we are not more concerned with BCD than with any other member of the family; or we might use the reciprocals of these three intercepts, also multiplied by some convenient arbitrary factors. The last choice is the standard one. The three intercepts are a, a, a; their reciprocals are $1/a, 1/a, 1/a$; we multiply these by a and obtain $(1, 1, 1)$ as the symbol for not only the plane or stratum BCD, but all the strata parallel to it, including for example any facets that may be formed upon such strata. The reader will easily identify three other families of planes for which the symbols are $(-1, 1, 1)$; $(1, -1, 1)$; and $(1, 1, -1)$. Some substances

with cubic lattices form crystals in the shape of octahedra; the surfaces of these are strata with such symbols. Usually the commas are left out of the symbols and the minus signs printed over the digits.

Now try the plane containing the atom-groups BCGH. Its intercepts on the x and y axes are both equal to a, but it is parallel to the z-axis, a fact which is described by giving its z-intercept as infinity. The reciprocals of its intercepts then are $1/a$, $1/a$, 0; we multiply by a and obtain (1 1 0) as the symbol for all the strata parallel to the one containing the atom-groups BCGH. The reader can easily identify members of the (0 1 0) and (0 0 1) families, and ascertain how many new families of planes he can get by reversing the signs of some or all of the indices. There are six altogether; one sometimes sees facets with these indices, beveling off the edges of natural cube-shaped crystals.

Next consider the plane containing CFHE. It is parallel to the yz-plane and distant from it by a, so that its intercepts are a, ∞, ∞, and its symbol is (1 0 0). One sees immediately that (1 0 0) and (0 1 0) and (0 0 1) are the symbols for the three families of planes which comprise these which we have taken for our coordinate-planes. When a substance with a cubic lattice forms crystals which are cubes, their facets belong to these families.

What could a symbol such as (2 1 0) imply? It would stand for a family of planes, one of which would have for its intercepts the values $a/2$, $a/1$, $a/0$, or $\frac{1}{2}a$, a, ∞. This plane would be parallel to the z-axis and would traverse the atom-groups B and G, and would slant across the cell in such a way as to pass through the wall ACDH midway between its vertical edges. Continuing it and the lattice in imagination, one sees that at the far angle of the *next* cell it would traverse another pair of atom-groups, and at the far angle of every second cell thereafter it would do the like. It is therefore a fairly "important" stratum, though not so populous as those of which the indices are zeros and ones exclusively. It might form facets, and would be likely to give a noticeable diffraction-

spot. But in general as the indices mount up, the importance of the plane declines.

It is evident that if any set of indices is multiplied by any constant, the new set thus obtained corresponds to the same family of planes. To choose one set definitely for each family, we may agree to adopt the triad of integers having no common divisor. Thus all three of the symbols (963), (642) and (321) refer to the same family of planes; we always choose the last one.

Given these the so-called "Millerian" indices of a family of planes, what are the direction-cosines of their common normal?

Denote the three indices by H_1, H_2, H_3. The question is: what are the direction-cosines $\gamma_1\ \gamma_2\ \gamma_3$ of the normal to a plane, of which the intercepts on the coordinate axes are a/H_1, a/H_2, a/H_3? The answer is given by a standard formula: the three direction-cosines are:

$$\gamma_1 = \frac{H_1}{\sqrt{H_1^2 + H_2^2 + H_3^2}};$$
$$\gamma_2 = \frac{H_2}{\sqrt{H_1^2 + H_2^2 + H_3^2}}; \qquad (7.14)$$
$$\gamma_3 = \frac{H_3}{\sqrt{H_1^2 + H_2^2 + H_3^2}}.$$

The common factor a has vanished, as it should.

Compare these expressions with those in equation (7.13). One sees instantly that the strata (H_1, H_2, H_3) are so oriented that these strata could serve as mirrors to reflect the primary beam towards the diffraction-spot of which the indices are $h_1 = H_1, h_2 = H_2, h_3 = H_3$; or towards the spots of which the indices are (nH_1, nH_2, nH_3) where n stands for any integer. Or: the diffraction-spot (h_1, h_2, h_3) may be conceived as due to a reflection of part of the wave-motion in the incident stream, by the atom-layers of which the symbol is $n_1/C, n_2/C, n_3/C$; C standing for the greatest common divisor of $h_1h_2h_3$.

This is part of the principle which Bragg deduced from Laue's theory, but not the whole of it.

It has earlier been mentioned in this book (Chapter I) that when a beam of plane waves falls successively on two plane parallel surfaces which reflect a part and transmit a part of it, the two reflected beams are in phase with one another and the resultant is maximum, when the following relation prevails between the wave-length λ of the light, the distance d between the mirrors, and the angle θ of incidence and of reflection:

$$n\lambda = 2d \cos \theta, \qquad n = 0, 1, 2, 3. \ldots \qquad (7.15)$$

When instead of a pair there is an endless or a very long sequence of mirrors spaced at equal intervals, the result is much the same as when a pair of atoms is supplemented by a very long row. The angles defined by equation (7.15) are now not merely the angles of maximum reflection; they are the *only* angles where reflection is at all appreciable. If a pile of parallel semi-transparent mirrors is to reflect to any notable extent, their thicknesses and the wave-length and the angle of incidence of the light must be very carefully adjusted according to equation (7.15) with some integer value for n.

Now the principle emphasized by Bragg is this, in full: *Any diffraction-spot results from reflection by the strata having the same indices as the spot, at the angle of selective reflection defined by equation* (7.15).

The proof of this statement depends on the following formula[1] for the distance between adjacent strata of the family having the Millerian indices $H_1H_2H_3$:

$$d = \frac{a}{\sqrt{H_1^2 + H_2^2 + H_3^2}}. \qquad (7.16)$$

[1] Let O and P stand for two planes of the family $(H_1H_2H_3)$, one being drawn through the origin, the other through any other lattice-point. The components of the vector r from the origin to this other lattice point must be integer multiples l_1a, l_2a, l_3a of the spacing a. The projection of this vector on the direction of the normal drawn from the origin to the plane P is equal to $(\gamma_1 l_1 + \gamma_2 l_2 + \gamma_3 l_3)a$; the values of $\gamma_1\gamma_2\gamma_3$ are to be taken from (7.14). This projection is equal to the distance D between the planes P and P, for which we therefore have:

$$H_1l_1 + H_2l_2 + H_3l_3 = \frac{D}{a}\sqrt{H_1^2 + H_2^2 + H_3^2}.$$

The least value (except for zero) which the quantity on the left can and does assume is unity; whence equation (7.16).

Substituting into equation (7.12) the value of a given by this formula, we find that:

$$\cos\theta = \frac{\lambda}{2d}\frac{\sqrt{h_1^2 + h_2^2 + h_3^2}}{\sqrt{H_1^2 + H_2^2 + H_3^2}}, \qquad (7.17)$$

and since the quantities H are integers without a common divisor, while the quantities h_1 are integers for which $h_1/H_1 = h_2/H_2 = h_3/H_3$, the ratio of the two radicals must be an integer.

The mirrors which I introduced at a previous page as a way of accounting for the spots do actually exist. They are the strata into which the groups of atoms fall. Each one by itself, however, reflects so little that in effect there is no reflection to speak of, unless and until an entire procession of parallel strata is brought into play. Earlier we located the diffraction-beams in the directions in which the scattered waves from four adjacent atom-groups enhance one another most by constructive interference. Multiplication of atom-groups beyond the first four merely made the beams sharper and more intense. Alternatively we may now locate these beams in the directions in which the reflected waves from two adjacent parallel strata enhance each other most. Multiplication of strata beyond the first two intensifies and sharpens them.

This theorem of Bragg's thus gives a remarkably helpful picture of "the way of a crystal with a beam of waves"; a picture most valuable, when one has a single large crystal of which the surfaces are natural facets. Suppose for instance one has a cubic crystal of rocksalt, one of its faces being parallel to the (100) strata. If one has a monochromatic beam of X-rays incident on that face and revolves the crystal so that θ varies steadily from zero to 90°, then for the several angles given by equation (7.15) with various values of n diffraction-beams spring forth. One may set up a photographic plate beside the crystal, and find the imprints of all the beams upon it after the rotation is completed; or alternatively one may revolve an ionization-chamber at double the angular speed of the crystal, so that whenever a beam shoots out the chamber

is in the right place to capture it, and then the curve of ionization-current versus angle θ shows a peak for every value of θ corresponding to an integer value of n. If the incident beam comprises many wave-lengths, one finds their spectrum spread out in the ionization-current curve; three examples are shown in Figure 14.4, where each of the sharp tall peaks is due to the first-order diffraction-beam of a monochromatic wave very intensely represented in the primary wave-mixture. Or one may hold the crystal and the collector still and vary the wave-length, obtaining a peak wherever λ is such that for some integer value of n the equation (7.15) is satisfied; the curve of Figure 7.7 was obtained in this way, using waves of negative electricity.

The process of measuring wave-lengths of X-rays is usually conducted by this method, using a crystal such as rocksalt for which the density is very accurately known. For if we know the density of the crystal we know how many atoms it contains in a given volume; and if we know in addition how many atoms constitute the atom-group which is repeated over and over again to form the crystal, we can compute by simple division what is the volume of the unit cell, and what therefore is the spacing from one atom-group to the next—the edge of the elementary cube. If we then set up the crystal so as to get reflections from the 100 face we know that the edge of the cube is the quantity d which figures in the equation (7.15); and measuring then the values of θ corresponding to several diffraction beams we can identify the corresponding values of n, and so evaluate λ. Diffraction of X-rays by ruled gratings can now be called upon in confirmation—or in correction, as certain recent data indicate.

The determination of the number of atoms in the atom-group is the delicate point of this computation; and perhaps it is to be accounted a piece of luck that with the first crystals used in the spectroscopy of X-rays the guess was easy and was rightly made. The routine of determining it is but a part of the general process of learning from the diffraction as much as possible about the atom-group; and this deserves a chapter to

itself, or many. Two examples of this process however are interesting, useful and very simple.

I remarked near the beginning of the chapter that while it would be the simplest possible thing to suppose that the particles arranged in a cubic lattice have full spherical symmetry, yet this supposition is as a rule too simple for the facts. In particular it is too restricted to explain the diffraction-pattern of any one of the numerous elements which crystallize on cubic lattices. One might then be forced to assume that the atoms themselves do not have spherical symmetry. But luckily this is unnecessary; for it happens that if with each lattice-point of the cubic lattice we associate a properly spaced and properly oriented *group* of spherical atoms—in some elements a pair, in other elements a group of four—the difficulties vanish. The diffraction-patterns are explained, and there is no outstanding conflict with the data assembled by the crystallographers; for both of these arrangements, like that in which each lattice-point is occupied by a single spherical atom, possess full cubic or isometric symmetry in the crystallographic sense of those words.

In the first of these permitted arrangements, the two atoms associated with each lattice-point are so placed and so spaced, that if we label them, say, A and B, the atoms A by themselves form one simple cubic array, and the atoms B by themselves form another simple cubic array with an atom B in the very center of each cube composed by atoms A—and vice versa. This is the "body-centered cubic" arrangement. The atom at the center of the cube may be associated with any one of the eight corner atoms to form a pair; this pair is then repeated over and over again on the cubic lattice to form the crystal. The alkali metals, iron, and several other elements are addicted to this arrangement.

In the second of the arrangements, the four atoms forming a group are so placed and so spaced that if we call them A, B, C, and D, the atoms of each letter form a cubic array; and these four cubic arrays are interlocked in such a fashion, that the cubes of any one of these arrays have in the centers

of all their faces atoms belonging to the others. The atom A at the corner of any cube may be associated with the atoms in the centers of the three faces of that cube which meet at that corner, and these four form the atom-group which is repeated over and over again on the cubic lattice to build the crystal. Many of the metallic elements have adopted this "face centered cubic" arrangement, the noble metals for example, and argon also.

How does one recognize from the diffraction pattern which of these arrangements exists in a cubic lattice? At this point I will not give an exact answer: but the principle is simple. Even as on an earlier page it was shown that for certain directions of diffraction adjacent atom-groups *reinforce* the scattering from one another, so it may be shown that for certain directions the different atoms of a single atom-group *destroy* the scattering from one another. One has only to write down the condition that the distance from source P to fieldpoint Q *via* one atom of the group differs by an odd-integer multiple of $\frac{1}{2}\lambda$ from the distance *via* the other; or if there are four atoms in the group, that the waves scattered to Q from the four are so balanced in phase that they annul one another.

Now if it should turn out that one or more of the diffraction-spots expected from the cubic lattice fall exactly where the effects of the atoms of each individual group cancel each other out, then those spots will be lacking. For the spots are due to amplification of the diffraction-pattern of the individual atom-group; but if at the location of a predicted spot this pattern sinks to a vanishing intensity, there is nothing to amplify. Well, owing to the neat and accurate way in which the spacings between atoms of a group are related to the spacings between the atom-groups, this sort of coincidence occurs for a respectable fraction of the diffraction-spots—or, in the powder method, it occurs for several of the diffraction-rings. From the missing spots or rings therefore one identifies which style of atom-group prevails in the cubic crystal.

Test of the first of the rules of correlation.

It having been established that beams of electrons of uniform speed have a wave-like character and ascertainable wave-lengths, the test of equation (6.32)—the Rule of Correlation between particle-momentum and wave-length—becomes the impending problem.

The test may be divided into three stages. The first consisted in finding whether (6.32), with a value of h taken from experiments of other kinds, is at least roughly correct for electrons moving so slowly that the relativistic dependence of mass on speed may be ignored. This stage (which really involves the confirmation that λ^{-1} is proportional to electron-speed v for low-speed electrons) was successfully passed in the very first experiments on electron-diffraction. The second consisted in finding whether as v is increased, λ^{-1} varies as $v/\sqrt{1-\beta^2}$. This test has been made by several during the last eight years. The third consists in tentatively accepting (6.32) i.e., in accepting the conjoint measurements of wave-length and particle-momentum as a technique for evaluating h, and finding whether the so-obtained value of h harmonizes fully with those obtained in many other ways: it is at present an urgent problem.

The third stage of the test is more complicated than I have just implied: for what the measurements actually yield (assuming the Rule to be right) are values not of h by itself, but of combinations of the fundamental constants h and m_0 (rest-mass of the electron) and e. Suppose for simplicity that the experiments are made with electrons slow enough to justify neglect of the relativistic dependence of electron-mass on electron-speed. Say first that the experimenter uses electrons of which the initial speed is negligibly small; accelerates them by a potential-rise V; writes $\sqrt{2m_0eV}$ for p; multiplies this quantity into the value of λ which he ascertains from a diffraction-pattern—it appears then that his measurements yield him a value for $h/\sqrt{m_0e}$. Say instead that he measures the electron-speed v by the method of the gates; writes m_0v for p; multiplies

this quantity into his λ-value—it appears that his measurements yield him a value for h/m_0. But how shall he "measure" the wave-length? To do it as for X-rays, by diffracting the beam with a ruled grating of assuredly known spacing, is seemingly impracticable! Even after ten years there is no reliable report of anyone having succeeded in getting a diffraction-pattern of electrons with such a grating. To do it by diffracting with a crystal means that the result is involved with the perplexity as to spacings and e-values which was recounted in Chapter I. Thus the full testing of (6.32) is not really an isolated enterprise, but a subdivision of the general task of obtaining (if possible, which is not to be taken for granted!) a set of values of e and m_0 and h which shall be harmonious with all of many experiments of various types; and as such it should be regarded.[1]

Diffraction of neutral atoms and molecules.

It may well seem that the Rules of Correlation "ought" to apply only to ultimate particles, such as electrons and photons may yet be presumed to be. There is however no such restriction implicit in the manner of their setting-up, and experiment has shown that the first of the Rules applies to neutral atoms and molecules, composite though they be.

Successful experiments have been made with beams of helium atoms He, of diatomic hydrogen molecules H_2 and of hydrogen atoms H (not to be confused with bare protons!) projected against crystals of various kinds, of which lithium fluoride appears for some reason to be by far the most suitable. The surface on which the beam impinges must be a cleavage-plane or natural facet of the crystal; that is to say, the surface-layer of atoms must be one of the populous or "important" strata of atom-groups, in the sense of page 217. The incident particles do not penetrate into the crystal with its three-

[1] The scheme of determining v by the method of the gates and λ by diffraction has been used by J. Gnan of Kirchner's school (AP 20, 361; 1934); the other scheme by several experimenters, lately by S. v. Friesen (Diss. Upsala, 1935). In Gnan's experiments β ranged from 0.3 to 0.4, so that he tested the relativistic dependence of λ on v.

dimensional lattice, and consequently whatever diffraction occurs must be due to the two-dimensional lattice of the superficial atom-groups. The diffraction-pattern must therefore be of the type described by the *two* simultaneous equations (7.3) and (7.4) and the text immediately following those equations. It should consist of diffracted beams shooting off in various directions from the place of impact of the incident beam upon the crystal surface; these beams should appear whatever the wavelength (so long as it is smaller than the spacing) but their directions should vary continuously with λ. Such diffraction-patterns have been observed by Stern and his colleagues at Hamburg, by T. H. Johnson of the Bartol Foundation, and by R. M. Zabel of the University of Iowa[1]; the last obtained diffraction-beams from helium projected against crystals of sodium chloride.

The technique differs remarkably in many ways from those which are practiced with electrons. Neutral atoms being practically non-accelerable, one has to take them with their speeds of thermal agitation: the beam emerges through slits from what is called an "oven" or "furnace," though sometimes it is much colder and sometimes much hotter than the environment. The atoms have a broad (Maxwellian) distribution of speeds, which is a disadvantage, partly overcome by the school of Stern as we shall see in a moment; the lowness of their mean speed is not a disadvantage at all, for their masses are such that if they were moving as fast as electrons are made to move in diffraction-experiments, the wave-length would be so small that the diffraction-pattern would be indistinguishable. Owing to their lack of charge, a collector connected to an electrometer is useless, and so are photographic plates. Stern used collectors which were chambers with narrow slits of entry, through which the diffracted beam (if any) entered, while the atoms or molecules of which it consisted could leak out only slowly;

[1] I. Estermann, R. Frisch and O. Stern, **ZP 73**, 348 (1932) and earlier papers there cited; T. H. Johnson, **PR 37**, 847 (1931); R. M. Zabel, **PR 42**, 218 (1936). This subject is well reviewed in C. F. Meyer's *Diffraction of Light, X-rays, and Material Particles* (University of Chicago Press, 1934).

the entering beam thus built up in the chamber a pressure proportional to its own intensity, and this pressure was indicated by a hot wire.[1] Johnson used a scheme practicable only with atomic hydrogen: the diffracted beams impinged on a plate covered with molybdenum oxide, a substance which undergoes a chemical reaction involving a striking change of color when atomic hydrogen encounters it.

The diffraction-pattern is duly that of a two-dimensional lattice, but it has a singular feature: the surface-plane of the crystal consists of lithium and fluorine atoms alternating, but the pattern is what would occur if only one of these two kinds of atoms were capable of forming wavelets, the other being effectively absent! This means a much more radical difference than would appear if the lattice were changed merely by doubling its spacing, as the student can convince himself readily.[2] This being accepted as a curious but definite fact, the full test of the Rule of Correlation consists in evaluating λ from the directions of the diffracted beams and the spacing of the lattice, and p from the masses and speeds of the atoms. The mean speed of the atoms in the beam is deducible from the temperature of the furnace from which the beam emerges. It can also be measured by the method of the gates, in the mechanical form mentioned on page 44 (two coaxial toothed wheels or indented discs whirling on a common axle with an adjustable speed). Estermann, Frisch and Stern used this method not only to measure the speed, but to select from the original beam a succession of flocks of particles moving with speeds comprised within a much narrower range than that of the original beam. They also diffracted atoms from two crystals in succession; the initial incident beam comprised particles of all speeds, the first diffracted beam was accordingly highly divergent with particles of different speeds going in different directions, a small

[1] Power was dissipated at a constant rate in the wire, and the resulting heat conducted away by the gas at a rate depending on the pressure, so that the temperature and hence the resistance of the wire varied with the pressure of the gas.

[2] Zabel's diffraction-beams also did not agree completely with what would be expected from the surface-lattice of his crystals, and varied from one crystal to another.

portion of it fell upon the second crystal and was re-diffracted as a much narrower and more nearly homogeneous beam than the first one. By these difficult experiments, the product $mv\lambda$ was determined as agreeing within one or two per cent with otherwise-obtained values of h. This corresponds accordingly to what I called, in speaking of the corresponding problem with electron-beams, the first stage of the test of the Rule of Correlation.

Attempts to observe the diffraction of protons have been ill-starred: beams of these particles have been projected against crystals, and sometimes the scattered protons have formed remarkable patterns, but never any that was intelligible.

CHAPTER VIII

Ionization

Ionization is the taking-apart of the atomic structure by the removing of one or more of the electrons which together with the nucleus compose it. Taking apart a thing is notoriously the best way of discovering how it is put together; and the removal of electrons from an atom supplies the physicist with items of fundamental knowledge as to how they are coordinated in that atom, which later he completes with numberless details derived from the study of spectra. Ionization is also all-important in the conduction of electricity through gases and through electrolytes, for it must occur in order to start the flow and must continually recur in order to sustain it; but these extensive topics lie mainly outside of the programme of this book. The study of the discharge in gases and of the current-flow in electrolytes made the pre-war physicists aware of many of the features of ionization, beginning with the charge-to-mass ratios of the resulting ions, from the values of which the content of the first sentence of this chapter is ultimately derived. Since 1911 the tables have been turned, and it is from the study of the ionization-processes under simplified conditions that physicists reason back to the principles of the discharge, as well as forward to the principles of atomic structure which are the concern of this book.

Onset of ionization

The classical experiment upon ionization—for I wish to call it classical, even though scarcely a quarter of a century has passed since it was first performed—consists in projecting a beam of electrons of sharply-definite and adjustable kinetic

energy into a rarefied gas, and observing whether or not positive ions appear in the gas. It is found that they do or do not appear, according as the energy of the electrons is or is not greater than a certain well-marked value characteristic of the gas. As the energy is increased from zero, it eventually attains this value I_0 at which there is sudden onset of ionization.

(But how can one be sure that the onset is really sudden? This is a delicate question, arising from the fact that " sudden " is one of those words with a tinge of absoluteness, which strictly have no place in an experimental science. The experimenter observes something which he believes to be either proportional to, or at worst a steadily-rising function of, the number N of ions appearing per second in the gas. He plots his readings against the electron-energy, and since all of them to the left of a certain abscissa are zero, he draws a curve which fits the rest as well as conveniently possible and rises from the axis at a point of which he then calls the abscissa I_0. If his apparatus were tenfold as sensitive, would the corresponding point lie nearer the origin, or would it not? Unless he has some theory to lead or mislead him, he can only make a guess based on whether his curve slants sharply down towards the axis, or bends toward an apparent tangency with it. This question arises in the study of the photoelectric effect of metals, in that of the " continuous " spectra of X-rays and of beta-rays, and in many other situations in physics; and often the answer which is given is really based upon a theory.[1])

The constant I_0 is called " ionizing-energy " or " ionizing-potential," the latter usage being due to the fact that the energy of the impinging electrons is communicated to them by a potential-rise. We are to think of it as the energy required to unbind or detach the most loosely-bound, or loosest, electron from the atomic structure; for in most atoms and molecules there are several electrons, and we shall later see that the

[1] The condition is sometimes even worse than this paragraph implies; for in some cases the curve has a long straight arc slanting sharply down towards the axis but not quite reaching it, being continued by a so-called "foot" which bends around into apparent tangency with the axis; when there is reason for expecting the "true" curve to strike the axis sharply, the foot is ascribed to defects in the apparatus.

removal of one is distinguishably not the same process as the removal of another. We may also think of it as the energy-difference between the normal neutral atom, and the system consisting of one free electron and one positive ion far apart and practically stationary. This may sound like a superfluous elaboration of a foregoing simple statement, but it has advantages which will later appear. As a matter of fact, when the impinging electron of energy I_0 strikes upon the atom it must communicate some of its momentum to the ion and the free electron which spring apart, so that these particles cannot be quite stationary; but the smallness of the ratio of electron-mass to atom-mass makes this a trivial effect.

The proper location of onset-of-ionization requires apparatus more elaborate than my words have yet suggested. The experimenter must guard against an effect which was not even suspected when the method was first employed.[1] Electrons having kinetic energy less than I_0 may cause the atoms which they strike to emit radiation (we shall study this effect in Chapter IX). Some of this radiation falls upon the electrode intended for collecting positive ions, and expels electrons from it. The field around the collecting-electrode, being such as to draw positive ions toward it, drives these electrons away; and so there is a continuous current of negative charge out of the electrode into the gas, which is quite indistinguishable from a current of positive charge out of the gas into the electrode. Thus the value of electron-energy at which positive charge first seems to be flowing from the electrode into the gas, is not I_0, but a "radiation-potential" or energy-value for the onset of radiation. Some of the earliest determinations of what were thought to be ionizing-potentials were vitiated by this effect.

To avoid or recognize the influence of radiation several schemes have been devised.[2] For example, if two collecting-

[1] The method dates from 1902, having been employed by Lenard; he applied it to several common gases, and got 11 volts as the value of I_0 for all of them, a strange result for which Franck has proposed the explanation that in every case Lenard was observing the ionization of mercury vapor mixed with the gas.

[2] For a detailed account of the methods developed up to 1924, consult K. T. Compton and F. L. Mohler: "Critical Potentials" (Bull. Nat. Res. Council, No. 48).

electrodes are used in alternation, one having a large area and the other being small, much more radiation will fall upon the larger one, and there will be a correspondingly great difference between the currents of negative charge out of the two; but if ions are being formed in the gas, the difference between the numbers of these which find their way to the large and to the small electrode will be much less pronounced. A slender collecting-electrode may record only a very small current due to radiation, but a very large one whenever ionization commences. This scheme has been adapted by K. T. Compton.

Another, and the most common, device for distinguishing ionization from radiation consists in surrounding the collector with a sheath of metal gauze, maintained at a potential slightly (say 3 volts) more negative than the electrode which it screens. Positive ions pass through its meshes to the collector, somewhat slowed down but not driven back. Radiation also passes through the meshes to the collector, but the electrons which it drives out are turned back by the adverse field and reenter the metal whence they came, so that the net result is the same as though they had never come out. Ionization thus produces a current of positive charge into the collector, and radiation none; or radiation may even produce a current of negative charge into the collector, thus accentuating the contrast, for electrons which are ejected from the metal gauze are drawn to the screened electrode. This is the scheme devised by F. S. Goucher.

Another, and possibly the best, method for measuring ionizing-potentials is quite insensitive to radiation. A hot filament is immersed in the gas, which may be supposed to be surrounded by connected metal walls so that its boundaries are all at the same potential. If the efflux of electrons from the filament is so plentiful that it is "limited by space-charge" (Chapter II), and this condition persists as the potential difference between walls and filament is raised to the value just sufficing to give to the electrons energy enough to ionize the gas, then at the moment of onset of ionization the space-charge limitation is partially or totally canceled, and the current

increases sharply. This is I. Langmuir's method. It is better to keep the potential-difference between the walls and the filament small and constant, and admit into the gas electrons with controllable energy from another source; when the energy of these auxiliary electrons is raised to attain the ionizing-potential of the gas, the current from the filament suddenly increases. This is the method of G. Hertz [1] and K. H. Kingdon.[2]

Most of the accurate measurements of ionizing-potentials have been made with a collecting-electrode sheathed by a gauze, according to the precept of Goucher. The apparatus is a complicated affair, for the parts already mentioned are by no means all that are required; in some cases the whole interior of the tube appears to be webbed with gauzes. A hot filament (occasionally an illuminated metal plate) is provided as source for electrons, and its potential—or the potential of its negative end—is taken as the zero from which the other potentials are measured. In the sketch (Figure 8.1) this is marked F. Close to the source there is a gauze (G_1) maintained at the controllable potential V and thus providing the potential-rise by which the electrons are accelerated. It is clearly desirable that the electrons should move at their known maximum speed over as long a path as possible in the gas; consequently a second gauze G_2 is set up beyond G_1, and maintained at nearly the potential V so that there is a nearly equipotential region between them. (Generally the potential of G_2 is raised a fraction of a volt

[1] **ZP 18**, pp. 307–316 (1923).

[2] **PR 21**, pp. 408–418 (1923). Experiments with helium and mercury in especially designed tubes having a slender wire as cathode and a wide surrounding cylinder as anode, and provided with extra electrodes for sucking out the positive ions so that their number could be independently estimated, showed that the effect of the ions in annulling space-charge limitation is much augmented by what is called their "imprisonment." That is to say: an ion released in the vicinity of the cathode often has an initial velocity of which the direction and magnitude are such that instead of falling inertly into the cathode wire it swings around in an orbit like a comet. Kingdon's experiments indicate that (in helium at 10^{-5} mm. pressure with an anode 51 mm. in diameter and a cathode 0.1 mm. in diameter) an ion not infrequently describes 350 loops around the filament, during which its presence enables 50,000 electrons to issue from the filament. The method has lately been revived by R. N. Varney (**PR 53**, 732; 1938).

above V so that there may be a slight impulsion of the ions formed between G_1 and G_2 toward the collector.) Beyond G_2 are the collector C and its protecting gauze G_3, maintained at potentials lower than the filmament so that no electrons may reach them. The current of which the sign indicates whether

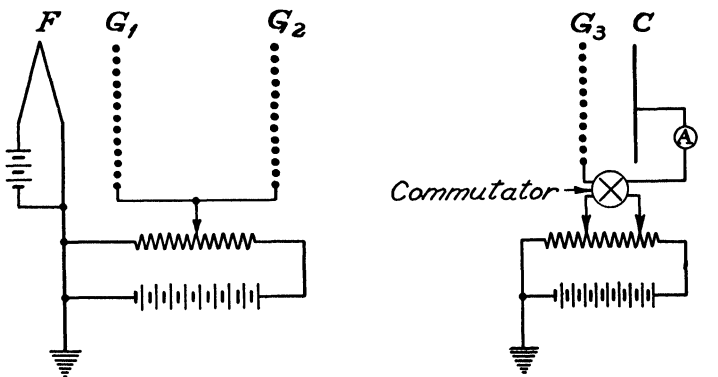


FIGURE 8.1—APPARATUS FOR LOCATING IONIZATION-POTENTIALS AND DISTINGUISHING THEM FROM RADIATION-POTENTIALS. (After F. S. Goucher.)

it is due to radiation or ionization, as was explained above, flows through the galvanometer at A.

In Figure 8.2 a pair of the earliest curves obtained with this arrangement is copied. The lower curve shows the current into C, plotted versus V, when C is slightly more positive than G_3. There is no current until V attains the critical value of 4.9 volts. At this point a current sets in, of which the sign shows that it consists of electrons ejected from G_3 by radiation from bombarded atoms and traveling over to C. Thus at 4.9 volts there is a critical potential for onset of radiation, a *radiation-potential* as the phrase is; we shall meet with it again in Chapter IX. The current now increases gradually, at a rate which itself suddenly increases at 6.7 volts, a second radiation-potential. Near 10.4 volts, the curve suddenly turns sharply upward and the current changes sign. Positive ions are now coming into C; the ionizing-energy is 10.4 equivalent volts. The upper

curve was obtained by making C slightly more negative than G_3; the reader will easily understand the difference.

With such an apparatus as this it seems to be easy enough to measure ionizing-potentials correctly within one or two volts. As soon as greater accuracy is sought after, the real troubles begin. The electrons do not all leave the source with negligible speed; their speeds are distributed over a finite range. The potential to which they climb in passing through the meshes of a gauze is not quite equal to the potential of the gauze-wires themselves. The potential-differences between the different electrodes are not accurately given by voltmeters, for there are contact-potential-differences superposed upon the values indicated. The filament is not an equipotential surface if it is heated by a current, although this difficulty can be overcome if the experimenter thinks it worth the trouble. Electric charges marooned upon the walls of the tube, electrons ejected by radiation from the gauze G_3 and accelerated backwards to G_2 with a final speed higher than the electrons from F ever attain, are capable of causing false conclusions. The third significant figure in the value of an ioniz-

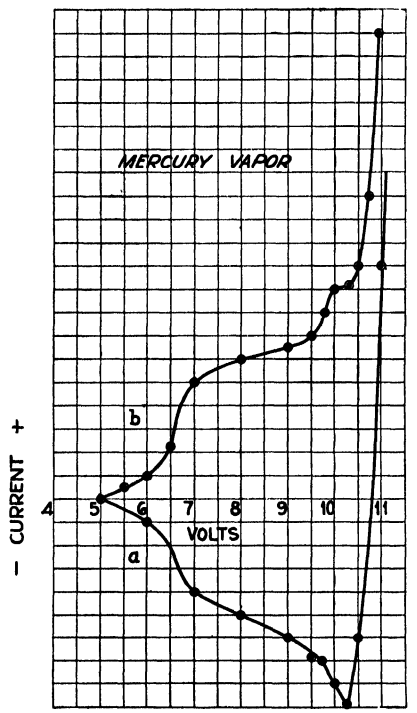


FIGURE 8.2—CURVE DISPLAYING IONIZATION-POTENTIAL, AS WELL AS RADIATION-POTENTIALS, OF MERCURY VAPOR. (Davis and Goucher: *Physical Review*.)

ing-potential is much harder to attain than the first two; and it is not surprising that many experimenters have chosen to mix some standard gas such as helium into the gases which they were studying, and to determine the differences between the ionizing-potentials of these gases and of their standard. However, the incentive for pressing onward to further significant figures has vanished, now that the spectroscopic method (which we shall later consider) is generally accepted; and it is unlikely that this technique will be developed further.

Values of ionizing-potentials

The ionizing-potential I_0 is one of the most strikingly periodic of all the atomic properties, as one expects from the fact that it relates to detachment of the most loosely-bound electron of the atom, therefore to a property of the outermost region of the atom (Chapter V). Its maxima occur in the column of the noble gases, its minima in the column of the alkali metals. The highest of all values is that for the first noble gas, 24.5 electron-volts for helium; one of the lowest, perhaps *the* lowest of all, is that of the last known alkali metal, 3.87 electron-volts for caesium. Its value for the simplest of all atoms, that of hydrogen, is 13.54; but since gaseous hydrogen is normally molecular, very special steps have to be taken to check this value which was originally derived from spectroscopic data. A partial tabulation will be given later (page 285).

Identification of ions produced by electron-impacts, and onset of new types of ionization

The methods described as yet in this chapter serve for locating the onset of ionization, but give no information about the nature of the positive ions: we have been assuming without proof that they are atoms deprived of a single electron. The methods of J. J. Thomson and of the students of isotopes who followed him revealed the charge-to-mass ratios of ions occurring in a gas sustaining a complex electrical discharge of moderately high voltage, but the discharge itself was (and is) not

well enough understood to enable them to infer the origin of these ions. A combination of methods of both kinds was first achieved by H. D. Smyth of Princeton.[1]

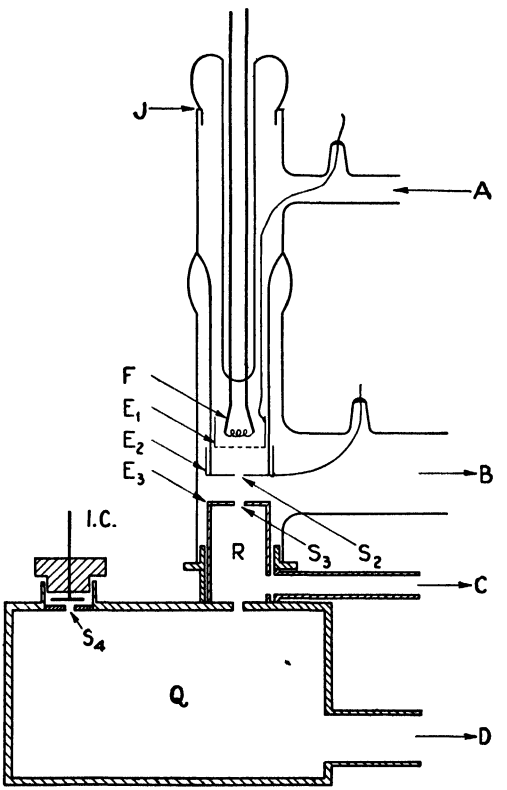


FIGURE 8.3—APPARATUS FOR MEASURING CHARGE-TO-MASS RATIO OF IONS OCCURRING NEAR ONSET OF IONIZATION. (After Smyth.)

One of Smyth's tubes is sketched in Figure 8.3. Electrons from the filament F are accelerated through the potential-rise V_1 to the gauze E_1, and then turned back by an adverse potential-fall V_2 before they reach the partition E_2 pierced by the

[1] **PRS 102**, 283 (1922); **RMP 3**, 347 (1931) and literature there cited.

slit S_2. Positive ions produced between E_1 and E_2 are drawn toward the latter; some emerge through S_2 and encounter an additional potential-fall V_3 which draws them to the partition E_3. Those among them which pass through the slit S_3 are now ready, after crossing the field-free region R, to be swung around in semicircular arcs by the magnetic field H normal to the plane of the sketch in the deflection-chamber Q; and so they arrive—when H is correctly adjusted—at the ion-collector behind the slit S_4. Differential pumping is required, the gas being fed in through A and drawn out by powerful pumps through B and C and D to keep the deflection-chamber as empty as possible. Varying H and plotting against it the current into the ion-collector, Smyth obtained curves with peaks—not sharp ones at first, but distinct enough to permit of recognizing the ions which produced them.

The apparatus may also be designed with the slit S_4 and the ion-collector in line with S_2 and S_3, and the chamber Q pervaded not only by the magnetic field but also by an electric field horizontal in the sketch. By properly adjusting the ratio E/H it is possible to make ions of any particular charge-to-mass ratio continue straight [1] (vertically, in the sketch) across Q. The accelerations in the horizontal direction due to these two fields are Hev/Mc and eE/M respectively, the symbols having their usual meanings; when they are equal and opposite, we have:

$$\text{charge-to-mass ratio} = ne/M = Ec^2/2VH^2, \qquad (8.1)$$

V standing here for the accelerating potential-drop to which the ions owe their kinetic energy (here expressed in electrostatic units). On plotting collector-current vs. either E or H, the other field being held constant, the experimenter finds a peak occurring when this condition is fulfilled for any kind of ion present. This is in essence the first method of W. Bleakney, who also introduced the scheme of drawing the ions sidewise out of the path of the electron-beam by a weak transverse electric field into a lateral region, where after passing through

[1] Cf. the text adjoining equation (2.18).

a slit they are accelerated by the voltage V. (The electrons are kept from wandering off sidewise by a strong magnetic field applied parallel to the original direction of the beam, and acting in much the way described on page 40). Bleakney, Tate and their colleagues[1] of the Minnesota school later retained this scheme, but reverted to the semicircular arcs in the deflection-chamber, varying, however, the kinetic energy of the ions while leaving H unchanged.

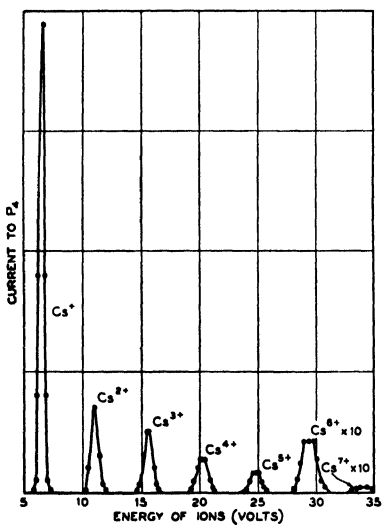


FIGURE 8.4—MASS-SPECTRUM OF THE IONS OCCURRING IN CAESIUM VAPOR BOMBARDED BY 700-VOLT ELECTRONS, SHOWING SEVEN STAGES OF IONIZATION OF CAESIUM. (Tate and Smith; *Physical Review.*)

We have seen in Chapter III that multiply-charged atomic ions, *i.e.* atoms lacking two or more electrons, emerge from self-sustaining discharges: can they be produced in this apparatus by controlled electron-impacts, and are there critical potentials for the onset of these modes of ionization? Such is indeed the case, as Figures 8.4 and 8.5 exemplify superbly. Mercury vapor bombarded by 350-volt electrons contains Hg atoms lacking from one to five electrons, and each of these five kinds of ion ceases, in the manner shown in Figure 8.5, to appear when the energy of the impinging particles drops below a corresponding onset-value. Tate and P. T. Smith made similar observations on no fewer than six kinds of xenon ion, and no fewer than seven stages of the ionization of caesium. To give the student an idea of the

[1] W. Bleakney, **PR 34**, 157 (1929); **35**, 139 (1930). J. T. Tate and P. T. Smith, **PR 46**, 773 (1934); Tate, Smith and A. L. Vaughan, **PR 48**, 525 (1935); A. L. Vaughan, **PR 38**, 1687 (1931), etc.

magnitudes involved, I quote their values of onset-potential for these last: 3.9, 27.4, 62, 113, 117, 275 and 410 for the first appearance of caesium atoms lacing 1, 2, 3, 4, 5, 6 and 7 electrons respectively.

But are these the onset-potentials for "single-stage" or for "cumulative" processes? In other words, do Cs^{++} ions (to take one example) appear at 27.4 because an electron of energy 27.4 *EV* is able to remove two electrons from a caesium atom in a single operation, or, on the other hand, because a 27.4-volt electron is able to remove a second electron from a Cs^{+} ion, *i.e.* from a Cs atom already once ionized by a previous impact? This is an exceedingly important question, which is attacked by ascertaining how the phenomena vary when the pressure of the gas and the density of the electron-stream are changed. Take the rate of production, *P*, of positive ions of a given category at an abscissa somewhat beyond the onset-potential. If it is a single-stage process which is occurring, *P* should be proportional to the vapor-density ρ and to the strength *i* of the electron-stream, since the total number of electron-against-atom impacts is proportional to the product of these. Suppose on the other hand that the process involves two consecutive hits by two different electrons against a single atom: *P* should then be proportional to the squares of ρ and of *i*. As a matter of fact, cumulative processes like this

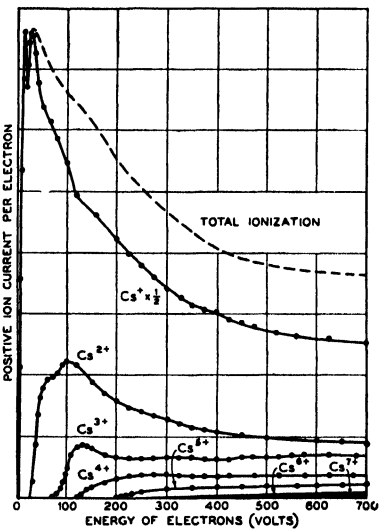


FIGURE 8.5—ILLUSTRATING THE APPEARANCE-POTENTIALS AND LIKELIHOOD OF IONIZATION FOR SEVEN STAGES OF THE IONIZATION OF CAESIUM BY ELECTRON-IMPACT. (Tate and Smith; *Physical Review.*)

last may occur in a variety of ways (we shall consider some in Chapter IX) so that the dependence on ρ or i may have some other form; but the proportionality of P to ρ and i is taken as a necessary, and usually as a sufficient, condition for assurance that the process in question is a single-stage process. Conversely, if at a particular value of electron-energy the ratio of the P-values for two types of ion varies with ρ or i, they cannot both have been produced by single-stage processes.

The foregoing test has been applied to most if not all of the onset-potentials for multiple ionization—as it may well be called—which are given in recent papers; and it affirms that multiple ionization, under the conditions of these experiments, occurs as a single process. Nobody contends, of course, that the cumulative process does not occur; nothing is more reasonable than to suppose that an atom may lose several electrons in succession by virtue of several successive impacts. The single process is, however, much the more interesting; as Tate has emphasized, how strange to think that as many as seven electrons can be detached from an atom by a single blow from a particle just like any one of them, and how difficult to visualize as a consequence of mere impact!

Ionization of molecules

We saw in Chapter III that experiments with the parabola-method revealed, a quarter of a century ago, that all sorts of things may happen to molecules in a self-sustaining discharge: removal of an electron (but removal of two or more is rare!), breakup into charged atoms or atom-groups in one or several ways, attachment of these to atoms or atom-groups derived from other molecules. One must therefore be prepared for a great variety of phenomena, when molecular gases are bombarded by electrons of controllable speed.

The earliest work of this kind was the already-cited research of Dempster,[1] who projected 800-volt electrons into hydrogen and determined the relative abundance of the ions H_1^+, H_2^+

[1] **PR 8**, 651 (1916).

and H_3^+ arriving at his collecting-electrode after travelling some distance through the gas. At gas-pressure 0.01 mm Hg, the H_3^+ ion was the most plentiful of all and the other two not far behind; at 0.0017 mm Hg, both H_1^+ and H_3^+ were definitely less abundant than H_2^+, and below 0.0005 mm Hg the H_2^+ ion was left almost alone upon the scene. The same trends are observed with much lower electron-energies, and imply that electron-impact results in detachment of one electron from the H_2 molecule, while break-up and attachment are not completed (as the latter of these two obviously cannot be!) until an impact

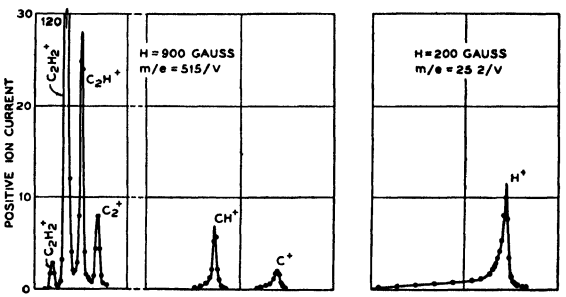


FIGURE 8.6—MASS-SPECTRUM OF THE IONS OCCURRING IN ACETYLENE BOMBARDED BY 100-VOLT ELECTRONS. THE SMALLER OF THE PEAKS LABELLED $C_2H_2^+$ IS BELIEVED TO INDICATE IONS CONTAINING THE RARER ISOTOPE OF CARBON. (Tate, Smith and Vaughan; *Physical Review.*)

of this ion with another molecule brings on the second stage of a cumulative process. The onset-potential of the initial ionization, by the way, is one of the most repeatedly-measured quantities in all of physics; Bleakney tabulated no fewer than twenty-four evaluations made between 1917 and 1932, ranging between 17.1 and 15.4 *EV*; to these he added the new value 15.37.

Not only diatomic, but polyatomic gases are now being studied: in acetylene, Tate, Smith and Vaughan located the onset-potentials or "appearance potentials" of the ions $C_2H_2^+$, C_2H^+, C_2^+, CH^+, C^+, and H^+, while in methane, Bleakney, Condon and L. G. Smith found eight different kinds

of ions each with its own appearance potential and each being formed by a single-stage process according to the test described in the section above. Perhaps the most striking case yet published is the formation of the ion C_6^+ when benzene (C_6H_6) is bombarded by electrons.[1]

A singular feature of some of these processes is, that when the impact of the electron breaks the molecule, or (to use a milder term) causes it to dissociate into charged and uncharged atoms or atom-groups, these fragments are moving with notable momentum and notable kinetic energy. A lucid case is afforded by hydrogen. I have mentioned evidence that an electron-impact may convert the molecule H_2 into the ion H_2^+ and a liberated electron, if the kinetic energy K_e of the impinging electron is greater than (roughly) 15.4 EV; and also evidence that the ion may be dissociated into H and H^+ by things that happen to it in its subsequent career in the electron-bombarded gas. There is evidence that if K_e is greater than (roughly) 18.0 EV, the conversion of H_2 into H and H^+ may occur as a single-stage process. This process has been studied by W. W. Lozier, who sent an electron-beam through hydrogen, and enclosed its path with a coaxial cylindrical electrode toward which the positive ions—the H^+ ions of the aforesaid process—started from the region of the ionization; he charged the cylinder positively until the ions just failed to reach it, and from its potential deduced their kinetic energy.[2] By conservation of momentum, the H atom and the H^+ ion resulting from the process ought to spring apart with very nearly the same speed and opposite directions; the kinetic energy of the latter is accordingly doubled, and I will call the resulting quantity K_f

[1] Tate *et al.*, *l. c. supra*; Bleakney, Condon and Smith, *J. Phys. Chem.* 41, 197 (1937); A. O. Nier, quoted in the article just cited.

[2] **PR 44**, 575 (1933). To give an idea of the conditions prevailing in a good recent experiment, I mention that the pressure of hydrogen was only 5.10^{-5} mm Hg and the strength of the electron-beam only 10^{-6} amp. A magnetic field parallel to the axis of beam and cylinder prevented the electrons from being drawn to the cylinder. Plates at right angles to the axis were disposed in the space between beam and cylinder, so that all ions not moving nearly at right angles to the axis should be caught by them before reaching the cylinder; nevertheless there seemed to be a wide-distribution-in-energy among the ions, and it is the upper limit thereof which appears as $\frac{1}{2}K_f$ in (8.2).

in order to direct attention to the fact that it is the kinetic energy of the fragments of the molecule. Lozier found that K_f is a linear function of K_e, and his results are embodied in the equation,

$$K_e = (15.4 + 2.6) + K_f, \qquad (8.2)$$

signifying that when the process in question occurs, we may divide the energy (greater than 18.0 *EV*) of the impinging electron into a quota 15.4 which goes to ionizing the molecule, a quota 2.6 which goes to dissociating the molecular ion into the atom H and the atomic ion H^+, and the left-over quantity K_f which becomes kinetic energy of these two fragments. One is not, however, to infer from this mode of description, that the process is to be visualized as a double one in which dissociation follows on ionization.[1]

Such a result as (8.2) permits of interesting and important tests. Thus, the value for the energy required to dissociate a H_2^+ ion is derivable by quantal theory from the model for this ion, and the agreement with Lozier's value of 2.6 *EV* is good. Again, if we subtract from the sum (15.4 + 2.6) the energy required to convert a H atom into a H^+ ion and a free electron *i.e.* the ionizing-energy of a H atom—the difference should be equal to the energy required to dissociate a H_2 molecule into two H atoms, which is determinable from thermochemical experiments and also from the spectrum of the H_2 molecule. Again the agreement is good. Incidentally, these agreements fortify the tacit assumption that in the process to which (8.2) refers, the kinetic energy of the freed electron is negligible. However, we are now going more deeply into molecular theory than the purpose of this chapter should admit.

Other agents of ionization

Of the other agents of ionization beside the electron, the one which most elucidates the fundamental matters is the photon.

[1] Equations of the class of (8.2) may be written in the general form $K_e - K_f = I_0 + D$, the meaning of D being obvious. Cases occur in which $K_e - K_f$ is actually less than I_0, and the fragments of the molecule are moving with appreciable speed even when K_e is adjusted to be equal to I_0 which is the onset-potential of the process.

Deferring for a moment the topic of ionization by light, I speak briefly of the remaining agents, referring elsewhere for fuller discussion.

Ionization by positive ions [1] is a much more extensive topic than we could foresee, if we were to imagine the ion as we have just been imagining the electron—that is, as a sort of a bullet which contributes more or less of its kinetic energy to ionizing the atom on which it impinges, and then flies onward otherwise unchanged. Such an action may be described by the form of words:

Ion A plus atom B become ion A + ion B + free electron. Such actions occur, notably when the impinging positive ions are bare nuclei (protons, deuterons, alpha-particles) but also when they are nuclei surrounded by incomplete electron-families, as most positive ions are. However, there are also actions of another category, thus described:

Ion A plus atom B become ion B plus atom A. The impinging ion completes its own electron-family by kidnapping one of the members of the family of the atom which it strikes! In this process of ionization, little or no drain at all need be made on the kinetic energy of the kidnapper. Take neon and helium, for instance; the ionizing-potentials of these gases being 21.5 and 24.5 EV respectively, it follows that if a neon ion encounters a helium atom, the transfer of an electron from the latter to the former requires only a supplement of $3EV$ which might be drawn from the kinetic energy of the impinging particle; while if a helium ion encounters a neon atom, the transfer not only requires no extra energy but actually releases $3EV$ to be disposed of elsewhere! As an ion-beam plunges onward through a gas, one after another of its members is neutralized by this process, and the current-density in the beam is steadily weakened; this is most marked when the ions are ionized atoms or molecules of the same kind as those of the gas which they are traversing.

Ionization by excited atoms.[2] In the foregoing case, the

[1] *Electrical Phenomena in Gases*, pp. 58–69.

[2] *Electrical Phenomena in Gases*, pp. 69, 297.

energy spent in ionizing an atom may be regarded as having been taken from the potential energy of an ion. May an atom, not itself ionized, have potential energy which it is able to employ in such a manner? We shall see in Chapter IX that an atom may indeed be put into an " excited state," or into any one of very many such states, in which it possesses potential energy of which it is able to dispose in various ways. There is indirect but forcible evidence that ionization is one of these ways. The process in question occurs mainly in mixtures of gases, of which the ionizing-energy of one member A is decidedly greater than that of the other member B, so that an atom of A may contain enough disposable potential energy to ionize an atom of B without itself becoming ionized.

Ionization by solid metals has already been mentioned in Chapter III (page 73). There is a partial analogy between this process and the second of the two processes in which a positive ion is the agent of ionization, the metal and its work-function (W_a, p. 167) taking the roles of the ion and its ionizing-potential respectively.

Thermal ionization is a term used to describe the state of affairs in an isolated gas in thermal equilibrium and so intensely hot that its atoms are continually being ionized, presumably by encounters with already-freed electrons, photons, ions, excited atoms, adjoining solid bodies if there are any—all the agents of ionization acting in single-stage or in cumulative processes or both. The process converse to ionization, *viz.*, the combination of positive ions with free electrons,[1] also continually occurs, and an equilibrium is established in which any particular feature—*e.g.* the proportion of atoms which at any instant are ionized—is a function of temperature. The situation can be analyzed by the methods of thermodynamics. The ideal condition of thermal ionization may be approached in flames and in powerful discharges in the laboratory, but manifests itself chiefly in the atmospheres of the stars.

We turn at last to ionization by light (but see also p. 294).

[1] More commonly known as recombination, and studied to some extent (cf. *Electrical Phenomena in Gases*, pp. 72–78).

Ionization by photons

Ionization is one of the actions in which light manifests itself in its corpuscular aspect, wherefore the title of this section. When a gas is being ionized by light of frequency ν, each separate act of ionization consists in the total absorption of a photon of energy $h\nu$ into an atom or molecule, which thereupon yields up an electron of kinetic energy T_i given by the formula

$$T_i = h\nu - W_i. \qquad (8.5)$$

Here W_i stands for a constant which shall be called " extraction-energy " or " binding-energy " of the electron in question.

But why not call W_i the ionizing-energy, and identify it with the quantity I_0 of the beginning of this chapter? Sometimes it can be so identified, but not in general; for if ν the frequency of the light and Z the atomic number of the irradiated element are not too low, the electrons springing from the atoms exhibit several different values of W_i—that is to say, they do not all have the same speed, but there are different speeds represented among them corresponding to different values of W_i independent of ν and distinctive of the element in question.

I prepared for this statement by describing the ionizing-potential, as defined on page 231, as the energy required to unbind or remove the most loosely bound electron from the atomic structure. Experiments with impinging electrons taught us the binding-energy of this loosest member of the atomic system, and now experiments with X-rays are going to teach us the binding-energies of various categories of more tightly bound electrons. We shall be visualizing an atom (preferably one of high atomic number, to begin with) as a system of electrons placed at various distances from the nucleus, or layers or shells of electrons of various widths disposed about the nucleus. To each of these corresponds a binding-energy, and the values of W_i supplied by the outspringing electrons according to (8.5) are these binding-energies. Better yet, in some respects, is it to visualize the ion

composed of the nucleus and $(Z - 1)$ electrons immediately after the departure of the evicted electron; but before dwelling upon these policies, I will indicate how the experiment is performed.

Let the element under study be irradiated by high-frequency light, and let a narrow beam of the emerging electrons be delimited by slits and admitted to a magnetic field perpendicular to the direction of its motion. Now the situation is like

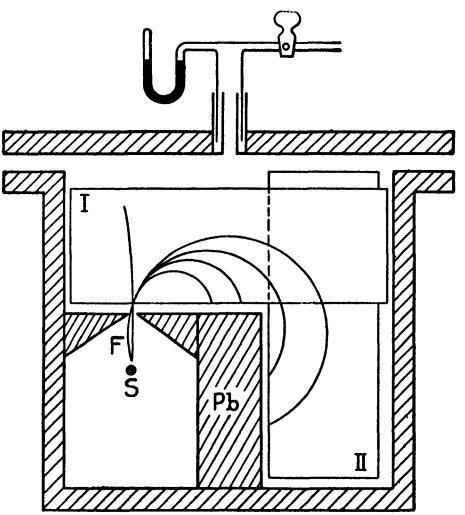


FIGURE 8.7—APPARATUS FOR INVESTIGATING THE SPEEDS OF ELECTRONS EXTRACTED FROM ATOMS BY HIGH-FREQUENCY RADIATION. (After J. Danysz.)

that which obtains when the speeds and charge-to-mass ratios of electrons and ions are ascertained by the method of electric acceleration followed by magnetic deflection (Chapter II). The differences are, that now the speeds v with which the electrons enter the magnetic field are imparted to them not by an applied electric field but by the photons which released them; and that the experimenter takes the value of e/m for granted and computes the values of v from the magnetic deflections alone. The electrons are swept around in circular arcs,

of which the radii yield their speeds according to the equation (2.7) of Chapter II.[1]

The apparatus by which such experiments are performed is of the type shown in Figure 8.7.[2] At S there is a long narrow rod or tube of the material to be tested; it is irradiated by X-rays proceeding from a source beyond the diagram to the left. A magnetic field, directed normally to the plane of the paper, sweeps the emerging electrons around in circular arcs, some of which pass through the slit. In Figure 8.8 electron-paths of this sort are reproduced from photographic films which were laid parallel to the plane of the paper, in the position of the rectangles marked I and II in Figure 8.7. These distinctly separated arcs reveal that the emerging electrons fall into several groups each characterized by a particular speed. The appearance of films laid along the top of the block Pb, normal to the plane of the paper, is shown by Figure 8.9. They suggest spectra; and though the lines are signatures of special electron-speeds rather than of special radiation-frequencies, the difference between these is not so radical as once it seemed, and we may without hesitation call them by some such name as *electronic spectra*.

Each line in such a spectrum is produced by electrons of a definite extraction-energy, extracted by radiation of a definite frequency. Continuing with the policy of referring to electrons with a definite extraction-energy as being definitely individualized within the atom, I will designate the electrons of greatest extraction-energy for any particular kind of atom as the K electrons; those of next greatest extraction-energy as the L electrons, and then the M and N electrons in due order.

[1] From equation (2.7) it is seen that, the value of e/m being taken for granted, the value of HR serves as a measure of the electron-speed v. More commonly nowadays it is regarded as a measure of momentum; the momentum-value corresponding to a given $H\rho$-value depends on the charge but not the mass of the particle. Values of HR, or (as it is more generally called $H\rho$, are frequently given instead of values of speed or momentum. A conversion table is given by Fournier (**JP 6**, pp. 29–32; 1925).

[2] This and the next three figures are from the articles of J. Danysz in the *Journal de Physique* for 1912 and 1913. I am indebted to the editor of the *Journal* for permission to republish them, Mr. Danysz having died in the war.

(Later on it will be necessary to subdivide these classes, but for the moment this may be avoided.) At other times I shall speak of these electrons as belonging to the K level or to the K shell, and so on. In a given electronic spectrum we may expect to find a set of lines due to K. L. M and other electrons.

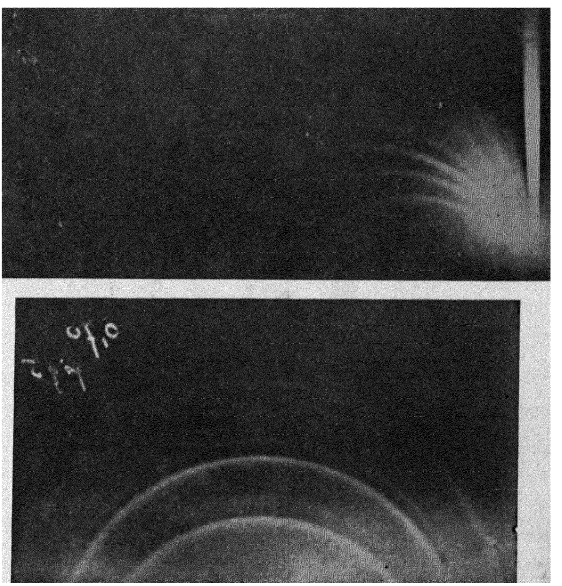


FIGURE 8.8—TRACES OF STREAMS OF ELECTRONS OF DISTINCTIVE DISCRETE SPEEDS ISSUING FROM IRRADIATED ATOMS, AND BENT INTO CIRCULAR ARCS BY A MAGNETIC FIELD. (J. Danysz; *Le Radium.*)

for each frequency represented in the incident radiation; unless there are some of these frequencies for which the quantum energy $h\nu$ is less than the extraction-energies of some of the electron-groups, in which case there will be no corresponding lines.

An ideally simple electronic spectrum would be produced by a single radiation-frequency; but this is impracticable, for even if one were to eliminate from the stream of X-rays proceeding out of an X-ray tube all but one frequency, the irradi-

ated atoms would themselves supply others.[1] As in the mass-spectra obtained in the study of isotopes, this unavoidable complexity is actually an advantage; it helps in identifying the several lines.

In Figure 8.11, a photograph taken in the manner already mentioned, there appears above the electronic spectrum due to silver atoms irradiated by the characteristic X-rays of tungsten.[2] To guard against the possibility that the photographs may lose in clearness by the process of reproduction, I will base the explanation upon the uppermost of the sketches in Figure 8.10, which is abstracted by de Broglie from similar pictures.

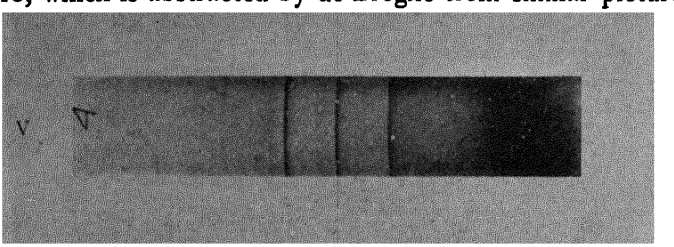


FIGURE 8.9—ELECTRONIC SPECTRUM. (J. Danysz.)

The electron-speeds corresponding to the lines increase from left to right. The irradiating X-rays consist of four characteristic frequencies from the X-ray spectrum of tungsten; in order of decreasing frequency they are known as $K\gamma$, $K\beta$, and the two members of the $K\alpha$ doublet. The four lines marked 4 and 5 in the electronic spectrum are made by electrons extracted by these four radiations from a single level—the K level of the silver atoms. The two following doublets, marked 6 and 7, are made by electrons extracted by the $K\alpha$ frequencies from two other levels of the silver atom, the L and M levels respectively. Line 8 is due to $K\beta$ extracting electrons from the L

[1] These are in fact especially efficient in ejecting electrons, as they originate within the atom-layer itself.

[2] I am greatly indebted to M. de Broglie for sending me the negatives of these admirable pictures, as well as those for the Figures 8.10 and 8.12.

level. At the other end of the spectrum, the three lines 1, 2, 3 are due to electrons ejected from the L and the M levels by two

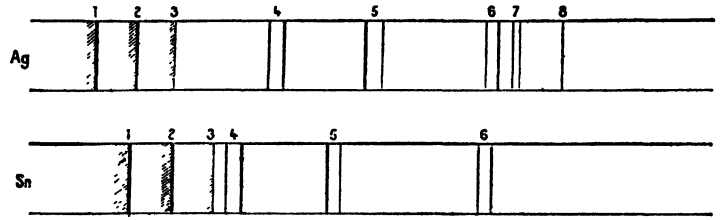


FIGURE 8.10—ELECTRONIC SPECTRA SKETCHED FROM PHOTOGRAPHS. (M. de Broglie; *Journal de Physique.*)

of the X-ray frequencies characteristic of silver, which the irradiating X-rays stimulate some of the silver atoms to emit. The rays responsible for these particular lines are the so-called

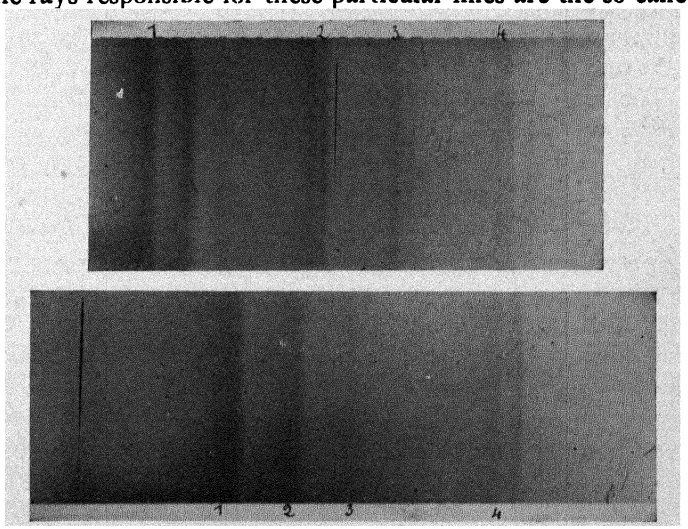


FIGURE 8.11—ELECTRONIC SPECTRUM OF SILVER IRRADIATED BY X-RAYS OF TUNGSTEN (above) AND OF SILVER (below). (M. de Broglie.)

Kα and Kβ rays of silver, which are so related to one another (as will be stressed in a later passage) that the electrons

extracted by the former from the M shell have very nearly the same energy as the electrons extracted by the latter from the L shell, so that the two frequencies acting on the two groups of electrons produce three (instead of four) distinct lines of the electronic spectrum.

Reverting now to the photographs: above in Figure 8.11, the pairs of lines marked 4, 3, and 2 are those designated respectively as 6, 5, and 4 in the sketch and in the foregoing explanation, while the lines to the left are those produced by characteristic X-rays of silver acting upon silver atoms. On a larger scale, this latter region of the spectrum is shown below in Figure 8.11; here the lines are marked by the same numerals as in the sketch; the pair at 4 is due to K electrons extracted by the two Kα rays of tungsten, the line 3 is due to M electrons extracted by the Kβ radiation of silver, the line 2 results jointly from L electrons extracted by the Kβ radiation of silver and M electrons expelled by the Kα radiation of silver, while the line 1 is due to L electrons ejected by the Kα rays of silver.

The resemblance and the differences between electronic spectra of elements not far apart in the Procession are illustrated by the two sketches in Figure 8.10, the lower relating to tin (atomic number 50) and the upper to silver (atomic number 47) irradiated by the same frequencies. Since the extraction-energy of each named class of electrons increases along the periodic table, the lines designated as 4, 5, and 6 in the electronic spectrum of silver reappear in that of tin, displaced in the direction of diminishing electron-speeds, that is, to the left. But, as to the lines 1, 2, and 3, both the extraction-energies of the electrons and the frequencies of the rays responsible for these alter as one passes from silver to tin, and the net result of the double alteration is that the lines are displaced to the right.

At this point it is suitable to insert a historical note. The electronic spectra earliest discovered were due to atoms irradiated, not by X-rays artificially generated and played upon them from without, but by gamma-rays proceeding from disintegrating atoms of radioactive substances interspersed among

the atoms supplying the electrons—in fact, the ejected electrons in many cases come out of the very atoms within which the ejecting radiation is generated. The figures introduced at the beginning of this section were reproduced from photographs made by J. Danysz in the course of research upon the electrons or " beta-rays " emerging from radium B and radium C. The grouping of these electrons into classes with sharply defined speeds results from their being extracted from the various levels by the several gamma-rays and the inevitable radiations which these stimulate within the atoms, in the same fashion as the X-rays from tungsten stimulate X-rays from silver atoms in the earlier-mentioned cases. The large number of distinct groups (Rutherford and Robinson distinguished sixteen from RaB and forty-eight from RaC) is due partly to the multitude of electrons possessed by these massive atoms and partly to the multitude of radiation-frequencies which participate. The earliest electronic spectra produced from atoms of non-radioactive elements were excited by playing gamma-rays upon these, and the earliest rule discovered was that these spectra resemble the electronic spectra of the elements responsible for the gamma-rays, in much the same fashion as the two sketched in Figure 8.10 resemble one another; being indeed identical, when the irradiated atoms and the atoms providing the radiation are isotopes of one and the same element. We revert to these matters in Chapter XVIII.

The energy-values of the various states of ionization of an atom—or, in terms of the customary picture, the extraction-energies of the various classes of electrons within the atom—may be determined with a certain degree of precision from experiments such as these. However, as in the case of the measurement of charge-to-mass ratios for individual ions by the methods of Aston and Dempster, there is little incentive to develop the accuracy of the method to the highest possible extent; for most of the energy-values in question can be determined with very great accuracy in another way, which we will next examine.

Absorption of radiation through ionization

When a beam of radiation of frequency ν is transmitted through a layer of matter, from the atoms of which it extracts electrons with an expenditure of energy $h\nu$ at each extraction, we should expect to find it correspondingly reduced in intensity when it emerges from the layer.

This effect is strikingly conspicuous with radiation high enough in frequency to detach the tightly bound electrons of massive atoms. Let a narrow beam of " heterogeneous " radia-

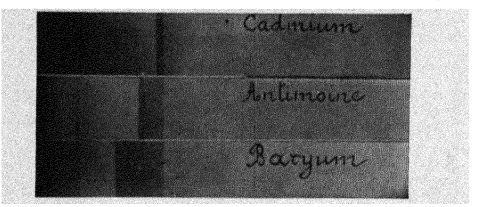


FIGURE 8.12—K ABSORPTION EDGES OF THE ELEMENTS $Z = 48, 51, 56$. (M. de Broglie.)

tion, containing all frequencies throughout the widest possible range, fall from an X-ray tube through slits and diaphragms upon a thin layer of such atoms; let the transmitted rays be dispersed by some appropriate spectroscope, and fall finally upon a photographic plate on which their spectrum—in the ordinary sense of the word, not in the sense of " electronic spectrum "—is outspread.

In Figure 8.12 there are three such spectra, of heterogeneous beams which have passed through layers of cadmium, antimony, and barium respectively. The frequency increases from right to left. The darkening at any point is a measure of the intensity with which the X-rays acted at that point.

Below a certain frequency identical for all three elements, the photographic films have evidently been little affected; as soon as this critical frequency is exceeded, the effect suddenly becomes enormous. This critical frequency is the one for which the quantum-energy just suffices to extract a K-electron from a silver atom; for the photographic film contains silver,

and it is the expulsion of electrons from the atoms in it which initiates the photographic process. Proceeding always toward higher frequencies, we see that presently the plates suddenly become whiter, at another critical frequency which, however, is not the same for the three elements. The photographic film is not responsible for these " absorption-edges " as they are called; each of them occurs at the particular frequency for which the quantum-energy just suffices to extract a K electron from an atom of the element which formed the absorbing-layer placed in the path of the beam before it reached the plate. To

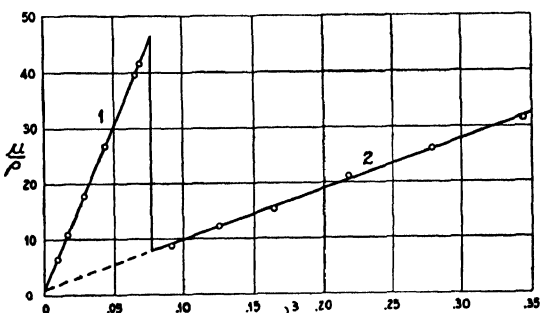


FIGURE 8.13—K ABSORPTION EDGE OF TIN, AND TREND OF ABSORPTION WITH WAVELENGTH IN ITS VICINITY. (Richtmyer; *Physical Review.*)

the right of the absorption-edge we have the lower frequencies, unimpeded by the cadmium (or antimony, or barium) atoms because unable to ionize them; to the left we have the higher frequencies, reduced in intensity by the intercalated matter because some of their energy was drawn off to detach electrons.

An absorption-edge of tin—the one corresponding to extraction of a K electron—is shown in another manner in Figure 8.13; here the absorption-coefficient of the element as determined by direct measurement is plotted against the cube of the wavelength of the radiation.

From the frequency ν_i at such an absorption-edge, the corresponding binding-energy is, of course, derived by the equation:

$$h\nu_i = W_i. \tag{8.6}$$

This is a much exacter way of measuring binding-energies than the observations on electronic spectra can supply. Nevertheless the measurements on these last are of extreme importance, for they show beyond cavil what is effected by the absorption of photons which sets in when one or another of these frequencies is overpassed.

Notice that the frequencies ν_i and binding-energies W_i are the examples *par excellence* of non-periodic and progressive functions of Z quoted in Chapter IV, because of their pertaining to inner electrons. It is true that the statement must be limited: take even the binding-energy W_K for the innermost shell of electrons—as we go down the procession of the elements from $Z = 92$ to $Z = 1$, even this innermost shell must in the end become the outermost. As a consequence of this, the linear relation between $\sqrt{W_i}$ and Z which was mentioned in Chapter IV extends from $Z = 92$ downward, not all the way to $Z = 1$ but roughly to the point where the shell in question is at the periphery of the atom.

Having in these last two sections concerned ourselves with ionization chiefly as an index to atomic structure, we begin anew to study it for its own sake.

Likelihood of ionization.[1]

We have seen that the power of corpuscles—be they electrons or photons—to ionize an atom in a specific way commences when their energy surpasses a specific definite amount. For each mode of ionization of each kind of atom there is a critical energy-value V_0 " of onset," such that the impinging particles can or cannot produce that mode, according as their transferable energy is greater or less than V_0. We have now some notion of how that statement is founded and how the many values V_0, beginning with that lowest of them all which has been called I_0, are sought for. Now we must pose and

[1] I am indebted to the Williams and Wilkins Company for permission to reproduce, here and elsewhere, certain passages *verbatim* from *Electrical Phenomena in Gases*.

seek to answer a question incomparably more difficult. We require knowledge more exact than the mere " can or cannot, according as " of the principle above.

The question may be roughly phrased as follows: how plentiful is the ionization of any particular mode produced by corpuscles of energy V, and how does its abundance vary with V? But to convert this into a precise question with an answer exact and definite, attainable by experiment and valuable for theory—this is a real task! The passage of a beam of corpuscles across a stratum of matter, say a gas, is formidably complex. If no corpuscle of the beam ever lost any energy nor ever was deflected except when performing an ionization, and every one of them vanished completely as soon as it performed an ionization, the problem would be a great deal simpler than it is. Then, a homogeneous beam of electrons or photons entering a gas with equal speeds along parallel lines would remain such a beam, and the quotient of its ionizing-power by its intensity would be the same all along its course. But as things are, the beam steadily becomes more and more non-uniform as it proceeds across the gas, owing to losses of energy and changes of direction incurred not only at ionizations but because of the other processes which we shall notice in Chapter IX.

Naturally, then, the experimenter tries to make so rare the gas, so thin the stratum thereof which the corpuscles traverse, that the beam is not thus modified. Conceivably this might be impossible. Conceivably the interactions of atoms and electrons might be such, that in most cases an electron would be deflected hundreds of times, losing energy at each deflection, before striking an atom in such a way as to ionize. If this were the case it would be hard indeed to devise an intelligible experiment. Often, however, it is not the case, and the experimenter can make the gas so rare and the stratum so thin that, in the common parlance, " most of the corpuscles hit only one atom on the way across." It is for this situation that the question is formulated, and the answer sought; and similar questions are asked not only as to ionization, but as to

every process which electrons, photons, atoms and molecules inflict on one another.

The suddenness of the onset of ionization spared physicists for a dozen years from coping with the greater problem. It is so much easier to set up a tube which gives a curve on which that onset is marked by an evident break, than to improve the design of that tube until the trend of the smoother parts of the curve and the heights of the ordinates thereof become significant! Now that the ionizing-potentials are so generally known or knowable from spectra, the relatively easy task is finished, the harder one is unavoidable; and high time that it should be! for quantal theory has been advanced to the point of being able to interpret such data, and also they are essential to the understanding of the electric discharge in gases.

Let us suppose that we have a slab of gas of area A and thickness D, containing N atoms or molecules per unit volume and NAD altogether. Let Q corpuscles of energy V be projected against the face of the slab in a beam of uniform density, so that Q/A of them fall on unit area of the slab. The number of atoms ionized in some specific way by single impact—call it I—should be proportional to NAD and to Q/A. Write it thus:

$$I = f(V) \cdot Q/A \cdot NAD = f(V) \cdot NQD. \qquad (8.10)$$

The coefficient of NQD in the right-hand member shall be the principal measure of the " likelihood " or " probability " of ionization-by-single-impact by corpuscles of energy V.

I wrote above " the number of atoms ionized . . . should be proportional to NAD and to Q/A." This is meant to mean that if I is not proportional to N, to Q and to D, the conditions are not right for experiments of this type. Conceivably in a certain tube at a certain pressure, the number of ions appearing might vary as the square of the strength of the beam. This might mean that ionization by double impact was occurring, or it might mean something else; but at any rate the conditions would not be proper. Or the ionization might not increase when the thickness D of the slab of gas was doubled; this would mean that the gas was already so thick that all the

corpuscles were swallowed up in it, which again would prove the experiment wrongly arranged for the present purpose. But of this, more later.

The quantity $f(V)$ has the dimensions of area; on which ground, there is some objection to calling it " probability " (though at present the notion of probability is so dominant in theoretical physics, that the name may well be an advantage!) It has another name, which I proceed to introduce.

Picture the atoms of the gas as tiny spheres, so that in looking down upon the stratum with the eye of the imagination you see a multitude of circles: the cross-sections of the various spheres, one for every atom. Imagine the spheres so small, or else the gas so rarefied or the slab so thin, that there are hardly any overlappings of the circles—hardly any atoms partially concealed by others getting in front of them. Denote by r the radius of a sphere, by $\sigma(=\pi r^2)$ the area of a circle. Since under unit area of the surface of the slab there are ND atoms, one may say that of unit area the part σND is covered by circles, the part $(1 - \sigma ND)$ is open. If now ionization occurs whenever a corpuscle strikes one of the spheres, the number of atoms ionized or the number of positive ions formed when Q corpuscles fall upon the slab bears the same ratio to Q as the area σND bears to unit area: it is σNQD. Comparing this expression for I with the one given above:

$$\sigma = f(V). \tag{8.11}$$

The function $f(V)$, which has been introduced as the principal measure of likelihood of ionization, is the cross-sectional area of the spherical model of the atom.

This last paragraph was written as though the size of the sphere, the values of r and σ, were predetermined. The fact of course is the opposite. The value of $f(V)$, being found by experiment, is given as the value of the " cross-section of the atom for ionization-by-single-impact of corpuscles of energy V." It is a long phrase, and people mostly shorten it to " cross-section of the atom "; but this is dangerous, for there

are many kinds of cross-section of the atom. Even the long phrase above is ambiguous as it stands. I have already spoken of two kinds of corpuscles, electrons and photons; the values of σ for the two are very different, even at equal values of V. The values for the various kinds of positive ions are still different. Further, there are numberless other kinds of cross-section, many of which we shall encounter later.

Another measure of likelihood is the ratio of this cross-section, $f(V)$ or σ, to some more or less arbitrary standard area. As standard, people usually take the "gas-kinetic cross-section." In the kinetic theory of gases, there is a famous method of computing the "size of the atoms" of a gas from the viscosity thereof; which is to say, a method of determining the size of a sphere, such that if the atoms were actually elastic spheres of that size, the gas would have the viscosity which it is actually observed to have. The radius and the cross-section of this sphere are called the "gas-kinetic" or the "kinetic-theory" radius and cross-section of the atom or molecule of the gas. Denote the latter by σ_0. The ratio $f(V)/\sigma_0$ or σ/σ_0 is frequently given as the measure of the likelihood of ionization, in place of σ by itself. But there is nothing compulsory about the choice of σ_0 as standard of reference. If some arbitrary area of convenient size were used instead, people would be less likely to accept the groundless idea that σ/σ_0 "ought to be" equal to unity, that any value markedly different from unit is a cause for wonder.

I interpolate here a little table of values [1] of the gas-kinetic cross-sections of gases common in discharges. Different authors may have chosen different values of σ_0 for the same gas, and in principle it would be better, in referring to an article in which a value of the aforesaid ratio is given, to quote the particular value of σ_0 on which it is based. But sometimes this is not explicitly stated; and in any case, the uncertainty in any given value of σ usually amounts to at least as much as

[1] Data for A, Ne, Kr, Xe from Rankine (see W. H. Keesom and W. H. Mooy, *Nature* **126**, 243; 1930); the others from Landolt-Börnstein, 2d supplement to 5th edition.

the disparities between the better values of σ_0. The table follows: the unit of area is 10^{-16}cm.^{2} ("square Ångstroms"):

He 2.83, Ne 4.14, A 6.46, Kr 7.54, Xe 9.16,
Hg 10.2,
O_2 6.9, N_2 7.8, H_2 3.7,
H_2O 5.8.

Another measure of the likelihood of ionization (notably found in German papers) is the product of σ by the number N_1 of molecules per cc of gas at 1 mm Hg and 0° C. N_1 is about $3.56 \cdot 10^{16}$. Still another—an inverse measure, this time—is the reciprocal of $N_1\sigma$, called "mean free path for ionization" at the stated pressure and temperature. The well-known concept of mean free path, useful when the results of these measurements are to be applied to actual situations, is not so recommendable when the results themselves are to be put on record.

It is evident that the stratum or slab of gas can safely be called sufficiently thin, only if hardly any of the circles aforesaid overlap, therefore only if σND is small compared with unity. In one case we can relax the severity of this demand: if the kinetic energy of the impinging electron is very high compared with the amount spent in ionization and if the electron is not sensibly deflected when it ionizes, overlapping of circles will not be serious, for if one atom is behind another the electron can ionize both. Evidently also, the mention of circles and spheres does not require the assumption that the atoms are really spherical; they may very well be non-spherical, in which case σ stands for the average cross-section.

Likelihood of ionization by low-energy electrons.

Most of the measurements of likelihood of ionization pertain to electrons of energy-values lower than 1000 ev, which by contrast with those to be later discussed may be called "low-energy electrons."

Since all modes of ionization result in freed electrons and positive ions, a mere enumeration of either gives no full answer

to the precisely formulated question. Few indeed are the experiments of which the data refer to a single mode of ionization. If the observer is undertaking to distinguish between single, double, triple and other degrees of multiple ionization, the neatest way by far is the method employed by the Minnesota school, and already mentioned when I was speaking of onset of ionization: the ions are drawn by an accelerating electric field out of the region of ionization, and separated by a deflecting magnetic field into beams according to their charges; the intensity I of each beam is measured, the tests for proportionality thereof to Q and N are made, σ is computed and is plotted against the electron-energy V. (Sometimes the experimenter plots I against V, not concerning himself with the computation of σ.) The typical shape to which most of such curves conform is distinctive: *a smooth swift rise from onset to a rounded maximum* (at a value of V two to four times as great as V_0) *and beyond it a smooth and gentler fall toward the axis of abscissae as the asymptote* (see Figure 8.5).

Even such data do not show how to discriminate between different modes of ionization in which the same number of electrons quits the atom, but the resulting ion is left in different states. Some of the curves, however, display humps or wrinkles which distort them out of the simple smooth shape aforesaid, some even displaying two maxima; it is reasonable to infer that these are signs of onset of new modes. One might expect to be able to detect better signs by observing the energy of the electrons which rebound after effecting the ionizations, but these are mixed up with the detached electrons, not to speak of others of the primary beam which have not ionized; something has been done along this line, but not very much.

If the experimenter (as always prior to 1930) measures the total charge of all the positive ions produced in a stratum of gas, determines a σ-value and plots it against the electron-energy, he gets a curve which resembles those more specific ones already mentioned in that it rises from zero at an onset-energy, bends over to a broad and flattish maximum, and thereafter slowly declines. The onset-energy is of course the ionizing-

potential I_0; over a short energy-range the curve refers altogether to detachment of the loosest electron, but long before the maximum is reached other modes of ionization must be commencing one after another. (It is odd that even so, most experimenters have found the curve sensibly smooth.) The interpretation is therefore obscure, but the curve is nevertheless of practical and some theoretical value.

Total ionization [1]

Now we turn to some experiments of an interesting type, leading to a result remarkably simple. Some were performed by projecting a stream of electrons (of uniform initial energy E) into a chamber filled with gas so dense and deep that they lost all of their energy except presumably that of thermal agitation —*sich totliefen*, "ran themselves dead," in the dramatic German phrase. The ions of either sign being drawn to a collector, their total charge was measured. Its quotient by the product of e into the number of injected electrons is called "total ionization" or "number of ion-pairs per electron" for the gas and the E-value in question. The latter name conceals the tacit assumption that each of the positive ions is singly charged,[2] which is certainly wrong; a better policy is to denote the quotient by a symbol such as S. In experiments of another type, the ionizing electrons were started off in the midst of the gas itself by X-rays, the number of photons absorbed in the gas being ascertainable; again the charge received by a collector was measured, and its quotient by the product of e into the number of absorbed photons was identified with S. This procedure involves the assumption that practically every absorbed photon has conferred its entire energy by the photoelectric process upon an electron initially

[1] See Eisl, **AP** (5) 3, 277 (1929), for his own experiments and for references to and a discussion of the others: or *Electrical Phenomena in Gases*, pp. 47–53, 70.

[2] Formally the situation may be saved by defining an ion-pair as composed of a freed electron and the corresponding charge on a positive ion, in which case a multiply-ionized atom would belong to two or more ion-pairs. Notice that it is nowhere guaranteed that all of the energy lost by the electrons is spent in ionization!

very loosely bound to an atom, which one would think too extreme; yet the method has been repeatedly employed and is well regarded.

The remarkably simple result is, that if E is greater than a couple of thousand electron-volts or so, *S is sensibly proportional to E.* With respect to air, to which nearly all of the data pertain, Eisl summarized the situation in the words: "it may be said that the true value [of E/S] should lie between 30 and 35 electron-volts per pair of ions, but with an uncertainty of twenty per cent." Of his own work however he said, that the value of E/S over the range from 9000 to 59,000 eV may be taken as constant and equal to 32.2, with an uncertainty of only a couple of per cent.

This result is widely generalized; for it has been shown by experiment [1] to be valid for alpha-particles of energies ranging up to eight millions of electron-volts, and is presumed to be valid for all charged particles electronic or nuclear traversing air. In the study of transmutation it is often invoked for estimating the energies of such particles, and even in the study of cosmic rays. The values of E/S for air accepted by various physicists in making such estimates range from 30 to 35 eV.

Ionization by very energetic charged particles

Let us now think of charged particles, electronic or nuclear, moving through air of ordinary density with speeds so high that over distances of the order of centimetres their tracks are sensibly straight. With respect to these it is possible to define, and permissible even to aspire to measure, such quantities as the amount of ionization or the number of ion-pairs *per unit length of track.* It is however necessary to analyze each measurement with care to see just what it gives.

Imagine first an ionization-chamber with plane-parallel walls a distance dx apart, traversed by a beam of particles which move along the direction normal to these walls, entering the enclosed air with energy E and leaving it with energy

[1] Gurney, **PRS 107**, 332 (1925) and literature there cited.

E–dE. The walls must of course be very thin to permit the particles to penetrate, and dx should be so small that dE/E is itself a small fraction of unity. What is then measured is akin to the "total ionization" just described, except for the fact that the ions are formed while the energy of the particles is dropping from E to E–dE rather than from E to zero. Let N stand for the number of molecules of air per unit volume in the chamber, QI the total charge received by the collector when Q particles traverse the chamber; if then we denote $I/eNdx$ by σ, the symbol σ will have as nearly as possible the same meaning as it had on page 261 where we were dealing with ionization by slow electrons without discriminating between single and multiple ionizations.

Suppose however that the instrument is the cloud-chamber, in which the tracks of individual particles are visible: can the observer estimate the ionization by counting the droplets? This would be quite impossible with nearly all of the pictures reproduced in this book, though the contrasts between the tracks of particles of widely different masses are often striking. One may however hope to distinguish the droplets with the aid of high magnification, combined at times with the scheme of observing the tracks of particles which passed through the chamber say half a second before the expansion, so that the ions had time to diffuse apart before the condensing water-vapor immobilized them. For alpha-particles and other nuclei of customary speeds the prospect is hopeless, the droplets being too congested; but for electrons of all but low speeds, and for nuclei of speeds closely approaching that of light, the feat is practicable—as Figure 3.2 has shown us already.

Looking at Figure 3.2 (page 68) we see that some of the droplets belong in pairs, others in groups of more than two members. The pairs are instantly intelligible. The groups are also explicable, for it is reasonable to assume that some of the electrons ejected from molecules of air by the traversing particle come forth with speeds so high that they go an appreciable way and do some ionizing of their own; indeed one often sees a track with a branch which is obviously the trail of such

a "secondary" electron endowed with initial speed abnormally high. By counting the groups rather than the individual droplets, one gets a measure of the number of acts of ionization performed directly by the traversing particle, the so-called "primary ionization." Theories of ionization generally make predictions of the amount of this primary ionization per unit length of path, and the expansion-chamber is thus the only instrument capable of testing them.

Such a picture as Figure 3.2 clarifies the role of the ionization-chamber, which is that of measuring the entire ionization produced directly and indirectly by the traversing particles; to this the quantity σ, as just defined, refers. The cloud-chamber seems capable of indicating the σ-value also, since the observer may count the individual droplets. Its drawback in this connection is, that no one can ever be sure that a single droplet is not enveloping two or more distinct ions, or a single positive ion of charge greater than $+e$; also there is the danger that many of the ions will annul one another by recombination in the interval between the passage of the particle and the condensation of the water. Only a lower limit for the σ-value may thus be asked of the cloud-chamber.

Figure 8.5 has already suggested to us the long steady decline of the typical σ-vs-E curve, as E is increased beyond the value at which the maximum of σ occurs. Electrons fast enough to make long trails in a cloud-chamber have energies so great that their σ-values lie very far out along the descending arc. As to the primary ionization per unit length of path, let us pause for a moment to consult the theories, beginning with the classical one exemplified in Chapter V. There, we were considering what happens when a stream of alpha-particles traverses the fields of a flock of charged nuclei of (practically) infinite mass: the traversing particles are deflected without transferring (practically) any energy. Here, the problem is to consider what happens when a stream of charged particles traverses the fields of another flock of charged particles—to wit, the orbital electrons of the air-molecules—which are no more massive than they (and indeed much less massive, if the

oncoming bodies are nuclei). The deflections are now attended by energy-transfers, the oncoming particles giving part of their kinetic energy to those which at first were stationary. The calculation of the energy-transfer occurring when one particle approaches another involves no new principles beyond those already used in Chapter V.[1] If Q electrons of sufficiently high velocity v traverse a stratum of thickness t containing b free stationary electrons per unit volume, the number of these last which acquire kinetic energy superior to an arbitrary amount J should be given by the formula,

$$It = Q \cdot 2\pi bte^4/mv^2J \qquad (8.12)$$

The problem thus solved is not however quite the one which confronts us, since the electrons in the air-molecules are neither stationary nor free, and we wish to know the number of these which the traversing electrons eject. Thomson, after deriving formula (8.12), interpreted it as giving the number of electrons which should be ejected if initially they were all bound to molecules with binding-energy J. When quantal theory is employed instead of classical, the formula is reminiscent of (8.12), for according to Bethe it takes the form,

$$I = Q \cdot 2\pi be^4/mv^2 \ (A \log_e Bmv^2/J) \qquad (8.13)$$

A and B here standing for constants dependent on the structure of the atom (given as 0.285 and 42 for atomic hydrogen). The logarithmic factor slows down the decline of I with increasing v. When the relativistic dependence of mass upon speed is taken into account, m in the logarithmic factor is to be replaced (according to E. J. Williams) by $m_0/\sqrt{1-\beta^2}$; and now I passes through a minimum at a rather high energy-value, about 1.5 Mev.

It is in this final form that the theory has been compared

[1] The procedure consists in referring the motions to the frame of reference in which the centre of gravity of the two particles is stationary; in this frame the problem becomes purely a problem of deflection, as in Chapter V; on reverting to the frame of reference in which the "struck" particle was initially stationary, one finds that in it this latter is moving.

with the countings of droplet-groups along the tracks of fast electrons in expansion-chambers; those of Williams, Terroux, Skramstad and Loughridge on electrons springing from radioactive substances, with energy-values ranging up to 2 Mev; and those of Corson and Brode on cosmic-ray particles, with energies ranging nearly twice as high. The first-named four ascertained that the decline of I with increasing v is much slower than the classical formula demands, while Corson and Brode succeeded in observing the minimum and the upturn beyond it.[1]

I revert to equation (8.12) in order, first, to emphasize that it is only an approximation—suitable for high values of v—to the proper "classical" formula which Thomson derived from the stated assumptions. The proper expression for I is a function which rises from zero at $v = \sqrt{2J/m}$ and passes through a maximum at $v = 2\sqrt{J/m}$: thus the classical theory is capable of depicting roughly, but only roughly, the trend of the σ-curves in the vicinity of the ionizing-potential as exemplified in Figure 8.5. A second and a very important point is this: m in (8.12), as well as in the proper formula to which (8.12) approximates, is the mass of the "struck" particle and not that of the "striking" one. The mass of the latter nowhere appears in the formulae: *the ionizing-power of a charged particle depends only upon its charge and its speed, not on its mass.* (After a moment's thought, this principle seems self-evident). Nothing has yet been said about the dependence of I on the charge of the ionizing particle, but this can easily be seen to be quadratic. If allowance is made for this, say by introducing Z^2 as a multiplying factor in such formulae as (8.12) and (8.13), the formulae claim equal validity for all types of ionizing charged particle (this incidentally is why v should always be used instead of E as independent variable). Thus an electron, a proton and an alpha-particle moving with identical speeds

[1] Classical theory: Thomson, **PM 23**, 449 (1912). Quantal theory: E. J. Williams, **PRS 135**, 108 (1932) and literature there cited. Cloud-chamber estimates of primary ionization: Williams and Terroux, **PRS 126**, 289 (1930); Skramstad and Loughridge, **PR 50**, 677 (1936); Corson and Brode, **PR 53**, 773 (1938).

through a gas perform ionizations at rates per unit length of path which stand in the ratios 1 : 1 : 4.

Combining the principle just printed in italics with the fact that any σ-curve for electrons has a maximum at a fairly low energy, we are led to anticipate a maximum in a σ-curve for alpha-particles at an energy-value some eighty-five times as great. Such a maximum is displayed by the curve in Figure 16.7 (page 521) which is a σ-curve for alpha-particles obtained with an ionization-chamber. The independent variable is a function of v (to be explained in Chapter XV) which increases from 3.0 to 8.5 as v decreases from $3 \cdot 10^9$ cm/sec to zero. Curves of this character, by the way, are known as "Bragg curves."

A Bragg curve with many more data-points could easily have been found, but this is one of peculiar interest. One of the prime objects of studying ionization by charged particles of high energy is to enable physicists to recognize such particles by observing their ionizing-power without having to take the (sometimes insuperable) trouble of measuring their charge-to-mass ratios by the methods of Chapter II. Such experiments as I have just been citing give I or σ as empirical functions, not indeed of v the speed of the particle but of qualities directly measurable, such as $H\rho$ of page 250 or the abscissa of Figure 16.7. For the particles "produced in the laboratory," as physicists say—*viz.* ejected from radioactive bodies, formed by transmutation, or accelerated by cyclotrons or high-voltage engines—these differ so much that it is easy to tell electrons, protons, alpha-particles and nuclei of greater charge apart. The curve of Figure 16.7 relates to particles born from a certain reaction of transmutation, which were identified as alpha-particles because it agrees with the curve appropriate to these.

This energy-range of facile discrimination runs up past 20 Mev. At very much higher energy-values where the speeds of all particles come close to that of light, it may be expected that protons and electrons will have sensibly the same ionizing-power while alpha-particles retain their four-

fold advantage. Many of the cosmic-ray particles show by their $H\rho$-values that they have energies in the hundreds, even the thousands or tens of thousands of Mev; these exhibit about the same density of droplets [1] along their courses as do the electrons of energy-values 2 Mev or thereabouts, by which equation (8.13) was tested; it is accordingly supposed that they bear charges $\pm e$ instead of $\pm 2e$ or greater, but no inference is drawn as to their masses. Occasionally one finds in the Wilson chamber the track of such a particle having a $H\rho$-value so small that it falls within the range of facile discrimination; the density of droplets may then be intermediate between the values proper to proton and electron, suggesting a body of identical charge and intermediate mass—these are the "mesotrons" (page 276).

The sudden termination of alpha-particle tracks as exhibited in many cloud-chamber photographs enables the physicist to define and employ the concept of "range," meaning the distance to the termination from either the point where the alpha-particle quits its source or some point along its path where its speed is known. "Range-vs-speed" relations play an important part in the study of transmutation (Chapter XIV). When the word "range" is used without qualification, it means range in air at normal temperature and pressure.

Likelihood of interception of high-energy photons.

Instead of "likelihood of ionization by photons," I choose the broader caption just above, using "interception" to connote anything that may happen to particles of light as they travel through matter. The adventures possible to photons are of four types at least, in the energy-range beyond (say) 10^4 ev to which this section shall be restricted. Ionization of the kind described on pages 248–258—total absorption of the photon and transfer of all of its energy to a bound electron—corresponds of course to interception in the strongest sense of

[1] Anderson (PR 44, 406; 1933) gives 62 and Locher (PR 39, 883; 1932) 72 droplets per cm, to be compared preferably with Corson and Brode's value of 50 (*l. c.*) for cosmic-ray particles presumed to be electrons of energy-values around 2 Mev.

the word. Frequently it is called "the photoelectric effect." Of the three other processes we made the acquaintance in Chapter VI. Of these the Compton effect is also an ionization-process, affecting usually the most loosely-bound electrons of the atom or else the "conduction electrons" which circulate in metals without being especially bound to any one atom; it does not entail the total absorption of the incident photon. Pair-formation resembles ionization, since electrons spring from the irradiated substance; it would indeed be included under some definitions of ionization, but not under the most meticulous, since the electrons are not detached from atoms but are newly created. Deflection or "scattering" of photons without sensible change of energy is a process which neither is nor resembles an ionization.

If one projects a well-collimated beam of photons against a stratum of matter, and measures the number of these which traverse the stratum and continue with undiminished energy in their initial direction, the extent to which this falls below the number of incident photons is a measure of all four of these processes conjointly; it yields what may be called a "σ-value for interception." Such a curve as that of Figure 8.13 is in fact made in this way. Of many such curves it can however be said that they are almost correct for one of the four by itself, because over the frequency-range and for the substance in question the other three processes are relatively rare. That of Figure 8.13 is a case in point; over the frequency-range of its data pair-formation does not occur at all, while Compton effect and scattering are comparatively rare events for an element of such high atomic number—the major features of the curve are features of the photoelectric effect alone.

When however a σ-curve for interception is followed so far toward higher frequencies that the K absorption edge is left well behind, it can no longer be said that the photoelectric effect is dominant. The Wilson chamber gives the clearest evidence. When the gas of such a chamber is irradiated with X-rays or with gamma-rays, the tracks of the electrons ejected from its molecules bear witness by their lengths as to whether

those electrons underwent the Compton process or the photoelectric one. It is found that when the photon-energies surpass 1 Mev, the photoelectric process sinks into insignificance. The yield of the Compton process itself is sinking with increasing ν, and it seems as though the cross-section for interception were steadily diminishing toward zero as a limit. This was indeed the belief till 1933; but in that year the quantal theory was applied by Oppenheimer [1] to the trend of the newly-discovered process of pair-formation, yielding the conclusion that from its threshold at 1 Mev the process should become steadily more probable with increasing ν. The σ-values therefore passes through a minimum, located by Oppenheimer at 3.5 Mev for lead and at 6.5 Mev for carbon; and the most penetrative of electromagnetic radiations lie between the extremes of low frequency and high.

Bremsstrahlung, the X-ray continuum, and shower-production

There is an adventure possible to high-energy electrons traversing matter, which I introduce into this place in default of a better, defying the title of this long chapter which now draws to its close. According to the classical electrical theory, a fast electron when sharply slowed down should radiate energy in the form of light; moreover even when a moving electron undergoes a deflection, it should radiate some of its energy and lose some of its speed. The consequences of this theory appear to be suspended for the electrons organized into the electron-families of atoms, as we shall later see; not, however, for those which are free and which are cutting through matter with speeds approaching that of light. Radiation of such an origin is called by the German word "Bremsstrahlung" which should be translated into the English "braking-radiation" but is not, probably because "braking" sounds exactly like a word of entirely different meaning. This manifests itself in two conspicuous ways.

First, *the X-ray continuum.* The spectrum of the high-

[1] Oppenheimer and Plesset, **PR 44**, 53 (1933); Crane *et al.*, **PR 46**, 531 (1934).

frequency light or X-rays emitted from a metal target bombarded by electrons of uniform speed is of the character shown in Figure 14.4. We shall later (Chapter XIV) study the lines or peaks, which consist of photons emitted by atoms from which K-electrons (page 250) have been evicted by the bombardment. The continuum on which these lines are superposed (or from which they are absent, as in the lowest curve) is the Bremsstrahlung aforesaid. Its shape has been the object of much theoretical and experimental study, but here we shall consider only its termination at the high-frequency end, the "upper limit of the X-ray continuum." At this upper limit, each photon is presumed to have the entire energy of an incident electron which was brought to a dead stop in a single "braking" operation. Denote by V the voltage by which this electron was accelerated from rest: then the presumption is expressed by the equation,

$$h\nu = hc/\lambda = eV(10^8/c) = eV/300 \qquad (8.14)$$

This at present is deemed the soundest way of evaluating the important constant or ratio-of-constants, h/e; for this it has given, in the hands of DuMond and Bollman,[1] the value $(1.3762 \pm .0003) \cdot 10^{-17}$.

Second, *the production of cosmic-ray showers.* As recently developed by several theorists, quantal theory predicts that for electrons of even the colossal energies of those detected among the cosmic rays—energies ranging up to 10^{10} eV or even further—the liability to loss of energy by radiation should be so great that even these electrons should scarcely ever be able to penetrate more than four or five centimetres of lead or similar metals. Moreover for photons of the energy-values generated under these circumstances, the phenomenon of pair-formation should become extremely likely. An electron of vast energy racing into a stratum of such a metal should

[1] **PR 51**, 400 (1937). In this citation it is tacitly assumed that the "grating wavelength" (p. 28) of the upper limit is adopted; if the "crystal wavelength" is used the value of $N_0(= F/e$, p. 19) is additionally involved, and the method may be regarded as giving the value of $h/e^{\frac{1}{2}}$.

generate a burst of photons and of electrons, and the newborn positive electrons merge in turn with negatives to become new photons. Thus the energy passes to and fro between the electronic and the photonic incarnation, and some of the corpuscles of each of these types emerge from the far side of the metal provided that this latter is not too thick. Such is the current theory [1] of the magnificent phenomenon of cosmic-ray showers, which manifest themselves in the cloud-chamber by simultaneous tracks of charged particles numbering up to the hundreds or even the thousands, accompanied by photons which attest themselves by ejecting electrons from the gas or the walls of the chamber. The theory has also the consequence, that the charged particles of the cosmic rays which are able to penetrate lead in thicknesses much greater than five centimetres must be other than electrons. They are presumed to be of charge e, for the reason stated in the last section but one; and of mass greater than that of the electron, since the liability of a particle to violent deflection and accordingly to loss of energy by radiation falls off rapidly with increase of mass. Here is a second argument for the existence of a new sort of particle, baptized by Anderson "mesotron." [2]

From the very highest energy-ranges known to man we now shift suddenly to phenomena involving the lowest amounts of energy which it is possible for atoms to receive: the phenomena of excitation.

[1] Oppenheimer, **PR 50**, 389 (1936); Carlson and Oppenheimer, **PR 51**, 220 (1937); Bhabha and Heitler, *Nature* **138**, 401 (1936), **PRS 159**, 432 (1937).

[2] Anderson and Neddermeyer, *London Conference on Nuclear Physics*, 179 (1934); **PR 54**, 88 (1938) and earlier papers there cited. For the literature of the mesotron I refer to Anderson and Neddermeyer, **PR 54**, 88 (1938), or to a forthcoming article of mine (**BSTJ** 1939). Other names for this particle are "barytron" and "heavy electron."

CHAPTER IX

Excitation and the Stationary States

We have considered at some length, in the chapter just previous, the ionizations which occur when a beam of electrons of adjustable energy is projected into a gas, and the electron-energy is raised past a critical value I_0. We now begin to consider what happens when the electron-energy is less than I_0 (and may also happen when it is greater than I_0, but is then obscured by the more striking effects resulting from ionization). To anticipate the conclusion: impacts of two classes occur, the "elastic" ones in which the atom simply recoils from the blow, taking no more energy from the electron than is necessary to assure conservation of momentum; and the "inelastic" ones in which more than this minimum of energy is taken from the electron and is used to alter the internal structure of the atom. We shall be concerned almost entirely with the inelastic impacts, and the inferences which are drawn from them; but I permit myself a short detour to give a very inadequate treatment of the large topic of elastic impacts, which comprises some of the most beautiful agreements between experiment and theory that are to be found in physics.

Elastic impacts.[1]

An elastic impact between two bodies is one in which not only momentum, but also kinetic energy, are conserved. If the bodies are not spherical, they may rotate more (or less) rapidly after an impact, there being an interchange between kinetic energy of translatory motion and kinetic energy of rotatory motion. Avoiding this complicated possibility, we

[1] *Electrical Phenomena in Gases*, Chapters V, VI, VII.

endow the atoms (and the electrons) with spherical symmetry, or, if the student does not object to so definite an image, regard them as elastic spheres. The equations for impacts between two such bodies, an initially moving particle of mass m and an initially stationary particle of mass M, have been given in Chapter III (equations 3.2, 3.3). From these it is easy to deduce the maximum possible transfer of kinetic energy in an impact, which occurs when the impinging particle rebounds straight backward. In the present case the impinging particle is an electron, the other one an atom moving so slowly by comparison as to be practically motionless. The ratio m/M is so small (always less than 0.0003) that for this maximum energy-transfer it is sufficient to use an approximate expression, $4(m/M)K$; here K stands for the kinetic energy of the electron. In a multitude of impacts the average deflection is 90°, the average energy-transfer ΔK half as great as the maximum, and given accordingly by the formula:

$$\overline{\Delta K} = 2(m/M)K \qquad (9.1)$$

We shall presently be considering experiments in which the energies of electrons are observed before and after single impacts, but they are not exact enough to reveal such a little difference between K before and K after; they show that the energy-losses cannot be many times greater than (9.1) proposes, and this in itself is an important result. If however the experiment is so arranged that the electrons make thousands of impacts, (9.1) can be tested. Such experiments require a statistical theory, far too long for this book. It must suffice to say that when electrons are poured into a sufficiently dense gas pervaded by an electric field (preferably not so strong that ionization occurs) they constitute an "electron-gas" characterized by an almost-random zigzag motion of its corpuscles. The mean kinetic energy $\overline{K}$ of the electrons of this "gas" is not the same as that of the atom-gas, but higher—often hundreds of times as high. This is because each electron is accelerated by the field between each impact and the next. The career of an electron is an alternation of free flights and sudden

collisions; on the average there are gains of its kinetic energy K during the flights, and always there are losses at the collisions, but the losses do not come into balance with the average gains until K has mounted to a value relatively high. It is obvious that the gains must increase with the field strength E and with the mean free path l of the electrons; and it can be shown that the average gain is $(eEl)^2/\bar{K}$, multiplied by a factor of the order of $\frac{1}{4}$, the exact value of which is uncertain. Equating this to the average loss as given by (9.1), we find:

$$\bar{K} = \sqrt{M/8m}\,(eEl) \tag{9.2}$$

so that if $\bar{K}$ and l can be measured, the hypothesis of elastic impacts can be tested. Now, values of $\bar{K}$ and l have been obtained in countless experiments by several different methods —by studying the gradual drift of the electron-gas through the atom-gas under the influence of the field, by studying the diffusion of the electron-gas through an atom-gas in which the electron-concentration (number of electrons per unit volume) varies from place to place, and by the method of the Langmuir probe. In some cases they confirm (9.2) as closely as can be asked, in view of the uncertainty of the numerical factor; such cases are offered by H_2, N_2, air and He when $\bar{K}$ is 0.5 EV or less, and by argon when $\bar{K}$ is 5 EV or less. In other cases the average loss of kinetic energy at an impact seems to be somewhat greater than $(2m/M)\bar{K}$; these invite further study.

The mean free path l of the electrons in a gas containing N atoms per unit volume is inversely proportional to N; we write:

$$l = (N\sigma)^{-1} \tag{9.3}$$

and here σ stands for what, if we visualize the electrons as mere points and the atoms as elastic spheres, is the cross-section of one of the spheres. Whether or not we so visualize, σ is called the "cross-section of the atom" with respect to electron-impact. In addition to the ways of estimating it which I mentioned just above, there are methods yielding it much more directly. In principle, they consist in setting up a device in which an

electron-beam traverses a known thickness D of a gas for which N is known, and moreover N is of such an order of magnitude that an appreciable but not a very large fraction of the electrons collides with atoms. Such circumstances are similar to those in which likelihood of ionization is measured (Chapter VIII), and here the fraction of the electrons which fail to reach the collector placed to receive the beam beyond the gas is $ND\sigma$. I inserted the words "in principle" to imply, of course, that the actual experiments are more intricate, and so they are indeed; it is scarcely extravagant to say that to each design of apparatus corresponds its own definition of σ! Referring again elsewhere for details,[1] I make here only some general statements. The cross-section is usually of the order of magnitude of the "gas-kinetic cross-section" (page 263). This vague remark is the most precise which can be made without much elaboration, for σ, far from being independent of electron-speed v as people used to assume or anyway to hope, may vary greatly and surprisingly with v. Forced to give up the assumption of constant σ, one would next hazard the guess that σ must increase steadily as v is diminished; but even this, while often correct, is not a universal rule—the cross-section may pass through a maximum or a minimum or both. Such remarkable variations go under the general name of "Ramsauer effect," after the man who first observed examples of them, while working with the noble gases in which they are particularly prominent. The fact that they have been verified by experiments in gases so dense that statistical theory is required is wonderful testimony to the validity of modern physics.[2] It must be added that curves of σ versus v are usually prolonged over ranges of electron-energy far surpassing the onset-values for inelastic impacts and even for ionization, and only over part of their course—if anywhere—do they pertain exclusively to elastic impacts; so that the statements in this paragraph are germane to the subject of this entire chapter, and even to that of the last chapter, rather than to elastic impacts alone.

[1] *Electrical Phenomena in Gases*, Chapter IV.
[2] *Ibid.*, pp. 258–252.

Inelastic impacts

When a monatomic gas is bombarded by electrons of which the kinetic energy exceeds a certain critical value, differing from one gas to another, occasional electrons lose much larger quantities of energy than those computed by (9.1). They are said to have made "inelastic impacts."

The sudden onset of inelastic impacts, when the electron-energy arrives at the critical value, is one of the most striking phenomena in physics. The various direct ways for observing it are all based on the same principle: an auxiliary electrode is charged to a potential so chosen that electrons can reach it from the electron-source if and only if they lose, in impacts with atoms on the way, no more than the slight amount of energy sacrificed in elastic impacts. The scheme of apparatus sketched in Figure 8.1 and used for detecting onset of ionization can easily be modified into one suitable for detecting onset of inelastic impacts. We have but to maintain G_3 and C both at a potential say one volt below that of F, and then the only electrons which can reach C and G_3 are those which lose less than one EV of kinetic energy in collisions as they are first speeded up and then slowed down by the fields. It is obviously better to suppress G_3 and bring C close up to G_2; in the apparatus of G. Hertz (Figure 9.1) G_2 and G_1 are so to speak united into the walls of a box with the filament D just beyond one side and the collector P just beyond another. One can also make measurements with a single gauze G_1 and a collector beyond it.

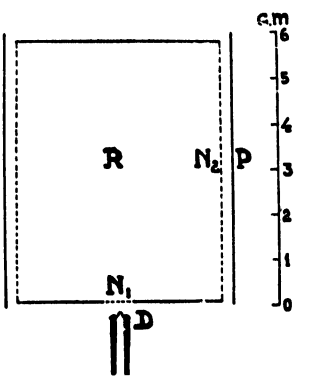


FIGURE 9.1—APPARATUS OF HERTZ FOR LOCATING RESONANCE POTENTIALS. (Hertz; *Zeitschrift fur Physik.*)

With a single-gauze apparatus and gas of the proper density,

mercury vapor, for instance, a curve such as Figure 9.2 is obtained when the electron current into the collector is plotted against V, the voltage-difference between G_1 and F. This curve signifies that as soon as the kinetic energy of an electron exceeds 4.9 EV it becomes liable to transfer that amount of energy to any mercury atom which it strikes. The first downturn of the curve occurs when V is raised to the value 4.9 where these transfers begin; the next upturn occurs when the electrons which have lost their first 4.9 EV at some point part way from the filament to G, gain enough kinetic energy in the

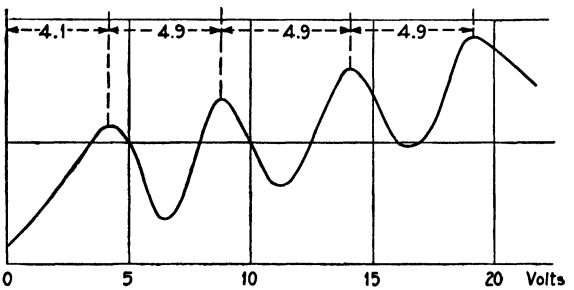


FIGURE 9.2—CURVES DISCLOSING THE RESONANCE POTENTIAL OF MERCURY, OBTAINED WITH A ONE-GAUZE APPARATUS. (After Franck and Hertz.)

rest of the way to G_1 so that they can go on to the collector; the next downturn when they suffer two such losses, and so forth. This experiment shows that mercury atoms are able to receive a definite and characteristic amount of energy, 4.9 EV, from an electron having so much energy.[1]

A curve like that of Figure 9.2 is obtained when the gas is fairly dense, so that an electron is likely to make a collision soon after attaining the critical energy-value. If there should be another such critical energy-value appreciably higher—if for instance 6.7 volt-electrons are able to transfer just 6.7 EV to mercury atoms—it would not be detected in such circum-

[1] The distance between successive maxima is a better measure of the critical amount of energy than is the distance from the first maximum to the axis of ordinates, as this latter is liable to be displaced by contact P. D. and other agencies.

stances, for electrons would seldom acquire so much kinetic energy. By working with rarefied gas and especially by adding the gauze G_2, additional critical potentials in fact are found. Hertz further refined the method by measuring the current to his collector P (Figure 9.1) alternately when the collector was at the same potential as the gauze and when it was 0.1 volt more negative, and plotting the difference between the two measurements against V. This difference is largely due to electrons which in inelastic impacts have sacrificed all but

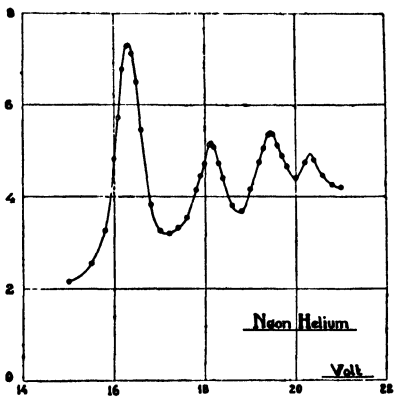


FIGURE 9.3—CURVE REVEALING THE RESONANCE POTENTIALS OF NEON AND HELIUM, OBTAINED WITH THE APPARATUS OF FIGURE 9.1. (Hertz.)

0.1 EV or less of their energy, and consequently the curves have maxima at the critical energy-values. The curve reproduced in Figure 9.3 was obtained with a mixture of neon and helium; the two left-hand maxima belong to the former gas, the two others to the latter.[1]

The maxima in such curves as that of Figure 9.3 show that an electron possessing more than the critical amount of energy which the atom is willing to receive transfers just that amount and retains the remainder. This is confirmed by experiments

[1] The maxima in such curves are rounded over for the usual reasons: the electrons do not all have the same speed when they leave the source, nor do they all pass through the gauze at right angles to its plane.

performed by the conventional method to determine the distribution-in-energy possessed by electrons which have started into the gas with speeds well above the critical value and come through a suitably thick layer of it; thus, among the electrons which issue from mercury-vapor into which they were projected as 16.5-volt electrons, there are some with kinetic energy amounting to (16.5–4.9) and others with energy of (16.5–6.7) EV.[1]

The critical energy-values so determined are known as *resonance-potentials* and are expressed in electron-volts.

The stationary states

The data cited above show, to take helium as an example, that an electron possessing less than 19.75 EV of kinetic energy cannot communicate any energy at all to a helium atom, save what it transfers as kinetic energy of translatory motion in order to guard conservation of momentum; that an electron possessing between 19.75 and 20.45 electron-volts can put 19.75 of these energy-units into the domestic economy of the atom, and that an electron possessing 20.45 or somewhat more can elevate the internal energy of the atom by precisely 20.45 EV. The helium atom can receive energy inferior to say 21 EV in these and only these amounts.

Another way of saying the same thing is this: the helium atom can exist in its normal state, and in either of two other states in which its internal energy is 19.75 or 20.45 EV greater than in the normal state, and in no others with energy-values intermediate among these three.

These are some of the *Stationary States* of the helium atom. Every measurement of a resonance-potential yields the energy-value of a Stationary State, referred to the energy-value of the Normal State as zero of energy. The Normal State is itself ranked among the Stationary States; the others are distinguished from it by the title *Excited States*. The process of

[1] J. A. Eldridge, PR (2) 20, pp. 456–475 (1922); and a difficultly accessible paper by Akesson (*Lunds Arsskrift*, 1916).

transferring an atom from its Normal State to any of the Excited States is known as *excitation*; the atom is said to become an *excited atom.*

As samples I give in Table VI the energy-values of some excited states detected by the method just described. There are other far more accurate and exhaustive ways of locating stationary states; these particular values are important because (together with some others[1]) they are known by direct experiment. The energy-values are referred to that of the Normal State as zero, as I shall continue to do until the end of this chapter; the value of I_0 from Chapter VIII is included under the name "state of the ionized atom," for a reason presently to be made known.

TABLE 9.1

	He	Ne	Na	Cs	Mg	Hg
Energy-value of the						
Normal state.....	0	0	0	0	0	0
First excited state.....	19.75	16.65	2.1	1.45	2.7	4.66
Other excited states.....	20.55	18.45	.....	.....	4 4	4 86
						5.43
						6.7
Ionized atom.....	24 5	21 5	5.12	3.9	7.6	10.4

Return from an excited state to the normal state

The Excited States are transient; for a gas subjected to excitation by electron-bombardment or any other agency reverts to its original condition after the exciting agency ceases. Now while the exciting agency is acting, the gas is usually emitting radiation or producing other effects. This radiation and these other effects are attributed to the return of excited atoms into the Normal State—to "transitions" of atoms from Excited States to the Normal State; or else to transitions between one Excited State and another of lesser energy. The

[1] There is a tabulation in the brochure by K. T. Compton and F. L. Mohler, *Critical Potentials* (Bull. Nat. Res. Council No. 48).

excited atom appears to dispose sooner or later of its energy of excitation, either through radiation or by transferring it to another atom or molecule.

We proceed to consider the evidence for these statements.

Electrons which are just able to effect inelastic impacts with atoms are also, it is frequently found, just able to stimulate radiation; onset of inelastic impacts often coincides with onset of radiation. The fact that when the energy of electrons bombarding a gas is gradually raised, radiation commences at a critical energy-value, has already engaged our attention in Chapter VIII. There it appeared as a source of deception in localizing ionizing-potentials by certain methods; these same methods may, however, be consciously used to determine these *radiation-potentials*. In the curve of Figure 8.2 on page 236, relating to mercury, it is obvious that radiation commences at 4.9 volts and that there is an upward bend in the current-curve at 6.7 volts. These values will be recognized as identical with the cited resonance-potentials of mercury vapor. *Radiation-potentials are resonance-potentials.*

The art of detecting radiation-potentials consists essentially in bombarding the substance under investigation with electrons of controllable kinetic energy V; measuring the "photoelectric current" i of electrons streaming out of an auxiliary electrode exposed to whatever radiations may come from the bombarded material, but shielded against primary electrons or ions; plotting i (or better the ratio of i to the bombarding current of electrons) versus V; and examining the curve to see whether it displays sudden changes of slope. If it does, one infers that at the corresponding values of V new radiations suddenly burst forth. A curve showing a number of instances of these abrupt changes in slope, or *breaks* as they are frequently called, is shown in Figure 9.4. Very many such curves appear in the literature of physics, showing more or less conspicuous breaks; some are as striking as the best instances in the figure, some require a good deal of care and experience to locate them properly, and some, one is driven to conclude, are visible only to the eye of faith. But it is hardly possible to doubt that such

a corner as those at 4.68 and 6.75 in the figure marks the *entrée en scène* of a new kind of radiation.

Now comes a principle of the first importance, foreseeable already, referring to the emitted radiation. In some cases *it is*

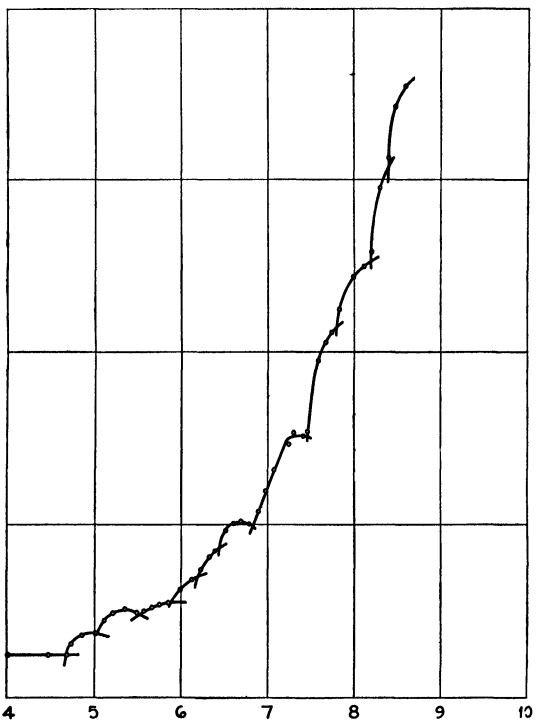


FIGURE 9.4—CURVE SHOWING "BREAKS" TAKEN AS SIGNS OF RADIATION POTENTIALS OF MERCURY VAPOR. (Franck and Einsporn; *Zeitschrift für Physik.*)

monochromatic radiation of frequency ν, given in terms of the energy-difference W between the excited state and the normal state by the equation

$$W = h\nu \tag{9.3}$$

Such is the case when mercury atoms absorb 4.9 equivalent volts of energy; the radiation emitted belongs to the well-known

spectrum line λ2536. Presumably this rule is always obeyed when the atom returns directly from the excited state to the Normal State; the energy released emerges in the form of a single photon. In other cases, however, the emitted radiation contains two or more frequencies inferior to W/h; this signifies that the atoms return from the excited state to the Normal State in two or more steps. Instances of this, like instances of atoms returning to the Normal State otherwise than by radiating, are most easily obtained when the atoms are excited by radiation.

Excitation by radiation

A mercury atom in its Normal State can be elevated to the aforesaid excited state by receiving 4.9 EV of energy, not necessarily from an electron having so much kinetic energy or more, but also from radiation of which the photon-energy $h\nu$ is equal to that amount; that is, from radiation of the wave-length 2536.

The direct evidence for this statement is furnished by *resonance-radiation*. Mercury vapor illuminated by radiation λ2536 responds in the same way as when it is bombarded by 4.9-volt-electrons; that is to say, it emits radiation λ2536 in all directions. Sodium vapor illuminated by the well-known D lines (until Chapter X we will treat them as one) re-radiates them profusely; now the photon-energy of these lines is 2.10 EV, and electrons having so much or slightly more kinetic energy are observed to make inelastic impacts with sodium atoms and transfer 2.1 EV to them. Sodium vapor illuminated by radiation λ3303 emits light of this very wave-length and likewise the D lines. This suggests that there is an excited state of the sodium atom of which the energy is greater than that of the normal state by 3.73 equivalent volts, the photon-energy of λ3303; that the incident radiation of this wave-length is continually transferring atoms into that state, and that some of them return directly to the normal state while others return in two or several steps via the 2.10 EV state. This is in fact the case.

Radiation of frequency ν thus is sometimes able to transfer an atom from its normal state to an excited state of energy-value $h\nu$, provided there is such a state. Radiation of more than sufficient frequency, unlike electrons of more than sufficient kinetic energy, cannot perform the transfer; it must give all or nothing (except when the "Raman effect," later to be described, occurs). This is shown by the fact that mercury vapor traversed by heterogeneous radiation absorbs the wavelength $\lambda 2536$ and nothing else nearby; sodium vapor absorbs the D lines and no other light in that region of the spectrum. It is obvious that one might take the absorption-spectrum of a gas, multiply the frequency of each absorption-line by h, and designate each product $h\nu$ as the energy-value of a Stationary State of the atom, thus disclosed. Now we have arrived at Bohr's interpretation of spectra; and it is expedient not to pass beyond this single allusion as yet.

In Chapter X we shall return, with additional information at our command, to this topic of excitation.

Passing allusion to the Raman effect

Imagine an atom or a molecule existing in its normal state, and capable also of existing in an excited state of energy W (referred to the aforesaid zero); and suppose that it suffers an impact from an electron of kinetic energy E or a photon of energy $h\nu$. According to what has thus far been said, excitation by the photon will not occur unless $h\nu = W$, but excitation by the electron may occur whenever $E > W$, the difference $(E - W)$ being carried off as kinetic energy by the recoiling electron. The contrast is not however quite so sharp, since for certain stationary states of *molecules* (not atoms) excitation may be effected by photons of energy superior to the W-values. When such a process occurs, the difference $(h\nu - W)$ is carried off by the recoiling photon, and the substance is said to "scatter" light of the wavelength appropriate to the new frequency $\nu' = (\nu - W/h)$. Whether the recoiling or scattered photons are to be regarded as new particles, or as the old photons rebounding with lessened energy

from the impacts, is a matter of taste. The latter choice affirms the analogy with inelastic impacts of electrons.

This process is known, after the discoverer of light thus scattered, as the "Raman effect;" in Germany as the Smekal-Raman effect, Smekal having been led by theory to envisage such a process. It is of great importance in the study of molecules, and is not by any means confined to the gaseous phase of matter; its importance in chemistry is considerable; in the size of its literature, it is one of the hugest of the subjects which limitations of space exclude from this book.

Direct transfers of energy from excited atoms to others[1]

An atom excited by radiation or by electron-impact may produce an effect, upon another atom or molecule in the gas, which the radiation and perhaps even the electron-impact could not themselves have produced; an effect, therefore, which is ascribed to a direct action of atom upon atom. The generic term "collisions of the second kind," a bad piece of nomenclature if ever there was one, has been applied to these processes.

Consider for instance the dissociation of hydrogen molecules into neutral hydrogen atoms, H_2 into $H + H$. I described in Chapter VIII (page 245) other processes of dissociation of hydrogen, from which the energy required for this one can be calculated: it is 4.4 EV. Yet when electrons of this or somewhat greater energy are projected into pure hydrogen, the process does not occur! nor is it produced in pure hydrogen by the photons of the line 2536, though their energy is ample. If, however, the hydrogen is mixed with a little mercury vapor and illuminated as before, the characteristic chemical effects of monatomic hydrogen make their appearance —tungsten oxide or copper oxide placed in the tube are reduced, the gas-pressure in the tube goes down because the water-vapor so produced is condensed on the walls, or (in the absence of oxides) because the monatomic hydrogen itself thus is fastened

[1] Discussion and bibliography of this subject by K. T. Compton, **JOSA 7**, pp. 955–971 (1923). Other references: G. Cario and J. Franck, **ZP 17**, pp. 202–212 (1923); O. S. Duffendack and K. T. Compton, **PR (2) 23**, pp. 583–592 (1924).

down. A low-voltage arc operated in hydrogen has the same effect, if and only if mercury is mixed into the gas. The mercury atoms receive the energy of the radiation and transfer it to the hydrogen molecules; the radiation falls upon the molecules directly but they are unable to receive the energy at first hand.

Again, radiation of wave-length 2536 is not absorbed by cadmium or silver or thallium vapors, nor does it excite light from these; but let mercury vapor be mixed with any of them and the mixture illuminated by radiation λ2536, and some of their characteristic spectrum lines appear. Those which come out are mostly due to transitions between Stationary States having smaller energy-values than the 4.9 EV state of the mercury atom; so that there is nothing contradicting the notion that the mercury atom, having acquired 4.9 EV and so entered into an excited state, transfers to another atom with which it collides just enough energy to raise it to a Stationary State of its own. As for the residue (when there is one) its fate is uncertain; probably it is converted into kinetic energy of the two atoms.[1]

Duration of excited states

I have already used, and shall presently again make use of, the assumption that an atom may remain in an excited state for an appreciable length of time: meaning thus far by "appreciable" a time so long that something else, such as a collision with another atom or an impact from an electron, can happen to the atom before it is over. This duration is appreciable also in the sense of being detectable and measurable.

The first detection and the first measurement were made by W. Wien in 1919, with apparatus such as we considered in

[1] If so the frequency of the radiation from the second atom, when it returns to its normal state, might be displaced by Doppler effect; evidence to sustain this was found by Cario and Franck (*l. c.* supra). They also found lines of thallium indicating that some thallium atoms had gone into a 6.3 EV state, which seems extraordinary, and is only to be accounted for by supposing that kinetic energy of translatory motion of the atoms had been converted.

Chapter III: a discharge-tube with a perforated cathode, through which a beam of canal rays enters a chamber behind. In this chamber the path of the beam is marked by a luminous trail. In usual conditions the light is mostly due to collisions occurring in the chamber itself, which frustrate the experiment. If however the gas-pressure in the chamber is made as low as possible, the trail is still visible though faint, and the brightness of any single spectrum-line thereof falls off exponentially with increase of distance from the channel through the cathode. Now the light is largely due to ions and atoms which were put into excited states while they were traversing the channel or even before they left the region of the discharge. The exponential fading signifies an exponential distribution of the durations or "lives" of the excited particles, and by dividing the particle-speed into the distance over which the intensity of the light fades by one-half, one computes the time during which one-half of the particles initially excited quit their excited state. The light must be analyzed and the spectrum-lines examined individually, since each of these corresponds to a transition out of one excited state, which is not true of the light as a whole. Contrariwise to what one would probably expect, this is not solely a method for observing the lines and measuring the states of ions. Neutral atoms are included among the canal-rays, for some of the ions in the discharge-tube neutralize themselves by capturing electrons before they reach or at least before they quit the channel; and Wien's first observations were made on neutral atoms of hydrogen.

The lives of excited states may also be estimated by observing the light emitted a brief time after the exciting agent ceases to act; this has been done by a clever arrangement of two Kerr cells subjected to a sinusoidal electric field, the light which excited the atoms entering through one of the cells and the light which they emit emerging through the other. Several other methods of great ingenuity have been devised and applied; for the description I must refer elsewhere.[1]

[1] *Electrical Phenomena in Gases*, pp. 128–132.

The values obtained by Wien for several excited states of hydrogen agree with one another, their common amount being $1.85 \cdot 10^{-8}$. Excited states of other atoms and ions have yielded values ranging from more than ten times as high to more than a hundred times lower. Such differences, large as they are, are small compared to the difference between any one of these values—all referring, as they do, to states which the atom may quit by radiating its energy—and those of the special category of states which are called "metastable." An atom in a metastable state is not likely to quit it at all, until the atom experiences something such as a collision of the second kind with another particle, or perhaps a collision with the wall of the tube. The duration of the state therefore depends not on an intrinsic quality of the atom, but on the length of time during which it manages to escape such denouements. This category of metastable states, and the manner in which the corresponding durations are measured, can be better described with the aid of the spectrum data which will be available in Chapter X.

Ionization of excited atoms

It was stated in Chapter VIII that when a rarefied gas is bombarded by a not-too-dense electron-stream of controllable electron-energy V, then as V is increased from zero ionization commences at the ionizing-potential I_0 corresponding to detachment of the loosest electron by single-stage process. It is however only too easy to transgress the necessary conditions, and find ionization commencing at the resonance-potential. We must then either give up the whole of the correlation with spectra and with atomic theory which is yet to be revealed, or suppose that in all such cases the ionization is by cumulative process: an atom is put by an impinging electron into its first excited state, and before it leaves that state it receives an extra supply of energy which is enough to convert it into an ion and a freed electron.

Onset of ionization by light[1] is peculiarly affected and confused by such events, on which account I expressly deferred it to this chapter. Mercury vapor affords a famous case, being undoubtedly ionized by photons of the line 2536 which are just those required to put the mercury atom into its first excited state (I leave out a metastable state) and do not have half the energy required to ionize in a single stage. Photons of energy only slightly greater are not able to ionize, as G. F. Rouse and G. W. Giddings showed in a notable experiment; this also speaks against a single-stage process, for we have seen that when ionization by electrons is being effected in a single stage the likelihood mounts steadily and smoothly upward as the energy of the particles is increased from the onset-value. The number of ions formed by incidence of 2536 increases as the square, or a yet higher power, of the intensity of the light—another point in favor of a cumulative process.

Knowledge of such instances is limited by the fact that for most monatomic gases (not to speak of molecular gases) the relevant part of the wave-length-range lies in a region where experimentation is difficult, because the light is so seriously absorbed in solid substances such as the walls of tubes must be made of. This is probably no mere coincidence; light of these wave-lengths is most easily absorbed, because it is most nearly attuned to the task of displacing the electrons loosely bound to the atoms both in the solid state and in the gaseous. There are other grave difficulties of technique, and our knowledge of ionization by light near the onset relates chiefly to one small but interesting group of gases, the alkali-metal vapors. For caesium in particular it has been shown, by F. L. Mohler and his colleagues of the Bureau of Standards, that ionization is produced by photons unable to ionize by single-stage process, but able to transfer the normal atom into one or another of several different excited states.[2] Subsequently it was shown that the amount of such ionization is directly proportional to the intensity of the light—a tough result, which has been inter-

[1] *Electrical Phenomena in Gases*, pp. 16–20.

[2] The (*nP*) states defined in Chapter X.

preted by supposing that an excited and a normal atom of Cs combine into a molecule from which an electron splits away spontaneously.

The stationary states and the state of the ionized atom

The fact that excitation is a partway step towards ionization suggests that the "state of the ionized atom" may well be listed as a Stationary State along with the Normal State and the Excited States of the neutral atom. This procedure will be tremendously fortified in the next chapter, in which we shall learn ways of detecting and locating Stationary States which are immensely more powerful than the direct way to which this chapter was devoted, and which reveal a much greater abundance of Stationary States than the direct way could ever have disclosed. From the arrangement and the inter-relations of the multitude of Stationary States so discovered, it is obvious that the State of the Ionized Atom belongs among them, and in fact is a more natural and suitable State from which to measure their energy-values than is the Normal State itself. As an introduction to the forthcoming data I therefore restate Table 9.1 as Table 9.2 shown on the opposite page, with the energy-values of the several states there listed referred to the energy-value of the ionized atom as zero of energy.

TABLE 9.2

	He	Ne	Na	Cs	Mg	Hg
Energy-value of the Ionized atom.....	0	0	0	0	0	0
Non-ionized atom.....	.....	.....	.....	.....	.....	.....
Excited states.....	− 3.95	− 3.0	.....	.....	−3.2	− 3.7
						− 4.97
						− 5.54
First excited state....	− 4.75	− 4.85	−3.0	−2.45	−4.9	− 5.74
Normal state.....	−24.5	−21.5	−5.1	−3.9	−7.6	−10.4

The fact that the Excited States are, so to speak, steps leading up to the State of the Ionized Atom, is extremely influen-

tial in shaping theories of atomic structure. Since ionization is complete detachment of an electron from an atom, excitation is naturally conceived as a partial detachment, as an outward displacement—in the strict geometrical sense—of one electron (or more) from its normal place. This is the best reason for adopting the idea, henceforward to figure so prominently in this book as it does in the whole system of contemporary physics, that each Stationary State of the atom is to be represented as a particular and specific arrangement of electrons in the atom-model.

CHAPTER X

Analysis of Spectra

From observations on electron-impacts we have learned that each atom possesses various Stationary States, its Normal State and divers "excited" States, each of them characterized by a distinctive energy-value. Also we have seen that sometimes, when atoms revert from an excited State to their Normal State, they emit radiation of a frequency ν which is given, in terms of the energy-difference ΔW between the excited and the Normal State, by the equation

$$\nu = \Delta W/h \qquad (10.1)$$

This radiation also constitutes one of the lines in the spectrum of the element.

The spectrum of any element contains very many more lines than have ever been associated, by direct experiments such as those described in Chapter IX, with particular transitions between particular pairs of Stationary States.

May we nevertheless make the assumption that every line in the spectrum of an element is associated with a transition between two Stationary States, according to equation (10.1)? and, having made that assumption, proceed to calculate energy-values for these postulated Stationary States? and, having thus mapped out a rich system of Stationary States for each element, find additional evidence which makes this whole procedure seem plausible?

That assumption is now universally made; it is the basis of the method of interpreting spectra which prevails everywhere today. In this chapter we shall consider the way in which systems of Stationary States are determined for the

various elements by analysis of spectra. In the course of the exposition, some of the reasons for accepting the fundamental assumption will gradually come to light.

Every spectrum-line yields the energy-difference between two Stationary States, or in other words the energy of one state referred to that of another state as zero. But how shall we learn to refer all of these to a common zero, such as the Normal State or the State of the Ionized Atom? To do this we must rely on phenomena of electron-impact, or on absorption spectra, or on line-series.

Line-series

Consider the spectrum of atomic hydrogen. In the visible region, this spectrum consists of a "line-series"—that is to say, a procession of lines converging upon a limit, falling at intervals ever narrower and narrower, these intervals so

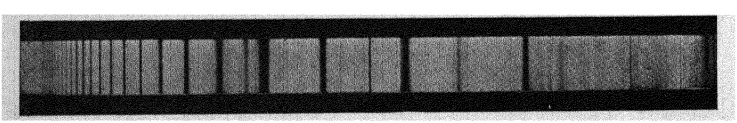


FIGURE 10.1—BALMER SERIES OF LINES IN THE SPECTRUM OF ATOMIC HYDROGEN. (Curtiss; from Foote and Mohler; *Origin of Spectra.*)

smoothly diminishing that they bear witness to a common character and a mutual relation of all the lines. This line-series, the "Balmer series," is shown in Figure 10.1. Not only to the eye is it of a wonderful regularity (so great that certain interpolating lines, due to other elements, are readily recognized); the frequencies of its consecutive lines are bound together in a simple numerical law ("Balmer's law"). They are equal consecutively to

$$\nu_{\text{lim}} - R/3^2,\ \nu_{\text{lim}} - R/4^2,\ \nu_{\text{lim}} - R/5^2,\ \nu_{\text{lim}} - R/6^2, \ldots \quad (10.2)$$

in which

$$\nu_{\text{lim}} = \text{frequency of the limit of the series} = R/4 \quad (10.3)$$

R standing for a certain constant. There is another series of

lines—the "Lyman series"—in the ultraviolet part of the same spectrum, whereof the frequencies are equal consecutively to

$$\nu = \nu_{\text{lim}} - R/4, \quad \nu_{\text{lim}} - R/9, \quad \nu_{\text{lim}} - R/16, \text{ etc.} \quad (10.4)$$

in which

$$\nu_{\text{lim}} = R, \quad (10.5)$$

R having the same value as before. The utter simplicity of the terms to be subtracted from ν_{lim} in each of these cases, not to speak of the related form of the expressions for ν_{lim}, suggests (like simple laws in other fields of physics) that in this formulation of the facts something highly important has been partially unveiled. There are certain other series in the spectrum of hydrogen, and inspecting them all one is led to the rule that *every frequency emitted by the hydrogen atom can be calculated by inserting different pairs of integers in the places of m and n in the formula*

$$\nu = R\left(\frac{1}{m^2} - \frac{1}{n^2}\right). \quad (10.6)$$

The case of the ionized-helium atom is quite as simple. Every frequency emitted by this atom can be calculated by assigning different pairs of integer values to the constants m and n in the formula

$$\nu = 4R\left(\frac{1}{m^2} - \frac{1}{n^2}\right). \quad (10.7)$$

Line-series have been found in the spectra of many other elements. Some of them are as strikingly outstanding as the line-series in the spectrum of hydrogen, and converge upon limits scarcely less easy to locate; for instance, the "principal" series of the spectrum of sodium (Figure 10.2). Most are by no means so obvious; often they are involved in the midst of a luxuriant jungle of unrelated or otherwise-related lines. Most spectra conceal their structures from the unpracticed eye, as a tone-poem of Strauss its themes or an opera of the Ring its *Leitmotiven* from the inexperienced ear. Long training and a skilled judgment are required in the deciphering of spectra,

except in the few untypically simple cases; and usually the arrangement of lines into series which the spectroscopist presents must be accepted by the theorist without question and without suggestion, for he is not competent to analyze the data for himself.

Having grouped a certain number of lines into a series, having guessed as well as possible the convergence-frequency ν_{lim} of this series, the spectroscopist has still the task of finding the numerical law to which the consecutive frequencies con-

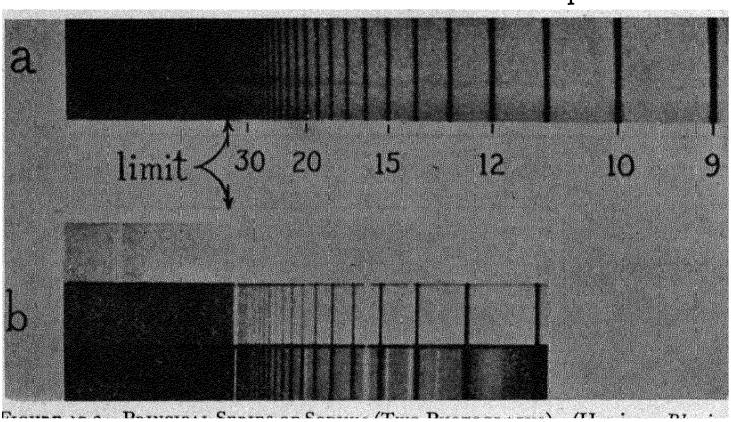


FIGURE 10.2—PRINCIPAL SERIES OF SODIUM (TWO PHOTOGRAPHS). (Harrison; *Physical Review.*)

form. As a matter of course, all the frequencies can be expressed by a formula generalized from (10.2) and (10.4):

$$\nu_i = \nu_{\text{lim}} - f(i) \qquad (10.8$$

in which i is the order-number distinguishing each line, and $f(i)$ is a different quantity for each of the lines, which approaches zero as we pass along the series to the limit. This means nothing by itself; the question is, does the function $f(i)$ have a simplicity comparable with the simplicity of the subtrahenda in (10.2) and (10.4) which suggested that they are the symbols of something profoundly important? In general, the function $f(i)$ is not so simple as the function which

occurs in the series of the spectra of hydrogen and ionized helium. In many cases, however, it is almost as simple, in others a little more complicated, in others a little more complicated yet, and so forth; so that the eventual result is this, that the formula (10.8) appears to be the proper way of describing the lines of series spectra, even in cases where the series is so irregular and the form of the function $f(i)$ so intricate that if it were the only series in existence, no one would attach any particular importance to it.[1]

To the physicists of a generation ago, who regarded the spectrum frequencies as natural vibration-frequencies of the atom, and tried to invent a mechanical model of which the vibration-frequencies should conform to the formula (10.6) or the more general formula (10.8), the character of these formulae was an insurmountable obstacle. Niels Bohr abandoned this procedure altogether; and taking equation (10.6), he multiplied both sides of it by the constant h.

$$h\nu = \underset{\sim}{h}R\left(\frac{I}{m^2} - \frac{I}{n^2}\right). \qquad (10.9)$$

The significance of this act depends on the meaning of h. Planck had found it expedient, in developing an adequate theory of radiation, to assume that solid hot bodies are populated with multitudes of oscillating electrons of all the various frequencies, possessing a very curious and inexplicable property; this being, that an oscillator vibrating with frequency ν can emit radiant energy of that same frequency ν only in units or *quanta* of amount $h\nu$. Einstein had found it expedient, in describing the photoelectric effect and other phenomena, to assume that radiant energy of the frequency ν goes about in units or quanta of the amount $h\nu$, emitted integrally, absorbed integrally, traveling integrally. Suppose then that we assume that the quantity $h\nu$, standing on the left-hand side of the

[1] As a matter of fact, the series-limit is not generally so obvious to the eye that it can be located at once; it is determined after and by means of a careful choice of the most suitable form for the function $f(i)$. This is one of the difficulties of the spectroscopist's task.

equation (10.9), represents the amount of radiant energy emitted by the hydrogen atom in the process of pouring out radiation of the frequency ν. The right-hand side is the difference between two terms. One term is the energy of the hydrogen atom before it emits the radiation of frequency ν; the other is the energy of the atom after the emission is concluded. *The radiation of frequency ν is emitted by reason of a transition between two stationary states of the hydrogen atom; the energies of these states are equal to the terms whereof the frequency ν is the difference, each term multiplied by h.* The terms of the spectrum formulae are the energy-values of the stationary states of the atom, when translated into the same units by multiplying them by h. When translated into proper units, *the terms are energies, and the energies are terms.* This is Bohr's great and memorable idea.

Since in the foregoing chapters the notion of the Stationary States was introduced by describing the direct evidence of experiment, and without mentioning the name of Bohr, it is now essential to do justice to his rank in contemporary physics by saying that they were discovered through his prediction. Not a single item of the evidence cited in Chapter IX was available in 1912, when Bohr looked at the line-series in the hydrogen spectrum, transformed the equation (10.6) expressing them into the equation (10.9), and inferred that the terms figuring in the equation (10.9) are the energy-values of Stationary States. He postulated these States and designed an atom-model to imitate them. The evidence for their existence derived from impacts of electrons against atoms was discovered by experiments intentionally performed to test his postulate. His prediction was one of the most notable in the history of science.

The values of the terms

Although I have said that the terms of the hydrogen spectrum as given in (10.6) become equal, when multiplied by h, to the energy-values of the Stationary States of the hydrogen atom, this statement is not sufficient to fix the energy-values

in question. It is necessary to decide which term, in such an expression as (10.9), corresponds to the initial and which to the final State of the transition associated with the line; and it is necessary to locate the zero from which the energy-values are measured.

While the form of equation (10.9) may seem to fix the choice of the terms absolutely, by requiring that Rh/m^2 be taken as the energy-value of the initial and Rh/n^2 as that of the final Stationary State involved in the transition, it is really setting a pitfall for the unwary. The transition is from an initial State of energy-value $(-Rh/n^2)$ to a final State of energy-value $(-Rh/m^2)$; and if spectroscopists had adjusted their practice to the contemporary theory, they would now be writing all such equations as (10.9) in the form:

$$h\nu = W_i - W_f = (-Rh/n^2) - (-Rh/m^2) \quad (10.10)$$

All of the energy-values indicated by the spectrum-terms for the Stationary States are negative. This signifies that they are measured, not from the energy of the Normal State as zero, but from some higher-lying zero relatively to which the energy of the Normal State itself is negative.

But where is this zero? It corresponds to what I have been calling "the State of the Ionized Atom"—the system composed of the valence-electron (with hydrogen, the *only* electron) and the remainder of the atom (with hydrogen, the nucleus) completely separated from each other, infinitely far apart and at rest. *The spectrum-terms with sign reversed, multiplied by h, are the energy-values of the Stationary States of the atom referred to the State of the Ionized Atom as zero* (Figure 10.3).

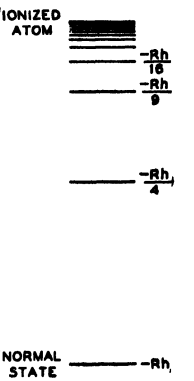


FIGURE 10.3—DIAGRAM OF THE STATIONARY STATES OF THE HYDROGEN ATOM AS INFERRED FROM ITS SPECTRUM.

This statement is neither a mere definition nor a self-evident fact. It is inherent in Bohr's model of the hydrogen atom, at which we shall arrive in the following chapter. Insofar as it is

applied to other atoms, it is derived by generalizing the inference from the hydrogen model. Its credit therefore is largely bound up with the successes of Bohr's model. Of these I now indicate two which bear directly upon it.

First, as to the relation between spectra and ionization-potentials. Every term, multiplied by h and reversed in sign, is the ionizing-energy of the atom in a particular state: this is just a restatement of the italicized sentence above. What is generally measured, in experiments by electron-impact, is the ionizing-energy of the atom in its Normal State; how shall we recognize this from the spectrum? Presumably the Normal State corresponds to the lowest energy and the greatest term. For hydrogen, from the facts and the interpretations of the last few pages, we have identified the greatest term as the one of which the value is R. Now, R is also the value of a series-limit, the highest in the hydrogen spectrum. The ionizing-energy of the atom in its Normal State should therefore be h times the highest series-limit in its spectrum. This result, derived from Bohr's model for hydrogen, is extended as a general principle to all other atoms.

Hydrogen does not lend itself willingly to the test, owing to the fact (so often hampering to physicists when they hope to test their theories!) that when it exists as a tranquil gas its atoms are paired into molecules. The test can however be made in conclusive fashion on several elements which in the gaseous phase are monatomic: the noble gases, and the vapors of such metals as sodium, potassium and mercury. The agreement between highest series-limit multiplied by h, and ionizing-potential determined with impinging electrons, is regularly good.

Second, as to the relation between spectra and resonance-potentials: if it is possible to identify from a spectrum the greatest term and the next-to-greatest term, corresponding to the Normal State and the nearest-to-normal excited State, then the difference between them (multiplied by h) should be the resonance-potential. The test upon elements monatomic in the gaseous phase is regularly favorable in its result.

From (10.4) and (10.5) it will be seen that the frequency of the first (lowest-frequency) line in the ultraviolet series of hydrogen is given by the same expression as that proposed for the resonance-potential. With many other elements, the first line in one of the most conspicuous line-series is associated with the transition into the Normal State from the nearest Excited State, and therefore the product of its frequency by h should be, and in fact is found to be, equal to the resonance-potential. This, however, is not so general a rule as the one connecting the ionizing-potential with the limit of a series.

A digression about units

Although in dealing with spectrum-lines I shall continually be mentioning their frequencies, I recall from Chapter VI that the frequency of a line is never measured or measurable. What is actually measured is the *wave-length* λ of the radiation constituting the line. This is expressed in various units: the X unit (symbol X; equal to 10^{-11} cm.), the Angstrom (A = 10^{-8} cm.), the millimicron ($m\mu = 10^{-7}$ cm.) and the micron ($\mu = 10^{-4}$ cm.). The reciprocal of the wave-length is called the *wave-number*, and is usually expressed in reciprocal centimeters. The wave-length, and consequently the wave-number, are often determined with extraordinary accuracy, even to eight significant figures. The *frequency* is the product of the wave-number by c, the phase-speed of light in vacuo. This latter factor not being exactly known beyond the fourth significant figure, values given for frequencies are to that extent less exact than values given for wave-numbers. The energy-value of a Stationary State deduced from a spectrum-term is composed of the three factors wave-number, speed of light, and constant h, the last of which is not exactly known beyond the third significant figure, if so far. Consequently it is customary to give the term-values in wave-numbers, rather than the energy-values in ergs, when referring to Stationary States deduced from spectra.[1] The wave-number there-

[1] The term-value of the Normal State is sometimes expressed in wave-numbers under the name of ionizing-potential.

fore serves in effect as an energy-unit hc (about $19.68 \cdot 10^{-17}$) times as great as an erg and 300 hc/e (about .0001237) times as great as an electron-volt.

The constant designated by R in equation (10.2) and thenceforth is the product of c into a constant more commonly designated by that symbol, and equal[1] to $109677.69 \pm .06$. Considering the uncertainty in the value of c, it is hardly profitable to give the value of the constant here called R to more than three significant figures:

$$R = 3.29 \cdot 10^{15}. \qquad (10.11)$$

The classification of stationary states by utilizing "rules of selection"

A spectrum-line corresponding to a transition between two stationary states is symbolized, on a diagram of stationary states, by an arrow connecting the dashes (or whatever marks are used) which symbolize the two levels. This is illustrated in Figure (10.4).

I pause at this point to remark that each of what I have been calling the "stationary states" is in fact usually a group of two or more stationary states, often but not always exceedingly close together; just as many seeming stars in the sky are in fact groups of stars too close together for any but an excellent telescope to discriminate. This will be discussed at length in a later section; at present it is expedient to regard each of these groups as one stationary state.

I have said that every line in a spectrum, at least of those arranged in series, may be represented by an arrow connecting two stationary states. If arrows are drawn from every one of the stationary states to every other, will every arrow correspond to a line actually observed in the spectrum? Every line has an arrow; does every arrow have a line? By no means; the answer is definitely and strongly negative. If the wavelengths deduced from all the possible arrows are sought in the

[1] F. Paschen, AP 50, pp. 901 ff. (1916).

spectrum, most of them are found unoccupied by lines. The great majority of the conceivable transitions either do not occur at all, or if they do occur, the energy which is liberated is disposed of in some other way than by radiation. There is reason for believing that the atom may embrace this last alternative if it collides with another atom or with an electron. Otherwise, it seems that if the atom cannot radiate the energy liberated in a transition, the transition itself cannot happen at all. If therefore, the line corresponding to an arrow is missing, the transition corresponding to the arrow must be inhibited by some agency as yet unknown. Many transitions must be inhibited, for many lines are missing.

These missing lines are precious to the student of spectra and to the architect of atom-models. Whatever explanation is devised for the stationary states must include a reason for the occurrence of some transitions and the non-occurrence of some others. This is good rather than bad fortune, since if such a reason is demanded, it may be found in one and not in another of two competing theories which otherwise would stand on an equal footing; the missing lines may even suggest a theory. At all events they suggest a system of classification; and, while the hardened theorizer may regard a system of classification as merely the forerunner of a theory, a theory is itself often nothing more than a classification stated in the language of an artificial analogy. It is, in fact, possible to arrange the stationary states, not in a single column as in Figure 10.3 but in several as in Figure 10.4; this arrangement being so contrived, that any transition can be identified in a moment as belonging among those which occur, or among those which are missing, whichever its case may be.

The mere fact that such an arrangement can be contrived shows that the missing lines are not distributed at random, but subject to some sort of a rule. Such rules are known as "principles of selection"; and the missing lines are commonly (and infelicitously) called "forbidden lines."

The families of stationary states

In the first edition of this book I said at this point that there is a far-reaching contrast between the spectra of hydrogen and ionized helium on the one hand, all other spectra on the other. The contrast, apparently as sharp as ever, has since been proved superficial. It results in part from the fact that lines which are far apart in the other spectra correspond to lines which are close together in these two, and in part from a strange and deceptive coincidence of levels which I will later describe (page 431). Sodium is a much better example to start with.

A few of the stationary states of the Na atom are exhibited in a single column on the left of Figure 10.4. The energy-value of every level, measured from our new zero at the state of the ionized atom, is affixed at the left. (The spacings between the lines which symbolize the levels are not proportional to the energy-differences, for if they were the diagram would be inconvenient.) Drawing arrows from each of the levels to every other, and ascertaining which of them correspond to actual and which to missing lines, one finds that the actual lines are so distributed that the stationary states may be sorted out in to several sequences or families. In Figure 10.4 these are shown in four parallel columns (there are other states falling into yet other columns, but we need not consider them). The feature of this arrangement is that *transitions between stationary states in adjacent columns correspond to actual lines; but the lines corresponding to all other transitions are missing.*

This is a principle of selection. It may be phrased in an equivalent but pregnant way, in this manner. Let me attach to the several columns the numerals 1, 2, 3, 4 . . ., as they are indicated at the bases; and let me use k as the general symbol for each and all of these numerals. Then this particular selection-principle may be phrased thus:

> *The only transitions which correspond to actual spectrum lines are those in which k changes by unity:* $\Delta k = \pm 1$.

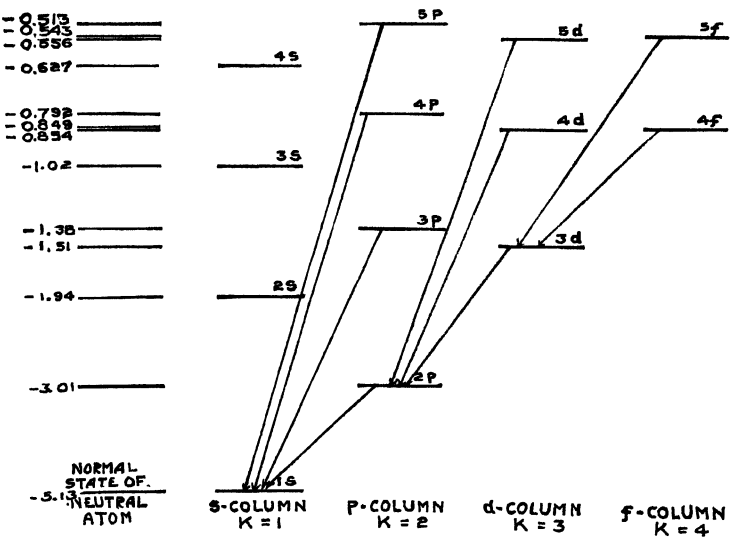


FIGURE 10.4—DIAGRAM OF THE STATIONARY STATES OF SODIUM, SORTED OUT INTO COLUMNS BY APPLYING THE SELECTION-PRINCIPLE FOR *k*. ARROWS REPRESENT LINES (BLUE FOR PRINCIPAL, YELLOW FOR SHARP, RED FOR DIFFUSE AND GREEN FOR BERGMANN SERIES).

To face page 308.

The numeral k bears the ponderous name of *azimuthal quantum-number*. This is a name derived from theory and not from experience, as will be made clear in due time. The principle of selection which has just been stated is the selection-principle for the azimuthal quantum-number.

Exceptions to this rule occur; the forbidden lines, like other forbidden things, occasionally evade the prohibition. This happens particularly when the atoms are subjected to intense electric fields, or to violent spasmodic electrical discharges in which strong transient fields are produced; in these circumstances great numbers of the missing lines leap suddenly into sight. Some lines corresponding to changes of k by two units or by none, which by the foregoing rule should be absent, do actually occur even when there is no obvious reason whatever for thinking that the atoms are subject to unusual stresses. The exceptions, however, are not numerous enough to jeopardize the rule.

Two other features of the columns must be noticed; first, that the successive levels in each column are not scattered at random, but form a converging series approaching the top of the column as limit (their energy-values form a sequence converging to zero); and second, that there is nothing arbitrary about the order of the columns, since the column at the extreme left admits of transitions to only one other and therefore is unmistakable, and all the others follow after it in an immutable order.

Notation

The symbol for a transition between two stationary states, and for the spectrum line which corresponds to that transition, consists of the symbols for the two states separated by an arrow, or a dash, or a semicolon, or any convenient mark. The final state (final with respect to emission) is commonly written first. Thus the line due to the transition from a state B to a state A is designated thus: $(A) - (B)$. Chess-players will be reminded of the "Continental" system of describing moves at chess, in which symbols for the squares from which

and to which the piece is moved are written down one before the other.

The notation for spectrum lines thus flows easily and naturally from the notation for stationary states or levels. The classification of levels which has just been described enables and requires us to give a twofold symbol to each level; the symbol must designate the column in which the level stands, and its order-number or serial number in that column. The columns are generally designated by the letters S, P, D, F, and thence forward by G, H, I and so on. Lower-case letters were formerly used, and there were other variants but the usage is now well standardized. The serial numbers are more troublesome.

One might expect that the level having the greatest (negative) energy-value in a particular column—the "deepest level" of the column, as it is sometimes called—would be called Number 1, and the successive ones Number 2, Number 3, and so forth towards the convergence-limit. For a while it was the practice to designate the first levels of the successive columns by the order-numbers 1, 2, 3 and 4, successively; so that their respective symbols were (1S); (2P); (3D); and (4F). These are the numerals I have affixed in Figure 10.4; but they are not the only (nor at present the usual) ones, as the order-numbers have jumped up and down several times to satisfy the exigencies of new atom-models. It would be unprofitable to confuse the beginner with further details, at least at this point. The important things to remember are three: that the symbol for each stationary state must contain one index for its column and another for its place in its column—that the former index is usually one of the specified letters—that the latter index is a number ascending along the column in unit steps.

Names and features of the most noted line-series

Every line in every series, according to Bohr's fundamental idea, corresponds to a transition or "combination" between two stationary states of the atom—to a transition from an

initial state to a final state. The atom possesses more energy in the initial state than in the final state (so long as we are considering emission as distinguished from absorption). Hence the energy-value of the initial state, reckoned as it usually is from the energy of the ionized atoms as zero, is algebraically lower and arithmetically higher than the energy-value of the final state.

The various lines of any one line-series have this in common: they correspond to transitions from various initial states which however all lie in one and the same column, into one final state which is the same for all and lies in an adjacent column. Each line-series thus belongs to one particular final state, and to one particular column of initial states.

The line-series consisting of transitions into the state (1S), or *terminating upon* (1S) as the phrase sometimes is, bears the name of *principal series*. Its consecutive lines are: (1S)–(2P); (1S)–(3P); (1S)–(4P) and so forth. They are signified by the blue arrows of Figure 10.4. The general symbol for this series is (1S–mP), which will be quite intelligible. The state is with many elements, though not with all, the Normal State; consequently, the various lines of the principal series correspond to transitions by which the atom regains its normal state after a temporary exile from it, and the limit of the series gives the ionizing-potential of the atom. It is probably for this reason that the series is prominent enough to have received the name *principal* from the spectroscopists.

Two series terminate upon the (2P) level. One of these consists of transitions from various levels of the S column. This is the *sharp* (or second) *subordinate* series, and its symbol is (2P–mS). The other series consists of transitions from various levels of the D column; it is the *diffuse* (or first) *subordinate* series, and its symbol is (2P–mD). Yellow and red arrows denote these series, respectively, in Figure 10.4. With sodium and many other elements, the (2P) state is among the Excited States the nearest to the Normal State; the energy-difference between it and the (1S) state is the resonance-potential, and the common limit of the two subordinate

series gives the ionizing-potential for atoms put into the (2P) state by electron-impacts or radiation. Of the two line-series terminating upon the (3D) level, only one has been endowed with a name; this is the series (3D–mF), known alternatively as the "Bergmann" or the "fundamental" series (neither name is appropriate) and symbolized by green arrows in Figure 10.4.

These series seem to be the only ones which impressed themselves strongly enough upon the minds of spectroscopic experts to receive names [1] from them. However, many other series have been identified, and emphasized, especially since Bohr's manner of thinking took root among the students of spectra; for instance, series terminating upon (2S) and (3S), which are conspicuous in the spectrum of helium, and such line-series as (3D–mP).

Several rules about line-series, which are very prominent in accounts of spectra (especially those written before 1914), become self-evident when the rules governing the stationary states are mastered (of course, this is only because the latter rules are based upon the former). For instance, there is a rule that the sharp and the diffuse series have the same limiting-frequency; and another, that the difference between this last and the limiting-frequency of the principal series is equal to the frequency of the first line of the principal series. The student may derive these by inspecting Figure 10.4.

Such rules are not of obvious application to the spectra of hydrogen and of ionized helium; and it is perilous to attach such names as principal or subordinate to the line-series of those two. Their stationary states are known by their energy-values, and the series by the names of their discoverers or interpreters; the series described by (10.2) is named after Balmer, that described by (10.4) after Lyman.

[1] The reader will recognize, in the initials of these names, the letters S, P, D, and F used to designate the several columns of levels.

Further analysis of the stationary states: multiplets

Let us consider the reasons for making these analyses in two stages. When the mid-Victorian physicist trained his spectroscope upon a tube full of glowing hydrogen, he saw the spectacle of Figure 10.1—the converging procession of distinct bright lines, of which the frequencies form that delightfully smooth numerical progression which we have already met. Later physicists with better instruments discovered that each of these "lines" was in fact a pair of lines. Now in strict truth, this discovery showed that the "lines" of the Balmer series were no lines at all; for a doublet is not a line. But the physicists continued to refer to the "lines" of the Balmer series, chiefly no doubt because to anyone equipped with an ordinary spectroscope the doublets do appear as single lines. By itself this is little reason; but the usage is not altogether faulty. Few people would hesitate to admit that each of these doublets is not a couple of casual neighbors, not two unrelated lines fortuitously close together, but a pair of lines sharing some deeply fundamental quality in common. This is indicated chiefly by the facts that the distance (measured in frequency) between the components of a doublet is the same for all the doublets, and very small compared with the distance between consecutive doublets. For this reason the doublets are treated as entities, and they require a name; which is what physicists have preserved for them, in continuing to call them "lines." "Doublet" would be better than "line," and "group" would be better yet; but we cannot ever be sure that even the apparently single lines are not very close groups, and yet it would be silly to call every line a group. Sirius appears as a double star in a few of the most powerful telescopes, but nobody would insist on calling it a double star when pointing it out in the night sky.

All this is not so trivial as it sounds. It is easy enough to speak of doublets when looking at lines which appear single except when viewed in the most powerful spectroscope, and then are resolved into components much closer together than the nearest similar line is to either. Such lines occur not in

the spectra of hydrogen and ionized helium only, but in the spectra of sodium and other elements generally. But the spectroscopist is constantly applying such names as "doublet" and "triplet" and "quadruplet," and the inclusive name "multiplet" to groups of lines which lie far apart in the spectrum, with scores of others intervening (as in Figure 10.5). Here his function is not to split apparent lines into narrow groups, but to unite widely-scattered lines into wide groups. This he does not because of propinquity of the lines, but because of resemblances or analogies or fixed intensity-relations between them, or because he finds it possible to construct a series of such groups with identical frequency-differences between corresponding lines within them, or because of analogies with other elements with more perspicuous spectra, or theoretical predictions, or intuitions or clairvoyance. Groups such as these are not generally termed lines, except in very abstract discussions; it is difficult to call a group a line, when it is clearly resolved by any instrument worthy the name of spectroscope. But they are like the lines of the Balmer series, treated as entities because their lines are believed to share some deeply fundamental quality in common.

What I have said about lines and groups of lines is transferable in substance to stationary states and groups of stationary states. Interpreting the groups of lines in the spectra of sodium and other atoms, we infer groups of levels. The levels in one of these groups are often far apart. They may be as many as eight in number, other levels may lie between; but by reason of the resemblances between the lines whence they were inferred, by reason of certain numerical relations between the levels themselves, they are believed to have some deeply fundamental quality in common. This quality we shall eventually interpret as the angular momentum of the electron-family of the atom; but this is for a later chapter.

The statements in the foregoing sections about the stationary states of sodium are now to be understood as relating to groups of stationary states. It is the *groups of stationary states which are arranged in parallel columns, designated by numerals k,*

such that no transition takes place unless in it k changes by one unit. It is the *group* of stationary states which is marked by a pair of numerals, one to designate its column and the other its place in its column; or by a letter to designate its column and

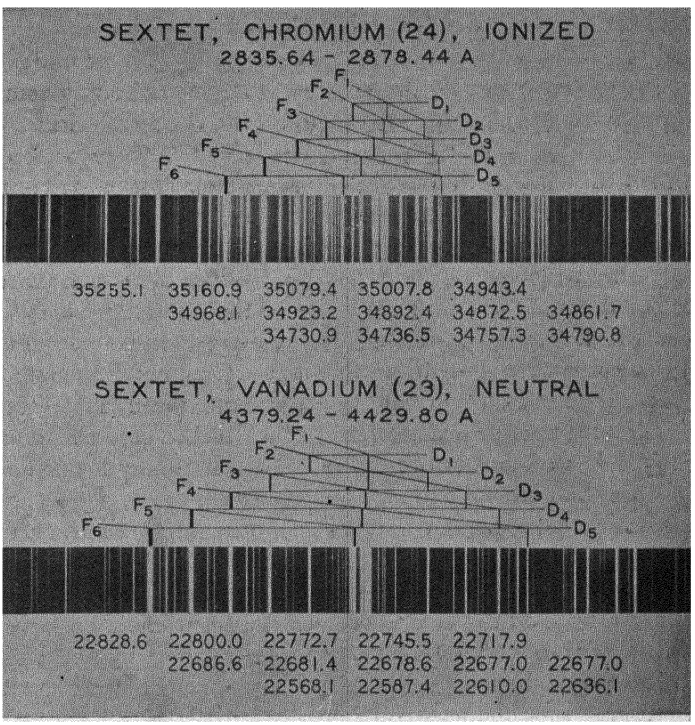


FIGURE 10.5—MULTIPLETS. FOURTEEN LINES RESULTING FROM THE PERMITTED TRANSITIONS BETWEEN A GROUP OF SIX F LEVELS AND A GROUP OF FIVE D LEVELS OF A SEXTET SYSTEM; BOTH PICTURES REFER TO ATOMS WITH 23 ELECTRONS, THE UPPER TO ONCE-IONIZED CHROMIUM AND THE LOWER TO NEUTRAL VANADIUM. (Meggers.)

a numeral to designate its place in its column. It is the *group* of stationary states which is denoted by (2P) or (5D).

To denote a particular stationary state we must add, to the symbols for its group, a third symbol for its place in its

group. This symbol is added as a subscript to the letter denoting the column. The most common general symbol for this numeral is j. Geometrically, the stationary states may be represented by lines or dots arranged, not in one row of several parallel columns as in Figure 10.4, but in several rows of parallel columns. Readers with three-dimensional imaginations in good working order may develop this idea *ad libitum*. The system for assigning the values of j is based upon a second selection-principle, which is co-equal with the first selection-principle.

This principle is derived in the same way as the first one. The groups of levels are established by inference from the groups of lines; then arrows are drawn from every level to every other, the corresponding spectrum-lines are sought, and most of them are not found. Some of these missing lines are those which would contravene the first selection-principle, as they correspond to transitions in which the numeral k changes by more than one unit, or not at all. Putting these aside, there are still a number of missing lines, to which the first selection-principle has offered no objection. Now it is found possible to choose the numeral j in such a manner that the only transitions which correspond to actual spectrum lines are those in which j changes by one unit or not at all ($\Delta j = 0, \pm 1$). Furthermore it is possible to adjust the values of j in such a manner that the lines corresponding to transitions, in which j is initially zero and remains unchanged, are missing. This was formerly called the selection-principle for the inner quantum number, for the numeral j was so entitled; now the title is almost forgotten, and j has so to speak become its own name, having practically attained to the dignity of a word.

As the two selection-principles are effective concurrently, the pair of them may be fused into this one:

Of the three numerals n, k and j, which specify a stationary state completely, two (k and j) may be so chosen that the only transitions which correspond to actual lines are those in which: first, $\Delta k = \pm 1$; second, $\Delta j = 0, \pm 1$; third, j is not zero both before and after the transition.

It remains to give some idea about the number of stationary states in the various groups. Now I have to make a premonitory allusion to the fact that we have not even yet exhausted the subdivision of stationary states; we are thinking about "fine structure of spectra," and there remains "hyperfine structure" to cope with. Deferring this last, I say that for sodium, as laid out in Figure 10.4, the groups in the S column are merely single levels (this sounds like a contradiction in terms, but may be borne for the sake of the generality); the groups in the other columns are pairs of levels, or "doublet terms." This is the common character of the alkali elements Li, Na, K, Rb and Cs, which share the first column of the periodic table with the "noble metals" Cu, Ag, Au; these last also exhibit doublet systems (together with others). For elements of the second column of the periodic table there are two complete systems of stationary states, each having its own S column, its own P column, its own D column, and all the rest. In one system, all the groups in every column reduce to single levels; it is a *singlet system*. In the other, all the groups in the S column are single levels, all the groups in the other column are triads of levels or "triplet terms;" it is a *triplet system*.

The complexity mounts up stage by stage as we cross the periodic table of the elements from left to right. To describe the various systems actually observed, it is necessary to have recourse to a table. To make clear this Table 10.1, I will

TABLE 10.1

Name of System	S	P	D	F	G	H
Singlet.....	1	1	1	1	1	1
Doublet.....	1	2	2	2	2	2
Triplet.....	1	3	3	3	3	3
Quartet.....	1	3	4	4	4	4
Quintet.....	1	3	5	5	5	5
Sextet.....	1	3	5	6	6	6
Septet.....	1	3	5	7	7	7
Octet.....	1	3	5	7	8	8

explain the fourth line; this line contains the statements that a "quartet system" of stationary states consists of an S column of single levels, a P column of groups of three levels each, and a D column, and F column, and additional columns of groups of four levels each. The type of system to which a level belongs is indicated by a superscript preceding its letter: ^4P refers to a level belonging to a P-column and a quartet system.

The "multiplicities," as they are called, depend in a striking way upon the number of electrons in the atom. Consider neutral atoms to begin with, as we have exclusively till now. For these it is a universal rule that elements of odd atomic number exhibit only even multiplicities, while elements of even Z exhibit only odd ones.[1] Here is a new type of dependence on atomic number, not mentioned in Chapter IV. Upon it however is superposed a periodic dependence: as one goes from left to right along a row of the periodic table beginning with an alkali metal, the systems found in the individual spectra increase and then later decrease in complexity, in a way which again can best be illustrated by a table. In this Table 10.2 I have inserted, for later reference, the systems discerned in the spectra of the once-ionized atoms of the element in question.

TABLE 10.2

Element	Z	Systems	Systems for Ion
K.....	19	2	
Ca	20	3, 1	2
Sc.....	21	4, 2	3, 1
Ti.....	22	5, 3, 1	4, 2
V.....	23	6, 4, 2	5, 3
Cr.....	24	7, 5, 3	6, 4, 2
Mn.....	25	8, 6, 4	7, 5
Fe	26	7, 5, 3	6, 4
Co	27	6, 4, 2	5, 3
Ni.....	28	5, 3, 1	4, 2

[1] Helium used to be thought an exception, displaying a singlet and a doublet system; but about 1927 it was shown that the latter is really a triplet system.

Periodic dependence, as we have seen in Chapter IV, is a sign that the property in question belongs to the electron-family, and indeed to the outermost members of the electron-family, of the atom. From this we suspect that insofar as Z controls the multiplicities of the spectrum, it is Z in its role of number of electrons in the family, rather than Z in its role of nuclear charge. This is confirmed by observations on the spectra of ionized atoms.

Spectra of ionized atoms

Positive ions—that is to say, electron-families of fewer than Z members surrounding a nucleus of charge Ze—have spectra of their own, exhibiting term-systems, multiplicities, and all the other features of spectra of neutral atoms. They are evoked by relatively violent agents: fast rather than slow electrons, dense rather than scant bombarding electron-streams, spark-discharges rather than arc-discharges. The name "spark-spectrum" still lingers, because as much as sixty years ago—long before the reason was realized—it was observed that a high-voltage electric spark in a gas (or between metal electrodes which it serves to vaporize) brings out whole flocks of spectrum lines which are not evoked if the same gas or vapor is introduced into an arc.

The first of the fundamental facts about these spectra is, that in regard to multiplicities they conform to the rule already given for those of neutral atoms, if for "atomic number" in that rule we read "number of electrons." To put this into a short phrase and introduce a new word of obvious meaning: *isoelectronic spectra are alike in their multiplicities.* Reverting for a moment to the notation of Chapter III for ions: the spectrum of Mg^+ displays a doublet system just as the spectrum of Na does, both being due to families of 11 electrons; the spectrum of Al^+ displays a singlet and a triplet system just as the spectrum of Mg does, both being due to families of 12 electrons. Introducing a new notation: the spectra Mg II and Na I are "11-electron spectra" and both have doublet

systems, the spectra Al II and Mg I are "12-electron spectra" and both have singlet and triplet systems. These instances can be amplified immensely, for the decade of the twenties was spent by spectroscopists largely in developing sparks and other agents of ionization and excitation so powerful; that spectrum-lines due to ions lacking several or even many electrons of the normal quota were evoked. Looking casually into the indexes of a few recent volumes of the *Physical Review*, I find notes or articles about the spectra P V to P XII inclusive (electron-families of 11 to 4 members, grouped around a nucleus of phosphorus $Z = 15$!), the "isoelectronic sequence" A I to Mn VIII inclusive (electron-families of 18 members, grouped around nuclei for which Z has the values 18, 19, . . . 25), and the isoelectronic sequence V I to Cu VII inclusive (electron-families of 23 members, grouped around nuclei $Z = 23$, 24, . . . 29), not to speak of many referring to sets of fewer spectra. It would doubtless be an exaggeration to say that the multiplicities of all these have been ascertained beyond question, but the statement must be true for many, and no violation of the fundamental rule has been demonstrated.

It must not be supposed that the main parts of all these spectra lie in or near the visible! if they did, perhaps even the skill of spectrum analysts might not have sufficed to distinguish them from each other. Here the value of the nuclear charge, as contrasted with the size of the electron-family, plays a part; for the greater the value of Z in an isoelectronic sequence, the further the spectrum moves off into the far ultraviolet. The progress in this field during the twenties was due not altogether to the perfecting of the agents of ionization and excitation, but also and largely to the improvement of the art of ultra-violet spectroscopy. A range of wavelengths which as lately as 1926 could still be called "the unexplored region" (*vide* the first edition of this book) is now as well explored as any. The general fact that the spectrum-frequencies in an isoelectronic sequence increase with Z, and many special details as to the variation of minor features of the spectra, are precious to atomic theory.

Systems of stationary states in the neutral atom converging upon distinct states of ionization

It having been shown that ions possess spectra and therefore systems of stationary states, the question again arises: did we, in Chapter IX and elsewhere, make any grave mistake in imagining only one "state of the ionized atom" to use as our zero of energy?

To this it can now be replied that the "state of the ionized atom" heretofore stressed—the state of which the energy is I_0 electron-volts greater than that of the normal state of the neutral atom, I_0 being the ionizing-energy defined in Chapter VIII—is indeed but one among many; but ambiguity is seldom to be feared, inasmuch as that particular state is usually of distinctly lower energy than any of the rest. Some of the others can however be inferred from the spectrum of the neutral atom.

We have seen that when the spectrum of an element displays converging line-series, the system of Stationary States or "terms" deduced from the spectrum-lines consists of a group of term-sequences, each converging upon the State of the Ionized Atom as limit. This is true of elements possessing only one system of terms, such as sodium. It is true also for many elements possessing two or more systems of terms; such as mercury with its singlet system and its triplet system, each comprising S and P and D sequences, all converging upon the State of the Ionized Atom which is 10.4 electron-volts "above" the Normal State in energy. However, spectra of ionized atoms show us that there are countless states and not one only; perhaps then detachment of an electron from an atom may be effected in several distinct ways, each involving the transfer of a distinct amount of energy to the atom, and leaving the ion in a distinct state. We may therefore expect to find evidence, in the spectrum of an element, for distinct systems of term-sequences converging upon distinct limits; and such evidence in fact exists.

The spectrum of neon affords a celebrated instance.

Some of its term-sequences converge upon one limit, others upon another limit differing from the first by 682 units of wave-number; which signifies, if the customary relation between sequence-limits and ionizing-potentials is valid in this case, that neon has two ionizing-potentials differing from one another by 0.085 electron-volt. The spectrum of calcium offers another instance; there are two families of terms, one comprising S and P and D sequences which converge upon a certain limit, and the other comprising P′ and D′ sequences (the primes distinguish one family from the other) which converge upon a limit of considerably higher frequency; indicating thus two states of ionization differing from one another in energy by 1.72 equivalent volts. There are many elements which possess systems of "unprimed" terms and systems of "primed" terms; among these are the alkaline-earth metals, and also many on the right-hand side of the Periodic Table, of which the spectra are complicated. Many of these, it must be admitted, do not display traceable line-series, or at all events the line-series have not yet been discerned; so that the term-sequences cannot be followed to their limits, and it is not certain that the different families of terms are associated with different ionizing-potentials.

Transitions may occur between a State belonging to a "primed" family and a State belonging to an "unprimed" family. There is a rule of selection governing such transitions: values of the azimuthal quantum-number k may be assigned to the various term-sequences of two such families, in such a way that the transitions occurring from one family to the other correspond to changes of zero or 2 in k, while the transitions within either family correspond to changes of one unit in k (Laporte's rule). Thus in the spectrum of calcium there are lines due to transitions between P states and S or D states; there are transitions between P′ states and D′ states; and there are transitions between P and P′ states and between D and D′ states.

Excitation of individual frequencies

So long as an atom is conceived as a belfry full of bells of various pitches, it would probably be argued that a shock to the atom would set all the bells to jangling, and a gas bombarded by electrons would emit all of its natural frequencies if any. The interpretation of spectra to which these pages are devoted leads to a very different idea. A spectrum-line of frequency ν is emitted when the atom passes from a stationary state B to a stationary state A. The energy-value of state B by itself does not determine ν; this is controlled by the difference between the energy-values of B and A, which is $h\nu$.

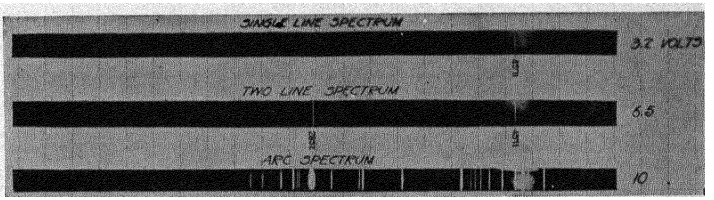


FIGURE 10.6—CONSECUTIVE EXCITATION OF SINGLE-LINE, TWO-LINE AND COMPLETE SPECTRUM OF NEUTRAL MAGNESIUM (Mg I) BY ELECTRON-IMPACTS OF ADJUSTED ENERGY. (Foote, Meggers and Mohler; *Origin of Spectra.*)

But the energy-value of B has everything to do with whether or not the frequency ν is emitted under given conditions; for it will not be emitted at all unless the atom is first put into state B. If the gas is bombarded with electrons of energy insufficient to raise an atom from its normal state to state B, then the line in question, and all of the other lines which result from transitions from B to other levels of lower energy-value, will fail to appear. If the energy of the electrons is raised past the critical value (the difference between the energy-value of B and the energy-value of the normal state) all of these lines suddenly appear.

This is illustrated by Figure 10.6, relating to magnesium. The Normal State of magnesium is an S state of the singlet system; the excited states differing least from it in energy [1]

[1] Not counting two states for which transitions into the normal state do not occur, being incompatible with the selection-principle for j.

are a P state of the triplet system and a P state of the singlet system exceeding it in energy by 2.65 and 4.42 equivalent volts respectively. These three states are denoted by 1^1S, 2^3P_2 and 2^1P. The transitions (1^1S–2^3P_2) and (1^1S–2^1P) correspond to the wavelengths 4571A and 2852A respectively. In Figure 10.6 the uppermost spectrum is that of magnesium vapor bombarded by 3.2 volt electrons. Only the line λ4571 appears. The middle spectrum was obtained with a bom-

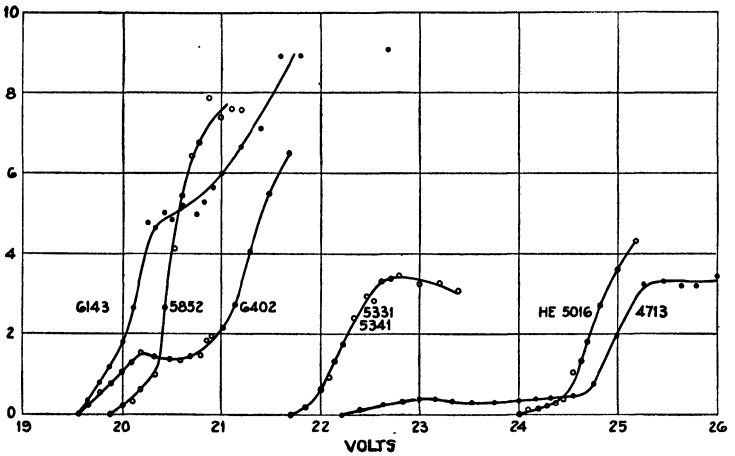


FIGURE 10.7—CONSECUTIVE EXCITATION OF INDIVIDUAL LINES OF THE NEON SPECTRUM. (Hertz; *Zeitschrift für Physik.*)

barding-voltage of 6.5. Both lines now are seen. The lowest spectrum is the spectrum of magnesium vapor bombarded by electrons having more than the energy (7.61 EV) required to ionize. All of the stationary states are now accessible, and the frequencies excited are therefore numerous.

Another instance is afforded by Figure 10.7, in which the curves represent the intensities of various lines of neon and helium in the spectrum of a mixture of these gases, bombarded by electrons of which the energy is the independent variable. The points at which these curves cut into the axis of abscissae stand at the energy-values of the initial states of the transitions

to which they correspond. The two curves meeting at the axes correspond to transitions out of the same initial into different final states.[1]

" Single-line " spectra like the uppermost one in Figure 10.6 are not difficult to obtain. Any curve revealing radiation-potentials, such as that of Figure 9.4, is evidence for the consecutive appearance of individual lines; although not by itself binding evidence, since the mere appearance of a " break " in a photoelectric-current-curve does not prove that the newly commencing radiation is a single line rather than a group of lines or a band. Furthermore, the appearance of a line corresponding to an initial state B is not always proof-positive that the electrons in the gas have kinetic energy equal to the energy-difference between B and the normal state; for the atoms in state B might have arrived there step-by-step through intermediate states. Only when the state B is the nearest to the normal state is this possibility excluded.

Excitation of a line by radiation would be expected to commence as soon as the frequency of the radiation is raised to the value at which its quantum-energy equals the energy-difference between the normal state of the atom and the initial state for the transition in question. The difficulty of producing intense radiation of any desired and continuously variable frequency prevents the test from being made. Instances of excitation by radiation have been given in Chapter IX; as another, I quote the fact that mercury vapor emits its visible spectrum when illuminated by the line λ2536 plus the rest of the spectrum from a mercury arc, but ceases to do so when λ2536 is cut off, though the frequencies of its visible spectrum continue to shine upon it. Photons of these frequencies cannot lift a mercury atom from its normal state into the states which are initial states for them. Neither, for that matter, can those of the line λ2536; but the consecutive absorptions of λ2536 and (say) λ4358 lift the atom first to the P state which is the initial state of λ2536 and the final state of λ4358,

[1] Hertz, de Visser; **ZP** 31, 470 (1925). Consult this paper also for citation of prior attempts on the problem, and description of the difficulties involved.

and then to the S state which is the initial state of $\lambda 4358$; which line thereupon is excited.[1]

Absorption-spectra

An atom which will emit a frequency ν when it is originally in a state B and passes over into a state A, will absorb light of the same frequency if it is initially in the state A. This has the important consequence that the lines which a gas absorbs, when lying at rest and unexcited, are those which it emits in passing from any and every other state *into the normal state*. The lines emitted when an atom passes from one of its stationary states into another which latter is not the normal state, are not absorbed by the gas lying quiescent and undisturbed. For this reason helium and neon and argon are quite transparent to all visible light, although they have many emission-lines in this region of the spectrum; for each of these lines corresponds to a transition into some other than the normal state, while the lines which correspond to transitions into the normal state lie far off in the ultraviolet But if such a gas is made the theater of a self-sustaining electrical discharge, the other lines likewise are absorbed; for the discharge puts the atoms of the gas temporarily but frequently into various abnormal states (Figure 10.8). This incidentally is one of the bits of evidence that an atom may sojourn for an appreciable time in another stationary state than the normal one. If the gas is heated, the same effect occurs; for the violent collisions between atoms in a hot gas occasionally bring atoms into excited states.

By observing the absorption-spectrum of a quiescent gas one learns which lines in the emission-spectrum correspond to transitions into the normal state—a valuable piece of information in the cases of elements of which the spectra are complicated and obscure.

If the central part of a volume of gas is bombarded by elec-

[1] R. W. Wood, **PRS A106**, pp. 679–694 (1924); Fuchtbauer, **PZ 21**, pp. 635–638 (1920); *Origin of Spectra*, pp. 99–108.

trons or otherwise excited, the radiation which the atoms there emit in returning to the normal state will be absorbed by atoms in the surrounding regions. The brightness of lines due to transitions into the normal state is sometimes enfeebled by this effect, other lines being unaffected and so appearing unduly bright in comparison. If the gas is dense or its central region

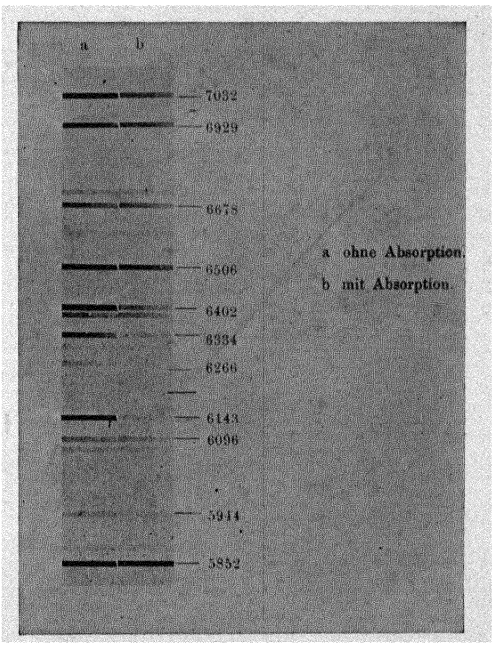


FIGURE 10.8—EMISSION-LINES OF NEON TRANSMITTED THROUGH A LAYER OF THE SAME GAS, THE LATTER (*a*) QUIESCENT AND (*b*) SUSTAINING A DISCHARGE. (Meissner; *Annalen der Physik.*)

is violently excited, it is often observed that an emission-line due to a transition into the normal state is quite broad and that its edges are brighter than its center, even that there is a dark band splitting it apparently into two lines. The dark band is due to the absorption; the brighter edges on either side testify that in the excited region of the gas the emission-frequency is

not always the same, that the energy-values of the Stationary States are slightly affected, so that the light emitted is not in every case of quite the right frequency to be absorbed by quiescent atoms. In the spectrum of a mercury arc surrounded by quiescent mercury vapor the line $\lambda 2536$ thus appears "self-reversed" as the phrase is, with its "core" removed.

CHAPTER XI

Original Atom-Model of Bohr

We have now ascertained that the atom is capable of existing in divers distinct stationary states marked by discrete distinctive energy-values, and is somehow inhibited from existing in the intermediate states which would correspond to intermediate energy-values. This feature must be imprinted upon our nuclear atom-model. It must be imprinted, that is to say, unless we can find it already implicit in the model. But we cannot find it implicit in the model, so long as we regard the electrons and the nuclei purely as we did in Chapters I to III—to wit, as particles which in an electric field behave exactly according to the laws of motion to which massive pieces of electrified matter conform, and heed no other laws. Take for simplicity the case of the hydrogen atom, for which the model consists of a nucleus of charge $+e$ and a single negative electron. Experiment has shown us that it has (among others) a stationary state of which the energy, referred to a certain clearly defined zero, is $-Rh$. Turning to the model, we can easily prescribe an orbit such that if the electron is following it according to the classical laws, the energy of the system is precisely $-Rh$. With equal ease, however, we can trace other orbits for which the energy differs from $-Rh$ by various amounts, indeed by any amount arbitrarily assigned. Nothing in the model thus far presented makes these orbits any less tenable than the one for which the energy is $-Rh$. If we retain the model we must make an explicit and superadded postulate that the orbits with the energy-values $-Rh/n^2$ ($n = 1, 2, 3 \ldots$, as stated in Chapter X) are the only ones which are tenable.

This postulate is an outright and a strange excrescence upon

the original model, or as it is usually called, " the classical theory." Actually, however, the situation is even stranger than has yet been intimated. Though I have implied that every classical orbit is equally possible with every other, it would be truer to say that every orbit is equally impossible with one another. According to electromagnetic theory, an electron cannot endlessly describe a closed ellipse about a nucleus as can a plant about the sun. In such an ellipse, it would be undergoing a continuous acceleration; now, a continuous acceleration of an electrified particle entails, by that theory, a continuous radiation of energy; the ellipse, therefore, would be transformed into a spiral, ending eventually at the nucleus, into which the electron would fall. The nuclear atom-model is, therefore, an unstable one. If we were to seek to avoid this impasse by imagining the electrons as stationary, nothing would be gained, for the new model would be as unstable as the old, and would have the extra disadvantage of being non-magnetic.

It was the multitude of stationary states identified from spectra which made the dilemma conspicuous; and yet, a little reflection shows that it would be equally marked, if no state excepting the normal state were known for any atom. The mere fact that an atom has a normal state, marked with a distinctive value of ionizing-potential—indeed the mere fact that each chemical element has sharply-defined features and properties of its own, is sufficient to prove that in the atom there must be a principle of identity and of stability, which is not implicit in the nuclear atom-model as I originally described it. Even if that difficulty about radiation did not exist, the atom would be a mere tangle of intricately interwoven and incessantly changing electron-trajectories, with no distinctive ionizing-potentials nor anything else distinctive about it, were there not this principle to regularize the paths and stabilize the system. To discover and express this principle is the first task of quantal theory.

The principle, as we shall presently see, is a consequence of the Rules of Correlation: it is the association of waves with

corpuscles which is responsible for the stability and the character of the atoms, qualities which corpuscles unconnected with waves could not by themselves provide. Yet even while one of the Rules ($p = h/\lambda$) was still unknown, and the other ($E = h\nu$) supposed to be applicable only to light, Bohr enounced the principle in a form necessarily incomplete, and nevertheless of immense value. I quote next the major features of his formulation: in small part because of its historical importance, in large part because its symbolism and its terminology still live, and chiefly because it signalizes the importance of angular momentum. For angular momentum is constant in any stationary state of the atom, as energy is constant; angular momentum has distinctive values, as energy has distinctive values; but the distinctive values of the angular momentum are so simple and so universal, that the stabilizing principle may be said to manifest itself in them primarily, and in the energy-values in a manner more secondary and remote.

Original atom-model of Bohr

Conceive the hydrogen atom as a nucleus of charge $+e$ and a negative electron. Assign to the electron its known rest-mass, m_0, and treat the system by the methods of ordinary as distinguished from relativistic mechanics. Assign temporarily to the nucleus an infinite mass, so that it may stand still while the electron moves; we shall soon correct the resulting error. Imagine the difficulty about radiation to be suspended for certain orbits, which we will call " permitted." Let these permitted orbits be circles, and let them be selected by the following rule bearing upon the angular momentum p of the electron in the orbit:

$$p = kh/2\pi \quad k = 1,2,3,4. \ . \ . \ . \qquad (11.1)$$

In this rule the principle of stability first manifests itself.

The distinctive or permitted energy-values corresponding to these permitted orbits are easily found. Denoting the orbit-radius by r and the speed of the electron by v, we have:

$$p = mvr, \ mv^2/r = e^2/r^2, \qquad (11.2)$$

two equations between which v is to be eliminated. Combining the resulting equation with (11.1), we find:

$$r_k = k^2h^2/4\pi^2me^2 \qquad (11.3)$$

for the radius of the kth permitted circle. The energy W_k of the atom with the electron revolving in this orbit, referred to the energy of the atom with the electron at rest and infinitely far away from the nucleus (for such is the " state of the ionized atom " which we have been using all along as standard of reference) is now to be computed. Were the electron at rest at a point of this orbit, the work required to remove it to infinity and leave it there at rest would be e^2/r_k. This suggests for W_k the value $-e^2/r_k$, to which, however, we must add the kinetic energy $\frac{1}{2}mv^2$ of the orbital motion of the electron; this last, as the student can easily show, is $+\frac{1}{2}e^2/r_k$. We have accordingly:

$$W_k = -\frac{1}{2}e^2/r_k = -2\pi^2me^4/h^2k^2. \qquad (11.4)$$

Now these energy-values will agree with those of the stationary states of the hydrogen atom as expounded in Chapter X (equation 10.16) provided that the value of the experimental constant there called R agrees with that of the following combination of fundamental constants:

$$R = 2\pi^2me^4/h^3. \qquad (11.5)$$

A close agreement *does* exist, and in 1913 when Bohr's famous first paper appeared in the *Philosophical Magazine* it made a sensation, which surely few physicists who were already acquainted with their science at that time can ever have forgotten. All the aid which the theory derived from that agreement was badly needed, so many and so universally accepted were the classical ideas which it flouted! The agreement at the time was not so impressive as it later became, for m and e and in particular h had not been measured so often or so carefully as they were soon to be. In the last few years the situation has become beclouded because of the uncertainty

in the e-value to which I alluded in Chapter I, and which here is of serious import.[1]

This theory finds further support when it is extended to the case of ionized helium. The once-ionized atom of helium—the ion He^+—consists of a single electron and a nucleus of charge $+2e$. In the second of the equations (11.2), the right-hand member must then be doubled. Selecting again the permitted orbits by the rule of (11.1), one arrives at the conclusion that all the permitted energy-values of (11.4) are to be quadrupled: the line-spectrum of once-ionized helium—the spectrum He II, in our recently learned notation—should then be a duplicate of the line-spectrum of hydrogen, with every frequency raised fourfold. In the previous chapter I implied this to be fact, by writing the formula (10.7) for the lines of He II.

The frequency-ratio of corresponding lines in the two spectra is, however, not exactly 4 to 1; and here appears an especially pleasant agreement. Discarding the assumption that the mass of the nucleus is infinite, let us put the actual mass-values M for the proton in the one case, for the alpha-particle in the other. Now the nucleus and the electron both revolve around the center of mass of the system. Thinking only of circular orbits as before, we see that the radii r and R of the circles described by the electron and the nucleus respectively, being the respective distances of these particles from the center of mass, must stand in the ratio of the masses M and m; and as the particles describe their circles in the same period (since the center of mass is at rest and always between them) their speeds v and V must stand in the same ratio:

$$r/R = v/V = M/m. \qquad (11.6)$$

I introduce the symbol μ to denote the equal quantities:

$$(M + m)/M = (r + R)/r = (v + V)/v = \mu. \qquad (11.7)$$

[1] The quantity equated to R in (11.5) may be written $2\pi^2(e/m)^{-1}(h/e)^{-3}e^2$. If for e/m and h/e we put the values which at the moment seem clearly the most reliable, agreement with R is attained by writing for e the square of the Millikan value but not by writing that of the newer value!!!

Were the electron and the nucleus at rest at the distance $(r + R)$ from one another, the work required to separate them infinitely far would be $Ze^2/\mu r$; here Ze stands for the nuclear charge, as heretofore. From this work must be deducted the sum of the kinetic energies of the two, $\frac{1}{2}mv^2$ and $\frac{1}{2}MV^2$. I point out that the centrifugal accelerations of electron and nucleus respectively are v^2/r and V^2/R, and that the product of each of these by the corresponding mass is equal to the reciprocal attraction $Ze^2/(r + R)^2$, and leave it to the student to complete the argument: in the course of doing which, he is to assume that it is the total angular momentum of the system, the sum of mvr and MVR, which is to be equated to an integer multiple of $h/2\pi$. The outcome is, that for the kth of the permitted energy-values the expression is not quite the same as that given in (11.4), but is rather:

$$W_k = -\ 2\pi^2\mu m Z^2 e^4/h^2k^2, \qquad (11.8)$$

in which for hydrogen we are to put $Z = 1$, for ionized helium $Z = 2$.

So little does μ differ from unity, and so relatively large are the uncertainties in the other fundamental constants in these expressions for W_k, that it would be quite impossible to show (11.8) preferable to (11.4) from observations on either of the two spectra by itself. If, however, one evaluates the ratio between the frequencies of corresponding lines in the two spectra, all the other constants drop out, and one is left with the ratio:

$$\nu_{He} : \nu_H = 4\mu_{He} : \mu_H. \qquad (11.9)$$

This opportunity of testing Bohr's atom-model existed already when the model itself was first made, for certain lines of the spectrum He II had already been discovered in stellar spectra. It is interesting to recall that these particular lines had been ascribed to hydrogen, owing to the simple numerical relations between their frequencies and those of the Balmer lines; and when they were first generated in mixtures of hydrogen and helium in the laboratory, they were still attributed to the

former of these gases. Bohr inferred from his own theory that they should belong to helium, an inference which was soon verified by someone seeking them and finding them in the spectrum of samples of helium gas made as pure as possible.[1]

Another opportunity for testing the model resides in the fact that hydrogen possesses two isotopes, for both of which $Z = 1$ while M differs in nearly the ratio 2 : 1, and in place of (11.9) we have:

$$\nu_H : \nu_D = \mu_H : \mu_D. \qquad (11.10)$$

It is most interesting to relate that the discoverers Urey, Brickwedde and Murphy of the second isotope (here indicated by the suffix D) availed themselves of this quality to achieve their discovery; they sought for faint companions of the Balmer lines, shifted toward higher frequencies. The samples of hydrogen which they used were obtained by liquefying large quantities of gas and letting it re-evaporate at a low pressure, they taking for their investigation the last two or three cubic centimeters of liquid out of several thousand; for they had foreseen, by the aid of thermodynamic theory, that if such an isotope existed at all it must be relatively more abundant in such a residue. The lines due to this isotope in the ordinary samples of gaseous hydrogen, exceedingly faint as they are, have since been detected; while on the other hand it is now possible to obtain samples of hydrogen in which the new isotope is so predominant, that its own lines outshine the others by far.

But what of the test itself? The μ-value for any of these systems—H atom, D atom, He^+ ion—depends upon the ratio of electron-mass to nucleus-mass. Our prior knowledge of these ratios depends on such experiments as those recounted in Chapters II and III, which are affected by uncertainties greater than those of measurements on spectra. It may be said at once that the test is successful, insofar as the quality of

[1] They belong to the series for which, in the notation of (10.7), m is 4; alternate members of this series lie between consecutive members of the Balmer series, while the remainder would coincide precisely with Balmer lines if it were not for the difference in the μ-values—the frequency-ratio between two such nearly-coincident lines should be the μ-ratio, which makes the test particularly easy.

that prior knowledge permits it to be made; by which I mean, that the very-exactly-measured values of the ratios on the left of (11.9) and (11.10) lie well within the comparatively broad limits of uncertainty allowed to the right-hand members of these equations by the deflection-experiments. Proceeding in their customary manner, physicists now reverse the argument, accept the theory, and employ the equations for determining the ratios of the μ-values and thence the mass-ratios; Shane and Spedding have lately given 1835.6 ± 0.2 as the ratio of proton-mass to electron mass derived with the aid of (11.10).[1]

Elliptical orbits in the Bohr atom-model

By speaking thus far of circular orbits exclusively, I have managed to emphasize angular momentum as a cardinal feature of the atom; and this is all to the good. Unfortunately, I have also created the impression that to each one of the thus-far-defined states of the hydrogen atom, there corresponds just one circular orbit and just one value of angular momentum: the p-value $kh/2\pi$, corresponding uniquely to the energy-value $-Rh/k^2$. This impression is wrong, and must be corrected even at the price of a temporary obscuration of that other principle which we are going to keep.

Think of any one of the circular orbits of which I have thus far been speaking. It can be shown that the energy-value corresponding to this orbit is duplicated by every elliptical orbit of which the major axis has the same length as the diameter of the circle.[2] There is in fact a continuous infinity of elliptical orbits, all having the same major axis and the same energy-value; at one extreme is the circle for which minor and

[1] *Cf.* R. C. Williams and R. C. Gibbs, **PR 48**, 971 (1935); C. D. Shane and F. H. Spedding, **PR 47**, 33 (1935). The method is there presented as a scheme for evaluating (e/m) for the electron; the ratio (m/M) is written as $(e/M)/(e/m)$, the value of e/M being then taken from experiments on electrolysis (Chapter I). Nevertheless it is (m/M) for which this particular combination of experiment with theory presents a value.

[2] It does not seem worth while to carry through the somewhat lengthy proof, which was given in the first edition of this *Introduction* (pp. 317–321).

major axes are equal, then comes the continuous succession of ellipses with continuously shortening minor axis, the last of which (if the mathematics still agrees with the physical situation) is a straight line. Corresponding to these ellipses there is a continuous succession of values of angular momentum ranging from that for the circular orbit, which is the greatest of all, down to zero.

This seems to make nonsense of our procedure of associating " permitted " values of angular momentum with the energy-values $-Rh/k^2$, and accepting them as fundamental: should we not have adopted the major axis of the orbit as the fundamental feature, and spent our time in ascertaining permitted values for that major axis? The answer would surely be " yes " if there were nothing to be explained except the spectra of hydrogen and ionized helium, and if these spectra were no more complicated than I have yet implied. There are, however, many other things to be explained, including the so-called " fine structure " of these two spectra (not yet mentioned in this book), and including also the classification of levels of other atoms than hydrogen into columns denoted by S and P and D and other letters. All these phenomena require us to say, not that the only permitted orbits are circular and not that all of the ellipses are permitted, but something intermediate. *The permitted orbits are those circles and also those ellipses, for which the angular momentum is an integer multiple of* $h/2\pi$. Thus if we choose a state of energy-value $-Rh/k^2$ for which k is sufficiently large, there will correspond to it not only a circular orbit of angular momentum $kh/2\pi$, but also an eccentric elliptical orbit of angular momentum $(k-1)h/2\pi$, and also a still more eccentric elliptical orbit of angular momentum $(k-2)h/2\pi$, and so forth until the possibilities are exhausted—all of these having the same major axis, its value being obtained by doubling r_k as given by (11.3). Incidentally these ideas were developed by Sommerfeld and his school, who made contributions so great that there is justice in speaking of the model with "quantized" elliptical orbits as the "Bohr-Sommerfeld atom-model."

This act of correlating a finite number of different orbits—or I might say, a finite number of different states of the atom-model—with a single energy-value is our first example of something encountered very often in theoretical physics. Its importance lies largely in the fact that if the model is slightly changed, *e.g.* by a small alteration in the field of force to which the electron is exposed, these several orbits or states exhibit energy-values which are slightly different from each other. Thus if in our model of the hydrogen atom we should postulate that the nucleus exerts upon the electron an attractive force which varies nearly but not exactly as the inverse square of the distance, the various orbits sharing a common value of the major axis would have energy-values all differing slightly from each other. Such is actually the case in (for example) the sodium atom, in which the outermost electron revolves in a field of force which is due to the positive charge on the nucleus and to the ten other electrons conjointly, and which, therefore, is not precisely an inverse-square field. It will later transpire that all of the levels in any one column of Figure 10.4 share a single k-value distinctive of that column, the value increasing by one unit from each column to its neighbor on the right. Were the force-field to become a strict inverse-square field, each level of any column would become identical—so far as the energy is concerned—with the corresponding level of every other column: the notation is meant to imply this, $4S$ and $4P$ and $4D$ and $4F$ being " corresponding " levels in this sense. Such a case—the case with which we started—is called " degenerate." When an alteration in the force-field does away with the peculiar situation in which so many levels have coincident energy values, it is said to "remove the degeneracy."

The foregoing " alterations " in the force-field should really have been called "departures from the inverse-square field," being intrinsic to the atom. Alterations, in the strict sense of the word, are made by the experimental physicist when he applies electric or magnetic fields to a gas engaged in absorbing or in emitting light. These fields, if strong enough, produce remarkable alterations in the spectrum, "splitting of lines;" and

these splittings in turn can be traced back by the theorist to separations of levels which were of identical energy before the field was applied. The generic names for the effects produced by applied electric and magnetic fields respectively are "Stark effect" and "Zeeman effect." The Stark effect is one of the greatest of the subjects excluded by limitation of length from this book; of it I shall say only that Epstein of Sommerfeld's school, in a celebrated paper written as early as 1916, explained it with great success, and it thereupon became one of the strongest supports of the Bohr-Sommerfeld atom-model and has served in the same role to the succeeding forms of quantal theory. The Zeeman effect will occupy us long in the next chapter but one. Now however we revert to the unmodified inverse-square force-field, and make the acquaintance of some of the doctrines of wave-mechanics.

CHAPTER XII

Introduction to Wave-Mechanics

I REPEAT that in its earliest form the nuclear atom-model offered nothing to represent the distinctive stationary states with their peculiar discrete energy-values. Instead it offered an infinity of orbits corresponding to a continuous range of permissible energy-values (I pass over the difficulty about radiation, so as to concentrate attention upon this one). Bohr imposed discontinuity upon that model, by fiat as it were; he imposed the condition that the angular momentum should have discrete distinctive values which he specified, he recognized no orbits but those which had such p-values, and he derived the discrete energy-values which experience confirms. He imposed discontinuity of angular momentum upon the atom-model by postulate, and he deduced discontinuity of energy. But is it possible to penetrate to some principle more fundamental, from which discontinuity of energy and of angular momentum are found to be implicit?

Such a principle was discovered by Schroedinger, from a study of the correlation of corpuscles with waves which L. de Broglie had formulated, and for which we have already witnessed some of the clearest evidence in Chapter VI of this book.

There is a certain degree of analogy—not perfect, but good enough to be helpful—between Schroedinger's principle, and that which leads to resonance in the phenomena of acoustics. An unbounded expanse of air will carry sound-waves of any frequency whatever, within a broad continuous frequency-range; no one frequency or wave-length is favored above its immediate neighbor; but if some portion of that air is enclosed in an organ-pipe, a box, a room, or any cavity with solid walls,

certain discrete frequencies are favored much beyond all the others. An endless stretched cord, if we could attain to such a one, would transmit a transverse vibratory motion of any frequency within a wide continuous range; but actual stretched cords have ends which are fixed, and this condition causes them to favor certain frequencies, a fundamental and its harmonic overtones. The continuous infinite substance has no preference for any frequency over any other, but once let it be limited or bounded, and it acquires " resonance frequencies " or " natural frequencies " depending on the boundary-conditions. A membrane tensed across a ring like a drumhead has natural frequencies, and these turn out to be due not only to the limitation which is caused by the fixing of its circular periphery, but also to the mere fact that the membrane has circular symmetry. An elastic sphere has natural frequencies due in part to its boundary and in part to its symmetry.

The waves which we considered in Chapter VI are associated with the corresponding particles, whatever they may be—electrons, for the purposes of this chapter—by Rules of Correlation, which I repeat:

$$p = h/\lambda \quad E = h\nu. \tag{12.1}$$

It will now be affirmed that a stationary state of energy E is correlated with a natural frequency ν according to the second of these Rules: the natural frequencies of an atom being calculable, in a way which shows them analogous to the natural frequencies of acoustics to a sufficient extent to justify them in bearing the same name.

In what follows, I first review the derivation of the natural frequencies of a stretched string, a stretched membrane and a ball of elastic fluid—examples drawn from acoustics; then make the first application of wave-mechanics to the simplest of possible cases, a model which has not yet figured in this book but is of great importance in molecular theory and elsewhere; and finally treat by the same technique the atom-model for hydrogen and ionized helium, of which we have just interrupted our study by the original scheme of Bohr.

Example of the stretched string

Imagine a stretched string, infinitely long, extended along the x-axis of a system of coordinates. Designate the tension in the string by T, the (linear) density of the string by ρ. To derive the differential equation governing the motion, conceive the string as a succession of short straight segments (Figure 12.1). Each segment exerts upon its neighbors a force, which is the tension in the string. When the string lies straight along the axis of x, each segment lies in equilibrium between the equal and opposite forces which its neighbors exert upon it. When, however, the string is drawn sidewise (remaining, we shall suppose, in the xy-plane) the neighbors of each segment are oblique to it and to one another, the forces which they exert upon it have components along the y-direction. These components are in general unequal, and their algebraic sum is a force urging the segment along the y-direction. Denote by dx the length of such a segment, by y its lateral displacement, by θ the angle between it and the axis of x; so that $dy/dx = \tan\theta$, and ρdx stands for the mass of the segment. The resultant force upon the segment is given by:

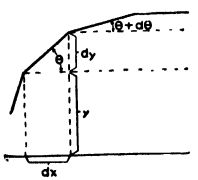


FIGURE 12.1—ILLUSTRATING THE EXAMPLE OF THE STRETCHED STRING.

$$F = T[\sin(\theta + d\theta) - \sin\theta] = T[\tan(\theta + d\theta) - \tan\theta]$$
$$= T \cdot d(\tan\theta/dx)dx = T(d^2y/dx^2)dx \qquad (12.2)$$

to the degree of approximation to which the difference between $\sin\theta$ and $\tan\theta$ may be neglected.[1]

Equating this to the product of mass by acceleration, we obtain:

$$\rho d^2y/dt^2 = T(d^2y/dx^2), \qquad (12.3)$$

[1] This is the degree of approximation all but universal in the theory of vibrating systems and sound. The conclusions from this theory are therefore strictly valid only in the limit of infinitesimal displacements or distortions.

or using dots to symbolize differentiation with respect to time, and dashes to represent differentiation with respect to space:

$$\ddot{y} = T/\rho \cdot y''. \tag{12.4}$$

This equation, a linear combination of a second derivative with respect to time and a second derivative with respect to space, is the first and simplest of our wave-equations.

It is called a wave-equation, because it may represent—it does not necessarily represent, but it may—a shape or a figure or a distortion of the string (whichever one may choose to call it) which travels continually and indefinitely along the string with a constant speed.

To illustrate this possibility, let us suppose that at the time $t = 0$ the string is distorted into a sinusoidal curve described by the equation:

$$y = A \sin mx \quad \text{at} \quad t = 0 \tag{12.5}$$

and that its points are moving parallel to the y-axis with speeds described by the equation:

$$\dot{y} = nA \cos mx \quad \text{at} \quad t = 0. \tag{12.6}$$

At any other moment t, the configuration of the string is described by the equations:

$$y = A \sin (nt + mx), \quad \dot{y} = nA \cos (nt + mx), \tag{12.7}$$

for these satisfy the differential equation which underlies the whole theory, and they satisfy also the "initial conditions" specified by (12.5) and (12.6). They satisfy these equations, that is to say, provided that a certain relation is fulfilled among the constants n and m, and the quantities T and ρ which describe the physical nature of the stretched string; this relation being:

$$n/m = \sqrt{T/\rho}. \tag{12.8}$$

If this relation is fulfilled, the condition of the string throughout all time is described by the equations (12.7).

Examining these equations, we perceive that they signify that the values of displacement and speed, which at the time

$t = 0$ existed at any point x_0 on the string, are at any other time t to be found at the point $x_1 = x_0 - (n/m)t$. These values are moving steadily along the string; the whole configuration of the string, its sinusoidal shape and its transverse velocities, is slipping steadily lengthwise in the direction of decreasing x—the shape of the string is being transmitted as a wave, with the ratio of the constants n/m for its speed of propagation u:

$$u = n/m = \sqrt{T/\rho}. \qquad (12.9)$$

This result justifies the title " wave-equation " for the differential equation (12.4), and the meaning " speed of propagation " or " wave-speed " or " phase-speed " for its coefficient $\sqrt{T/\rho}$.

The reader will scarcely have failed to notice, however, that the result was obtained only by prescribing very sharply defined physical conditions. The string was supposed infinitely long; it was supposed distorted into the form of a sine-wave; the transverse speeds of its successive particles at the instant $t = 0$ were preassigned as rigorously as their positions. Were we to alter this last specification, we should arrive at very different results. If for instance we should make the assumption that at $t = 0$ the string is distorted into a sine-wave *and is stationary*, the equations (12.7) would not be adequate to describe what happens. We should then be forced to have recourse to a more general solution of the differential equation:

$$y = C \sin (nt + mx) + D \sin (nt - mx) \qquad (12.10)$$

and to adjust the constants C and D so as to conform to the newly prescribed initial conditions, which are:

$$y = A \sin mx, \quad \dot{y} = 0 \quad \text{at} \quad t = 0. \qquad (12.11)$$

The adjustment is attained by making $C = D = \frac{1}{2}A$, whereupon we get:

$$y = A \sin nt \cos mx, \qquad (12.12)$$

an equation which describes not a wave advancing perpetually

along the string, but a stationary oscillation with nodes and loops of vibration, like those which a violin-string properly bowed exhibits, those in the air-column of Kundt's tube which the hillocks of dust reveal. One would hardly detect by instinct in this stationary wave-pattern the superposition of two oppositely gliding wave-trains each traveling with the speed $u = n/m = \sqrt{T/\rho}$. Yet the one is always equivalent to the other, and in the equation (12.12), the coefficients n and m are linked to one another through the wave-speed characterizing the string, and the equation may be written

$$y = A \sin umt \cos mx, \quad u = \sqrt{T/\rho}. \qquad (12.13)$$

From now on I discontinue the use of n in the foregoing sense (frequency times 2π) as it will be urgently needed for other purposes.

Although the tension and the density of the string thus determine the frequency when m is preassigned (or vice versa), nothing so far brought upon the scene compels any limitations upon the coefficient m. The infinitely long wire can sustain vibrations of any wave-length, or vibrations of two or any number of wave-lengths simultaneously, with any interrelation whatever among their several amplitudes and phases. On this fact rests our freedom to impose any initial conditions whatsoever on such a wire (subject to the usual restrictions of continuity and finiteness). For, if it be demanded that at $t = 0$ the displacement y shall vary along the wire according to any totally arbitrary function $f(x)$, and the transverse speed y according to any totally arbitrary function $g(x)$, then we have only to expand these functions f and g into Fourier series, or if need be, Fourier integrals; and each term in such an expansion corresponds to such a solution as (12.10), with a specific value of m and such specific values of C and D as the initial conditions require; and the configuration of the wire forever before and after is described by the sum of all these solutions. In such a case we should not see an unchanging distortion of the wire slipping steadily along its length with a con-

stant speed, nor a stationary pattern of nodes and loops. All the obvious features of wave-motions would be blotted out; and yet the infinitely complicated and variable figure of the string would be equivalent, in the last analysis, to a multitude of sinusoidal wave-trains perpetually gliding to and fro with the same uniform speed.

As soon, however, as we imposed *boundary-conditions*, the vibrations which the string can execute are severely restricted.

As a simple and familiar example of boundary-conditions, I will assume that the string is clamped at the points $x = 0$ and $x = L$, and concern myself only with the finite length of string, L, comprised between these two fixed extremities.

As a preparation for future developments, it is advisable to restate the underlying differential equation, and solve it *ab initio*. We have:

$$\ddot{y} = u^2 y'', \tag{12.14}$$

in which u stands for the speed of propagation of a sine-wave along an infinite wire. We essay a tentative solution, in the form of a product of a function of t only by a function of x only:

$$y = g(t) \cdot f(x). \tag{12.15}$$

The differential equation subjects the functions g and f to the condition:

$$f''/f = \ddot{g}/u^2 g = -m^2, \tag{12.16}$$

for, since the first member of this triple equation does not depend on t, and the second does not depend on x, each of the two must be independent of both t and x, and equal to a constant which (for the sake of consistency with prior notation) I denote by $-m^2$. Solutions of these differential equations into which the underlying one was broken up are these:

$$f = A \cos mx + B \sin mx, \quad g = C \cos mut + D \sin mut. \tag{12.17}$$

So far, there is no limitation upon m.

Now come the boundary-conditions, formulated thus:

$$f(0) = f(L) = 0. \tag{12.18}$$

We have now encountered, in its simplest example, the peculiar and characteristic problem of the Theory of Acoustics, which supplies the best of the classical analogues for the peculiar and characteristic problem of the type of Atomic Theory which is embodied in wave-mechanics. This is not the question which we meet in the theory of moving particles, where we are asked what path a particle will follow through all future time if its position and velocity at a single moment are given.

To adjust the function $f(x)$ to the boundary conditions, it is evident that we must set $A = 0$ and $\sin mL = 0$; therefore we must assume that m has one of the values:

$$m = k\pi/L \quad k = 1, 2, 3, 4 \ldots . \tag{12.19}$$

The boundary-conditions have compelled the coefficient m to choose among a rigidly defined series of values. The wave-lengths, and consequently the frequencies, of the permitted vibrations are strictly determined.

The permitted values of m may be called, in anticipation of the quantal usage, the " eigenvalues " of the differential equation for the boundary-conditions in question. To each corresponds a value of the vibration-frequency $mu/2\pi$; it is a resonance-frequency or a natural frequency.

To each eigenvalue corresponds also a solution of the differential equation, which will be called an " eigenfunction." Here the eigenfunction corresponding to $m = k\pi/L$ is the following:

$$y_k = \sin \frac{k\pi}{L} x \left(C_k \cos \frac{k\pi u}{L} t + D_k \sin \frac{k\pi u}{L} t \right). \tag{12.20}$$

It represents a sinusoidal stationary oscillation of the wire, with nodes at the ends and at $(k - 1)$ points spaced evenly between the ends—a case not difficult to realize with a violin-string, if k be not too great. The constants C and D specify the amplitude of the oscillation, and its phase at any given instant.

It is, of course, not necessary that the motion of the wire

should conform to a single eigenfunction. Any number of eigenfunctions, corresponding to different permitted values of m—different integer values of k—might coexist simultaneously, each with its particular values of C_k and D_k; the actual distortion of the wire would be the superposition of all. It would in fact be necessary to adjust the initial distortion of the wire and the initial velocities of its points with infinite accuracy, to cause its future motion to conform exactly to a single eigenfunction. On the other hand, any choice whatever of initial distortion and initial velocities would entail a future motion compounded out of the various eigenfunctions with suitable values of C_k and D_k, which could be computed.

Example of the tensed membrane

The differential equation of the tensed membrane is:

$$\nabla^2 z = \frac{d^2z}{dx^2} + \frac{d^2z}{dy^2} = \frac{1}{u^2}\frac{d^2z}{dt^2}. \qquad (12.21)$$

The coordinate-axes of x and y lie in the equilibrium plane of the membrane, and z stands for the displacement of any point of the membrane normally from this plane. The symbol u stands for the speed of a sine-wave traveling in an infinite membrane of the same tension T and surface-density ρ as the actual one, and is determined by the equation:

$$u^2 = T/\rho, \qquad (12.22)$$

which is derived by an obvious extension of the method employed in deriving the like equation for a stretched string. In an actual bounded membrane the motion may be tremendously complicated, but it can be analyzed into a multitude of wave-trains traveling to and fro with the speed u.

The symbol ∇^2 (to be read *del* or *nabla-squared*) stands for the Laplacian operator which in rectangular coordinates is d^2/dx^2, or $(d^2/dx^2 + d^2/dy^2)$, or $(d^2/dx^2 + d^2/dy^2 + d^2/dz^2)$, according as we are dealing with one, two or three dimensions. In other coordinates than rectangular, it naturally assumes

other forms. Now in these problems of two and three dimensions, the choice of coordinate-system and the imposition of boundary-conditions are two decisions which cannot be separated from one another. Were we to decree that the membrane should be square or rectangular with its edges clamped, the suitable coordinate-system would be the rectangular. The problem would then be extremely simple (the reader can easily solve it for himself by using the method adopted for the stretched string, and will arrive at very similar results) but not so instructive to us as the problem of the circular membrane with clamped edge. For this we must adopt polar coordinates (with the origin at the center of the membrane, naturally). In these, the Laplacian operator assumes the form:

$$\nabla^2 = \frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} + \frac{1}{r^2}\frac{d^2}{d\theta^2}. \qquad (12.23)$$

We restate the fundamental differential equation (12.21) in this fashion; we essay a tentative solution in the form of a product of a function $f(r)$ of r exclusively, a function $F(\theta)$ of θ exclusively, and a function $g(t)$ of t exclusively; and we discover as before that each of these functions is subjected to a differential equation of its own. The procedure is like that already used in the case of the stretched string clamped at its ends. First we have

$$\frac{1}{f}\frac{d^2f}{dr^2} + \frac{1}{rf}\frac{df}{dr} + \frac{1}{r^2F}\frac{d^2F}{d\theta^2} = \frac{1}{u^2g}\frac{d^2g}{dt^2} = -m^2, \qquad (12.24)$$

for, since the first member of this triplet does not depend on t, the second not on r nor on θ, both must be independent of all three variables and equal to a constant which, as before, I denote by $-m^2$. The differential equation for the factor dependent on t has the solution:

$$g(t) = A \cos mut + B \sin mut. \qquad (12.25)$$

Our experience with the stretched string suggests that m will be restricted to certain eigenvalues, derived from the boundary-conditions; and this is true; but before arriving at these, we

must attend to the differential equation governing the functions f and F. This assumes the form:

$$\frac{r^2}{f}\frac{d^2f}{dr^2} + \frac{r}{f}\frac{df}{dr} + m^2r^2 = -\frac{1}{F}\frac{d^2F}{d\theta^2} = k^2, \qquad (12.26)$$

both members of the equation being, by the familiar reason, equal to a constant which I denote by k^2. It follows that the function $F(\theta)$ is of the form:

$$F(\theta) = C\cos k\theta + D\sin k\theta \qquad (12.27)$$

and the coefficient k thus far seems to be unrestricted. But it carries its own restrictions in itself; for the coordinate θ is a cyclic coordinate, like longitude on the earth; whenever it is altered by 2π, we are back at the same place. The function $F(\theta)$ must therefore repeat itself whenever θ is altered by 2π; but this will not occur, unless k is an integer:

$$k = 0, 1, 2, 3 \ldots . \qquad (12.28)$$

These are the eigenvalues, and the functions (12.27), with one or another of these values assigned to λ are the eigenfunction of the equation (125). In this case we have obtained eigenvalues for the parameter and eigenfunctions for the solutions of a differential equation, not out of boundary conditions but out of the simple fact that the independent variable is by its nature cyclic. Such cases will occur in the undulatory mechanics, and in particular when we investigate rotating systems, for rotation also is cyclic.

We arrive at the third and last step of the problem: the determination of the function $f(r)$. It is governed by the differential equations:

$$\frac{d^2f}{dr^2} + \frac{1}{r}\frac{df}{dr} + \left(m^2 - \frac{k^2}{r^2}\right)f = 0, \qquad (12.29)$$

a distinct equation for each of the permitted integer values of k. As the solution of such an equation as (12.16) is a sine-function of the variable mx, so the solution of such an equation as (12.28) is a function of the variable mr; not, however, a sine-function, but a Bessel function. For the values 0, 1, 2, . . .

of k, the solutions of (12.28) are the Bessel functions of order 0, 1, 2, . . ., denoted by $J_0(mr)$, $J_1(mr)$, $J_2(mr)$, and so forth. (Equation (12.29) also admits of solutions which are infinite at $r = 0$, but these we discard.)

Like the sine-function of mx, the Bessel functions of mr oscillate back and forth between negative and positive values as their variable increases from zero to infinity, and pass through zero at an infinite number of discrete values of mr. These do not lie at equal intervals, as do the values of mx at which sin mx vanishes. Their values may be found in the tables; I shall designate them as b^1, b^2, b^3, . . . in order of increasing magnitude, using the superscripts not as exponents, but as ordinal numbers so that I may reserve the subscripts to distinguish the various Bessel functions from one another. The function

$$Z = J_k(mr)(C \cos k\theta + D \sin k\theta)(A \cos mut + B \sin mut) \quad (12.30)$$

represents a stationary oscillation of an infinitely extended membrane, in which k lines intersecting one another at the origin are nodal lines, and an infinity of concentric circles centered at the origin are nodal circles. These lines and circles are motionless while the sections of the membrane which they delimit vibrate with the frequency $mu/2\pi$. The k lines are spaced uniformly in angle; the radii r_1, r_2, . . . of the infinity of circles are obtained by dividing m into the roots b_k^1, b_k^2, b_k^3, . . . of the Bessel function of order k, $J_k(mr)$.

How then does the boundary-condition upon the finite membrane enter in? Obviously, if a membrane of radius R be clamped at its edge, and if it is vibrating in the manner described by (12.30), then the edge must coincide with one of the nodal circles; the radius R must be equal to one of the quantities b_k^i/m. Or rather, since the nodal circles are to be adjusted to the size of the diaphragm and not the size of the diaphragm to the nodal circles, the coefficient m must conform to one of the equations:

$$m = b_k^1/R, \text{ or } b_k^2/R, \text{ or } b_k^3/R, \ldots \quad (12.31)$$

These equations define eigenvalues of the parameter m *in the differential equation of the tensed membrane.* There is a double infinity of these—an infinite series of them for each of the eigenvalues of the parameter k. To each corresponds a natural frequency of the membrane, and to each corresponds an eigenfunction, the one written down in (12.30) with the proper value of m taken from (12.31). The constants A, B, C, and D in the eigenfunctions specify the amplitude of the oscillation, the phase of the vibrations at any given instant, and the orientation of the nodal lines with respect to any given axis. Any number of eigenfunctions may coexist simultaneously; the actual distortion of the membrane will be the superposition of all. Any initial conditions imposed on z and $\dot{z}$ (and not involving discontinuities or infinities) could be satisfied by adjusting the constants.

Example of the ball of fluid

Among the familiar vibrating systems, the ball of fluid presents the closest analogy to the atom-model for the hydrogen atom in wave-mechanics, the wave-patterns in the two cases being strikingly alike.

In three dimensions and in polar coordinates (those appropriate to the boundary-conditions which we shall impose) the wave-equation assumes the somewhat alarmingly intricate form:

$$\ddot{\Psi} = u^2\nabla^2\Psi = u^2 \frac{\operatorname{cosec}\theta}{r^2}\left[\frac{d}{dr}\left(r^2 \sin\theta \frac{d\Psi}{dr}\right) + \frac{d}{d\varphi}\left(\operatorname{cosec}\theta \frac{d\Psi}{d\varphi}\right) + \frac{d}{d\theta}\left(\sin\theta \frac{d\Psi}{d\theta}\right)\right]. \tag{12.32}$$

The argument Φ can no longer be visualized as a displacement perpendicular to the equilibrium-position of the undistorted medium, since all three dimensions are already used up. The reader may visualize it, if he will, as a condensation or a rarefaction, after the fashion of sound-waves. Perhaps not to visualize it at all would be a better preparation for the study of wave-mechanics.

In the familiar way, we essay a solution in the form of a product of a function of time $g(t)$, a function of radius $f(r)$, a function $\Phi(\phi)$ of the longitude-angle ϕ and a function $\Theta(\theta)$ of the colatitude-angle θ. As before, we find that the time-function is of the form:

$$g(t) = A \cos mut + B \sin mut \qquad (12.33)$$

and, as before, we shall find that the boundary-conditions confine the coefficient m and the frequency $mu/2\pi$ to certain "permitted" values. The angle-functions and the radius-function are governed by the differential equations:

$$\frac{1}{f}\left[\frac{d}{dr}\left(r^2\frac{df}{dr}\right) + m^2r^2f\right]$$
$$= -\frac{1}{Y}\operatorname{cosec}\theta\left[\frac{d}{d\varphi}\left(\operatorname{cosec}\theta\frac{dY}{d\varphi}\right) + \frac{d}{d\theta}\left(\sin\theta\frac{dY}{d\theta}\right)\right] = k, \qquad (12.34)$$

in which Y stands for the product of Θ and Φ, and k for a constant which seems to be arbitrary, but as a matter of fact is constrained by the same circumstance as arose in the case of the membrane; for, whenever φ is altered by 2π and θ by π, we are back at the same place as before, and the function Y must have the same value as before; and this will occur only if

$$k = n(n+1), \quad n = 0, 1, 2, 3 \ldots, \qquad (12.35)$$

these being the eigenvalues for the differential equation in (12.34) for the angle-function.[1] The corresponding eigen-functions are spherical harmonics. To each value of n belongs a "spherical harmonic of order n," which itself is a sum of $(2n+1)$ terms, each multiplied by a constant which is at our

[1] This and the following statements about the functions Y_n are proved by writing Y in the second of equations (12.34) as the product of a function of θ and a function of ϕ, and so dissolving the equation into two in the manner which I have already used five or six times; the values of the constant s in equation (12.36) are the eigen-values of the latter of these two. I thought it desirable not to overload the exposition by carrying through all stages of the process of solution, especially as the splitting of $Y_n(\theta, \phi)$ into the two functions is of secondary importance in the atom-model to which all this leads up; nevertheless the reader may find it advantageous to supply the lack.

disposal and can be adjusted to fit initial conditions or to emphasize particular modes of vibration. These terms are products of sine-functions of φ by peculiar functions, the *Legendrian functions* $P_{n,\,s}$, of the variable θ; so that the eigenfunction for a permitted value $n(n+1)$ of the parameter k has this for its most general form:

$$Y_n(\theta, \varphi) = a_{n,\,0}P_{n,\,0}(\cos\theta) + \sum_{s=1}^{n} a_{n,\,s}\cos(s\varphi)P_{n,\,s}(\cos\theta) + \sum_{s=1}^{n} b_{n,\,s}\sin(s\varphi)P_{n,\,s}(\cos\theta). \quad (12.36)$$

Each term by itself describes a particular mode of vibration of the fluid; the sum represents a superposition of divers modes of vibration. If we isolate one of these modes by giving to n some particular value n_1, and to s some particular value s_1, and causing all the constants a and b in (12.35) to vanish except $a_{n_1,\,s_1}$ and $b_{n_1,\,s_1}$; we then find that Y, and consequently Φ, and consequently the motion altogether, vanishes at s_1 values of φ and at $n_1 - s_1$ values of θ. If we draw a sphere centered at the origin, we find that its surface bears s_1 nodal meridian-circles, and $n_1 - s_1$ nodal latitude-circles, along which there is perpetual rest. If we consider all the spheres at once—if, that is to say, we consider the entire volume of the fluid medium—we see that when the fluid is vibrating in the mode distinguished by the integers (I had almost said "quantum-numbers"!) n_1 and s_1, it is divided into compartments by s_1 nodal planes intersecting along the axis $\theta = 0°$, and $n_1 - s_1$ double-cones having that axis for their axis and the origin for this apex.

We have not yet considered the dependence of the wave-motion on the radius r; but the close analogy between this and the corresponding stage of the problem of the tensed membrane will make the task easy. The differential equation (12.34) for $f(r)$ resembles Bessel's equation (12.29), and has the somewhat similar solution

$$f(r) = \frac{1}{\sqrt{r}} J_{n+\frac{1}{2}}(mr). \quad (12.37)$$

This function vanishes, entailing the vanishing of the wave-motion, at an infinity of discrete values of the variable mr:—the roots of the function, which I denote in order of increasing magnitude by B^1, B^2, B^3, . . . In an infinite medium we could assign any value whatever to r, and then there would be an infinity of nodal spheres, their radii given by B^1/m, B^2/m, B^3/m, . . . If the medium is bounded by a rigid spherical wall of radius R, the coefficient m must possess one of the values B^i/R, so that one of the nodal spheres may coincide with the wall. These are the eigenvalues of the constant m, and the natural frequencies of the corresponding vibrations are given by $B^iu/2\pi R$. The eigenfunctions are given by the equation (12.37) with the various values B^i/R substituted for the parameter m.

The eigenfunctions of the fundamental differential equation for the fluid sphere are, therefore, each a product of a radius-function given by (12.37), with a "permitted" value for the constant m determined by the boundary-condition; an angle-function given by (12.36), with "permitted" values for the constants n and s, determined by the fact that the angles are cyclic variables; and a time-function given by (12.33), with a "permitted" vibration-frequency determined by the boundary-condition. Each eigenfunction with the indices m, n, s describes a mode of vibration, in which the fluid sphere is divided into compartments by s meridian planes, $(n - s)$ double-cones, and a certain number of spheres, upon each of which the fluid is perpetually at rest; within the compartments, it vibrates with a prescribed frequency.

Case of a "string" for which the wave-speed is variable, or even imaginary

Thus far I have used the images of the stretched string, the tensed membrane, and the elastic fluid to illustrate the behavior of the differential equation

$$u^2\nabla^2\Psi = d^2\Psi/dt^2, \qquad (12.38)$$

when the coefficient u^2 is a positive constant. In these exam-

ples u^2 is interpreted as the ratio of the intrinsically positive quantities "tension" (or "pressure") and "density," and turns out to be equal to the square of the speed of propagation of sine-waves in the string, membrane, or fluid. In certain problems of undulatory mechanics we encounter just such an equation. In some of the most important applications of Schroedinger's theory, however, one meets with differential equations of the type of (12.38), in which, however, the coefficient u^2 depends on the coordinates and even assumes negative values! Such equations need not be more difficult to solve than the conventional wave-equation in which u^2 stands for a positive constant; but the image of the elastic medium becomes unsatisfying. In the one-dimensional case, so long as u^2 remains a positive function of x, one can visualize a string of which the density varies along its length; but when u^2 passes through zero and becomes negative, the wave-speed attains zero and is superseded by an imaginary quantity. One may speak, in such a case, of a "string" or a "fluid" characterized by an "imaginary wave-speed." So speaking, one comes perilously close to the verge of using words devoid of physical meaning; but otherwise, there is no verbal language with which to relieve the monotony of the procession of equations.

The differential equation of the type of (12.38), with a constant negative value of the coefficient u^2, is not a difficult one. Confining ourselves to one dimension, we find for one of the solutions of the equation for a "string with constant imaginary wave-speed" this expression:

$$\Psi = (A \cos mUt + B \sin mUt)(Ce^{mx} + De^{-mx}), \qquad (12.39)$$

in which U stands for the (real) square root of $-u^2$. This is a much less tractable function than the product of sine-functions which serves when u^2 is positive. One cannot, for instance, find eigenvalues for the constant m whereby the function can be made to vanish at all times at two distinct points upon the "string"; or rather, one can find only the value $m = 0$, which fulfils this familiar boundary-condition by destroying the function. Similarly, one cannot force Ψ to remain finite

everywhere except by annulling either m or else both A and B, again destroying the function. Vibrations which are sine-functions of time are, however, permitted by the differential equation.

Consider now the equation

$$d^2y/dx^2 = (a - bx^2)d^2y/dt^2, \qquad (12.40)$$

which may be regarded as the wave-equation of a string of which the wave-speed varies with x along its length as the function $(a - bx^2)^{-\frac{1}{2}}$, being therefore real over the central part from $x = -\sqrt{a/b}$ to $x = +\sqrt{a/b}$, and imaginary from each extremity of this central range outward to infinity. In the usual way, we derive the equations:

$$y = f(x)g(t), \qquad g = A\cos 2\pi\nu t + B\sin 2\pi\nu t,$$

$$d^2f/dx^2 + 4\pi^2\nu^2(a - bx^2)f = d^2f/dx^2 + (C - x^2)f = 0. \qquad (12.41)$$

It is no longer expedient to write $mu/2\pi$ for the frequency, since m would be a function of x; I have written ν instead. It is now our problem to solve the equation [1] for $f(x)$.

Essay a solution in the form of a power-series, multiplied by $e^{-(\frac{1}{2})x^2}$:

$$f(x) = e^{-(\frac{1}{2})x^2}\sum_{n=0}^{\infty} a_n x^n. \qquad (12.42)$$

Substitute this into the differential equation, and group all the terms involving the same power of x. For each such group, we have

$$a_{n+2}(n + 1)(n + 2)x^n - a_n(2n + 1 - C)x^n, \qquad (12.43)$$

and equating each group separately to zero, we arrive at the relation

$$a_{n+2}/a_n = (2n + 1 - C)/(n + 1)(n + 2). \qquad (12.44)$$

Put $a_0 = 0$, thus causing all the even-numbered coefficients to vanish; assign any arbitrary value to a_1, and calculate the odd-numbered coefficients a_3, a_5, a_7, and so onward. Or, put a_1

[1] The constant $\nu^2 b$ has been equated to unity, a procedure which entails no loss of generality.

and all the odd-numbered coefficients equal to zero, assign any arbitrary value to a_0, and calculate the even-numbered coefficients a_2, a_4, a_6, and so onward. Either way we shall get a solution of (12.41), whatever the value of the parameter C; but there are certain specific values of C which admit a peculiar sort of solution. It is, in fact, evident from (12.43) that we shall arrive at two entirely distinct results, according as C is or is not equal to some value of $(2n + 1)$—according, that is to say, as C is, or is not an odd integer. For, if C is equal to an odd integer $(2n + 1)$, the chain of coefficients will come to an abrupt end at the member having that particular value of n; it and all the succeeding members will be zero; the power-series in the tentative (and adequate) expression (12.42) for the unknown function $f(x)$ will consist of a finite number of terms. But, if C is not equal to an odd integer, the power-series will go on forever.

Here we have a new kind of eigenvalue. If the parameter C, in the differential equation for the curious kind of "string" which I have just defined, has for its value one of the numbers:

$$C = 2n + 1, \qquad n = 0, 1, 2, 3, 4, \ldots, \qquad (12.45)$$

the equation enjoys a special sort of solution. If the parameter does not have one of these eigenvalues, the solution of the differential equation is altogether different.

Let us see what difference these eigenvalues make in the general solution (12.42) of the differential equation. If the parameter C had some other value than one of these, the series $a_n x^n$ goes on forever and as x approaches infinity, the value of its summation increases at such a rate as to overwhelm the steadily declining factor $e^{-(\frac{1}{2})x^2}$, so that the function $f(x)$ is infinite at both ends of the range $-\infty < x < \infty$. If, however, C is equal to one of the eigenvalues, the series $a_n x^n$ comes to an abrupt end; and as x approaches infinity, the decline of the factor $e^{-\frac{1}{2}x^2}$ overpowers the increase of the summation, and $f(x)$ remains finite at infinity. *The values* 1, 3, 5 . . . *of the constant C are therefore the eigenvalues which permit solu-*

tions which remain finite all through the range of values of the independent variable from zero to infinity.

The corresponding eigenfunctions are,

$$f_m(x) = e^{-\frac{1}{2}x^2} H_m(x), \qquad (12.46)$$

the symbol $H_m(x)$ standing for the finite series $\Sigma a_n x^n$ constructed according to the rules of the foregoing paragraphs, and terminating at the mth term. These [1] are known as the "polynomials of Hermite."

It is evident that the analogy between this curious pseudo-string, and the real mechanical string of pages 342–348, is far from being exact. There, the wave-speed was real and constant all along the string, and independent of frequency; and the clamping or immobilization of the string, at two points a fixed distance apart, determined the natural frequencies. Here the wave-speed varies along the pseudo-string and in how singular a manner! it is indeed imaginary, for part of the way; but this peculiar mode of variation is capable of determining a natural frequency, without any clamping or other geometrical condition being needed. Whether or not it determines a natural frequency depends upon its exact form, *i.e.* upon the value of the constant C which controls in detail the mode of variation. One might say that there are as many distinctive pseudo-strings as there are distinctive values of C and distinctive natural frequencies, each having its own peculiar dependence of u upon x. One may also say that the pseudo-string has a wave-speed depending both on x and on ν, and this makes the analogy seem rather better. At best, however, it is an analogy which cannot be made complete.

Schroedinger's wave-equation, being a non-relativistic wave-equation of quantal theory.

We now take from (6.26), or from elsewhere in this book, the expression for the wave-speed of the waves associated with

[1] The first five are written down by Schroedinger (**AP** 79, 515; 1927). An arbitrary numerical multiplier remains at disposal.

particles of energy E and momentum p, and introduce it into the wave-equation typified by (12.32):

$$\frac{\partial^2 f}{dt^2} = (E^2/p^2)\nabla^2 f. \qquad (12.47)$$

I repeat the Rules of Correlation in order to point out that we may not use *every* pair of values of ν and λ having the correct product E/p, but only the particular pair:

$$\nu = E/h, \qquad \lambda = h/p. \qquad (12.48)$$

Next, I point out, for the sake of completeness, the self-evident fact that if p as well as E is constant—if the particles are in a region where they suffer no forces and no refraction—then u is constant, and one of the solutions of the equation is,

$$f = A \cos \frac{2\pi}{h}(Et - px), \qquad (12.49)$$

the description of the ideal sinusoidal wave train which we have already correlated with beams of photons or of electrons advancing with uniform energy along the axis of x.

If, however, the particles are in a region pervaded by a field of force, u will depend on the coordinates, and the situation will be drastically changed; how drastically, we are now about to learn. Suppose the field derivable from a potential-function, such that the potential energy of any particle may be represented by $V(x, y, z)$. Using ordinary as distinguished from relativistic mechanics, we have:

$$E = \tfrac{1}{2}mv^2 + V = \tfrac{1}{2}(p^2/m) + V, \quad p = \sqrt{2m(E - V)}, \qquad (12.50)$$

substituting which into (12.47), we get:

$$d^2f/dt^2 = (E^2/2m(E - V))\nabla^2 f. \qquad (12.51)$$

This is one of the wave-equations of quantal theory.

Another, also called a wave-equation, is obtained by treating (12.51) as we did its predecessors: putting for f the

product $g(t)\Psi(x, y, z)$ of a function of time only by a function $\Psi(x, y, z)$ of the coordinates only, we find:

$$\ddot{g}/g = (p^2/E^2)(\nabla^2\Psi/\Psi) = \text{const.} \qquad (12.52)$$

Assigning to the constant a negative value denoted by $-4\pi^2\nu^2$, we see that we are now describing a stationary wave-pattern exhibiting the frequency ν and a distribution-of-amplitude conforming to the equation

$$\nabla^2\Psi + 4\pi^2\nu^2(p^2/E^2)\Psi = 0. \qquad (12.53)$$

Putting E/h for ν and availing ourselves of (12.50), we get:

$$\nabla^2\Psi + (8\pi^2m/h^2)(E - V)\Psi = 0, \qquad (12.54)$$

an equation with which we shall have much to do.

Obviously (12.53) is the equation,

$$\nabla^2\Psi + (4\pi^2/\lambda^2)\Psi = 0, \qquad (12.55)$$

with λ expressed as h/p by virtue of the Rules of Correlation. This is a good point of departure for the succeeding argument.

These two wave-equations—(12.51) and (12.54)—are known by the name of Schroedinger who first conceived them and applied them. The latter of the two will serve as basis for so much of the theory as is reproduced in this book.

The simple-harmonic linear oscillator in wave-mechanics

Natural as it would seem to make the first application of wave-mechanics to the case of the hydrogen atom, there is a case which rightly takes precedence; largely indeed because of its unique simplicity, but partly also because of its historical seniority, and partly because of its serviceableness in the theory of molecules. It is that of the "simple-harmonic linear oscillator"—the entity to which Planck first ascribed by postulate at the turn of the century, the power of receiving and retaining and disbursing energy in fixed finite amounts. Planck's oscillator was intrinsically charged, but this property will not as yet enter into our reasoning.

Conceive a particle of mass m, constrained to move along

the x-axis, attracted to the origin by a force $-k^2x$ proportional to its displacement, and consequently prone to oscillate to and fro across the origin with frequency $\nu_0 = k/2\pi\sqrt{m}$. Its potential energy is the following function of x:

$$V = \tfrac{1}{2}k^2x^2 = 2\pi^2 m\nu_0{}^2x^2. \qquad (12.56)$$

The wave-equation (12.54) assumes the form

$$\frac{d^2\Psi}{dx^2} + \frac{8\pi^2 m}{h^2}(E - 2\pi^2 m\nu_0{}^2x^2)\Psi = 0. \qquad (12.57)$$

A simple change of variable ($q = x\cdot 2\pi\sqrt{m\nu_0/h}$) transforms this into the equation (12.41):

$$d^2\Psi/dq^2 + (C - q^2)\Psi = 0; \quad C = 2E/h\nu_0. \qquad (12.58)$$

According to wave-mechanics, *the Stationary States of the linear oscillator are distinguished by the energy-values which cause this equation to have a solution finite at all values of the variable, infinity included.*

These are the values of the constant C which cause the parameter C to take one of the eigenvalues set down in (12.45).

The energy-values of the Stationary States should therefore be

$$E_n = \frac{h\nu_0}{2}(2n + 1),$$

$$= \frac{1}{2}h\nu_0, \frac{3}{2}h\nu_0, \frac{5}{2}h\nu_0, \ldots \qquad (12.59)$$

The successive permitted energy-values of the linear simple-harmonic oscillator of frequency ν_0, the energy-values of its consecutive Stationary States, are therefore specified by wave-mechanics as the products of the fundamental factor $h\nu_0$ by the consecutive "half-integers" 1/2, 3/2, 5/2, and so onward.

The linear simple-harmonic oscillator thus furnishes an instance of what are called "half-quantum-numbers." In most of the earlier theories it was either assumed or inferred that this "Planck" oscillator displayed "whole quantum-numbers"—that its permitted energy-values were the products of $h\nu_0$ by

the successive integers 1, 2, 3, 4, . . . However, in the interpretation of certain features of band-spectra by the assumption that the two atoms of a diatomic molecule vibrate as linear oscillators along their line of centers, the half-quantum-numbers sometimes led to better agreement with experience than did the whole-quantum-numbers.

The eigenfunctions corresponding to the consecutive Stationary States are these:

$$\Psi_n(x) = \text{const} \cdot e^{-2\pi^2 m\nu_0 x^2/h} H_n(2\pi x\sqrt{m\nu_0/h}). \qquad (12.60)$$

Interpretation of the hydrogen atom by wave-mechanics

The hydrogen atom is conceived as a system endowed with the potential energy $V = -e^2/r$. This form for the potential energy, I recall, is obtained by imagining an electron and a nucleus, or more precisely two point-charges $+e$ and $-e$, separated by a distance denoted by r. The image of the electron and the nucleus does not come over explicitly into the new theory; but in spirit it does come over, for the potential-energy-function derived from that image is the basis for the new theory.

Polar coordinates for the wave-equation are imperiously suggested by a potential-energy-function of this form, and consequently it is thus expressed:

$$\frac{E^2}{2m(E + e^2/r)} \nabla^2\Psi = \frac{d^2\Psi}{dt^2}, \qquad (12.62)$$

and putting E/h for the vibration-frequency, we attain

$$\nabla^2\Psi + \frac{8\pi^2 m}{h^2}\left(E + \frac{e^2}{r}\right)\Psi = 0. \qquad (12.63)$$

The resemblance of these equations to those laid down for the ball of fluid is as unmistakable as the resemblance of the wave-equation for a linear oscillator to that of a stretched string. Here we have the case of a fluid in which the wave-speed varies from point to point, according to the law

$$u^2 = E^2/2m(E + e^2/r), \qquad (12.64)$$

and we meet the problem of finding modes of vibration and stationary wave-patterns.

If E is supposed positive, the wave-speed is everywhere real. Boundary-conditions of the usual sorts (*e.g.*, the prescription that the fluid shall be confined within a rigid spherical wall of given radius) might be imposed, and then eigenvalues of the constant E could be calculated, and from these the wave-patterns and natural frequencies of the fluid. If no such boundary-conditions were prescribed, the equation (12.63) could be solved with any value of E.

If E is supposed negative, the whole state of affairs is changed. The wave-speed is now real within the sphere of radius $-e^2/E$, zero over this sphere and imaginary beyond it. This recalls the case of the "string" proposed as an analogy for the linear oscillator, for which the wave-speed was real along its central segment and imaginary from each end of its central segment onwards to infinity. There are important differences: in the present case, the variable r assumes positive values only, and the wave-speed at $r = 0$ is infinite though real.

In the case of the "string" with variable and at some points imaginary wave-speed, we found that the law of variation of wave-speed could be so chosen that the "string" enjoys a natural mode of vibration with a stationary wave-pattern and a natural resonance-frequency. This was done by selecting any of a series of eigenvalues for a parameter of the differential equation. Here we shall do likewise.

Essaying for the function Ψ in (12.63) a solution in the form of a product of a function of θ and φ exclusively by a function of r exclusively, we arrive in the familiar way at differential equations:

$$\operatorname{cosec}\theta\left\{\frac{d}{d\theta}\left(\sin\theta\frac{dY}{d\theta}\right)+\frac{d}{d\varphi}\left(\operatorname{cosec}\theta\frac{dY}{d\varphi}\right)\right\}=-KY. \qquad (12.65)$$

$$\frac{d}{dr}\left(r^2\frac{df}{dr}\right)+\frac{8\pi^2mr^2}{h^2}\left(E+\frac{e^2}{r}\right)f=+KF. \qquad (12.66)$$

The equation (12.65) is identical with the one which we encountered in the case of the ball of fluid. Here, as there,

the fact that the variables θ and φ are cyclic requires eigenvalues of the constant K:

$$K = l(l+1), \qquad l = 0, 1, 2, 3, 4, \ldots \qquad (12.67)$$

Equation (12.66), however, is not the same as the corresponding equation (12.34) of the prior case; here we find the difference between the fluids of actual experience and the "imaginary fluid" which is to serve as material for the atom-model supplied by wave-mechanics for hydrogen.

If in that equation (12.66) one were to assign an arbitrarily chosen negative value to the parameter E, one would usually not be able to find a solution which is finite both at the origin and at infinity. This is the same situation as occurred in the theory of the linear oscillator, where an arbitrary choice of a value for the parameter C would usually have led to a solution implying infinite amplitude at both ends of the "string."

Schroedinger, however, discovered that there is a series of eigenvalues for the parameter E of (12.66), each of which (subject to a limitation to be introduced below) entails a solution which is single-valued, continuous and finite [1] over the entire range of the variable r.

These eigenvalues are no other than the sequence,

$$E_n = -2\pi^2 me^4/h^2 n^2; \qquad n = 1, 2, 3, 4 \ldots, \quad (12.68)$$

the very energy-values found for the stationary states of hydrogen by Bohr with his earlier methods of "quantization" or of selecting permitted orbits! This method of quantization (for Schroedinger called his first paper by the name *Quantisierung als Eigenwertproblem* = quantization as an eigenvalue-problem) has won its spurs by achieving the same formula as, when Bohr attained it, became the point of departure for all subsequent atomic theory.

It is a long business to make the proof unimpeachably, and

[1] The exact and universal formulation of the requisite condition is still a matter of debate among theorists; for the beginner it is perhaps best not to demand more than the termination, after a finite number of terms, of the power-series figuring in (12.42) and (12.70).

to disprove all the contrary possibilities; but I will epitomize the argument as given by Sommerfeld. The following symbols will be used for brevity:

$$A = (8\pi^2 m/h^2)E, \qquad B = (4\pi^2 m/h^2)e^2, \qquad C = -l(l+1),$$
$$\rho = 2r\sqrt{-A},$$

so that equation (12.66), with the admissible choices for the parameter l, assumes the form,

$$\frac{d^2f}{dr^2} + \frac{2}{r}\frac{df}{dr} + \left(A + 2\frac{B}{r} + \frac{C}{r^2}\right)f = 0. \qquad (12.69)$$

For this we essay the solution,

$$f = e^{-\frac{1}{2}\rho}\rho^c \Sigma a_s \rho^s, \qquad (12.70)$$

the summation being taken over all positive integer values of s from zero onward, so that in effect we are trying out the product of an exponential by a power-series beginning with the power c and proceeding by integral steps toward greater s-values. Substituting this into the differential equation, we group by themselves the terms containing each power of ρ, and undertake the annulling of each group separately. The first of the groups is $[c(c+1) - l(l+1)]a_0\rho^{c-2}$; and it vanishes if $c = l$. (It would also vanish if c were equated to $-(l+1)$, but this is excluded by another consideration). Accepting this fixing of c, we set about the annulling of the other groups in turn, and find the same situation as in the case of the "string": the ratios $a_{s+1} : a_s$ are now rendered determinate, according to the equations:

$$[(s+l+1)(s+1) + 2(s+l+1) - l(l+1)]a_{s+1},$$
$$= [s + l + 1 - (B/\sqrt{-A})]a_s. \qquad (12.71)$$

Again there are two distinct results, according as a parameter has or does not have one of a certain rigidly-prescribed sequence of definite values. Usually the series $\Sigma a_s\rho^s$ will go on forever; but if the parameter $B/\sqrt{-A}$ happens to be an integer which is greater by $(l+1)$ than some positive integer S, then the

series will come to a stop when s becomes equal to S. In the former and usual case, the function (12.70) will become indefinitely great as r increases without limit; in the latter and exceptional case, that function will vanish at infinity. The student will see that if l may have any positive-integer value from zero upwards, then the condition for the latter case is precisely (12.68).

Implications of the symbol Ψ

Many probably find it disconcerting that no name has been attached to that which the symbol Ψ denotes; neither displacement, nor condensation, nor electric field strength, nor any other. These I will remind that many of the successes of the wave theories of sound and light are not in any way dependent on the fact that the function which figures in their wave equations is called condensation in the one case, electric (or magnetic) field strength in the other. The explanation of diffraction and interference of sound or light is based on (*a*) a suitable wave equation, in which the dependent variable is denoted by any convenient symbol, and (*b*) a relation between the squared amplitude of the dependent variable and that-which-is-observed, loudness of sound and brightness of light, respectively. To explain reflection and refraction at a surface of discontinuity, it is necessary to have in addition (*c*) a set of boundary conditions. When the wave theory of sound or light is " verified " by observations on diffraction, interference, refraction or reflection, it is the combination of the wave equation (*a*) and the relation (*b*) and the boundary conditions (*c*) which is verified. True, in the cases of sound and light the specific form of the postulates has been suggested by pictures of the dependent variable as condensation and field strength, respectively. The same postulates might however have been derived from other pictures, or from none at all. When in the middle of the nineteenth century the elastic-solid aether was in fashion, it supplied the postulates (*a*) and (*b*) in the same form as the electromagnetic theory now provides them. The

diffraction patterns of light thus do not speak for either of these conceptions against the other, but solely for the wave equation and for the relation between squared-amplitude and observed intensity which are compatible with both.

In the wave theory of matter and electricity we have the postulates without the picture. The lack of a picture has not precluded the invention of a wave equation, a relation between squared amplitude and that-which-is-observed, and a set of boundary conditions—the formulation, that is to say, of as much of a wave theory as it is possible to test by experiments on diffraction and reflection.

The significance of Ψ is therefore wholly contained in these statements:

(*a*) Ψ is the solution of one or another of certain wave equations with certain restrictions, examples of which have been given.

(*b*) The square of the amplitude of Ψ is proportional to the number-per-unit-volume of corpuscles. (This statement will presently be rephrased.)

(*c*) At a surface of potential-discontinuity, Ψ and its gradient are continuous.

It must of course be admitted that in the theories of light and sound, there is incontrovertible reason for giving the names of field strength and condensation to the solutions of the wave equations. In observing sound, it is quite possible to perceive (e.g., with a revolving mirror) the actual vibrations of solid bodies whence the sound waves stream through the air, and possible also to calculate the speed of sound correctly from the elastic constants and the density of the medium. With light there are various powerful arguments (for instance, the agreement between the speed of light *in vacuo* and the ratio of the electrical units) for deeming that it is of the same nature as the familiar electrical waves of which the field strength and the frequency are observable. From the physicist's point of view, the wave theory of electricity and matter is inferior in this respect to those of sound and light. It seems likely that the inferiority will be permanent.

Postulate (*b*) will now be restated in the customary way. Hitherto I have been writing the solutions of such differential equations as 12.3 in such forms as 12.7; that is to say, as sinusoidal functions $C \cos (nt + \gamma)$, with C and γ standing for the constants named amplitude and phase angle. The complex expression,

$$A \exp (int) = (R + iS) \exp (int)$$
$$= (R + iS) (\cos nt + i \sin nt)$$

is a pair of such functions, one with and one without the factor i. They have equal amplitude $(R^2 + S^2)^{\frac{1}{2}}$; the radicand may be written as the product of $(R + iS)$ by $(R - iS)$, that is to say, of A by its conjugate A^*. (Mathematicians will object that a superimposed bar instead of a star is commonly used in mathematics for a conjugate; but in physics it signifies an average.) Their phase angles differ by 90°, but this is not important in quantum mechanics. If therefore we insert the complex expression $A \exp (int)$ into a differential equation such as 12.3, we are in effect inserting a sine-function of squared amplitude AA^*. Postulate (*b*) then takes the following form:

In combining wave theory and corpuscle theory, the wave function Ψ *and the concentration* N *of the corpuscles are related as follows:*

$$\Psi\Psi^* \propto N \qquad 12.72$$

Ostensibly this proportionality might be converted into an equation by multiplying either side with some constant factor, but such a step would be futile and misleading. It is affirmed merely that at different points of a particular wave train or diffraction pattern or stationary wave pattern, the values of $\Psi\Psi^*$ stand to those of N is a constant ratio. Nothing is affirmed about the magnitude of this ratio, nor about any relations between its values for different wave trains or different patterns.

Another of the difficulties peculiar to the wave model now makes its appearance. When the relation between squared amplitude and concentration of corpuscles was first introduced

in this book, the reference was to infinite or at any rate to very long wave trains, associated with clouds of very many (non-interacting) corpuscles spread uniformly through the train. Concentration of particles was then a permissible concept. Latterly, however, we have been considering wave patterns in peculiar force fields, designed for representing atoms; for example, the inverse-square central field proposed for the hydrogen atom. It is not permitted to suppose that there are enormously many corpuscles in such a field; for the hydrogen atom, we are obliged to content ourselves with a single electron for the entire wave pattern. The right-hand member of Eq. (61) now is meaningless.

Here is, indeed, another incapacity of the wave model: it is unable to represent a single corpuscle in a definite place at a given time. Earlier we found it incapable of representing a flock of corpuscles all close together in space and having identical momentum, and we saved it only by the drastic step of surrendering our intuitive idea that such a flock can exist in Nature. To surrender again is impossible, for we cannot yield up our belief in the existence of single corpuscles. We are constrained to a policy which may be stated as follows:

Take some accepted atom model, and a particular proper value of E and the corresponding proper function Ψ. For definiteness, choose the already-described model of the hydrogen atom—the central field with potential-energy function $-e^2/r$—and the energy value, Ψ-function, and stationary wave pattern corresponding to the normal state thereof. Choose some point P in the field; describe a volume element dv around it; evaluate Ψ and Ψ^* at P. The question is: what is $\Psi\Psi^* dv$? We cannot say outright that it is the number of electrons in dv, there being only one to the entire atom. We can however say that the stationary wave pattern pertains not to a single atom, but to all the hydrogen atoms in the universe (or anyhow, to all those in a particular tube) which at a given moment are in the normal state. Or, we can say that it pertains to a single atom not at one instant of its career but over the whole of all the intervals of time during which it is in its normal

state. Then $\Psi\Psi^* dv$ will be proportional, in the former case, to the number of atoms for which the single electron is located at the given moment in dv (that is to say, in a volume element shaped and placed, with respect to the nucleus of its atom, as the element dv is shaped and placed in the model). In the latter case we must imagine the atom at very many different moments of its sojourn in the normal state; then $\Psi\Psi^* dv$ will be proportional to the number of moments at which the single electron is located in dv.

The former of these policies is on the whole the more serviceable, since the Ψ-functions are chiefly employed for explaining phenomena which are known by observations not on single atoms but on great multitudes. The laws of the absorption of light by a billion billion atoms of gaseous sodium in a tube, or the response of the gas to a magnetic field, are just as well interpreted by the proposed idea of $\Psi\Psi^*$ as they would be if one could assign great numbers of electrons to each individual atom. The latter policy has advantages of its own. Both are embodied in the statement that $\Psi\Psi^* dv$ is a measure of the *probability that the electron is located in the volume-element dv.* Similarly, if for Ψ we take one of the functions specified in Eq. (12.63) and evaluate it at some value of x lying within some interval dx, then $\Psi\Psi^* dx$ is taken as the measure of the probability that an oscillating atom in the state characterized by that particular Ψ-function is located in dx. Neither of these statements is a general formulation of the doctrine concerning Ψ, but they must suffice for this book.[1]

Degeneracy and perturbations

I now conclude this chapter by taking up the role and the implications of the constant l, and the " degeneracy " which the stationary states of hydrogen exhibit according to the wave-mechanical theory.

[1] The definition of $\Psi\Psi^*$ as "probability" implies that the integral of $\Psi\Psi^*$ over the entire wave pattern is unity. We can always arrange this by adjusting a numerical coefficient, except when the integral is infinite, as in the case of the endless sinusoidal wave train; in such cases it is best to evade the difficulty by returning to Eq. (12.72).

Combining the statements made in the foregoing section, the student will see that each of the E-values listed in (12.68) is an eigenvalue—or a permitted energy-value—if and only if l has one or another of a certain limited number of permitted values. E_1 is an eigenvalue only if $l = 0$; E_2 is an eigenvalue only if $l = 0$ or $l = 1$; in general, E_n is an eigenvalue for n values of l ranging from 0 to $(n - 1)$. From the opposite viewpoint, whatever l-value we may choose to assign will be compatible with all the permitted energy-values from E_{l+1} onward, but not with those from E_l downward.

Resorting again to the language of page 338, we may say that each of the states characterized by the values 1,2,3 . . . of n the " total quantum-number " (for such is the name given to n) and by the permitted energy-values E_n, exhibits a degeneracy of order n. This however is far from being a full description of the degeneracy of these states. We have already encountered the solutions Y of that equation-in-the-angles which is common to the real ball of fluid and to the Schroedinger atom-model for hydrogen (and to many other cases), having appeared in this chapter first in (12.34) and then as (12.65); and I have remarked, in connection with the expression (12.36), that for a given l-value the eigenfunction Y_e consists of $(2l + 1)$ terms each involving an arbitrary multiplying constant. Another, and for the present purpose a preferable, way of stating the situation is: to each l-value correspond $(2l + 1)$ distinct eigenfunctions, the general solution of the equation for that l-value being a linear combination of these.[1] Now, choose any value for n, and choose for l any value compatible with that n-value, and consider the situation. There will be an eigenfunction $R_{n,l}$ of equation (12.66), and there will be the linear combination of those $(2l + 1)$ eigenfunctions of (12.65). The full solution of the underlying wave-equation (12.63) will be the product of these, and again there will be an eigenfunc-

[1] This becomes an obvious policy when the student follows through the procedure of writing Y as the product of a function of θ by a function of ϕ and ascertaining the eigenfunctions of the resulting two differential equations, as recommended in footnote 1 (page 353).—In this section, l corresponds to n of page 353.

tion of $(2l + 1)$ terms preferably to be called a linear combination of $(2l + 1)$ eigenfunctions. This is a degeneracy of order $(2l + 1)$. Now envisage *en bloc* the totality of l-values compatible with the chosen n-value, and count the eigenfunctions. It is easily shown that the number thereof, *the degeneracy*—or to use an alternative name, *the statistical weight—of the state characterized by n and* E_n *is* n^2.

Actually the degeneracy of the state n is twice as great —$2n^2$—owing to a quality of the electron which I have not yet mentioned, the electron-spin. It is this quality which prevents me from proceeding much further along the lines of this chapter without making statements too seriously false to be admissible even temporarily. The most that can be risked at this stage is to mention, that if in the atom-model a different sort of a field is superposed upon the inverse-square electrostatic field thus far assumed—if in the wave-equation the potential-energy function V is altered from the form $-e^2/r$ hitherto postulated, then the state n will be superseded by a group of not more than $2n^2$ states distinguishable from each other because of small differences in their energy-values. Better is it to say, that in the general case of an arbitrary field there are these groups of $2n^2$ $(n = 1,2.3 \ldots)$ distinguishable states, and the energies of the members of any one thereof slip into coincidence at the values (12.68) as the prevailing forces are reduced to the ideal limit of the perfect inverse-square force-field.

This case of the inverse-square field thus plays a very peculiar role in Nature. Apparently it entails a smaller number of permitted energy-values than almost any other. Yet in a profound sense, all of the much more numerous stationary states evoked by any other kind of field are latent in it; for each of its E_n-values really belongs to a whole aggregate of states, distinguishable from one another by their eigenfunctions though not by their energy-values; and when the case is altered by alteration of the field or the potential-energy function, the states of such an aggregate become distinguishable in energy as well.

The student must now become acquainted with the words "perturb" and "perturbation," taken over from celestial mechanics into quantal theory. A remarkable and distressing feature of quantal theory is the fewness of the cases—*i.e.*, of the force-fields or potential-energy-functions—of which the eigenvalues and the eigenfunctions can be computed exactly. Those which I have given in this chapter are almost the only examples. Almost every other case is treated by some method of approximation, in which the postulated potential-energy-function V is written down as the sum of one of the tractable functions—usually $-e^2/r$, for atom-models as distinguished from molecule-models—and a "perturbation," and the results are more nearly accurate the smaller the perturbation. The force-field, the eigenvalues and the eigenfunctions are all said to be "perturbed," to distinguish them from those of the tractable "unperturbed case." The description of the mathematical technique required for perturbed cases constitutes a very large and difficult part of atomic theory as a whole, and even so it is frequently not good enough to be reliable.

To use this technique to prove the following statements is beyond the scope of this chapter, but the terminology will help in expressing them briefly. First think of the cases in which an atom of nuclear charge $+Ze$ has two or more electrons, any one of which is exposed to the inverse-square field of the nucleus and to the fields of all the other electrons. The V-function for any electron is the sum of $-Ze^2/r$ and of terms due to the other electrons, these last being the perturbation. Discouraging as the prospect looks in view of the presumably-incessant interweaving motions of the electrons, it turns out often—surprisingly often—that the perturbation acts approximately like an extra central field superposed upon the inverse-square one. On attacking the wave-equation with V written as $-e^2/r + f(r)$, the mathematician finds that the effect of the perturbation $f(r)$ is to supplant each of the states n of the unperturbed case by a flock, not indeed of n^2 but of n states of distinguishable energy-values. The perturbation has separated, in respect of their energies, all of the states sharing a

common n-value but differing in l; it has not however separated the states sharing a common l-value as well as a common n-value. It is said to effect a " partial removal " of the degeneracy.

This partial removal of degeneracy it is, which is responsible for the classification of stationary states—those of sodium, for instance, which in Chapter X we took for an example—into the columns marked by the letters S and P and D and F and so on. What is common to the states of a column is the l-value, the correlation being as follows:

symbol	S	P	D	F	G	H	.....
l-value	0	1	2	3	4	5	.....

What distinguishes the different states of any column one from another is the n-value; for the order-number introduced on page 310 either is the same as n, or else differs from n by an integer which is the same all along the column.

Full removal of degeneracy is attained, if to the foregoing fields there is added an applied magnetic field. I give to this important statement the emphasis of italics, being unable to give it that of detailed presentation until the theoretical picture shall have been completed—or at any rate improved—by the introduction of the electron-spin. We must also wait upon this introduction before we can assess any more of the numerical agreements between experiment and theory, which fortify the latter in its present form. To introduce the electron-spin is the first of the tasks of the following chapter; in this one I have yet to restore the concept of angular momentum.

Angular momentum in the wave-mechanical atom-model

The angular momentum of the circulating electron, which was so prominent in the original atom-model of Bohr as well as in the later specializations thereof, is entirely missing from the wave-mechanical theory at the stage to which we have now come. In the technique of computing the permitted energy-values E_n by wave-mechanics, there is nothing to suggest even

the revolution of an electron in an orbit, much less the numerical value of the angular momentum thereof. The theory must be broadened to make a place for angular momentum. For the present I give only the result:

In the wave-mechanical theory of the hydrogen atom, *an l-value corresponds to an angular momentum of magnitude* $\sqrt{l(l+1)}\,(h/2\pi)$.

We have seen that in the ideal limiting case of the perfect inverse-square field, any one of the permitted energy-values E_n ($n = 1,2,3\ldots$) is associated with n different l-values ranging from $l = 0$ to $l = (n-1)$, and with the corresponding eigenfunctions. These have the angular momenta 0, $\sqrt{2}$, $\sqrt{2.3}$, $\sqrt{3.4}$, $\ldots$ $\sqrt{n(n+1)}$ times $(h/2\pi)$. The energy-value E_n thus belongs to an aggregate of states, all differing in angular momentum and all identical in energy. When the atom is perturbed by an additional central field, these states retain their angular momenta unchanged, while the identity of their energy-values ceases.

This formulation suggests the elliptical orbits of page 337, and is meant to suggest them. What was there achieved by introducing elliptical orbits is here achieved by discriminating between the different l-values compatible with an n-value, and correlating them with angular momentum. (This by the way is what first caused l to be accepted as a measure of angular momentum.) Consider a state which by wave-mechanical theory is marked by the quantum-numbers n and l, by the energy-value E_n and the angular momentum $\sqrt{l(l+1)}(h/2\pi)$: according to the former model it is represented by an orbit having the proper n-value and energy-value, and the angular momentum $kh/2\pi$. The relation between k of the former model and l of the wave-mechanical theory is

$$k = l - 1 \qquad (12.73)$$

a relation which some day may be of interest to historians only, but for the present must be remembered, as k is often to be found in articles and books too recent to have been superseded.

It is disconcerting, of course, that different values of angular momentum have to be assigned to any state, according as one uses the Bohr-Sommerfeld model or the wave-mechanical theory. The contrast is greatest for S states, for which k is unity, while l is zero and accordingly $\sqrt{l(l+1)}$ likewise is zero. The normal state of the hydrogen atom—like that of all alkali-metal atoms and many others—being an S state, this is a conspicuous case; and happily it is a case, as to which experiment speaks clearly.

It would be too bad to end this chapter on a note of clash. I remind the student therefore, that by either theory we have attained to the correct energy-values for the stationary states of hydrogen—we have learned to regard each of these " states " as an aggregate of states, coincident in energy when the intra-atomic field is of perfectly inverse-square character, but becoming distinct from each other in energy when an appropriate perturbing field exists—we have learned (borrowing, it is true, from an argument to come) to recognize angular momentum as a distinctive feature of a stationary state. Further we cannot go, without taking account of the electron-spin; and to this we now address ourselves, reverting meanwhile to the original atom-model with the orbits.

CHAPTER XIII

Magnetism, Electron-Spin and Spectra

We have seen that an atom with revolving electrons is *ipso facto* magnetic; and we have learned that spectra are interpreted, in modern atomic theory, by supposing that an atom has stationary states which are characterized by special electron-orbits and special values of angular momentum. May we then infer that an atom, in its normal state or in any other of its permitted states, has a magnetic moment of which the value is characteristic of the atom and of the state, and can moreover be deduced from the spectrum of the atom? Indeed we may! and the deductions can be tested by magnetic measurements, and the ensemble of these verified deductions is one of the most important items, perhaps *the* most important item of the evidence for the general rightness of contemporary theory. Beyond these checks of the numerical values of atomic magnetic moments, there are other striking services which the theory owes to magnetic experiments; for instance, the direct and visible proof of one of the most incredible of all the quantal properties of Nature, to wit, the fact that an atom having a magnetic moment and angular momentum, and placed in a magnetic field, is constrained to accept one or another of a severely limited number of orientations in the field. In this chapter we shall address outselves alternately to magnetic and to spectroscopic phenomena, unifying them by means of theoretical principles which are exemplified as well by the ones as by the others.

We begin with one of the most important theorems in the whole range of atomic theory.

Ratio of magnetic moment to angular momentum, for an electron-orbit

According to the fundamental principles laid down more than a hundred years ago by the first analyst of the magnetic effects of electric currents—Ampère—a current I flowing in a circuit enclosing a plane area A is equivalent, in respect of the magnetic field which it creates at sufficiently large distances, to a magnet of magnetic moment μ given by the formula,

$$\mu = IA. \tag{13.1}$$

Imagine an electron revolving in a circle of radius r with linear speed v. Its charge e is carried $v/2\pi r$ times per second through any point of the circle, so that to this extent it is equivalent to a current equal in electrostatic units to $ev/2\pi r$, and in electromagnetic units—those of (13.1)—to $ev/2\pi rc$. The area of the circle being πr^2, we find:

$$\mu = evr/2c. \tag{13.2}$$

For the angular momentum, we have:[1]

$$p = m_0 vr \tag{13.3}$$

and for the quotient, the fundamental formula:

$$\mu/p = e/2m_0c. \tag{13.4}$$

This formula refers to an electron-orbit, or to the magnetic moment and the angular momentum which an electron acquires by virtue of revolving in an orbit. These somewhat clumsy phrasings are essential, for (as we shall presently see) an electron also possesses a magnetic moment and an angular momentum inherent in itself. I ignore these last a little longer for the sake of the exposition, but it is never safe to ignore them even for a moment.

The formula (13.4) is valid not only for circular orbits, but for elliptical ones in an inverse-square field, and generally for

[1] Henceforth the electron-mass will be denoted by m_0, the symbol m being reserved in this chapter for the magnetic quantum-number.

all orbits in a central field.[1] One sees that while μ and p differ for different orbits and states, their ratio does not; its peculiar value $e/2m_0c$ is common to all orbits.

The Larmor precession

Imagine a magnet of moment μ, placed in a magnetic field of strength H and inclined at angle θ to the direction of the field: it may be visualized as a bar pointing in the direction θ, with poles of strength M separated by distance μ/M. The field exerts upon it a torque of strength $\mu H \sin\theta$, which if the bar had no angular momentum would cause it to oscillate to and fro in the plane containing its initial direction and the field-direction. If however the magnet has an angular momentum p about its own axis, all this is changed; the magnet is now a gyroscope exposed to a torque, and as such it precesses around the field-direction at a uniform rate, the angle θ between its axis and the field remaining unchanged while the plane containing these two directions revolves around the latter. The angular velocity of this plane, the "rate of precession" ω, is given by a theorem of classical mechanics:

$$\omega = \mu H/p \qquad (13.5)$$

into which we substitute from (12.4), obtaining the rate of precession appropriate to an electron-orbit in a magnetic field H:

$$\omega_L = (e/2m_0c)H \qquad (13.6)$$

This rate is known as the "Larmor precession."

The Larmor precession is independent of θ, and retains its value even in the interesting extreme cases in which $\theta = 0$ or π. If the orbit is circular and at right angles to the field, the angular velocity ω of the electron is augmented (or diminished,

[1] The general expression for p is $m_0 v_t r$, in which v_t stands for the velocity-component of the particle perpendicular to the "radius vector" r drawn from the centre of force to the particle. Now, $\frac{1}{2}v_t r$ is the rate dA/dt at which the radius vector sweeps out the area of the orbit; thus $dA/dt = p/2m_0$, and A is the integral of this constant quantity over a complete period T of the revolution, which is $pT/2m_0$; the magnetic moment is e/cT times this quantity, whence formula (13.4).

according to which case we have) so that the alteration in it is ω_L and the alteration in the frequency is $\omega_L/2\pi$. The angular momentum and the magnetic moment of the orbit are altered in the proportion $(\omega \pm \omega_L)/\omega$; these changes are the origin of diamagnetism, but are negligible in most of the phenomena which we are about to consider. Whether or not the orbit is circular, and whether or not its plane is at right angles to the field-direction, the frequency $\omega_L/2\pi$ is added to or subtracted from the other frequency or frequencies originally present; one may say that these are "modulated" by the frequency of the Larmor precession. This circumstance is the basis of the classical theory of the "normal" Zeeman effect, to which we now turn for a moment.

The "normal" Zeeman effect

When a gas which is emitting or absorbing light is subjected to a magnetic field, its ordinary spectrum-lines are superseded

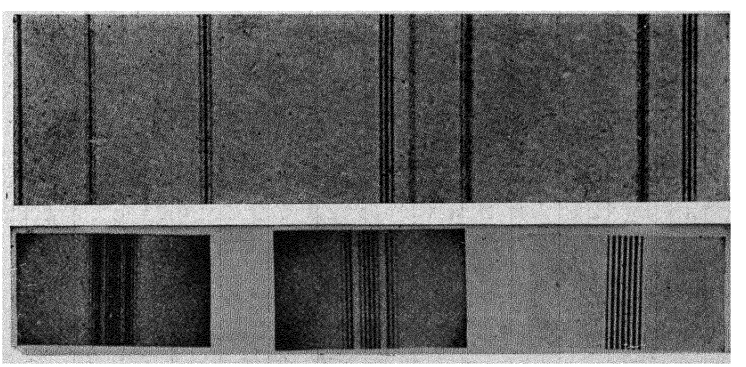


FIGURE 13.1—ILLUSTRATIONS OF THE EFFECT OF MAGNETIC FIELD ON SPECTRA. (P. Zeeman; *Journal of the Franklin Institute.*)

by groups of lines, the "Zeeman patterns." Over wide ranges of field strength H the number of lines to a pattern remains the same and the spacing (frequency-difference) between any two of them is proportional to H. The phenomenon is called "Zeeman effect," after the man who discovered it (in 1898).

Certain lines of certain elements are changed, or "resolved" as the usual phrase is, into patterns of three: the middle member has the frequency of the original line unchanged, while in the others this frequency is diminished and increased, respectively, by the amount $\omega_L/2\pi$. Such are known as examples of the "normal" Zeeman effect; not because they are the most usual, which they fall very far short of being, nor because they are fundamentally the simplest, as we now know very well they are not—but because they are the ones which conform to the classical theory of the emission of light from an oscillating electron.

Forgetting for a little while about orbits and stationary states, imagine an electrified particle vibrating to and fro with frequency $\omega/2\pi$ along a straight line, inclined at angle θ to the z-direction which is destined to be that of the field. The motion may be resolved into two linear components, one parallel to the z-direction, the other in the xy-plane. The latter of these linear components may in turn be resolved into two circular motions, described in the xy-plane with equal and opposite angular velocities $+\omega$ and $-\omega$. Now let the field be applied. The linear component along the z-direction will be unaffected, the angular velocities of the two circular components in the xy-plane will be altered by the Larmor precession into $(\omega + \omega_L)$ and $-(\omega - \omega_L)$. The waves radiating from the electron (remember that we are now thinking in the classical fashion!) must have these three frequencies imprinted upon them: $\omega/2\pi$, $(\omega + \omega_L)/2\pi$, $(\omega - \omega_L)/2\pi$.

This inference as to frequencies is verified in the "normal" Zeeman effect, as I have just said; but there is much more. If the oscillator happened to be moving in the z-direction at the start, we should expect it to be emitting plane-polarized light with the electric vector parallel to z, and this emission would not be isotropic—the intensity would be greatest in the directions lying in the xy-plane, and would be zero along the z-direction, falling off according to a simple law from the one value to the other. Now in whatever directions the oscillators may be moving, we should expect these qualities to be displayed

by the light of unaltered frequency ascribed to the first of the linear components aforesaid—and they *are* so displayed, in the normal Zeeman effect! Similarly the light corresponding to each of the lateral members of the Zeeman pattern would be expected to be circularly polarized when going off in the z-direction, and plane-polarized when going off in the xy-plane; the sense of the circular polarization should be right-handed for one of the lateral members, left-handed for the other. All this is confirmed.

The normal Zeeman effect is thus a phenomenon to which the classical theory is remarkably, almost uniquely, adapted. Nevertheless we have to substitute a quantal theory for it. The process of adapting quantal theory to the explanation of polarization and of distribution-of-intensity-indirection is too advanced to be expounded in full in this book; whatever is said about it here will have to be said dogmatically. The process of adapting it to the explanation of the frequencies is simpler, and very important.

Quantal theory of the "normal" Zeeman effect; space-quantization

In adopting the viewpoint of quantal theory, the first stage consists in interpreting spectrum-lines as transitions between stationary states. We must therefore interpret a Zeeman pattern as a sign that the stationary states, the transition between which gave the original single line, have been "resolved" or "split up" by the magnetic field into groups or clusters of states. Here the terminology again becomes difficult: I will refer to these last as groups or clusters of sub-levels.

Let us designate by W_1 and W_2 the energy-values of the original two states, therefore by $(W_1 - W_2)/h$ the frequency of the original line. Let us then assume that the magnetic field H resolves the former state into $(2l_1 + 1)$ sub-levels having the energy-values,

$$W_1 \pm m\omega_L h/2\pi, \qquad m = 0,1,2, \ldots l_1 \qquad (13.7)$$

and the latter state into $(2l_2 + 1)$ sub-levels having the energy-values,

$$W_2 \pm m\omega_L h/2\pi, \qquad m = 0,1,2, \ldots l_2 \qquad (13.8)$$

Here l_1 and l_2 stand for integers of which, for the moment, I make no further specification.

It is evident that (unless both l_1 and l_2 are made zero!) we have here two systems of sub-levels which between them will produce the observed lines of the normal Zeeman pattern. The middle line of the pattern corresponds to a transition in which m does not change; the lateral lines correspond to transitions in which m changes upward by one unit and downward by one unit, respectively. The two systems appear capable of producing even more lines; thus, if l_1 and l_2 are both unity or greater, a change of two units in m might conceivably occur, entailing a line displaced from the original frequency by $2\omega_L$. Such lines are not observed: we accordingly infer a selection-principle, according to which m may change in a transition by zero or by one unit, but never by two or more units. This is the "selection-principle for the magnetic quantum-number," for such is the name of the integer m.

When we postulate the energy-values (13.7) and (13.8) for the sub-levels, and in addition the selection-principle just stated, the normal Zeeman pattern is correctly reproduced by our theory. Transitions in which m does not change then give plane-polarized light, of the character described on page 382; transitions in which m increases by one unit give circularly-polarized light of one sense, transitions in which m decreases by one unit give circularly-polarized light of the opposite sense. We may impose these rules by postulate upon our theory.

Now we turn back to the model, and envisage the atom with its circulating electron as a magnet of magnetic moment μ and angular momentum $p = (2m_0c/e)\mu$, not necessarily (nor generally!) the same in both the states; and *we interpret the sub-levels with their different energy-values as different orientations of these atomic magnets in the field.* The torque upon such a magnet being $-\mu H \sin\theta$ when the axis of the magnet is

inclined at θ to the field-direction, the atom possesses by virtue of its being in the field an extra magnetic energy ΔW which may be written,

$$W = -\mu H \cos\theta \qquad (13.9)$$

the additive constant being so chosen that ΔW vanishes when the magnetic moment is at right angles to the field.

Identifying this extra magnetic energy with the train of energy-values specified in (13.7) or (13.8), we see that in effect we were there prescribing discrete " permitted " orientations to the atomic magnet, according to the rule:

$$m\omega_L h/2\pi = \mu H \cos\theta \qquad (13.10)$$

This rule becomes quite striking when we invoke the fundamental value of the ratio μ/p, for now

$$m\omega_L h/2\pi = p(e/2m_0c)H\cos\theta = p\omega_L \cos\theta,$$

$$p\cos\theta = mh/2\pi \qquad (13.11)$$

The axis of the electron-orbit has permitted orientations, such that the projections of the angular momentum upon the field-direction are integer multiples of $h/2\pi$. This incidentally shows the meaning of the magnetic quantum-number m.

To this conclusion have we been led by the study of the " normal " Zeeman effect! The theory now speaks for itself, and places a limitation on the values of the integer called l_1 and l_2 in (13.7) and (13.8). Since the projection of the angular momentum upon the field-direction cannot be greater than the angular momentum itself, that integer cannot be greater than the l-value of the orbit. It can however be *as* great, and there is no point in supposing it smaller and so denying some of the possible orientations which the model admits. Moreover, the Zeeman patterns of the soon-to-be-mentioned " anomalous " types show directly that in their cases, the corresponding integer does have the greatest possible value.

We therefore accept this inference, which is tantamount to saying that *for an electron-orbit of any l-value there are* $(2l + 1)$ *permitted orientations and* $(2l + 1)$ *permitted values, ranging*

from $-l$ to $+l$, of the magnetic quantum-number. But $(2l + 1)$ is also the order of the degeneracy, as defined in Chapter XII, of the state represented in that chapter by the Ψ-function and in this chapter by the electron-orbit with the l-value in question. The sub-levels which manifest themselves as separate when the magnetic field produces the Zeeman effect—these sub-levels retain their separateness, at least to the mathematician's mind, when the magnetic field is withdrawn and their energy-values become identical; they are distinguished by the distinctions in the stationary-wave-pattern to which allusion was made on page 375.

The student may now expect me to say that the " normal " Zeeman effect is exhibited in the simplest possible cases, to wit, wherever there is a single electron revolving about a nucleus—in the spectra H I and He II. This however is definitely not so: in such cases, the spin of the electron plays an ineradicable part. The normal Zeeman effect occurs only when there are two or more electrons, so oriented that their spins annul one another, while their orbits are so interlocked as to behave like a single orbit. It is exceedingly inconvenient for exposition that this should be so, since it means that the electron-spin is never really absent but only hidden, and a treatment which disregards it from the start is to that extent deceptive. I now make haste to rectify the omission, and introduce the electron-spin on the basis of a purely magnetic phenomenon.

Gyromagnetic effects

Imagine a cylinder of some strongly magnetizable solid substance, hanging freely from a suspension attached to one end. Suppose it to be unmagnetized at first; this signifies that the atomic or subatomic magnets are so oriented that the resultant of all their magnetic moments, and accordingly that of all their angular momenta, are both zero[1]. Now let a strong axial

[1] It is being assumed that magnetization (exception being always made for diamagnetic substances!) consists in re-orienting the atomic or subatomic magnets without altering their μ-values, and that angular momentum is parallel to magnetic moment—two assumptions justified in the event.

magnetic field be applied, as by a solenoid wrapped around the cylinder. The substance is suddenly magnetized; let us for convenience suppose it magnetized to saturation. All of the tiny magnets now point axially, and the sum of their moments $\Sigma\mu$ is the magnetic moment M of the rod. Their angular momenta also point axially and have a resultant Σp, the sudden advent of which ought to entail the appearance of an equal and opposite angular momentum—call it $-P$—to assure the conservation of angular momentum in the world as a whole. This *does* appear; it appears in the mass of the cylinder, and manifests itself as a sharp twist of the rod upon its suspension occurring when the solid is magnetized.

This is one of the most important phenomena of physics, for it demonstrates that that which is called angular momentum in the uttermost refinements of atomic theory *is* really angular momentum in the most ancient and literal sense of the word—something which the student might well be disposed to doubt, so strange are some of the features of the former! It was first observed in 1915 by Einstein and de Haas, who accentuated the effect by applying an alternating field which caused the hanging rod to oscillate about its axis, the frequency of the field being so adjusted that advantage is taken of resonance. Of the experimenters who followed—including S. J. Barnett of Pasadena, and Chattock, Bates and Sucksmith of the Bristol school—some have developed that scheme, some have employed a null method which consists in counteracting the tendency to oscillation by a measurable torque applied in opposition to it, some have simply reversed the magnetization and so doubled the effect as I at first described it. For years the results were confined to ferro-magnetic metals and alloys, a fact which shows how hard and delicate the experiment is; latterly Sucksmith has refined it so much as to be able to make the measurement on strongly paramagnetic salts.

If the experiment is hard, the interpretation at least is easy: the ratio M/P of magnetic moment to angular momentum of rod-as-a-whole is equal to the ratio μ/p for the elementary magnets. There are not many qualities of matter of which it

can be said that the values measured upon visible and tangible pieces are appropriate to the elementary particles, but this is one of them. It is found, first, that M/P has the same value (for a given substance) whether saturation is attained or only approached; this strengthens the inferences from the experiment. It is found, next, that the sense of the twist is correct—that is to say, the twist is in the proper sense to be interpreted as recoil, the elementary magnets themselves being circulating *negative* charges (not positive). This is an important item in the evidence that negative electricity is more mobile than positive, little as this doctrine may now seem in need of further proof.

As for the numerical values of M/P, I pause before giving them in order to mention the other of the two " gyromagnetic effects,"—magnetization produced by continuous rotation, an effect discovered by Barnett in 1914 and still mainly associated with his name. The rod is revolved as rapidly as possible upon its axis, and is found to acquire a feeble magnetization; the torque is urging the elementary magnets to turn their axes more nearly into parallelism with the axis of revolution, as can be foretold from the fact that they are gyroscopes. We may circumvent the details by realizing that as a field H evokes a Larmor precession $\mu H/p$, so an angular velocity ω imprinted upon a flock of gyroscopic magnets has the same effect upon them as would a field $\omega p/\mu$. The procedure then is, to measure the magnetization produced by a known angular velocity ω, to measure by a separate experiment the field strength H capable of producing the same magnetization, and to interpret the ratio ω/H as μ/p. To simulate as small a field strength as one millionth of a gauss requires 1.4 revolutions per second.

Now for the sensational result: *the ratio μ/p for ferromagnetic substances is about twice as great as $e/2mc$*; or, introducing a new symbol later to be much used,

$$\mu/p = ge/2m_0c, \quad g = \text{about } 2 \qquad (13.12)$$

a result not to be accounted for by any peculiarity in shape, size, or frequency of the orbits, since we have seen that irre-

spective of all these parameters the theory prescribes $e/2m_0c$ for the ratio μ/p, or unity for the g-factor.

I have said that g is about 2 for ferromagnetics; actually, Sucksmith and Bates of the Bristol school find values agreeing with 2 within the statistical uncertainty, which is a few promille for iron and cobalt and nickel, a few per cent for a Heusler alloy and for magnetite. Barnett on the other hand finds g four or five per cent smaller than 2 for iron and permalloy. For various paramagnetic bodies, g differs markedly from 2, being greater for most but smaller for some; of these cases, more later.

It will now be assumed that the g-value 2 *multiplied by* $e/2m_0c$ *is the ratio of a magnetic moment and an angular momentum which are intrinsic to and inherent in the electron itself*, apart from those which it possesses by virtue of its revolution in an orbit. Like the earth, the electron has both a revolution and a rotation, though it is dangerous to press analogy too far. The revolution changes whenever the electron executes a transition, but the rotation—or to use its common name, the *spin*—is apparently immune from anything that we (or Nature?) can do, except for the direction in which its axis points.

If a magnetizable body exhibits a g-value of 2 exactly, it is assumed that magnetization thereof consists in orienting the electron-spins, the orbits being as it were extinguished; either all the orbits are of the type $l = 0$ for which the orbital angular momentum and magnetic moment vanish, or else they are interlocked into a rigid non-magnetic whole in a manner which we shall presently study. If however a magnetizable body exhibits a g-value which is neither 2 nor 1, it is assumed that the orbital angular momenta and the spin angular momenta are interlocked in such a way that the g-value represents a compromise. Disconcerting as this may sound, it is in the laws of compromise that physicists find their best confirmation of the idea of electron-spin. Now however we turn to a remarkable experiment applied to another atom-type in which the orbital angular momenta are concealed and the spin is manifested by itself. This experiment displays to us the permitted orientations and

reveals to us the magnitude of the magnetic moment of the spinning electron.

The Gerlach-Stern technique

The method whereby the magnetic moments of atoms are measured, and the permitted orientations are shown with a vividness defying skepticism, was invented and first applied in 1922 by Gerlach and Stern. The latter then developed it, and he and his pupils—notably Rabi and Estermann—have continued to perfect it throughout all the subsequent years.

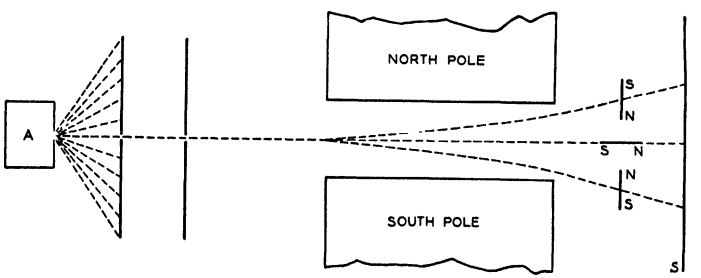


FIGURE 13.2—LONGITUDINAL SECTION OF THE APPARATUS OF THE GERLACH-STERN TECHNIQUE IN ITS SIMPLEST FORM.

Metals such as silver and sodium are vaporized in a small electric oven, box-shaped and having a small orifice. A beam of atoms, issuing through the orifice and delimited by it and by a series of diaphragms, passes across a magnetic field in a good vacuum. Imagine the atoms to be tiny magnets of magnetic moment μ; visualize them, if need be, as tiny bar-magnets of pole-strength M and distance r between the poles, the product Mr being denoted by μ. Evidently if the magnetic field is uniform the net force upon each atom will be nil, the forces upon the two poles being equal and oppositely directed. (We are not now concerned with the torque exerted by the field upon the atomic magnet, which evokes a precession only.) In the Gerlach-Stern experiment, however, the field is *not* uniform: in the early days it was due to pole-pieces shaped as in

Figure 13.3, and was much stronger near the wedge-shaped than near the hollowed-out pole-piece.

In Figure 13.3 let the z-axis be laid vertically, in the plane of symmetry of the pole-pieces; let H represent the field-strength (parallel to z) at any point upon it, μ the magnetic moment and M the mass of an atom. Consider an atom which enters the field along the x-axis (perpendicular to the plane of the paper) with its magnetic axis inclined at angle θ to the z-axis. The force upon it is $\mu(dH/dz)\cos\theta$, and if it spends time t in flying through the field from end to end, it is drawn sidewise along the z-direction by the amount,

$$z = \frac{1}{2}\frac{\mu}{M}\frac{dH}{dz}t^2\cos\theta. \qquad (13.14)$$

For the value of t the experimenter takes the length of the path of the atoms across the field, divided by the average speed which atoms of the mass in question have when in thermal equilibrium at the temperature of the oven.

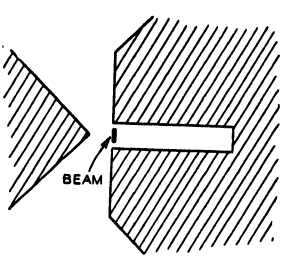


FIGURE 13.3—CROSS-SECTION OF THE MAGNET-POLES PRODUCING THE INHOMOGENEOUS FIELD IN THE GERLACH–STERN TECHNIQUE.

By classical ideas one expects that the atoms entering the field from the oven should be oriented quite at random, and therefore that the beam should be spread out into a fan in the xz-plane. The atoms of which the paths mark out one edge of the fan should be oriented parallel to the field ($\theta = 0°$), the atoms travelling along its other edge should be anti-parallel to the field ($\theta = 180°$) while atoms of intermediate inclinations should fill out the center of the fan. What is observed is different, and quite remarkable. The beam which enters the field is not spread into a fan, but divided into two or more divergent beams. With atoms of such metals as sodium and potassium, silver and gold, there are two beams bent symmetrically up and down, as though half of the atomic magnets were oriented

parallel and the other half anti-parallel to the field. The imprints of these two, on a photographic plate placed normal to the axis of x beyond the field, are shown in Figure 13.4; the z-direction is vertical; the linear shape of the imprints is due to the slits which shape the initial beam as it departs from the furnace. The picture refers to atomic hydrogen, so that we find the simplest of all atoms behaving in this respect like those of the alkali metals and those of the noble metals. This seems

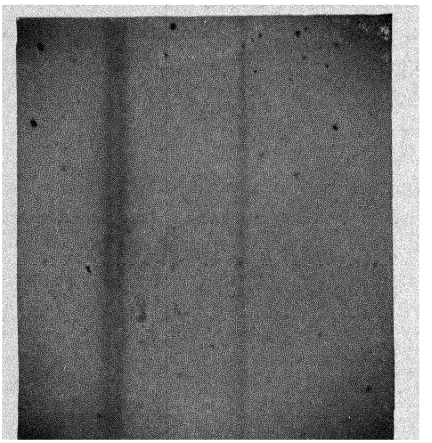


FIGURE 13.4—PATTERN FORMED BY HYDROGEN ATOMS SUBJECTED TO THE GERLACH-STERN TECHNIQUE, INDICATING TWO ORIENTATIONS OF THE ATOMS IN THE MAGNETIC FIELD. (Kellogg, Rabi and Zacharias; *Physical Review.*)

a self-evident statement, and nevertheless in view of the curious history of the theory of the hydrogen atom (to be related in Chapter XIV) the experiment is by no means superfluous.

Figure 13.5 exhibits curves obtained with sodium vapor, which at high temperatures consists almost entirely of atoms and yields the double-peaked curve of the picture, while at lower temperatures many of the atoms are compounded in pairs into molecules Na_2 which have no magnetic moment (nor angular momentum) and go undeflected through the field as the other curve reveals.

Astonishing as this result appears, it was foreseen, and even provided the incentive for the experiment! for this was undertaken by Gerlach and Stern in order to test the all-but-incredible inference, drawn from the Zeeman effect in the manner adopted on pages 383–386, that magnetic atoms in a magnetic field are constrained to a few discrete permitted orientations. The inference was confirmed by the division of the atom-beam into discrete deflected pencils; and one of the most

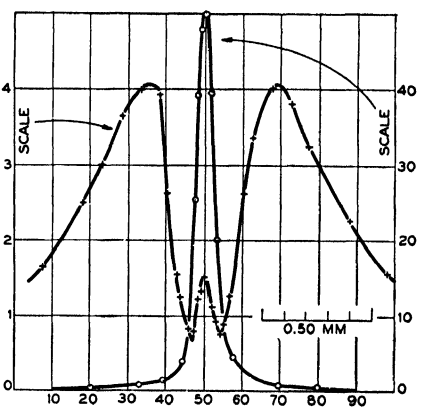


FIGURE 13.5—PATTERNS FORMED BY THE GERLACH-STERN TECHNIQUE WITH SODIUM VAPOR CONSISTING IN ONE CASE CHIEFLY OF ATOMS WITH A SMALL PERCENTAGE OF Na_2 MOLECULES (THREE-PEAKED CURVE), IN THE OTHER CASE ENTIRELY OF MOLECULES. (Lewis; *Zeitschrift für Physik.*)

unbelievably-radical features of the quantal theory was thus converted into one of the most surely established.

Yet in the cases of the atoms which I just listed, the number of deflected pencils was *two*. Now, there is no one of the electron orbits for which this is the permitted number of orientations; whatever the value of l, that of $(2l + 1)$ is always some odd number. Moreover according to wave-mechanics, the normal state of the hydrogen atom has the l-value *zero*; while in the alkali-metal atoms the same is true for the normal orbit of the outermost electron or "valence-electron," and as for all the rest of the orbital electrons they form a group which

(from many separate items of evidence) is believed to have zero for its net magnetic moment and net angular momentum. Wave-mechanics thus implies that all of these atoms should be non-magnetic in their normal states, and a beam formed out of any of them should be undeflected and undivided by a magnetic field. But the beam *is* divided; and if it is assumed that the two pencils consist of atoms with their magnetic moments oriented parallel and anti-parallel, respectively, to the field, and if then the magnetic moment is computed by means of (13.14), it is found to agree[1] with the quantity

$$\mu_B = eh/4\pi m_0 c \qquad (13.15)$$

which is known as the "Bohr magneton," being the magnetic moment which would be associated with an electron-orbit of angular momentum $h/2\pi$ if there were no such thing as spin.

All of this strangely-assorted assemblage of facts is harmonized by attributing to the spinning electron the magnetic moment μ_B, derived from the Gerlach-Stern experiment; *the value 2 for the g-factor*, derived from the gyromagnetic effect; *and the angular momentum* $\frac{1}{2}h/2\pi$, deduced by combining these two. The magnetic moment of the hydrogen atom, or any alkali-metal atom, or any other atom of the first column of the periodic table, is purely that of the spinning valence-electron. The orbit of the valence-electron contributes nothing, being of the type for which $l = 0$; the orbits and spins of the other electrons are so oriented as to cancel each other out. Atoms of this type (in their normal states!) are practically inert carriers of the spinning valence-electrons. As for the fact that the permitted orientations are two in number, it follows from an easy generalization of the rule for electron-orbits given on page 385. In words, the rule is generalized by saying: *for electron-orbits and spin alike, there are as many permitted orientations as are compatible with the condition that for any two consecutive ones,*

[1] Probably the best experiments are those of Meissner and Scheffers (**PZ 34**, 48, 245; 1933) on lithium and potassium; the results are concordant within one or two promille; the agreement is such as to establish this as one of the methods worthy of being invoked to help in settling the question as to the right values of h, e, and m_0.

the values of p cos θ shall differ by h/2π. While thus the wording of the rule is somewhat altered, the formula is not: the number of orientations is $(2l + 1)$, with l put equal to $\frac{1}{2}$ for the electron-spin.

The spinning electron and the fine-structure of the alkali-metal spectra.

We have seen in Chapter X how the stationary states of the sodium atom (as of any other alkali-metal atom) are sorted out into columns denoted by the letters S,P,D,F, . . . ; and we have seen in Chapter XI that in the atom-model these are correlated with l-values of 0,1,2,3 . . . for the valence-electron. I recall now that it was also said in Chapter X that the "fine-structure" of the sodium spectrum discloses that what have just been denoted as "states" are really pairs of closely-adjacent states, all along the P-column, the D-column and all the columns which follow; only the states of the S-column remain single, at this stage of the analysis. This is an instance of a doublet term-system (page 317).

It will now be affirmed that the two states of any such pair are distinguished from each other by the orientation of the spin with reference to the axis of the electron-orbit, the two being parallel in one of the cases, anti-parallel in the other. The electron-spin is dominated by the field due to the circulation of the electron, just as in the Gerlach-Stern experiment it is dominated by the applied magnetic field; and under this domination it is constrained to take one or the other of two discrete orientations. The field which operates in this instance may seem difficult to understand, though it is specified by classical electromagnetic theory. A rough-and-ready way of picturing it consists in imagining an electron endowed with a magnetic moment and revolving around a nucleus—then, reversing the situation and imagining the nucleus revolving around the stationary electron, treating the revolving nucleus as a circulating current just as the electron was treated *supra*, and realizing that it must produce a magnetic field at the loca-

tion of the electron. For the moment, however, we are not interested in the exact value of the magnitude of this field. Its mere existence shows us why the P-column, the D-column and all the rest consist of pairs of states, while the S-states—an essential detail!—are single; for in the S-states the l-value is zero, there is no angular momentum of the electron-orbit with reference to which the spin can orient itself, and accordingly no occasion for the subdivision.

In the S-states of a doublet system, then, the angular momentum of the atom is purely that of the electron-spin; they are therefore denoted by the symbol $^2S_{\frac{1}{2}}$, and the subscript $\frac{1}{2}$ is the j-value of Chapter X. In the P-states, the spin is compounded in two ways with the l-value ($l = 1$) of the orbit; the symbols are $^2P_{\frac{1}{2}}$ and $^2P_{\frac{3}{2}}$. In the D-states, the spin is compounded in two ways with the l-value ($l = 2$) of the orbit; the symbols are $^2P_{\frac{3}{2}}$ and $^2P_{\frac{5}{2}}$; and so it goes onward. What clinches this theory is the character of the Zeeman effect, when an applied magnetic field acts upon the atom. An S-state is resolved into two sub-levels, since the angular momentum of the atom is solely that of the valence-electron, and accordingly has two permissible orientations in the applied magnetic field. Of the paired P-states, one member of each pair is likewise resolved into two sub-levels, the other however into four. This is our first example of what is miscalled the "anomalous Zeeman effect." Now however the situation must be formulated more exactly, in order to give precision both to the foregoing statements and to those which are to follow.

Composition of angular-momentum vectors in the atom

We must immediately recognize and take account of the prolific source of trouble already once mentioned in Chapter XII, where the angular momentum of an orbit was designated both by l and by $\sqrt{l(l+1)}$. In general, it is necessary to assign two different numbers to each of the angular momenta occurring in an atom or a molecule. This is one of the impediments which are unavoidable in fitting visualizable atom-

models to the results of quantal theory, and which to avoid, some theorists would be willing to forego models altogether. Were it not for this impediment, I could say quite simply that in every P-state of the sodium atom, the valence-electron has a spin, with angular momentum $\frac{1}{2}(h/2\pi)$, and is moving in an orbit with angular momentum $(h/2\pi)$; that as regards the two members of each of the aforesaid pairs of P-states, these two angular momenta are oriented parallel for one member and anti-parallel for the other, so that the net angular momentum of the atom-as-a-whole is $\frac{1}{2}(h/2\pi)$ in one case and $\frac{3}{2}(h/2\pi)$ in the other; that when an atom with a net angular momentum of $\frac{1}{2}(h/2\pi)$ is exposed to an applied magnetic field, it orients itself either parallel or anti-parallel to the field, so that the projection of its angular momentum upon the field-direction is either $+\frac{1}{2}(h/2\pi)$ or $-\frac{1}{2}(h/2\pi)$; that when an atom with a net angular momentum of $\frac{3}{2}(j/2\pi)$ is exposed to an applied magnetic field, it orients itself in one or another of four permitted ways so that the projection of its angular momentum upon the field-direction is either $+ 3/2$ or $+ 1/2$ or $- 1/2$ or $- 3/2$ times $h/2\pi$.

This sort of thing *is* frequently said in the literature, and one must realize its limitations. The trouble is, that quantum mechanics prescribes for these angular momenta (but not for their projections on the field-direction!) magnitudes which differ from those which I have been giving. For the spin s of the electron, it substitutes $\sqrt{1/2 \cdot 3/2}(h/2\pi)$ for $\frac{1}{2}(h/2\pi)$; for the orbital motion l in the P-state, it substitutes $\sqrt{1 \cdot 2}$ for the factor unity whereby $(h/2\pi)$ was multiplied; in the two values $j'\ h/2\pi$ and $j''\ h/2\pi$ of angular momentum of the atom-as-a whole, it substitutes $\sqrt{1/2 \cdot 3/2}$ and $\sqrt{3/2 \cdot 5/2}$ for the factors $1/2$ and $3/2$. In order that these new values of j' and j'' may be resultants of s and l, it is necessary that the angular momenta of orbital motion and of spin should be neither quite parallel nor quite anti-parallel to one another! The two permitted values of the angle between them must differ from 0° and 180°; and this also is admitted by quantum mechanics.

The last two clauses of the foregoing paragraph remain, however, unchanged.

Then why not say from the outset that the spin of the electron has the magnitude $\sqrt{1/2 \cdot 3/2}(h/2\pi)$, and that all the other angular momenta occurring inside an atom have the magnitudes assigned to them by quantum mechanics? Inertia of habit, sanctified by years of earlier theories, is itself a mighty obstacle: after so long a time of saying that the electron-spin is 1/2, the world of physics could scarcely get accustomed to saying that really it is $\sqrt{1/2 \cdot 3/2}$ (in terms of $h/2\pi$ as unit). There is also a difficulty of printing to be considered: these numbers have often to be used as subscripts; it is bad enough to print 1/2 or 3/2 as a subscript, without resorting to their quantum-mechanical substitutes. The most serious reason, however, is, that the original 1/2 and 3/2 (and, of course, their analogues in the many other kinds of atomic states) lend themselves uniquely well to stating how many permitted orientations there are. Thus in the present example, I quoted 2 and 4 respectively as the number of permitted orientations in a magnetic field, of certain P states for which the angular momenta were designated as $j(h/2\pi)$ and j had the values 1/2 and 3/2 respectively. I was reminded of those numbers by the rule that they are equal to $(2j + 1)$. Had I kept in mind only the magnitudes $\sqrt{3/4}$ and $\sqrt{15/4}$ assigned by quantum mechanics to these angular momenta, the rule would not have been available.

One should therefore keep in mind both the "marker" or "quantum-number" of an angular momentum, of which the foregoing 1/2 and 1 and 3/2 are examples; and the numerical value assigned by quantum-mechanics to the magnitude of that angular momentum. Fortunately this is rendered easy by the fact that there is a general formula for the latter in terms of the former, with which we can now make acquaintance in the course of taking a deeper plunge into the lush notation of spectroscopy; as follows:

The angular momentum of electron-spin has the quantum-

number s and the magnitude $\sqrt{s(s+1)}$, and s is always equal to $1/2$.

The angular momentum of the orbital motion of an electron has the quantum-number l and the magnitude $\sqrt{l(l+1)}$, and l may have the various values 0, 1, 2, 3 . . . .

The angular momentum of the atom-as-a-whole has the quantum-number j and the magnitude $\sqrt{j(j+1)}$; for the sodium atom j may have the various values $1/2$, $3/2$, $5/2$ . . . .

I next give the rule for compounding the last of these three angular momenta out of the first two, it being as I earlier said a "sort of resultant" thereof.

Rule for compounding l and s into j. If l is greater than s, start with the numerical value of $(l+s)$ and write down all the sequence of numbers spaced at unit intervals from $(l+s)$ to $(l-s)$ inclusive: to wit $(l+s)$, $(l+s-1)$, $(l+s-2)$, . . . $(l-s)$. These $(2s+1)$ numbers are the permitted values of the quantum-number j. If on the other hand s is greater than l, start with the numerical value of $(s+l)$ and write down all the sequence of numbers spaced at unit intervals from $(s+l)$ to $(s-l)$ inclusive; to wit, $(s+l)$, $(s+l-1)$, $(s+l-2)$ . . . $(s-l)$. These $(2l-1)$ numbers are the permitted values of the quantum-number j.

(It will be noticed that this rule is much more generally phrased than is required for the case of sodium, where $s = 1/2$, and it suffices to say that $j = 1/2$ for $l = 0$ and $j = l \pm 1/2$ for $l > 0$. If, however, we were dealing with an atom having more than one valence-electron, s might be replaced by a quantum-number different from $1/2$—not because the individual electrons would have new values of spin, but because the spins of two or more of them would be compounded—and the general phrasing of the rule would then be required.) Now we turn to the rule which defines how a magnetic field resolves a state characterized by j.

Rule for enumerating sub-levels in a magnetic field. Imagine the atom immersed in a magnetic field of strength H. It may then take any of several distinct permitted orientations, these

being denoted by various values of a quantum-number m_j. In any such orientation the projection of the angular momentum of the atom-as-a-whole upon the field-direction is equal to $m_j(h/2\pi)$. To ascertain how many permitted orientations there are and what are the corresponding projections, start with the numerical value of j and write down all the sequence of numbers spaced at unit intervals from $+j$ to $-j$, inclusive: to wit, $j, j-1, j-2, \ldots -j$. These $(2j+1)$ numbers are the permitted values of the quantum-number m_j.

We now start out upon a train of reasoning which leads to the remarkable verifiable formula already once alluded to, the successfulness of which speaks more powerfully than any other single test for the rightness of this elaborate hypothetical structure which has been devised for the atom.

Note, in the first place, that the rule given above for the permitted orientations of the atom in the applied magnetic field is in accordance with a major fact of experience: the energy-values of the atom in its various "magnetic levels," as I shall hereafter call them, are known from spectroscopy to follow upon one another in a uniform evenly-spaced sequence. This is just what the rule requires: for if M denotes the magnetic moment of the atom and θ_j the angle which it makes with the z-direction, the energy due to the field is equal to $MH\cos\theta_j$, which is $MH(m_j/\sqrt{j(j+1)})$, and this changes by uniform steps as m is changed from each of its permitted values to the next.

Now recalling the importance of the ratio of magnetic moment to angular momentum, and noticing that $M/\sqrt{j(j+1)}$ is none other than this ratio, and introducing the g-factor of equation (13.12), and the precession ω linked with it, we may write for ΔU the energy-difference between one magnetic level and the next:

$$\Delta U = g(e/2mc)H = \omega \qquad (13.16)$$

so that a measurement of the energy-difference or separation of two magnetic levels of an atom gives immediately the value of g for that atom. One sees at once how to make a special test of the value 2 assigned to the g-factor of the electron-spin; for

when the sodium atom is in any state for which $l = 0$, it is the spinning valence-electron which contributes the whole of the magnetic moment and the angular momentum of the atom; and when from the spectrum of sodium vapor in the magnetic field the value of ΔU is determined for these states, it is precisely the value $2(e/2mc)H$ which is found.

To get a notion of what actually happens in the general case, it is best to take a sheet of paper and make a graphic composition of the angular momenta. These three—s, l, and j, to denote them by their quantum-numbers—are to be laid down as a triangle having sides of the lengths $\sqrt{s(s+1)}$, $\sqrt{l(l+1)}$, and $\sqrt{j(j+1)}$; for convenience I drop out the common factor $h/2\pi$ for the next few lines. The cosines of the three angles are obtained in terms of the sides by applying the well-known trigonometric formula and getting three equations of which here is one,

$$s(s+1) = l(l+1) + j(j+1) - 2\sqrt{l(l+1)}\sqrt{j(j+1)}\cos\theta_{l,j}. \quad (13.17)$$

The magnetic moment of the orbital electron-motion is a vector parallel to l and of the length $g(e/2mc)\sqrt{l(l+1)}$, with *unity* put as the value of g. The magnetic moment of the electron-spin is a vector parallel to s and of the length $g(e/2mc)\sqrt{s(s+1)}$, with *two* put as the value of g. Owing to the inequality of these g factors, the resultant of the magnetic moments is not parallel to j. We could easily calculate its magnitude and direction, but they are not relevant. It is to be presumed that s and l are constantly revolving or precessing around the direction of j,—the triangle aforesaid is constantly revolving around its side j as a fixed axis, while retaining its size and shape unchanged. If we resolve the resultant angular momentum into a component parallel to j and another component perpendicular to j, the latter will be forever changing in direction and its average will vanish, leaving only the former as a perpetual constant. But this former component we can evaluate by projecting separately upon the j-direction the magnetic moments associated with orbital motion and spin, and adding the two

projections. Thus for the magnetic moment of the atom-as-a-whole we get an effective average which is a vector parallel to the angular momentum of the atom-as-a-whole, and it is of the magnitude:

$$M_a = [\sqrt{l(l+1)}\cos\theta_{l,j} + 2\sqrt{s(s+1)}\cos\theta_{s,j}]\,\frac{e}{2mc}\frac{h}{2\pi} \qquad (13.18)$$

(where I have restored the factor $h/2\pi$). If we work this out with the aid of (13.17) and the similar equation for $\cos\theta_{s,j}$, and then divide it by the angular momentum $(h/2\pi)\sqrt{j(j+1)}$ and by $e/2mc$, we get *the g-factor for the atom-as-a-whole*—commonly denoted by g_j—in terms of the quantum-numbers s, l and j:

$$g_j = 1 + \frac{j(j+1) + s(s+1) - l(l+1)}{2j(j+1)}, \qquad (13.19)$$

and this is the celebrated *g-formula*, which is tested by applying magnetic fields H to atoms, splitting their stationary states into clusters of levels, measuring the separation between successive levels of a cluster, equating it to $\omega = g(e/2mc)H$, evaluating g and comparing it with the value which the right-hand member of (13.19) assumes when in it s, l and j are given the values appropriate to the state from which the cluster of levels was formed. So great is the variety of atomic states, so great the number of different triads of values of s, l, j represented among them, that the study of even a single element like sodium produces many different checks of the validity of (13.19); and since many different elements have been studied, the total of the available verifications of the g-formula, and therefore of the intricate network of its underlying ideas, is considerably impressive.

Revision of the values of magnetic moment

It will be noted that the quantal difficulty of page 396 has intruded itself into the specification of magnetic moments: in particular, a distinction must hereafter be drawn between the

"true value" of magnetic moment on the one hand, and on the other hand the maximum permitted value of the component thereof along the direction of any applied field. The former is $g(e/2m_0c)$ times the magnitude of the angular momentum; therefore it is $g(e/2m_0c)(h/2\pi)\sqrt{J(J+1)}$, where J now stands for the quantum-number of the angular momentum. The latter is $g(e/2m_0c)(h/2\pi)J$, and therefore smaller than the former in the ratio $J:\ \sqrt{J(J+1)}$. This raises an acute terminological problem; for the former deserves the conventional symbol μ on the ground that it is really (or so we now believe) *the* magnetic moment; but the latter also has a claim, on the ground that it is what is measured in the Gerlach-Stern experiment when the magnetic moment of the atom is turned as nearly as is possible into parallelism with the direction of an applied field. For the present I will deal with the situation by letting μ continue to stand for the "true value" of magnetic moment, and revising the statements of page 394 as to the spinning electron.

As for the electron-spin, we are now to assign to it:

a g-value of 2;

an angular momentum of magnitude $\sqrt{\frac{1}{2}(\frac{1}{2}+1)}(h/2\pi)$, the projection of which upon the direction of an applied field is $\pm\frac{1}{2}(h/2\pi)$;

a magnetic moment of magnitude $2(eh/4\pi m_0c)\sqrt{\frac{1}{2}(\frac{1}{2}+1)}$, the projection of which upon the direction of an applied field is $\pm eh/4\pi m_0c$ or one Bohr magneton.

Interpretation of the multiplicities of multiplet term-systems in atomic spectra

Of the various multiplicities or term-systems described in Chapter X, the only one which we have yet interpreted is the doublet system. There, the spin of a single electron is compounded with the orbital angular momentum thereof, and this brings about a doubling of all save the S-terms, which remain

single because the orbital angular momentum is zero. We have not however avoided the making of a tacit assumption about the composition of the spins and the orbits of electrons greater in number than one; for all of the instances of doublet spectra, excepting those exhibited by H I and He II, are furnished by atoms with many electrons; and we have in effect been assuming that in every such case, the spins and the orbits of all the electrons save one compensate one another so well as to lead to a net magnetic moment and a net angular momentum of zero. Such is the property of so-called "closed shells" of electrons. There are however very many other cases, and of these there is a category to which the other multiplet term-systems are due.

These are the cases of "Russell-Saunders coupling," or "LS-coupling," in which it appears that on the one hand the electron-spins conspire with each other to form a resultant of their own, and on the other hand the electron-orbits conspire with each other to form a resultant of *their* own, and finally these two resultants conspire with one another. The simplest cases occur when there are two electrons—either because there are no others in the atom (the spectrum He I) or because all but two are integrated into closed shells (the spectra of the alkaline-earth metals and of other elements in the second column of the Periodic Table). Both singlet and triplet term-systems are found in each of these spectra: in every state of the former character the two electron-spins are anti-parallel, in every one of the latter the spins are parallel.[1] Singlet states have a peculiar interest, inasmuch as they and they alone exhibit the "normal" Zeeman effect. The spins, so to speak, have driven one another into hiding, and the atom thus may conform both to the early form of quantal theory—that of Chapter XI—in which electron-spin is disregarded, and to the classical theory; but the mutual cancellation of the spins would

[1] *N.B.* In both atomic and nuclear theory, when a model involves two angular momenta each of quantum-number ½, it is customary—for brevity, I suppose—to speak of it as being in a singlet or in a triplet state, according as the angular momenta are antiparallel or parallel.

not by itself suffice to achieve this, were it not that the angular momenta of the two orbits compound themselves into a resultant of which the L-value—for capital letters are used instead of lower-case letters, in referring to these resultants—is taken from the same train of integer numbers 0, 1, 2, 3 . . . as are permitted to the individual l-values. Owing also to this cancellation of the spins, the individual states of a singlet system are single however great the L-value, as the name implies. Triplet states, on the other hand, are clusters of three all along the P-column, the D-column, and the others of still greater L-value, for there are three permitted orientations which the paired electron-spins—frozen together, as it were, into a single vector of quantum-number unity—may assume relatively to the resultant of the orbital angular momenta. Only in the S-column does this last resultant vanish, so that orientation ceases to matter, and the states are single.

In the general case (of Russell-Saunders coupling of two or more electrons) the orbital angular momenta l form a resultant of magnitude $\sqrt{L(L+1)}$ and quantum-number L, which last is permitted the integer values 0, 1, 2, 3 . . . . This quantum-number L it is, which is zero for all S-states, one for P-states, two for D-states and so onward. The electron-spins s form a resultant of magnitude $\sqrt{S(S+1)}$ and quantum-number S. If there are x electrons apart from the closed shells, their spins may compound themselves so that S is equal to $\frac{1}{2}x$, or to $\frac{1}{2}(x-2)$, or to $\frac{1}{2}(x-4)$, and so onward as far along the sequence as these numbers remain positive. (In other words, the electron-spins may all agree, or an even number of them may cancel each other in pairs.) The two resultants now form a super-resultant, the "angular momentum of the electron-family as a whole," having magnitude $\sqrt{J(J+1)}$ and quantum-number J. The rule for compounding L and S into J is the very one which I have lately given (page 399) with lower-case letters instead of capitals. If S is smaller than L, we may consider it as being dominated by L, and orienting itself in as many different ways with respect to L as it would with

respect to an applied field, to wit, $(2S + 1)$ ways; if L is smaller than S they are to be interchanged in the foregoing statement; if L and S are equal, it makes no difference which we consider as dominating the other. The outcome is, that to the quantum-number J are permitted either $(2L + 1)$ or $(2S + 1)$ different values, whichever of these numbers is not the smaller. These J-values are integers or "half-integers" (*i.e.*, numbers chosen from the list $1/2, 3/2, 5/2, \ldots$) according as S is an integer or a half-integer.

Now the student can readily prove for himself that from these rules of composition follow the strange and striking sequences of numbers which distinguish the eight known multiplicities, and which were exhibited in Table 10.2. The degree of the multiplicity is determined by the quantum-number S; it is $(2S + 1)$. If there are x electrons apart from the closed shells, the multiplicity is odd or even according as x is even or odd. Now, closed shells always comprise even numbers of electrons, as I shall presently repeat; therefore an atom or an ion displays only odd or only even multiplicities, according as it has an even or an odd number of electrons—the principle stated, as a fact of experience, on page 318.

This success in interpreting the laws of the multiplicities is one of the most important points in favor of the atom-model and the laws of composition adopted in these pages. Another is the success of the generalized g-formula. Generalizing from the case of the single electron with its spin and its orbital angular momentum, we find—or rather, it is found from quantal theory—that the vectors L and S precess around the direction of their resultant J; to L we assign the g-factor unity, to S the g-factor two; and retracing the reasoning of pages 400–402, we arrive at the g-formula (13.19) with capital letters replacing the lower-case ones. Most of the verifications of the g-formula pertain to this generalized form thereof, and the ensemble of them is notable.

So well established is the theory in general and the g-formula in particular, that the accurate measurement of Zeeman patterns—that is, the wave-length-differences between the lines

of such patterns—in accurately-known magnetic fields has been for several years one of the most highly-esteemed of the ways of evaluating e/m_0. It is the one to which allusion was made on page 49; and it provides a remarkable illustration of the occasional superiority of "indirect theory" over "direct experiment"; for, as I there pointed out, the value which it gave for e/m_0 was for years in definite disagreement with that derived from the deflection of free fast-moving electrons by electric and magnetic fields, and when agreement came at last it was through emendation of this latter.

This seems as good a place as any to mention that the Russell-Saunders coupling, wherein the spins on the one hand and the orbital angular momenta on the other hand compound themselves into separate resultants before these combine themselves into a super-resultant, is not universal. There are also examples of so-called "*jj*-coupling," in which the spin and the orbital angular momentum of each electron separately compound themselves into a resultant j, and these resultants for the several electrons combine themselves into the super-resultant which is the angular momentum of the electron-family as a whole. There are also innumerable gradations between the two types, in the pursuing of which the ultimate efforts of such theoretical physicists as still concern themselves with spectra are now being expended. These difficult cases infest the atoms toward the right-hand side of the Periodic Table. We turn however to the case in which the coupling is abolished by the applied magnetic field.

Concluding remarks on the Zeeman effect; decoupling by a strong magnetic field

Choose any values for L and S, write down the permitted values for J (in Russell-Saunders coupling), enumerate the $(2J + 1)$ sub-levels corresponding to each of these in an applied magnetic field, and add; it will be found that the sum is $(2L + 1)(2S + 1)$.

Suppose however that the field were able to act upon the

two angular momenta L and S as though they were entirely detached from one another, each capable of orienting itself according to the laws expounded in preceding pages, in complete disregard of the other. There would then be $(2L + 1)$ permitted orientations of the former and $(2S + 1)$ of the latter, and there would be $(2L + 1)(2S + 1)$ conceivable pairings of the two—the same number as that of the sub-levels aforesaid.

It would be futile indeed to point out this agreement, were it not that both of the cases are realized in Nature, and for the same atoms: the former, which thus far we have alone considered, in weak magnetic fields, and the latter in strong ones. The adjectives "weak" and "strong" must be explained. When H is zero, each of the J-values permitted by the given L and the given S corresponds to what I have been calling a "state." As H is gradually increased from zero, each state is resolved into a set of sub-levels according to the laws which we have just been studying, and at first the separation between the extreme members of any set is proportional to H and small compared to the initial separation between two states. So long as this is true, the field is called "weak." Now, if the separation between members of a set were to remain rigorously proportional to H however much this is increased, the different sets would eventually overlap one another. This however does not happen; instead, the phenomenon undergoes a gradual change as H increases, and finally it approaches a form which reveals a new state of affairs prevailing within the atom. This limiting state is that of "decoupling" in the sense of the previous paragraph. Though the electron-family of the atom is by no means torn apart, yet the internal forces—supposedly magnetic—which in weak fields achieve the Russell-Saunders coupling are now overpowered by the applied magnetic field, which dominates each of them by itself. When this limit is closely approached the field is called "strong"; and H is strong in this sense, if it is so great that had the internal coupling persisted, the individual sets of sub-levels would have extended over a frequency-range several times greater than does the cluster of the original stationary states when there is no field.

All this is inferred from Zeeman patterns. There will be space for only one example: I will take the D-lines of sodium, which are transitions between two states $^2P_{\frac{1}{2}}$ and $^2P_{\frac{3}{2}}$ respectively, and a state $^2S_{\frac{1}{2}}$. This last is the normal state with which we have had so much to do, and since in it the orbital angular momentum vanishes, there is no coupling even in the weak field and nothing novel is to be expected in the strong. Its m-values are $+\frac{1}{2}$ and $-\frac{1}{2}$. The P-states in the weak field have respectively two sub-levels ($m = +\frac{1}{2}, -\frac{1}{2}$) and four ($m = +\frac{3}{2}, +\frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}$). The D-lines would be resolved into four and eight Zeeman components respectively, were it not that the selection-principle for m prohibits two of the latter group, leaving four and six: these are duly observed, and the predicted g-values are verified. So much for the weak-field case! Now, in the strong field, there are still six sub-levels corresponding to the pair of P-states, but they are not to be particularly assigned two to the one and four to the other. For three of these six, the spin is parallel (in the sense of page 397) to the field, and the orbital angular momentum ($l = 1$) assumes its three permitted orientations: by a self-explanatory notation, m_s is $\frac{1}{2}$ in all three cases, m_l has the values $+1$, 0, -1. For the other three, the spin is anti-parallel to the field, and l assumes the same three orientations: m_s is $-\frac{1}{2}$, m_l is 1, 0, -1. As for the S-state, m_s is $\frac{1}{2}$ for one sub-level and $-\frac{1}{2}$ for the other, m_l also zero. Now finally comes the selection-principle for these "strong-field quantum-numbers": in a transition, m_s does not change, m_l changes by one unit or not at all. The student will deduce the principle that *in strong fields the Zeeman pattern simulates the normal Zeeman effect.*

These phenomena in the strong magnetic field, and also the gradations by which they are approached as the field passes from weak to strong, are known by the collective name of "Paschen-Back effect," after the physicists who (in 1912) made the first observations in strong fields, and many thereafter. The study of the gradations, by the way, involves some of the most difficult and masterful work in theoretical physics.

It is too advanced for this book, but an analogous case will be mentioned in the account of hyperfine-structure.

As to the polarization and the intensity-distribution of the components of a Zeeman pattern: they were roughly described, on page 384, for the cases $\Delta m = 0$ and $\Delta m = \pm 1$ in the "normal" Zeeman effect; well, the same descriptions hold for the same cases in the general Zeeman effect, and if m_l be substituted for m they hold also for the strong-field effect.

Closed shells and the periodic table

I recall that the atomic numbers of the elements in the "zero" column of the periodic table—the noble gases, the elements which terminate the periods and which are distinguished above all others by their chemical and physical inertness—form the sequence 2, 10, 18, 36, 54, 86. There are 2 elements in the first period, 8 in the second, 8 in the third, 18 in the fourth, 18 in the fifth, 32 in the sixth.

I recall also from Chapter XI that the degree of degeneracy of an orbit having the quantum-number l is $(2l + 1)$; and that to each value of the total quantum-number n, which we encountered in the listing of the stationary states of hydrogen, there correspond the l-values which are the integers from 0 to $(n - 1)$ inclusive, so that to the n-values 1,2,3,4, . . . there correspond 1, 4, 9, 16 . . . sub-levels respectively. Or rather, there *would* correspond so many sub-levels, were it not for electron-spin which was disregarded in the previous sentence. Now, the effect of electron-spin is to replace each of those sub-levels by two; and consequently[1] to the n-values 1, 2, 3, 4, . . . , there correspond 2, 8, 18, 32, . . . sub-levels respectively.

This sequence of numbers resembles that of the numbers-per-period of the elements in the table, and yet is not altogether the same. Suppose however for a moment that both of these sequences were the same—2, 8, 18, 32, . . . : we might

[1] To each l-value now are associated the two j-values $(l + \frac{1}{2})$ and $(l - \frac{1}{2})$, with $2l + 2$ and $2l$ sub-levels respectively, or $2(2l + 1)$ altogether.

then adopt the following conception of an atom of atomic number Z. Around the nucleus there would be first, two electrons of total quantum-number unity, corresponding to the two sub-levels compatible with $n = 1$; then, eight electrons of total quantum-number two, corresponding to the eight sub-levels compatible with $n = 2$; then, eighteen electrons of total quantum-number three, thirty-two for which $n = 4$, and so forth. This would go on, until the total number of electrons amounted to Z. If Z happened to be 2 or 10 or 28 or 60, then any value of n which was represented at all in the electron-family would be represented by as many electrons as could simultaneously have it, and we might speak of such an atom as consisting of "closed shells" of electrons. We should expect these particular Z-values to belong to elements like the noble gases. An alternative way of describing the idea would be the following: imagine the consecutive atoms of the table of the elements to be formed, each in turn by adding to its predecessor one electron in the electron-family (and the necessary charge and mass to the nucleus!); then, this step-by-step formation would consist in putting two electrons into a shell for which $n = 1$, then eight into a shell $n = 2$, then eighteen into a shell $n = 3$, then thirty-two into a shell $n = 4$, and so forth. Whenever a shell was just completed, the element in question would be a noble gas.

Well, this conception does not adequately represent the reality, since the Z-values of the noble gases are not 2, 10, 28, 60, but 2, 10, 18, 36, 54, 86; yet it comes so close to the actual conception, that I have thought it worth introducing as a prelude. Before amending it, I mention that the shells bear the names already used in Chapter VIII: K-shell for $n = 1$, L-shell for $n = 2$, M-shell for $n = 3$ and so forth. Now employ the image of the building-up of the consecutive atoms by addition of consecutive electrons. It is necessary to suppose:

that the K-shell is begun at hydrogen ($Z = 1$) and completed at the first noble gas, helium ($Z = 2$);

that the L-shell is begun at lithium ($Z = 3$) and completed at the second noble gas, neon ($Z = 10$);

that the M-shell is begun at sodium ($Z = 11$) but *not* completed at the third noble gas, argon; for this element occurs at $Z = 18$, and the first eight electrons of total quantum-number 3 are thus able to constitute a closed shell without awaiting the remaining ten;

that the N-shell is begun at potassium ($Z = 19$) and continued for a while; but before long the process of building the M-shell recommences(!) and is carried through to its finish, whereupon the N-shell is brought up to a population of eight and the fourth noble gas krypton is reached at $Z = 36$, where the M-shell is complete and the N-shell has only eight of its electrons—these eight also are able to constitute a closed shell without awaiting the remaining twenty-four;

that the O-shell is begun at rubidium ($Z = 37$) and continued for a while; but before long the process of building the N-shell recommences and is carried on, not however to its finish but only until the N-shell comprises eighteen electrons and fourteen vacancies, whereupon the O-shell is brought up to a membership of eight and the fifth noble gas xenon is reached at $Z = 54$;

that the P-shell is begun at caesium ($Z = 55$) and continued for a while; but before long the process of building the N-shell is resumed and carried to completion, and next the O-shell is brought up to a membership of eighteen by the addition of ten electrons, and then the construction of the P-shell is resumed and it is brought up to a membership of eight and the sixth noble gas (emanation) is reached at $Z = 86$;

that the Q-shell is then begun, but the procession of the elements comes so soon thereafter to an end (owing to the non-existence of stable nuclei for which $Z > 92$) that its progress cannot be followed far.

So, the odd combination of partial agreement and partial disagreement between the sequence 2, 8, 18, 32, . . . of the sub-level numbers and the sequence 2, 8, 8, 18, 18, 32 of the period-lengths is not after all a mere accident! Insofar as it is agree-

ment, it signifies that the electron-families of atoms are classifiable into shells, which correspond to the values 1, 2, 3, 4, 5, 6, 7 of the quantum-number n and which individually are completed when there are as many electrons to a shell as there are sub-levels to its n-value—that is to say, $2n^2$ electrons. Insofar as it is disagreement, the disagreement means that the building-up of the consecutive atoms of the periodic table, by consecutive addition of electrons, is not so simple as a mere completion of the consecutive shells in turn. One shell may be brought to an intermediate stage, a new one begun, the old one carried on to completeness or merely to another intermediate stage, yet another new one begun—such are the implications of the ordering of the periodic table.

Moreover it appears from that ordering, that a shell is closed when it contains eight or eighteen electrons, even if for its completion a greater number is required. Now it becomes necessary to define "closed shell" with greater care. Thus far I have been tacitly defining it as a system of electrons, such that if the electron-family of an atom consists of nothing but closed shells, the element to which the atom belongs is a noble gas. For all such atoms in their normal states the angular momentum and the magnetic moment of the electron-family are zero; the symbol of the normal state is $^1S_0$. There are however other atoms of which the same is true, notably those of the elements which lie two steps beyond the noble gases, such as beryllium ($Z = 4$), aluminium ($Z = 13$) and the rest.[1] Let me re-define so as to say that if the electron-family of an atom consists of nothing but closed shells, the normal state of the atom is $^1S_0$, devoid of angular momentum and of magnetic moment. It turns out then that shells may be closed if they contain 2 or 8 or 18 or 32 electrons, irrespective of whether they are completed. It may be inferred that every closed shell by itself has an L-value of zero and an S-value of zero, and this

[1] I do not mean to imply that if μ and p both vanish the state of the atom is necessarily $^1S_0$; there are other possibilities, one of which is realized in the normal states of several atoms—it is $^3P_0$, in which L and S are both equal to 2 and the corresponding vectors are antiparallel.

is borne out by observations on yet other atoms and by quantal theory.

It seems excessive to bother the student with yet another terminology, especially since in this book so little use is to be made of it; yet he will scarcely be able to avoid it in the further study of physics, and happily it does not require much explanation. The two electrons of the K-shell are known as $1s$ electrons; the eight of the L-shell are classified as a pair of $2s$ electrons ($l = 0$) and a set of six $2p$ electrons ($l = 1$); the eighteen of the M-shell comprise two $3s$, six $3p$ and ten $3d$ electrons ($l = 0, 1, 2$, respectively); the thirty-two of the N-shell comprise two $4s$, six $4p$, ten $4d$ and fourteen $4f$ electrons, and so it goes onward. I now restate a little of the content of page 411, to show how this terminology is used and how it permits of a description both briefer and fuller than the earlier one. Imagining again the building-up of the atoms by adding one electron after another; the first period of the table (H, He) displays the successive addition of the $1s$ electrons; in the second arrive the $2s$ electrons (Li, Be) and then the $2p$ group; in the third arrive the $3s$ electrons (Na, Mg) and then the $3p$ group; at potassium arrives the first member of the N-shell which is a $4s$ electron, at calcium arrives the second $4s$ electron, and now at scandium commences the addition of the $3d$ electrons which are required to complete the M-shell. The famous iron triad occurs while the $3d$ group is being inserted, the palladium triad while the $4d$ group and the platinum triad while the $5d$ group is being installed. The fourteen rare-earth elements from cerium ($Z = 58$) to lutecium ($Z = 71$) inclusive correspond to the installation of the fourteen $4f$ electrons one after another, the $5s$, $5p$ and even the $6s$ electrons being already in place: these elements differ but little in their chemical properties, and this quality is ascribed to the fact that the outermost shells are the same in all, the distinctions being localized in an inwardly-lying shell.[1]

[1] One must not however go so far as to imagine each shell as lying completely within all those which follow it; if a geometrical model is to be used at all, a good deal of overlapping of shells must be admitted.

The electron-family of an atom (or an ion) is described symbolically in a manner which I will illustrate by applying it to the case (already described in words) of neutral potassium in its normal state: $1s^2 2s^2 2p^6 3s^2 3p^6 4s^1$. The superscript following the symbol for each set of electrons shows how many there are in the set. An electron-family consists entirely of closed shells, in the sense of the foregoing paragraphs, if every exponent is 2 or 6 or 10 or 14 according as the letter which it follows is *s* or *p* or *d* or *f*. Every excited state has a description of its own. In the cases which we have considered almost exclusively in this book—such as those which are responsible for the line-series in the spectra of potassium and other alkali metals—these differ from the description of the normal state only in that the portion referring to the last electron ($4s^1$, in the case of potassium) is replaced by symbols appropriate to the various permitted orbits of this electron; but this simplification is by no means universal. One may say that as far as the theorist is concerned, the object of spectroscopy is to fix the descriptions of the normal and the excited states of all the kinds of atoms and ions. For the normal states of the neutral atoms this task is already almost completed, and the results are embodied in tables.[1]

The principle underlying most of this reasoning is known by the name of W. Pauli. It is also called "Pauli's prohibition" (*das Pauli'sche Verbot*), as it is considered as "forbidding" more than $2n^2$ electrons from sharing any *n*-value, and more than $(2l + 1)$ from sharing any *n*-value and *l*-value simultaneously. One may go further, and assign *m*-values—*m* being the magnetic quantum-number of page 384—to the several electrons which share a common *n*-value and a common *l*-value; then, the principle forbids more than two electrons to share a given set of values of *n*, *l*, *m*. Finally one may assign a pair of "spin quantum-numbers," *e.g.*, $+\frac{1}{2}$ and $-\frac{1}{2}$, to differentiate between the two of such a pair; and the most frequent formulation of the Pauli prohibition is phrased accordingly; *no two electrons of an atom may share the same values of all four quantum-numbers n, l, m, and s.*

[1] See for instance E. C. Stoner's *Magnetism and Matter*, pp. 546–9.

CHAPTER XIV

Miscellany as to Spectra

The alkali-metal atoms again

We have viewed the structure of the level system of the alkali metals in Chapter X, and we have interpreted that structure by invoking our general conceptions of electron-spins and orbital angular momenta and the laws of composition thereof in Chapter XIII. If I return yet again to the topic, it is to fill in some details which bear on the scale of that structure —on the actual energy-values of the various levels, which as yet we have not touched.

The levels indicated in Figure 10.4 are supposed to correspond to different orbits (in the Bohr model) or different $\Psi\Psi^*$-functions (in the wave-model) of the valence-electron. If the closed electron-shells of an alkali-metal atom of nuclear charge $+Ze$ were really continuous enclosing walls of charge amounting altogether to $-(Z-1)e$, then beyond the outermost wall the field would be the same as that of a hydrogen atom. If for any particular level the orbit or the Ψ-waves lay entirely beyond such an outermost wall, that level would be identical with one of those of the hydrogen atom, and its energy would have one of the values $-Rh/n^2$. This over-simplified notion should come closer to the truth, the greater the values of n and l. Thus for the levels in any one column of Figure 10.4, we should expect the approximation to the energy-values $-Rh/n^2$ to improve from the bottom toward the top. For corresponding levels in different columns, we should expect the approximation to be worst for the S-column, better for the P-column, better yet for the D-column and so forth.

These expectations are confirmed.[1] Graphically, the justification is shown in Figure 14.1, where the horizontal dotted lines mark some of the energy-values $-Rh/n^2$ (the one for $n = 1.5$ is inserted merely to show the scale). The most left-

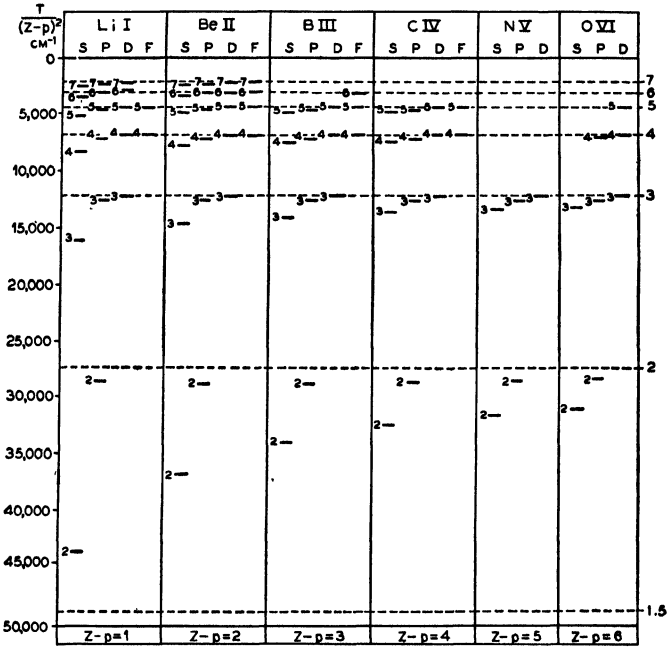


FIGURE 14.1—STATIONARY STATES OF THE ATOMS OF AN ISO-ELECTRONIC SEQUENCE (3 ELECTRONS) BEGINNING WITH NEUTRAL LITHIUM. HORIZONTAL DASHED LINES INDICATE THE VALUES $-Rh/n^2$ (n MARKED ON RIGHT). G. Herzberg; *Atomic Spectra and Atomic Structure*; Prentice-Hall.)

ward of the compartments refers to lithium, and the horizontal dashes mark the actual energies of various S, P, D and F levels. As for the other five compartments, they refer to ions of an

[1] Algebraically, the justification is shown by the fact that spectroscopists are in the habit of using a formula $-Rh/(n + a)^2$ for the actual energy-values of the levels; here a stands for a constant which has characteristic values for the several columns, being greatest for S, next greatest for P and so on. This is the so-called "Rydberg approximation" to the actual terms; there are others which are closer.

"iso-electronic sequence" (page 319), having in common a closed shell of two electrons, but differing from each other in the value of nuclear charge which runs from $4e$ to $8e$. It is easy to see that to make the like comparison for them we ought to divide the energy-values of their levels by $(Z - 2)^2$ before plotting in Figure 14.1; this has been done. One sees that for given n and l the approximation to $-Rh/n^2$ is better the greater the Z-value. This is not at all surprising, as with increasing nuclear charge the closed shell is expected to shrink and the region of nearly inverse-square field to begin closer in.

To calculate the actual energy-values is a task of advanced atomic theory, involving postulates as to the field which are tested by the results. The task is most important and most difficult for the lowest levels, of which the energy-values deviate the most from $-Rh/n^2$. Hitherto I have spoken as though the values of total quantum-number n were independently known for all the levels. Quite the contrary! for all but the lowest levels, it is by the use of such diagrams as Figure 14.1 that the n-values are nowadays fixed, being so chosen as to minimize the discrepancies between the actual energies and the values of $-Rh/n^2$. For the lowest S-level of lithium, and for several low levels of each of the other alkali metals, the discrepancies are so large as to cast doubt on this procedure, and the calculations are invoked. At present the n-values accepted for the lowest levels of the S, P, D, F columns are: in lithium, 2,2,3,4 respectively; in sodium, 3,3,3,4; in potassium, 4,4,3(*sic*), 4; in rubidium, 5,5,4,4; in caesium, 6,6,5,4. One sees that these are far from the order-numbers introduced at page 310 before the concept of quantum-number was encountered.

Having thus far in this section disregarded fine structure, I now recall that everything thus far called a "state" or "level," the S-states alone excepted, is actually a pair of levels; that the two members of any such pair are alike in having a common n-value and a common l-value, and differ in having respective values $(l + \frac{1}{2})$ and $(l - \frac{1}{2})$ of a third quantum-number j; and that the physical difference between the two members is con-

ceived to lie in the orientation of the electron-spin with respect to the electron-orbit, the two angular momenta being parallel for one and anti-parallel for the other. (In the last clause, "parallel" and "anti-parallel" are used in the special senses of page 397). This conception leads to a testable formula for the energy-difference between the members of the pair, the so-called "spacing of the doublet." I will presently quote the formula as quantal theory supplies it, after having derived an approximate form of it for the simplest conceivable case.

Reviving the primitive Bohr model, imagine the circular orbit of radius r_n described by an electron of linear velocity v_n and angular momentum $p = nh/2\pi$ around a nucleus of charge $+ Ze$. We shall require the value of p/r_n^3, which with the aid of equations 11.1 to 11.3 the student may work out to be $(32\pi^5 m^3 e^6/n^5 h^5)Z^3$. Now interchange the motions, supposing the nucleus to be describing the circle about the electron conceived as stationary. Imitating the argument of an earlier passage (page 395), we identify the circulating nucleus with a current of strength $Zev_n/2\pi r_n c$, producing at the site of the electron a magnetic field of which the strength is Zev_n/cr^2, equal to $(Ze/mc)(p/r_n^3)$. Now impose upon the electron the property of spin, with the accepted values of magnetic moment, angular momentum and gyromagnetic ratio. In this field it should precess with an angular velocity e/mc times as great as the fieldstrength. Now transform back to the original motion, with the nucleus stationary and the electron revolving. In effecting this transformation we are to halve the rate of precession, as Thomas[1] has shown; it thus becomes $(Ze^2/2m^2c^2)(p/r_n^3)$. Next we should multiply this quantity by $\frac{1}{2}(h/2\pi)$ $\cos\theta$, to get the value of the extra magnetic energy correlated with this precession when the spinning electron of angular momentum $\frac{1}{2}(h/2\pi)$ is inclined at angle θ to the plane of the orbit. We are however interested only in the energy-difference ΔU between two orientations which, while using this primitive model, we shall visualize as being exactly opposite to one another and exactly at right angles to the plane of the orbit:

[1] Thomas, *Nature* 117, 514 (1926).

$\theta = 0°$ and $\theta = 180°$. We may therefore content ourselves with multiplying the precession by $(h/2\pi)$; and after substituting the aforesaid value for p/r_n^3, we shall find:

$$\Delta U = (8\pi^4 me^8/h^4c^2)n^{-5}Z^4 \qquad (14.1)$$

Before commenting on this formula I regroup the multiplying factors, introducing two combinations of the fundamental constants denoted by R and α:

$$R = 2\pi^2 me^4/h^3, \qquad \alpha = 2\pi e^2/hc \qquad (14.2)$$

R is the (theoretical) Rydberg constant which we have already met in Chapter XI, while α is called "the fine-structure constant." (I mention in passing that α is dimensionless, and that it plays a great and peculiar role in many recent theories of physics). The formula now becomes:

$$\Delta U = (R\alpha^2 h)n^{-5}Z^4 \qquad (14.3)$$

Having been derived for circular orbits only, it is applicable as yet only to states for which n and k are equal, or l is smaller by one unit than n. This is a grave limitation, which must be removed. Invoking once more the Bohr model and thinking of the elliptical orbits, we see that ellipticity must matter a good deal, and that for a given n-value the dependence of energy on orientation of electron-spin must increase with decrease of k, because of electron and nucleus being closer together at all but one point of the orbit (page 337). It is found that $n^{-3}k^{-2}$ must be substituted for n^{-5}; but the time has come again for us to forsake the original model of Bohr, and accept the formula derived from wave-mechanics, which is

$$\Delta U = \frac{R\alpha^2 h}{n^3 l(l+1)} Z^4 \qquad (14.4)$$

This is the form in which the theory is submitted for verification.

The principal test of this formula is made upon the isoelectronic sequence cited earlier in this section (Figure 14.1). If we could assume that the valence-electron lies always and

entirely beyond the closed shells, we might anticipate a verification not indeed of equation 14.4 as it stands, but of this equation with $(Z - 2)^4$ substituted for Z^4. This however is too simplified an assumption, as we have just been seeing. The next step consists in assuming a formula like 14.4, except that Z^4 is replaced by $(Z - s)^4$; here s is regarded as a constant to be determined by measuring ΔU. For the six isoelectronic types ranging from Li I to O VI, Bowen and Millikan found values 2.019, 1.937, 1.884, 1.858, 1.838 and 1.816 for s; these differ from one another and from 2 by amounts sufficiently small to justify us in attributing the discrepancies to the difference between the actual field in the regions where the valence-electron circulates, and the idealized field which we have been postulating.[1]

Curious history of the theory of the hydrogen atom

Considered from the viewpoint of contemporary theory, the hydrogen atom seems to have been designed by Nature with malice, in order to mislead. Physicists have wanted to regard it, ever since the earliest days of the nuclear atom-model, as the limiting case to which the alkali-metal atoms are approximations—the very idea which was developed in the section preceding this one. In that section we saw how support is found for the idea by comparing the energy-values of the stationary states of lithium with those of the sequence $-Rh/n^2$ characteristic of hydrogen. No regard was paid to multiplet structure in making that comparison. It is in regard to multiplet structure that the strange deception occurs; for the structure of the alkali-metal spectra is explained by assuming for each value of the total quantum-number n a set of $(2n - 1)$ levels classified as S, P, D, and so forth; but that of the hydrogen spectrum is interpreted by assuming n levels for each of the n-values.

This notion of the hydrogen structure dates from 1915, when Sommerfeld made the theory already mentioned in Chapter XI,

[1] Gibbs and White, **PR 33**, 157 (1929); Bowen and Millikan, **PR 27**, 144 (1926) and earlier papers there cited.

according to which for every n-value there are n permitted elliptical orbits having values $kh/2\pi = (1,2,3 \ldots n)h/2\pi$ of angular momentum. By ordinary mechanics their energy-values coincide at $-Rh/n^2$; but by relativistic mechanics (as Sommerfeld showed) these differ slightly from $-Rh/n^2$ and from each other. Each of the so-called "lines" of the hydrogen spectrum thus far considered should then be a cluster of closely-set lines; or, in the customary phrase, each line should have components. Omitting the supposedly-forbidden transitions (all except those for which k changes by one unit) one readily finds how many components should go to any line; each of the Balmer series should have three. This however was but a minor part of Sommerfeld's theory, for he also computed the frequencies and the (relative) intensities to be expected for the components. In the Balmer-series lines one of the three components should be insignificantly faint, leaving the other two to appear as a doublet of spacing about 0.335 cm^{-1}—that is to say, a doublet resolvable in a very good spectroscope. Now, the Balmer lines do appear as doublets in very good spectroscopes, and the agreement of actual with calculated spacing is such that for a decade it was listed as one of the strongest supports of the atom-model of Bohr.

In 1925 this theory still persisted, but so also did the realization that for the alkali-metal atoms there are not n but $(2n - 1)$ different levels for each and every n-value. By this time the contrast was growing so uncomfortable that almost at once several savants were driven to make the suggestion that for hydrogen also there may be $(2n - 1)$ levels, $(n - 1)$ of them practically coinciding in energy with $(n - 1)$ of the others.[1] The suggestion was more specific. We have seen how the levels corresponding to any given n-value in an alkali-metal spectrum are classified into one of S-type, two of P, two of D and as far on as the n-value permits. The conjecture was that in hydrogen, the lower P-level may coincide with the S-level,

[1] Slater, **PNAS** 11, 732 (1925); Uhlenbeck and Goudsmit, *Physica* 5, 266 (1925); Bichowsky and Urey, **PNAS** 12, 80 (1926); Uhlenbeck and Goudsmit, *Nature* 117, 264 (1926); Thomas, *ibid.* 514.

the upper P-level with the lower D-level, the upper D-level with the lower F-level, and so on. The coinciding members of each pair would then have l-values one unit apart, and (as the student may convince himself) identical j-values. I put the matter more succinctly by using symbols (n, k, j) to denote the levels. Any two levels $(n, l, l + \frac{1}{2})$ and $(n, l + 1, l + \frac{1}{2})$ are separate in the alkali-metal term-systems, but it is conjectured that in the limiting case of hydrogen they may become coincident in energy.

At the time when this idea was put forth, it was a leap in the dark, and a leap away from the solid ground of numerical agreement between theory and data on which the Sommerfeld conception had succeeded in planting itself. The solid ground was soon to be regained. In the spring of 1926, Uhlenbeck and Goudsmit introduced the spinning electron into atomic theory, employing non-relativistic mechanics to begin with. With it they interpreted the alkali-metal spectra in the fashion which we have followed, and then they turned to the great question: in the limiting case of hydrogen, did it lead to term-values coinciding in pairs in the desired way, and agreeing with the data as well as had the Sommerfeld values? By itself, it did not; but there remained to be taken into account the effect of replacing ordinary by relativistic mechanics. Had this last continued to be done in Sommerfeld's way the desired end would not have been attained; but presently it was done anew by the methods of quantum mechanics.[1] When thus the concepts of relativity and wave-mechanics and electron-spin were all integrated with each other, it was found that the energy-values of any two levels $(n, l, l + \frac{1}{2})$ and $(n, l + 1, l + \frac{1}{2})$ do indeed coincide—and not with one another only, but also with the value computed by Sommerfeld for that particular elliptical orbit (in the Bohr model with the spinless electron) having n for its total quantum-number and $(l + 1)$ for its k-value.

Remembering how for any n-value k runs from 1 to n while l runs from 0 to $(n - 1)$, one may gather that Sommerfeld's theory and the new one make exactly the same predictions in

[1] Heisenberg and Jordan, **ZP 37**, 263 (1926); Richter, **PNAS 13**, 476 (1927).

all respects, so that nothing which can be observed has any bearing on the choice to be made between them. This is however not so, for the theories differ in their predictions about the relative intensities of the components. Statements about intensities lie beyond the scope of this book, excepting the selection-principles which may be paraphrased as statements that certain lines have vanishing intensities and others do not; but, even when the inquiry is limited to these, distinctions can be drawn. The selection-principle for the new theory is that of the alkali-metal spectra, and may be regarded as a fact of experience: we have already seen (page 316) that it permits the transitions for which $\Delta l = \pm 1$, $\Delta j = \pm 1$ or 0. The selection-principle for Sommerfeld's theory may be divined from the correspondence[1] between k and l: it permits the transitions for which $\Delta k = \pm 1$. The new principle looks as though it might kill more components than the old; but here the student had better make a pair of level-diagrams—one by each theory—for (say) the first line of the Balmer series, H_α. (I recall to him that in all the transitions the initial n-value is 3 and the final n-value is 2; the initial l-values are 2,1,0 and the final ones are 1,0; the j-values associated with each l-value are $l \pm \frac{1}{2}$; the initial k-values are 3,2,1 and the final ones are 2,1). He should put down all the conceivable transitions before using any selection-principle. He will then notice that in several cases a single component derived from the old theory has the same frequency as two or even four components derived from the new. On applying the selection-principles he will find that in some of those cases the single component of the old theory is forbidden, while one or two of the components of the new theory escape the axe. The outcome is, that for this line (as for all the others of the Balmer series) the old theory permits components at only three different frequencies, while the new one allows of five. These last five are exhibited in Figure 14.2.

It seems as though we should need only to ask a spectroscopist to look at a Balmer-series line and report to us whether

[1] Both selection principles may be derived from quantal theory, but the derivation is omitted from this book.

he sees five components or three; but this is too good to be true. From the predicted spacings the observer knows in advance that the Doppler effect, even when reduced by using deuterium instead of the light isotope of hydrogen and by cooling the radiant gas to liquid-air temperature, will prevent

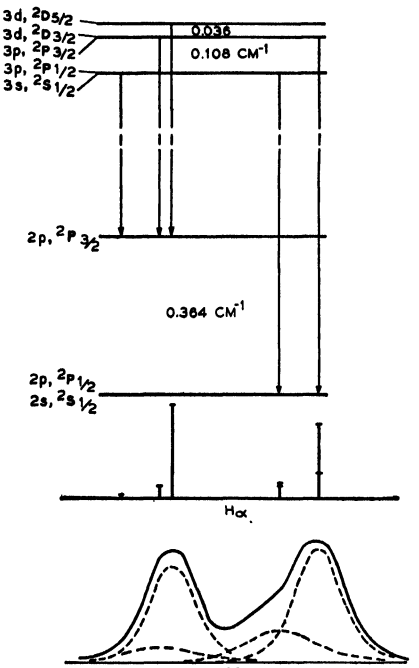


FIGURE 14.2—LEVEL-SYSTEM RESPONSIBLE FOR THE FIRST BALMER-SERIES LINE OF HYDROGEN, AND RELATIVE INTENSITIES OF THE COMPONENTS OF THE LINE, ACCORDING TO THE "NEW" THEORY. (After H. E. White.)

him from resolving all of the five components of the new theory. He may expect a curve of distribution-of-brightness-in-wavelength exhibiting two rounded peaks about 0.33 cm^{-1} apart, and this in fact is what the best prism or grating spectrographs show. He may even hope to distinguish a minor peak in the

valley between, and such a one has indeed been observed with deuterium at liquid-air temperature. The question now takes the form: do the finer details of shape of the principal peaks and the valley between them show the presence of the five components?

Such a question can scarcely be answered without pre-assuming intensity-ratios among the five, plotting the profiles of the individual peaks (as broadened by Doppler effect) at the right places on the wave-length-axis, and combining them to obtain the theoretical distribution-of-brightness-in-wave-length to be compared with the actual. The procedure is suggested at the bottom of Figure 14.2, which also shows the relative intensities given by the new theory for the five components (one turns out to be so very small that it is always disregarded in making the comparison). Notice how the observer is testing the detailed consequences of the new theory, and not the selection-principle only!

This difficult and exacting problem has brought back to modern physics, in a day when nuclear research is sweeping almost all before it, a welcome incentive for using the spectroscopic art with fullest skill and accuracy for the settling of a fundamental question. In the last four years it has been tackled by spectroscopists[1] of six institutions at least, with instruments (Fabry-Pérot interferometer, Lummer plates) better suited to such problems than the grating. Three years ago the situation was perplexing, for the authorities were disagreeing slightly but definitely with each other and slightly but definitely with the new theory. At the moment of writing it may be better, for the two latest reports refer to four components in "roughly" and in "approximately" the right places—as to their intensities, there is said to be rough agreement in the one case, partial agreement and partial disagreement in the

[1] Kent *et al.*, **PR 51**, 990 (1937); Houston, *ibid.*, 148; Williams and Gibbs, **PR 48**, 971 (1935); Spedding Shane and Grace, **PR 47**, 38 (1935), and earlier literature there cited; Kopfermann, **NW 22**, 218 (1934); Heyden, **ZP 106**, 499 (1937). In the days of the Sommerfeld theory and in the earliest days of the new one, observations by Paschen on the components of certain He II lines were often cited in evidence.

other.[1] Should the discrepancies be removed for good, there will be thorough confirmation of the analogy between alkali-metal and hydrogen atoms, and justification of what I have been calling "the new theory" involving electron-spin, relativity, and quantum-mechanics. Should they persist, it can pretty safely be forecast that the situation will be handled by modifications of the new theory, leaving its basis intact. The general desirability of the analogy between hydrogen and the alkalis, fortified as it is by the manner of splitting of atom-beams in the Gerlach-Stern experiment (page 390), will make this policy almost imperative. Nevertheless the original theory of Sommerfeld should remain unforgotten, to remind the physicists of the future that an elaborate, exact and verifiable formula may be derived from either of two sets of assumptions which are basically unlike.

Having with the aid of the spinning electron established an analogy between the spectrum of hydrogen and those of the alkali metals, we now turn to the broader analogy among all these and the X-ray spectra.

Structure of X-ray spectra

In Chapter VIII we considered the X-ray absorption-spectra and what they suggest. They imply the existence of definite Stationary States of the atom, each characterized by the absence of an electron from one of the inner shells—characterized, so to speak, by a vacancy. The simplest and most conspicuous cases are those in which the vacancy is in what, but for it, would be a closed shell. To them we shall restrict ourselves, and the states which we shall consider will be distinguished by the symbol of the shell in which the vacancy occurs.

The absorption-spectra and their associated electronic spectra show (to revert to Chapter VIII) that absorption of an X-ray photon by an atom results in the eviction of an electron —in the making of a vacancy—in the establishment of one of these states. Now the emission-spectra can be interpreted.

[1] I am orally informed that the situation has worsened again (August 1938). Latest paper: R. C. Williams, **PR 54**, 558 (1938).

They consist of lines, the principal ones of which are due to transitions between these states. Any such transition may be visualized as the passage of an electron from one shell to another, or as the passage of a vacancy from the other shell to

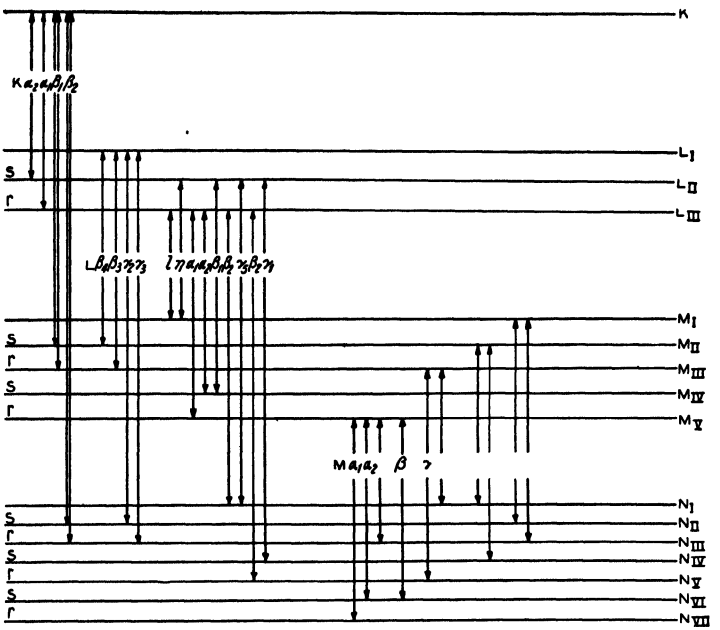


FIGURE 14.3—ILLUSTRATING STATIONARY STATES, TRANSITIONS, AND NOTATION FOR THE X-RAY SPECTRUM OF AN ATOM OF HIGH ATOMIC NUMBER (URANIUM). (Siegbahn, after Coster.)

the one—Kossel's idea. The one shell ceases from being closed, the other becomes so.

The lines of an X-ray emission-spectrum lie in groups along the frequency-scale. The groups are called "series"; there are the K-series, the L-series, the M-series, and others in order of diminishing frequency. This is a piece of bad luck linguistically, for these so-called "series" are opposed in an important respect to those of the "optical" spectra which we have just been considering. The transitions of an optical series in an

emission-spectrum share a common *final* state, but the transitions of the "K-series" of X-ray emission share a common *initial* state. This last is the K-state, characterized by a vacancy in the K-shell. The various K-series lines or K-lines correspond to the falling of an electron into such a vacancy from various other shells, or to the passage of the vacancy from the K-shell to various other shells. For the other X-ray "series" the phrasing must be more detailed, for though there is but one K-state, there are three states sharing the letter L, five sharing the letter M, seven the letter N. These states manifest themselves in absorption as well as emission: there are three distinct L-absorption-edges (page 257), and the "L-series" lines may be divided into three sets, each having one of the L-states for the initial state common to all of its members. (For consistency I should perhaps be speaking of three L-shells and five M-shells; this is the same linguistic difficulty as recurs so often in discussion of stationary states, and I shall leave the student to extricate himself with the aid of the remarks about shells on page 414.) The notation of X-ray emission spectra is not good.

When an element is bombarded by electrons of kinetic energy W, and W is raised continuously, the K-series lines do *not* appear severally as W/h attains their several ν-values: they appear suddenly and altogether when W/h attains the ν-value of the K absorption edge. Figure 14.4 shows them absent when W/h has already passed their ν-values but has not yet reached the K absorption edge, present after W/h has exceeded this last. This was a mystifying result when D. L. Webster first demonstrated it in 1916.

It may seem extremely strange that in the ordering of this chapter we should pass directly from the case of the hydrogen atom with its single electron, to the case of an atom with many closed shells, one of which is lacerated by the extraction of an electron. We are not however passing from a very simple to a very complicated case, but rather from one case of moderate simplicity to another which in a sense is the mirror-image thereof. The connection has been foreshadowed by the reap-

pearance, in the paragraph just above, of the sequence 1,3,5,7 . . . of which we have been seeing the significance in the term-systems of the alkali-metal atoms and of hydrogen. The states which are responsible for X-ray spectra are grouped as are the states of a doublet term-system. Here we find an analogy between shells which would be closed were not one electron absent, and shells consisting of one electron only.

The analogy is readily justified from our knowledge of shells as quoted in Chapter XIII. When an electron is reft from a closed shell, and carries away its own angular momentum, it leaves the remainder of the shell endowed with equal and opposite angular momentum. To say "equal and opposite" is to say "equal" when speaking of quantum-numbers; the lacerated shell has the same j-value as the departed electron. What I have just said of j may be said as well of l and of s. In the closed shell the orbital angular momentum of any one electron is equal and opposite to the resultant of the orbital angular momenta of those which it leaves behind when it goes away, and the same may be said of the spins. The lacerated shell has therefore the same l-value as the departed electron, and a j-value which is either greater by $\frac{1}{2}$ or less by $\frac{1}{2}$ than the l-value. Recall now how in Chapter XIII we built up (in imagination) the successive atoms of the Periodic Table by contriving shells with total quantum-numbers $n = 1,2,3,4 \ldots$, and populating each of them with electrons displaying all of the combinations of l and j compatible with the respective n-values. In so doing we were preparing for the theory of X-ray spectra; for to every combination (n, l, j) corresponds one of the states responsible for these spectra, one of the absorption-edges, and a set of emission-lines.

We see these rules exemplified in the table which follows:

TABLE 14.1

	K	L_I	L_{II}	L_{III}	M_I	M_{II}	M_{III}	M_{IV}	M_V	N_I
n.....	1	2	2	2	3	3	3	3	3	4
L.....	0	0	1	1	0	1	1	2	2	0
J.....	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$1\frac{1}{2}$	$1\frac{1}{2}$	$2\frac{1}{2}$	$\frac{1}{2}$
	S	S	P	P	S	P	P	D	D	S

Here the topmost row exhibits the notation now commonly used for these states.[1] The second, third, and fourth are self-explanatory: capital L and J have been substituted for lower-case to stress the fact that these are quantum-numbers of the whole electron-family of the ionized atom. The fifth row adds nothing to the content of the Table, but is meant to remind the student of the letters used to designate states of these l-values in describing optical term-systems. Were the symbols in the fifth line completed according to the precept of page 318, the first three would be $1^2S_{1/2}$, $2^2S_{1/2}$ and $2^2P_{1/2}$. I take occasion to point out that the electrons enumerated in Chapter XIII are more numerous than the states here listed; each combination (n, L) corresponds to $2(2l + 1)$ electrons in the closed shell, but to not more than two distinguishable states.

By themselves the data thus far described might hardly support so definite a theory; but there are many more. There are, for instance, those from which comes the selection-principle for X-ray emission lines. To abide by the analogy, the permitted lines should be those for which $\Delta L = \pm 1$, $\Delta J = 0$ or ± 1, and the rest should be forbidden. This is in accord with experience.[2]

The analogy with hydrogen is greatly strengthened by the energy-values of the states, the binding-energies or extraction-energies for the several shells. Instead of stressing these I follow the historical order by speaking of the line-frequencies, and notably that of what used to be called the "Kα line." Of the transitions[3] (KL_I), (KL_{II}), and (KL_{III}), the first is forbidden, the second and the third give rise to lines (now known as

[1] In an alternative and now disappearing notation, Arabic numerals were used as the subscripts instead of Roman, and they increased with increasing term-value, so that L_1, L_2, L_3 corresponded to L_I, L_{II}, L_{III} respectively.

[2] The selection-principle was given in this form by Landé, **ZP 16**, 391 (1923) and **24**, 88 (1925). Formerly it was said that all lines for which $\Delta n = 0$ are also forbidden; it appears that some have since been discovered, but so many are missing that there must be some basis for this additional rule. The analogy between the system of K, I., M . . . states and the doublet term-system is said to have been noticed by Grotrian.

[3] In these symbols I give first the letter of the final state and second the letter of the initial state, as is the practice for optical spectra.

$K\alpha_2$ and $K\alpha_1$) which in the earliest X-ray spectroscopes were not resolved. Moseley, taking the first photographs of X-ray line-spectra, made his setting upon the apparently single line which was actually the pair; and collating the data for many elements, he found them concurring with the formula,

$$\nu = R(Z - 1)^2 \left(\frac{1}{1^2} - \frac{1}{2^2}\right)$$

which he obtained by generalizing according to Bohr's precepts from the formula for the frequency of the first line of the Lyman series in hydrogen—the line due to the transition from $n = 2$ to $n = 1$. The obvious way of generalizing leads to the factor Z^2 instead of $(Z - 1)^2$, but this amounts to supposing that the atom emitting the $K\alpha$-line has only one electron; the discrepancy is just about what one expects from the presence of the other electrons. By the same argument, any transition (LM) should concur with the formula $R(Z - s)^2(\frac{1}{2}^2 - \frac{1}{3}^2)$, with an appropriate value for s regarded as a disposable constant: for a certain line of the L-series Moseley found fairly good agreement when for s he put 7.4, an acceptable value.[1] Values of s determined in this way are now material for advanced atomic theory.—Figure 4.5 (page 110) exhibits these laws and the corresponding ones for several other lines.

From dependence of energy-values on n we turn now to dependence on L and J. Here we find remarkable contrasts among the three cases of alkali-metal spectra, hydrogen spectrum and X-ray spectra. Consider all the levels sharing a given n-value, and for brevity let me call a pair of such levels having the same L-value and different J-values a "spin-doublet," a pair of levels having the same J-value and different L-values an "azimuth-doublet." For alkali-metal atoms the spin-doublets are narrow enough to be called "fine-structure," the

[1] Moseley, **PM 26**, 1024 (1913), **27**, 703 (1914). These were the historic researches whereby atomic number was promoted to precedence over atomic weight. In the first, $K\alpha$ was observed from $Z = 20$ to $Z = 30$ (21 excepted); in the second, K-lines were observed for most of the elements from 13 to 47, and several L-lines for most of the elements from 41 to 79.

azimuth-doublets relatively so broad that generally they are not called doublets at all. For the hydrogen atom the spin-

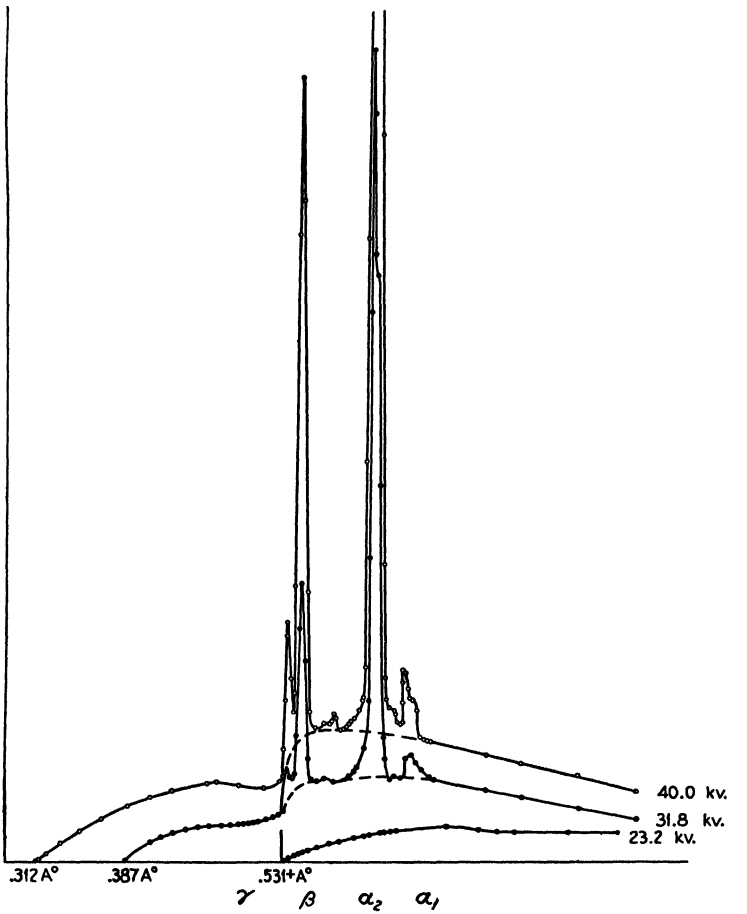


FIGURE 14.4—THE K-SERIES OF RHODIUM SUPERPOSED UPON A CONTINUUM, EVOKED BY ELECTRONS OF THE KINETIC-ENERGY-VALUES MARKED AT THE ENDS OF THE CURVES. (Webster; *Physical Review.*)

doublets again are narrow, but the azimuth-doublets shrink to a vanishing spacing, the members of each coinciding. Among

the states responsible for X-ray spectra, spin-doublets and azimuth-doublets are of the same order of magnitude in spacing—there is no general rule to the effect that either type is broader than the other.

Spin-doublets are admirably fitted by the formula (14.4), provided that Z^4 be replaced by $(Z - s)^4$ and s be adjusted. The best examples are furnished by the L_{II}–L_{III} and the M_{II}–M_{III} doublets; for the nine elements ranging in Z from 41 to 50 (43 not included), the values of s which give the best agreements lie between 3.45 and 3.50 for the former, between 8.28 and 8.58 for the latter. Azimuth-doublets follow—not very accurately—a quite different law, according to which the difference between the square roots of the energy-values is the same for a given doublet whatever the value of Z; this was discovered by Hertz.[1]

Auger effect (*compound photoelectric effect, internal conversion*)

Now we take up an effect the connection of which with X-ray emission-spectra is, that it hinders them from coming into being.

Imagine an expansion-chamber containing a gas (krypton or xenon, for instance) of atoms massive enough to possess several closed electron-shells, and think of it exposed to X-rays of frequency ν high enough for extracting K-electrons. When the chamber is in operation, tracks of ejected K-electrons should appear, having the length appropriate to the energy-value $h(\nu - \nu_K)$. We have learned to visualize the atom whence

[1] The history of the topic of these two last paragraphs deserves to be told along with that of the theory of the hydrogen atom, but space forbids. The two kinds of doublets have borne various names, of which "relativity" versus "screening" and "regular" versus "irregular" are some. Formula 14.4 was applied by Sommerfeld and Wentzel to the spin-doublets at a time when the former had derived it from a quite different theoretical basis, *viz.* the Bohr atom-model with electron-orbits modified by the relativistic dependence of mass on speed, and without any allowance for electron-spin; hence the name "relativity-doublets." Sommerfeld's formula involves additional terms in higher powers of $(Z - s)$, but these seem not to be of serious magnitude in the cases thus far tested. References: Sommerfeld and Wentzel, **ZP 7**, 86 (1921); Wentzel, **ZP 16**, 46 (1923).

such an electron came as having a vacancy in the K-shell, destined presently to be filled by an electron dropping inward from one of the shells labelled L, M, N . . . (I disregard here such distinctions as that between L_I, L_{II} and L_{III}). When one of these processes takes place, energy in one of the amounts $h(\nu_K - \nu_L)$, $h(\nu_K - \nu_M)$, . . . is made available for release. We have been thinking of this energy as going forth in the form of a photon belonging to one or another of the K-series emission-lines, and having the ability to extract electrons from other atoms. Suppose however that this photon were to be absorbed before ever it left the atom of its birth, and to extract an electron from that very atom? Conceivably for instance a photon of energy $h(\nu_K - \nu_L)$ might extract from the atom of its birth a L-electron, of binding-energy ν'_L. Here I write ν'_L instead of ν_L, because in this atom there will already be a vacancy in the L-shell which will affect (increase) the binding-energy for those L-electrons which remain. (In practice the distinction between ν'_L and ν_L is usually ignored). If this process ever occurs, there should appear in the expansion-chamber pairs of tracks starting from the same point, their lengths corresponding to the energy-values $h(\nu - \nu_K)$ and $h(\nu_K - \nu_L - \nu'_L)$.

The foregoing paragraph contains a description of some well-known phenomena, and a theory which enables the student to remember them readily, but perhaps has no other merits. To think of a photon as having an existence while passing from one point to another inside of an atom causes more of a strain than we should put upon our imaginations. The phenomena are adequately described by saying that, in such a case as we have just been viewing, the atom passes directly from a state characterized by one vacancy in the K-shell to a state characterized by two vacancies in the L-shell, and at the same time the originally ejected free electron of kinetic energy $h(\nu - \nu_K)$ is followed forth into outer space by a second free electron of energy $h(\nu_K - \nu_L - \nu'_L)$. No photon fares forth into space. The process has a resemblance to the "collisions of the second kind," mentioned in Chapter IX.

Such phenomena were first observed in the expansion-chamber by Auger, from whom the effect takes one of its names.[1] Figure 14.5 shows two such pairs of tracks, springing from atoms of krypton. Here it chanced that the ν-value of the irradiating X-rays was such, that $h(\nu-\nu_K)$ and $h(\nu_K-\nu_L-\nu'_L)$ were nearly equal. The first test of the theory consists in altering ν, and seeing whether the pairs are then composed of members one of which retains the constant length corresponding to the latter energy-value, and the other varies in the proper way with ν. Auger applied this test (roughly) in the early days, and so verified his theory, which is the one just described. It

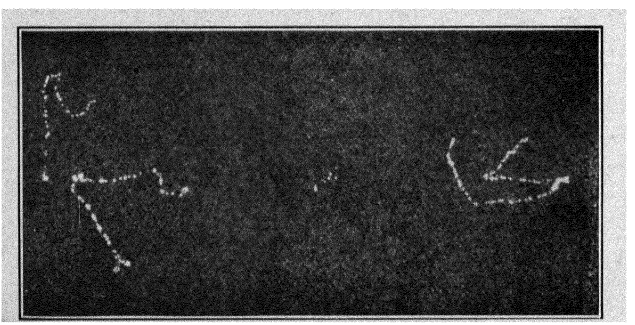


FIGURE 14.5—ILLUSTRATING THE AUGER EFFECT CONSEQUENT ON EXTRACTION OF A K-ELECTRON FROM A KRYPTON ATOM. (Auger; *Journal de Physique.*)

is possible so to choose ν that the track of the ejected K-electron which indicates the initiating process is almost invisibly short, while that of the L-electron (which Auger calls "the tertiary") retains its constant and considerable length. The two small projections from the point of origin of the two long tracks in Figure 14.5 may be the trails of what Auger defines as "quaternary" electrons, presumed to be driven from the M-shell into space when energy is made available by the dropping of another M-electron into one of the vacancies left behind in the L-shell.

[1] Auger, JP 6, 205 (1925), and *Annales de Physique*, 6 ,183 (1926); Locher, PR 40, 482 (1932).

Much more distinct trails of quaternaries were in fact obtained by Auger with xenon.[1]

Far from being a rare phenomenon, this Auger effect may be more frequent than the emission of photons to which it is the alternative! An obvious way of measuring the ratio r of photon-emissions to sum of photon-emissions and Auger transitions (r is called "fluorescence yield," and $(1-r)$ "coefficient of internal conversion") consists in counting the trails of ejected K-electrons which are paired in the aforesaid way and counting also those which are unpaired, and forming the appropriate quotient. To illustrate from Auger's pioneering work: with krypton he found 156 out of 320 tracks of K-electrons to be paired, and with the heavier element xenon only 57 out of 206, and with the lighter element argon no fewer than 210 out of 254.

This admirable method of the expansion-chamber unluckily serves for a few elements only. Most of what is known about fluorescence-yield has been found out by comparing the intensity of an X-ray beam falling upon a metal plate with the intensity of the X-rays emerging from the plate at nearly a right angle to the beam, ionization-chambers being used for measuring these intensities.[2] The primary beam is made as nearly monochromatic as possible, consisting usually of a single series of emission-lines and sometimes of a single line. Allow-

[1] In Auger's earliest work the trails of the tertiaries were shrunken into mere blobs: he noticed some of these at the points of origin of the tracks of ejected K-electrons, and had then the idea of mixing the gas of the chamber with a large excess of hydrogen in order to lengthen the blobs into trails. The obvious query "How can we know that two tracks spring from the same atom?" is based on the alternative idea of a process in which (1) an electron is evicted from an atom A, (2) an X-ray photon springs from A and (3) the photon is absorbed in another atom B and expels an electron from B. If, however, many of these pairs were due to this process, then the absorption-coefficient of these X-rays in the gas would have to be much higher than it is well known to be.

[2] Recent experiments: Stephenson, **PR 51**, 637 (1937); Backhurst, **PM 22**, 737 (1936). Earlier data collated by Stephenson. Work before 1935, and description of the method in Compton and Allison, *X-rays in Theory and Experiment* (Van Nostrand) pp. 477 ff. Theory: Massey and Burhop, **PRS 153**, 661 (1936). Remarkable anticipations of contemporary knowledge occurred as long ago as 1909; for the early history see Barkla and Dallas, **PM 47**, 1 (1924).

ances have to be made for primary photons scattered with or without change of wave-length, or absorbed with other results than the ejection of K-electrons; and account should be taken of the fact that even when a K-electron is expelled, the ensuing

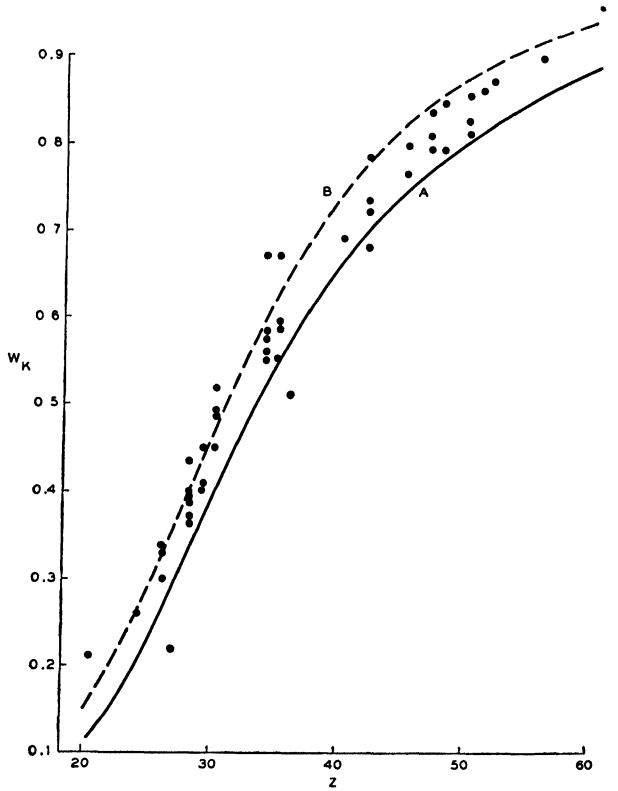


FIGURE 14.6—FLUORESCENCE-YIELD SUBSEQUENT TO EXTRACTION OF A K-ELECTRON. THE CURVES ARE TRACED BY QUANTAL THEORY INVOLVING ORDINARY MECHANICS FOR THE UPPER, RELATIVISTIC FOR THE LOWER. (Stephenson; *Physical Review.*)

chain of consequences is not uniquely fixed (the vacancy may be filled from any one of several shells, photons of various frequencies may be emitted and Auger transitions of various kinds occur).

Values of the fluorescence-yield are now available for about a score of elements, ranging in atomic number from 18 to 56. In confirmation of Auger's early inference, there is an upward trend of r with Z which is definite and strong. Around $Z = 20$ (the data are fewest and least concordant in this region) r is about 0.2, above $Z = 50$ it is rapidly approaching 0.9. Quantal theory supplies a formula, or formulae rather (depending on whether ordinary or relativistic mechanics be used) for r as function of Z; there is one disposable constant; this being adjusted the agreement with data, as Figure 14.6 displays, is better than fair. Apparently no one yet has made use of primary photons energetic enough to extrude the K-electrons from elements beyond 60; but for three of the most massive of all ($Z = 82, 90, 92$) Stephenson measured the fluorescence-yield ensuing upon expulsion of L_{III} electrons, and found the values 0.32, 0.42 and 0.44.

In the final chapter of this book, we shall come to abundant examples of an effect like this one, except that the energy comes from the nucleus and the electron-family passes from its normal state of completeness to a state in which there is a vacancy in one of the shells: it is called "internal conversion of gamma-rays."

Molecular spectra and nuclear masses

The topic of this section has attained to such a vast voluminous expansion that there is a great deal of presumption in attempting even to "introduce" it in less than the scope of a very long chapter. Yet rather than leave it unmentioned, I will give a very cursory description of the feature of molecular spectra which is the most conspicuous, and which also exemplifies the quantal properties of angular momentum and vibration in the most unquestionable way.

A diatomic molecule is a pair of nuclei, environed by electrons numerous enough to balance the charges of both. Owing (it is presumed) to the interplay of electrical forces among the electrons and the nuclei, there is a "distance of stable equilib-

rium" r_0 between these last. The actual internuclear distance r may vary cyclically around r_0, in a vibratory motion. The system may also revolve around any axis perpendicular to its axis of symmetry; we are therefore to be concerned with its moment of inertia about such an axis, I. There may be several or many different arrangements of the electrons, comparable with the various stationary states of atoms. They are called "electronic states" or "electronic levels," and a distinction is drawn between the normal and the excited ones. To explain these states in full is the last and greatest task of molecular theory; but in this place we must content ourselves with realizing that the balance of forces which controls the motions of the nuclei, and with it the values of r_0 and I, differs from one to another of them.

Shortly after introducing the notion of Stationary States of atoms in Chapter IX, I was obliged to say that each of the entities to which I first applied that title is actually a cluster of a few distinguishable stationary states. Later on it was explained that different members of such a cluster differ in respect of the orientations of the various angular momenta localized in the electron-family. All this may be repeated as to molecules, and still we have not taken account of the complication which converts what should be a small cluster of electronic states into a cluster of literally hundreds of states. This arises from the rotations and from the vibrations of the molecules. Both of these are quantized; and to each combination of electronic state, permitted angular momentum of the molecule, and permitted amplitude of the vibration, there corresponds a distinct permitted energy-value.

By the method of wave-mechanics (Chapter XII) we have already found, for the vibrator or oscillator, the permitted energy-values

$$E_v = vh\nu_0, \qquad v = \tfrac{1}{2}, \tfrac{3}{2}, \tfrac{5}{2} \ldots \qquad (14.5)$$

ν_0 standing here for the frequency of the vibrations, which we there conceived to be simple-harmonic. Actually the forces in a molecule are such that the vibrations depart perceptibly from

the simple-harmonic ideal, but we will ignore the discrepancy except in one respect. By the method of wave-mechanics one may also obtain, for the rotatory motion, the permitted energy-values,

$$E_j = (h^2/8\pi^2 I)j(j+1), \qquad j = 0, 1, 2, 3. \ldots \quad (14.6)$$

The symbols v and j stand for quantum-numbers, the latter being chosen for analogy with the quantum-number bearing the same symbol in the theory of atoms. They distinguish the different states into which the so-called electronic state is resolved; all of these have in common a system of quantum-numbers corresponding to the electron-family, which I denote by e but do not analyze further. Denote by E_e the energy of the electronic state as it would be in absence of rotation and vibration (but notice that if the values just assigned to v are right, there never is absence of vibration !). The energy-values E represented in the cluster are given thus:

$$E = E_e + E_v + E_j. \quad (14.7)$$

There is no upper limit to the permitted values of v and j except that enforced by the fact that the molecule is liable to fly apart when either rotation or vibration exceeds a certain amount. Sometimes a spectrum will indicate the presence of molecules having all of the values up to this limit. More often the proportion having a given v-value or a given j-value diminishes so rapidly with increase of v or j as to become imperceptible before the limit is reached. The distribution of molecules among j-values is likely to be dominated by the temperature of the gas, for the difference between successive E_j-values may be small compared with the average energy of thermal agitation $\frac{3}{2}kT$ even when T is as low as room-temperature. In a tranquil gas at room-temperature the lowest value of v may be practically the only one occurring; but as for rotation, the state $j = 0$, far from being the only one represented or even the most abundantly populated, may actually be inhabited by

fewer molecules than some of the others! the most popular state will be the one for which E_j is most nearly equal to $\frac{1}{2}kT$.

We are going too rapidly in discussing the states so much before taking up the selection-principles which limit the observable spectrum-lines from which the states are deduced. * If the vibrations were ideally simple-harmonic, the selection-principle for v would be the one already encountered in Chapter XII: to wit, $\Delta v = \pm 1$. Transitions with greater values of Δv do however occur, and this is one of the two things which most directly prove that the vibrations are not purely harmonic (the other being the deviations of the ratios between the E_v-values from the sequence $1 : 2 : 3 : 4 \ldots$). The selection-principle for j requires a fuller statement, and indeed it requires some theory which there is not space to provide. In general the permitted transitions correspond to $\Delta j = 0, \pm 1$; but when the electronic state does not change (and sometimes even when it does), $\Delta j = 0$ is also forbidden. For the permitted lines, then, we have the stipulations:

$$h\nu = \Delta E_e + \Delta E_v + \Delta E_j, \qquad \begin{array}{l} \Delta v \text{ any value,} \\ \Delta j = \pm 1 \text{ if } \Delta E_e = 0 \\ \Delta j = \pm 1, 0 \text{ if } \Delta E_e \neq 0 \end{array} \qquad (14.8)$$

Even with these restrictions there may be hundreds of lines corresponding to the transitions between two electronic states. A line-cluster of this type is known as a "band" owing to its appearance in spectroscopes so weak as were formerly used; and "band-spectrum" is synonymous with "molecular spectrum."

The simplest cases of all are those in which only the rotation changes, the electronic state remaining the same and the vibration remaining at its lowest permitted amplitude. Such spectra may be observed in absorption, by selecting gases such as the hydrogen halides of which the moment of inertia is low and the factor $(h^2/8\pi^2 I)$ accordingly is high. Even with such a favorable choice of molecules, the lines must be sought in the

extreme infra-red.[1] Those permitted by the selection-principle for j have the frequencies,

$$\nu_{j,\,j+1} = 2B(j+1), \qquad B = h/8\pi^2 I \qquad (14.9)$$

j standing for the quantum-number before the absorption-process. The frequencies of consecutive lines of such a "rotation-band," as it is called, should stand to one another as the 1, 2, 3, ... This has been verified, and serves as a test between the newer and the older forms of quantal theory (according to the older ones, the frequencies should stand as 1, 3, 5 . . .).[2] From the coefficient B as adjusted to fit the data, the value of I is deduced. Values of I are derived from many bands of many molecules, and from them are computed values of internuclear distance. These in turn may be checked against values derived from phenomena of remarkably different kinds, such as the diffraction-patterns which the molecules impress upon an electron-beam, or the molecular diameters calculated by the methods of the kinetic theory of gases. The agreements are good, and they sustain the quantal theory of rotation as strongly as the doctrine of the spinning electron is sustained by the measured values of the magnetic moments of atoms.

The next simplest cases are those in which vibration changes and so does rotation, the electronic state, however, remaining the same. Take any particular transition Δv as an instance, and let ΔE_v represent the energy-change with which it would be attended if there were no change in rotation. In an emis-

[1] It is interesting to notice why the ordinary diatomic hydrogen molecule, for all its low value of I, does not figure on the list of molecules of which these lines have been observed. Owing to its symmetry in the normal electronic state, its electric moment is nil; and this signifies that if it is revolving around an axis perpendicular to its axis of symmetry, the circulating current constituted by the charges of its nuclei is equal and opposite to that constituted by the charges of its electrons. By classical theory such a revolving system of charges would not be expected to radiate, and from the so-called "correspondence-principle" one draws the conclusion that no more can it radiate when making transitions; experience confirms this.

[2] Czerny, **ZP** 34, 227 (1925); his results have been restated by Sponer (Molekülspektren, p. 46) in terms of the contemporary theory.

sion-spectrum, one might then expect a sequence of lines having the frequencies,

$$\nu_{j+1,\,j} = \Delta E_v/h + 2B(j+1) \qquad j = 0, 1, 2, \ldots \quad (14.10)$$

j standing now for the quantum-number after the emission-process; for these are the lines which should occur, if the emission is always attended by a slowing-down ($\Delta j = -1$) of the rate of rotation. This sequence appears, but not by itself; there also appears its counterpart at frequencies lower than $\Delta E_v/h$, *viz.*,

$$\nu_{j,\,j+1} = \Delta E_v/h - 2B(j+1) \qquad j = 0, 1, 2, \ldots \quad (14.11)$$

signifying that the emission may equally well be attended by a speeding-up ($\Delta j = +1$) of the rate of rotation. There is after all nothing startling about this, for the energy gained by rotation is more than matched by the energy lost from vibration. The two sequences are called "branches of the band"—the R branch and the P branch respectively (Figure 14.7).

The frequency-difference between consecutive members of either branch has the constant value B, but that between the initial members of the two is twice as great, leaving a striking gap. The center of this gap—sometimes designated as "the missing line" or "zero line"—fixes for the observer the value of ΔE_v, and is quoted as the distinctive frequency of the band. A group of such bands may lie close together (usually in the near infra-red), its various members corresponding to various transitions Δv. It will usually be observed in absorption, and the band for which v passes from $\frac{1}{2}$ to $\frac{3}{2}$ will then be much the most prominent. Such a group is known as a "vibration-rotation band-system."

The case of full generality involves a passage from one electronic state to another. The presence of ΔE_e in the expressions for all of the line-frequencies often promotes the resulting bands into the visible part of the spectrum, or even the ultraviolet. More important is a set of new complications, due to the already-mentioned fact that I has different values for different electronic states. It must be admitted that even for

any one electronic state I will depend on v and j; on which account, corrections relatively minor must be made to the foregoing statements when the treatment is fuller; but the effect to which we now turn is major.

Let the transition Δe occur in emission, and let us take any one value of Δv as an instance, thus isolating one band from a

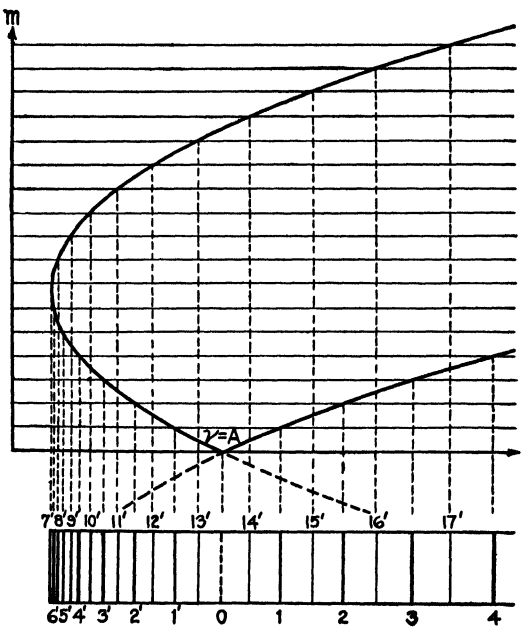


FIGURE 14.7—ELUCIDATING THE ARRANGEMENT OF LINES IN THE P AND R BRANCHES OF A BAND ("FORTRAT DIAGRAM"). (After Sommerfeld.)

band-system. The general formula for the lines of this band is (14.8), but we concern ourselves only with the values of $\Delta E_i/h$ which are the displacements of these lines from the basic frequency or missing line at $(\Delta E_e + \Delta E_v)/h$. Let us put B' and B'' for the values of B in the initial and final electronic states; j' and j'' for the initial and final values of j for any one line; j for the smaller of these two, which will be j' in the

P-branch and j'' in the R-branch. For these branches we now find the expressions,

$$\text{P: } \Delta E_j/h = (B' - 3B'')j - 2B'' - (B'' - B')j^2$$

$$\text{R: } \Delta E_j/h = (B' + B'')(j+1) + (B' - B'')(j+1)^2 \quad j = 0, 1, 2, \quad (14.12)$$

which now contain quadratic as well as linear terms in j.

So long as the moment of inertia of the molecule was unaffected by the transition, linear terms alone appeared in these expressions, and accordingly the spacing was constant along each of the branches. Now the quadratic terms import a distortion, and one which is capable of producing a strange and beautiful effect. In one branch, we find that the spacing increases as we step from one line to the next while going always in the sense of increasing j; but in the other branch, we find that the spacing decreases until it passes through zero and we have to step backwards, the sequence turning back upon itself and its higher members overlapping its lower!

This is a situation which is made clearer by diagrams and photographs than by any number of words. Figure 14.7 is a diagram in which frequency is plotted horizontally (the missing line is at A) and order-number m, equivalent to j, is plotted vertically; primes distinguish the lines of one sequence from those of the other. Figure 14.8 shows us a panorama of bands displaying how they gain in beauty as well as intricacy by this effect. It is easy to guess that in the neighborhood of the turning-point of a branch that bends back, an instrument of low resolving-power will show an apparent continuum with one sharp edge: hence the name of "band-head" for the turning-point. A further addition to intricacy is sometimes made by the presence of a so-called "Q-band," made up of lines due to transitions in which j does not change; it departs less from the missing line than do the two others.

In entering upon molecular spectra, we were on the verge of going further away from nuclear physics than at any other time since Chapter IV; but our brief journey has been so

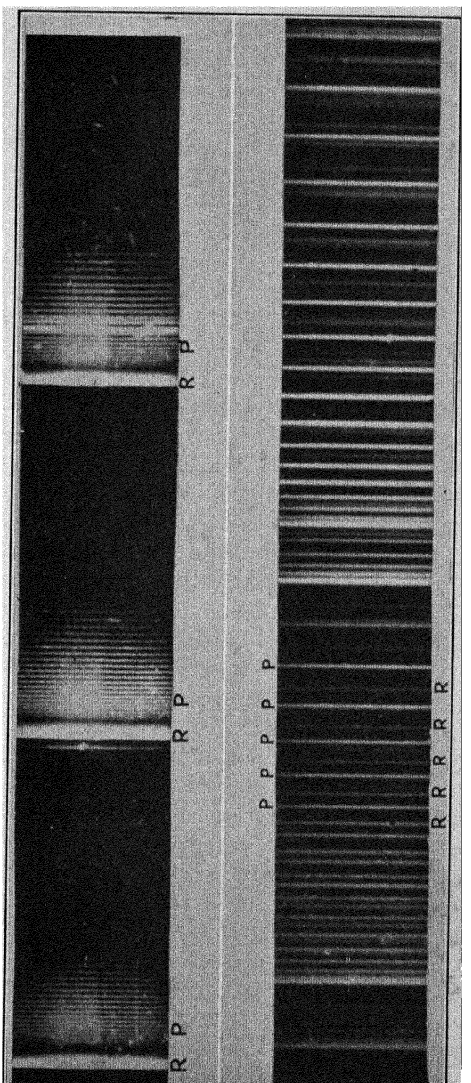


FIGURE 14.8—PHOTOGRAPHS OF BANDS COMPRISED IN A BAND-SYSTEM OF THE ONCE-IONIZED CARBON-MONOXIDE MOLECULE.
(C. M. Blackburn.)

directed that actually we have been returning towards the nucleus, by restricting ourselves to phenomena in which nuclear masses affect the spectra in a vivid and drastic way. We leave unexplored the huge subject of the electronic states of molecules, and from this point onward in the book we devote ourselves to nuclei only. Molecular spectra have one more contribution to make to our knowledge of these; or rather, a refined consequence of the contributions which we have studied already.

Though for any state of a molecule ΔE_v and ΔE_j depend on the nuclear masses, yet we cannot compute these last from the former because it depends also upon the forces within the molecule, of which we do not have accurate independent knowledge; and though we might compute them fairly well from the latter by virtue of our independent knowledge of r_0 mentioned on page 443, yet this knowledge is not good enough to give us mass-values rivalling those obtainable by mass-analysis. Compare however molecules which are identical in nuclear charges and differ in nuclear masses, such for example as the set of four hydrogen-chloride types H^1Cl^{35}, H^1Cl^{37}, H^2Cl^{35} and H^2Cl^{37} Everything said about atomic theory thus far in this book implies that the forces within the molecule, and therefore the values of r_0, will be the same for all four of these. One readily works out, then, what the difference in the masses between any two of them will do to the vibration-frequencies and the rotation-rates, to the E_v-values and the E_j-values. From one to another, there will be a shift in each vibration-band considered as a whole (or one may say, a shift in the missing line) and superposed on this there will be additional shifts of the individual bandlines, varying from one to another. If the comparison is made between two molecules differing in respect of one nucleus only, then the shifts depend upon the ratio of the masses of the nuclei which differ in the two; and when they are measured (preferably those of the missing lines) this ratio may be computed.

Ordinary hydrogen chloride consists almost entirely of H^1Cl^{35} and H^1Cl^{37} molecules, the former about three times the

more abundant. Photographs of one of its vibration-rotation bands, which Imes had taken in 1919, displayed each line resolved into a doublet; and in the following year this unexpected semblance of fine-structure was recognized by Loomis and independently by Kratzer for what it was, the isotope-effect.[1] In the ensuing score of years the presence of two or three isotopes has been detected thus for about a score of elements, and sometimes made doubly sure by detection in the spectra of two or more molecular compounds of a single element. Discoveries soon arrived to enliven what otherwise might early have become a rather dull sort of research. Giauque and Johnston examined Babcock's photographs of the absorption-spectrum of the thickest layer of oxygen ever traversed by a beam of light on its way to a spectroscope: the atmosphere itself, absorbing the light of the nearly setting sun. In certain of the band-systems they found two faint repetitions of each band, attributable to a pair of isotopes O^{17} and O^{18} very much rarer than the familiar O^{16}. Next, C^{13} and N^{15} were discovered from bands, the former by Birge and King and the latter by Naudé.[2] At this moment the optical spectroscope seemed to be proving itself a more sensitive device for finding scanty isotopes than did the mass-spectrograph, which as yet had not shown any of these four. Subsequently all but O^{17} were found by mass-analysis, and since then the spectroscope seems not to have scored any more beats over its rival.

The ratio of the intensities of corresponding bands due (for example) to H^1Cl^{35} and H^1Cl^{37} in ordinary hydrogen chloride is the ratio of the abundances of the isotopes of chlorine in this gas, which appears also to be the ratio of their abundances in Nature generally: this illustrates a method of determining abundance-ratios which has been used occasionally.[3]

[1] Imes, *Astroph. Jl.* **50**, 251 (1919); Loomis, *Nature* **106**, 179 (1920); Kratzer, **ZP 3**, 460 (1920), **4**, 476 (1921).

[2] Giauque and Johnston, *Nature* **123**, 318, 831 (1929); King and Birge, **PR 34**, 376 (1929); Naudé, **PR 36**, 333 (1930).

[3] See for instance Brosi and Harkins, **PR 52**, 472 (1937).

Next arises the question: can the ratio of masses be determined with an accuracy which admits this method to rivalry with mass-analysis and with the study of transmutation (Chapters XV–XVIII) as a way of ascertaining nuclear masses in terms of one another and of a common standard? The requirement is severe, for both of the other procedures permit of very close measurements; but so also, of course, does the spectroscopic procedure. We are however obliged to be much more critical of our assumption, hitherto made without detriment, that in a molecule the substitution of one nucleus for another of the same charge but different mass has no effect upon the intra-molecular forces or the value of r_0. When this assumption is made as to the substitution of a H^2 nucleus for a H^1 nucleus in the molecules of various hydrides and of diatomic hydrogen gas itself, it turns out that the values for the ratio of masses of H^2 and H^1 derived from various molecules (and even from various transitions of the same molecule) differ slightly from one another. The assumption is thus not quite right; a considerable amount of theorizing has been done on this point. Unexpected results occurring in the study of the Li_2 molecules (Li^6Li^7, Li^6Li^6, Li^7Li^7) may or may not be traceable to that source.[1] It is believed that the faults in the assumption are negligible for molecules with yet heavier nuclei. If this may be taken for granted, such a value as that obtained by Jenkins and Wooldridge for the ratio of the masses of C^{13} and C^{12} with a statistical uncertainty of 4 parts in 10^5 is to be listed among the sources of information for the mass of C^{13}. Otherwise the excellent agreement between this and the other estimates is to be taken as evidence for the assumption.[2]

From a quality of molecular spectra which indicates nuclear masses, we pass now to a quality of atomic spectra which indicates the angular momenta and the magnetic moments of kernels.

[1] Cases of lithium and of hydrogen compounds: Almy and Irwin, **PR 49**, 72 (1936); Jenkins and McKellar, *ibid.* 205; McKellar, **44**, 155 (1933), and literature there cited; for the theory see Van Vleck, *J. Chem. Phys.* **4**, 327 (1936).

[2] Jenkins and Wooldridge, **PR 53**, 137 (1938). For the situation as to O^{18} and N^{15} see Mattauch, **PR 50**, 622 (1936).

Hyperfine structure and nuclear spins

In the seemingly endless process of resolving apparent stationary states into clusters of states, we now arrive at one more stage and the last. Taking again the alkali metals and hydrogen as temporary examples: in saying that all their S states are single and all their P, D, F . . . states double, I did not exhaust their complexity—nor did I exhaust it even when listing the sub-levels into which these states are broken up by a magnetic field. There is yet one more subdividing of terms, which manifests itself in the "hyperfine-structure" of spectrum-lines; a structure which (as the name implies) usually cannot be seen without an extremely good spectroscope. Hyperfine structure is not confined to alkali metals and hydrogen, but ranges all over the Periodic Table. It is definitely *not* a periodic function of atomic number, and therefore must be controlled or anyhow influenced by the nucleus; but neither is it a monotonic or nearly monotonic function of atomic number, and therefore it must depend on some nuclear property other than charge and mass. This property turns out to be *the spin of the nucleus* (Pauli's idea).

In advance of the data let us indulge in some guesses as to what might be observed. The H^1 nucleus being the lightest and presumably the simplest of all nuclei, we might conjecture that it is like the electron in having $\frac{1}{2}$ for its spin quantum-number. In making such an assumption we are of necessity making the inseparable assumption that in a magnetic field or relatively to a neighboring electron or electron-family, the proton will have two permitted orientations. It does not seem so reasonable to assign the same magnetic moment to both proton and electron, since that of the latter is given by the formula $eh/4\pi mc$ in which m stands for the electron-mass. We conjecture rather that the magnetic moment μ_p of the proton may be $eh/4\pi Mc$ with M standing for the proton-mass. Since M/m is about 1835, we thus are making the guess that the magnetic moment of the proton is only about $^1/_{1835}$ of that of the electron. (This is quite compatible with the fact that so far we

have got into no trouble through identifying the magnetic moment of an atom with the resultant of those of its electrons only.) Now consider a stationary state of a hydrogen atom; revive the picture of an electron revolving in a quantized orbit around a proton, and producing a magnetic field in which the proton (endowed with the aforesaid values of moment and spin) is capable of orienting itself and precessing. Repeating the argument of page 419 with the obvious changes, we infer that the stationary state is divided into two which correspond to the two orientations of the proton, and that the energy-difference between the two is given by the right-hand member of equation 14.4 multiplied by $2m/MZ$. The effect of proton-spin on the term-system of the hydrogen atom should consist in the conversion of each term into a pair—a "hyperfine doublet"—and this hyperfine structure should be on a scale about a thousandth of that of fine structure. Similar statements should hold for any alkali-metal atom of which the nucleus has $\frac{1}{2}$ for its spin quantum-number.

Of these guesses, those which pertain to spin are verified by the data. Those as to the value of μ_p, and as to the scale of the hyperfine structure, are verified only in order of magnitude. I postpone the topic of magnetic moment, and continue with that of spin.

It turns out that for many nucleus-types, the spin is different from $\frac{1}{2}$. We will see how this is ascertained. For the ensuing argument I adopt the symbols: I for the quantum-number of the nuclear spin; J for that of the angular momentum of the electron-family surrounding the nucleus; F for the resultant of I and J, which now takes over from J the role of angular-momentum-of-the-atom-as-a-whole. The relation between quantum-number and magnitude for all of these angular momenta is the same as in previous cases (page 399). The rule of composition of I and J is copied from that of page 399: the number of permitted mutual inclinations is either $(2I + 1)$ or $(2J + 1)$, whichever is not the greater, and the F-values are set by the rule. Throughout this section I use the word "level" for each of the states of the so-called "hyperfine mul-

tiplet" into which the nuclear spin resolves what but for it would be a single state.

In Figure 14.9 the upper row refers to various stages of resolution of any S state of the atom H^1, of nuclear spin $\frac{1}{2}$. On the extreme left the state appears unresolved, and so it does in the second column ("fs") owing to the absence of fine structure from S states. In the third column ("hfs") it is resolved into a hyperfine doublet, the two levels corresponding to parallelism and antiparallelism of I and J (which in the case of H^1 are proton-spin and electron-spin respectively); the F-values are 1 and 0. We disregard the other columns for a while and look at the lower row, which refers to any P state of

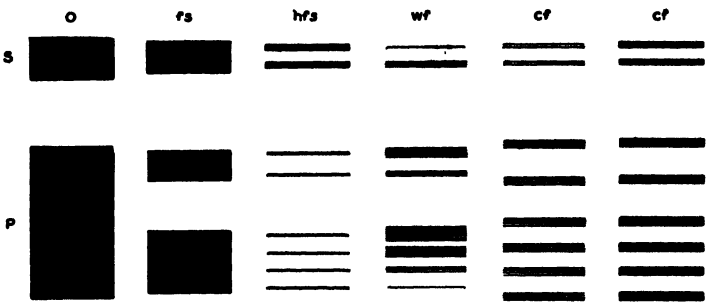


FIGURE 14.9—TWO STATIONARY STATES, SHOWING SUCCESSIVE DEGREES OF RESOLUTION DUE TO ELECTRON-SPIN, NUCLEAR SPIN, AND MAGNETIC FIELDS OF VARIOUS STRENGTHS.

the atom Na^{23} of nuclear spin $\frac{3}{2}$. In the first column, the P state appears unresolved. In the second, we see the fine-structure resolution into the states $P_{1/2}$ (above) and $P_{3/2}$ (below); the J-values are $\frac{1}{2}$ and $\frac{3}{2}$. In the third, these states have undergone the further subdivision due to nuclear spin, $P_{1/2}$ into two levels and $P_{3/2}$ into four.

One of the ways of identifying an I-value is suggested—but only suggested—by the previous sentence. If by analyzing the hyperfine structure of an optical spectrum (and this is often a very difficult task, hyperfine structure being the last frontier of spectroscopy!) the observer finds that a $P_{3/2}$ state is split into

four levels, this is compatible not only with the value $\frac{3}{2}$ for I but with any greater value. Suppose that the analysis should also show that states with higher J-values, such as $D_{5/2}$ or $F_{7/2}$, exhibit a four-fold resolution: this would establish $\frac{3}{2}$ as the I-value. One way of settling an I-value thus consists of finding a state which is split into x levels, x being less than $(2J + 1)$; the observer then identifies x with $(2I + 1)$.

Another way depends on the energy-differences or "intervals" between consecutive levels belonging to the same state. The existence of these intervals is proof that there are forces between the spinning nucleus and the electron-family which surrounds it. If these forces are properly to be called magnetic, they may reasonably be expected to vary as the sine of the angle θ between the vectors I and J; for either the magnetic moments associated with these two will be parallel to them, or else (by the reasoning of page 401) their non-parallel components will presumably change so rapidly in direction as to be ineffective, leaving only the parallel components to act. The energy due to the interaction should then be proportional to $cos\ \theta$, and the spacings ΔW between the levels should be proportional to the differences $\Delta\ cos\ \theta$ between the consecutive permitted values of $cos\ \theta$. I leave it to the student to derive these permitted values from the rule of composition of I and J with the aid of the formula for $cos\ \theta$ modelled after equation (13.17). The outcome is, that the consecutive spacings of the levels of a hyperfine multiplet should stand to one another thus,

$$\Delta W_m : \Delta W_{m-1} : \Delta W_{m-2} \ldots :: (I+J) : (I+J-1) : (I+J-2). \qquad (14.13)$$

The biggest interval may be that between the highest and the next-to-highest energy-value, or that between the lowest and the next-to-lowest; the choice depends on whether magnetic moment and angular momentum of nucleus are parallel or anti-parallel, but unfortunately not on this alone. If this "interval-rule" is verified, it enables one to determine $(I + J)$ and therefore I from the analysis of a single hyperfine multiplet of known J-value. Much use has been made of this method, and there

are a few cases in which a fairly accurate measurement of a chain of intervals has shown that it closely agrees with a chain of consecutive integers or half-integers, though more usually the intervals are small and the measurements rough and it is merely assumed that there is perfect agreement with that particular succession of integers or half-integers with which there is the closest apparent agreement.

Relative intensities of spectrum-lines may also be used in attempts to find the value of I. In atomic spectra it is necessary to invoke quantal formulae of appalling complexity and uncertain validity, while intensity-measurements are peculiarly difficult when the observer is working so near the limits of the possible as in the study of hyperfine structure. In molecular spectra there is a beautiful phenomenon known under the name of "alternating intensities." It is deplorable that owing to the omission from this book of statistical theory, the effect must here go unexplained.[1] It occurs in the P and R branches (page 444) of bands of diatomic molecules of which the two nuclei are identical. Of any two consecutive lines in either branch, one is brighter and the other fainter: the intensities of the two stand to each other in a ratio which the theory gives as $(I + 1)/I$. Notice that if I should be zero, this expression would require alternate lines to vanish. In the bands of He_2, the spacing of the observable lines is greater than one would expect from independent estimates of the moment of inertia of this molecule, but satisfaction is regained if one assumes that every other line is missing.

Finally the nuclear spin may be ascertained by observing the sub-division of the levels into sub-levels by applied magnetic fields; and this brings us to theories and experiments remarkable in their own right.

Return to Figure 14.9 and to the third column thereof. I leave it to the student to assign the F-values to the various levels there shown. He then will see that the fourth column exhibits the sub-levels into which these are resolved by an applied magnetic field, according to the rule of enumeration of

[1] See for instance Mulliken, *Rev. Mod. Phys.* 3, 149 (1931).

page 399 with F substituted for J. There are $(2F + 1)$ sub-levels for a level characterized by F.

At this point it looks as though the whole theory of the Zeeman effect given at such length in Chapter XIII has been repudiated. Indeed we assumed in that chapter that the angular momentum of the atom is J, and we derived such consequences as are shown on the extreme right of Figure 14.9: two sub-levels for S, two for $P_{1/2}$ and two for $P_{3/2}$. Now it has just been asserted that the angular momentum of the atom is F, and this assertion has led to the very different consequences shown in the fourth column. Yet of these two apparently-so-contradictory ideas, the earlier turns out to be a special case of the later, appropriate to a particular range of field-strengths which happens to be the one most usual in spectroscopic work.

To see how the two are interrelated, imagine the field-strength to be gradually increased from zero. At first, in what are technically called "weak fields"—fields so weak that even the best of spectroscopes will not show any change in the aspect of spectrum lines—the grouping of the sub-levels is as shown in the fourth column of Figure 14.9. (No attempt was made, in drawing this figure, to proportion the spacings accurately to the energy-differences; but in respect of grouping it is correct.) As the field-strength is further increased, the sub-levels depicted in column "wf" draw further apart in energy, but not in a uniform fashion. They tend to regroup themselves, and when the field-strength has attained the values customary in spectroscopic studies of the Zeeman effect, they have attained the grouping shown in the fifth column ("cf," customary field). In most though not in quite all spectroscopic analyses of Zeeman patterns, the scale and the resolving-power are such that each of the groups is taken for a single sub-level. Of these simulated sub-levels, there are two to an S state and two to a $P_{1/2}$ state and four to a $P_{3/2}$ state. They are indeed (as the fifth and the sixth columns of Figure 14.9 are intended to imply) the very sub-levels of which the theory of the Zeeman effect as given in Chapter XIII affords an adequate description.

In the customary field the number of sub-levels per group is $(2I + 1)$, and accordingly would reveal the nuclear spin if it could be counted. There are in fact a very few cases (found in the spectra of bismuth and thallium by Back, Goudsmit and Zeeman) in which the separation of the sub-levels is great enough to be discerned in spectrum lines. With a spectroscope, of course, one counts not the number of sub-levels but the number of line-components due to transitions from them into some other set of sub-levels; and experience does not justify us in expecting that the two numbers will be the same. Yet, by a rare if not an unique coincidence, in this case the two numbers *are* the same. The selection-principle which is involved is such, that, if the influence of the nuclear spin resolves every state of the atom into $(2I + 1)$ sub-levels, it also resolves every line connecting two such states into a group of $(2I + 1)$ components. There are in the literature wonderful photographs of the spectrum lines of bismuth exposed to a magnetic field, each line under high resolution exhibiting ten components and attesting an I-value of $9/2$.

This brings up to five the number of methods for finding nuclear spin. There is no element to which all five have been applied, and very few for which so many as three have been used; many of the recorded I-values are based upon data by two, many on only one. With these five and one other[1] we have come as far as the spectroscope can lead us, but another recourse remains.

Rabi's refinement of the Gerlach-Stern technique, and detection of nuclear spins

Nothing has yet been said about the sub-levels in the condition of "weak field," except for a statement about their grouping in the extreme case of vanishingly small field-strength; and nothing at all has been said about the conditions when the field-strength is somewhere between "customary" and "weak."

[1] Due to Ellett, and involving the observation of resonance radiation produced by polarized light acting on atoms of gases in magnetic fields.

Yet everything that has been said so far is the drawing of an analogy between I, J and F on the one hand, L and S and J of Chapter XIII on the other; many a statement in these last few pages has been copied from a statement in pages 397–400 with mere replacement of L by I, S by J and J by F. We therefore continue by pursuing the analogy further. It suggests that the passage from the weak-field situation to the customary-field situation is to be compared with the passage from the Zeeman effect to the Paschen-Back effect. There, we learned to think of the progressive "decoupling" of the angular momenta L and S from one another, as the field-strength passes from "customary" to "very strong"; here, we are to think of the progressive "decoupling" of the angular momenta I and J from one another, as the field-strength passes from "weak" to "customary."

Quantal theory not only ratifies this way of visualizing the situation, but leads to formulae for (*e.g.*) the energy-values of the sub-levels of a given state as functions of magnetic field-strength H. These formulae were originally worked out for the progressive decoupling of L and S in high fields; having found verification there, they have been carried over to the case which now engages us. They permit the physicist to follow (in theory) every detail of the passage from the weak-field grouping to the customary-field grouping of the sub-levels. All this however would not be much more than idle speculation, if the spectroscope were still our only instrument; for at the requisite field-strengths, the alteration of spectrum lines is nearly always on too small a scale to be observed. Rabi transformed the situation by refining the Gerlach-Stern technique to a degree so extraordinary that it is capable of giving the data desired.

In dealing with these experiments it must always be kept in mind that the apparatus treats every atom not according to its energy-value but according to its M_z-value, M_z standing for the component of the magnetic moment of the atom along the field-direction (which I always take as the z-direction). One is concerned therefore with the quantal formulae not for

the energy-values but for the M_z-values of the several sub-levels as functions of H. I quote them (after Breit and Rabi) for the various sub-levels of any S state, such as the Normal State of the atoms of hydrogen, the alkali metals and the noble metals—almost the only atoms which can be studied thus. (They are valid also for a $P_{1/2}$ state.)

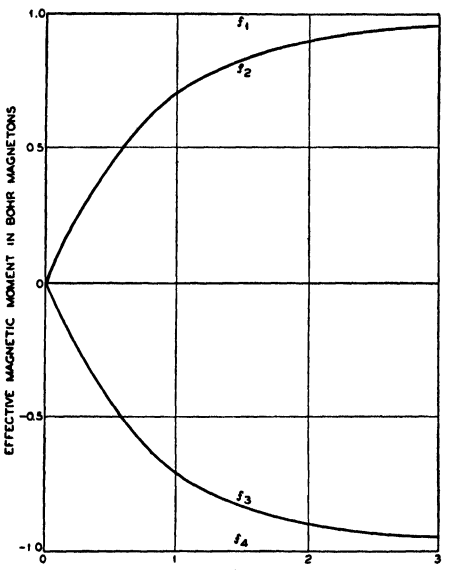


FIGURE 14.10—INFLUENCE OF MAGNETIC FIELD ON SUB-LEVELS IN A CASE LIKE THAT OF THE NORMAL STATE OF THE H^1 ATOM ($J = I = 1/2$). (After Rabi, Kellogg and Zacharias.)

When there is no magnetic field, any S state is subdivided into two levels for which F is $(I + \frac{1}{2})$ and $(I - \frac{1}{2})$ respectively. Call them N′ and N″, and denote by b the energy-difference between them. When there is a weak magnetic field, N′ and N″ are sub-subdivided into $(2I + 2)$ and $2I$ sub-levels respectively. Mark them by their m-values, m standing for the projection upon the field-direction of the angular momentum of the atom-as-a-whole, in terms of $(h/2\pi)$ as unit. The per-

mitted values of m are $(I + \frac{1}{2})$, $(I - \frac{1}{2})$, $(I - \frac{3}{2}), \ldots,$ $-(I - \frac{1}{2})$, $-(I + \frac{1}{2})$. The first and the last of these values are attached each to a single sub-level, belonging (in the weak field) to the level N′. Each of the rest is attached to a pair of levels, one belonging to N′ and the other to N″.* Now let a stand for $2m/(2I + 1)$; let g stand for the g-factor associated with the extra-nuclear electron-system and the angular momentum J; let x stand for $(g/b)(eH/2m_0c)(h/2\pi)$ or $(gH/b\mu_B)$ (μ_B standing for the "Bohr magneton," page 394). The formulae of Breit and Rabi are as follows:

$$M_z = \pm \frac{a + x}{2(1 + 2ax + x^2)^{\frac{1}{2}}} g\mu_B, \qquad (14.14)$$

and they describe a singular flock of curves, which can hardly be appreciated without looking at such examples as appear in Figure 14.10 ($I = \frac{1}{2}$, the case of H^1) and Figure 14.11 ($I = \frac{7}{2}$, the case of Cs^{133}). In these diagrams the unit of length along the vertical axis is μ_B.

For the two sub-levels characterized by the extreme values of m (*viz.*, $\pm (I + \frac{1}{2})$) and belonging in the weak field exclusively to N′, the first factor is equal to $\frac{1}{2}$ and the two sub-levels are distinguished by the choices of sign. M_z for these is independent of field-strength; the curves degenerate to the horizontal straight lines at ordinates ± 1. For all the rest, there is dependence of M_z on H, and of a curious sort. Any value of m, say $(I - \frac{1}{2})$, corresponds to two different values of M_z which are equal in magnitude and opposite in sign. The opposite value of m, in this case $-(I - \frac{1}{2})$, likewise corresponds to two different values of M_z which are equal and opposite, and each of which at zero field-strength coincides with one of the previous two. Now consider any two of these M_z-values which correspond to equal and opposite values of m and coincide with each other at $H = 0$. As H is increased from zero, these two diverge from each other, and one of them actually passes through zero and changes in sign while the other moves in the opposite sense and never vanishes. A special case is

that of $m = 0$ (occurring only when I is half of an odd integer); it comprises two M_s-values which coincide at zero when $H = 0$, and depart more and more from zero to equal extents in opposite senses as H is increased.

The pictures help to make vivid the regrouping which characterizes the passage between "weak field" (near the left-hand edge of each diagram) and "customary field" (beyond the

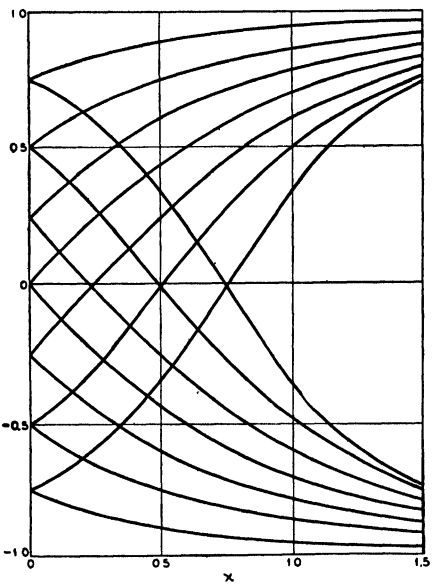


FIGURE 14.11—INFLUENCE OF MAGNETIC FIELD ON SUB-LEVELS IN A CASE LIKE THAT OF THE NORMAL STATE OF THE Cs^{133} ATOM ($J = 1/2$, $I = 7/2$). (After Cohen.)

right-hand edge of each diagram). In the early days of the Gerlach-Stern technique, H and x were so great that each of the two customary-field groups made a single line upon the photographic plate, the position of these being controlled by equation (14.14) with a set equal to unity. What the observers accordingly measured was $\frac{1}{2}g\mu_B$, here equal to μ_B the magnetic moment of the spinning electron, as we saw in Chapter XIII.

When H and x are reduced so that x lies (roughly speaking) in the right-hand part of the diagram or not very far beyond it, the sub-levels in the groups may be distinguishable when the technique is refined to the degree attained by Rabi and his school. Figure 14.12 shows that there are four of them in one of the groups for sodium, and thereby fixes $\frac{3}{2}$ as the nuclear spin of Na^{23}. Another way of testing the theory in general and of measuring I in particular may be readily grasped by noticing that in Figure 14.11 there are certain values of x for which certain M_x-values vanish, so that when H is adjusted to

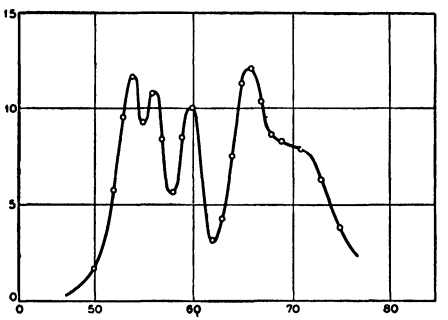


FIGURE 14.12—CURVE SHOWING FOUR OF THE EIGHT SUB-LEVELS OF THE NORMAL STATE OF Na^{23}, IN A FIELD SOMEWHAT BELOW THE "CUSTOMARY-FIELD" RANGE. (After Rabi and Cohen.)

any one of these there will be atoms traversing the inhomogeneous field without suffering any deflection. Let then the detector be placed where undeflected atoms will reach it, and let H be varied: the curve of detector-reading *vs.* H will exhibit three peaks (not counting the one at $H = 0$) if the nuclear spin has the value $\frac{1}{2}$ for which Figure 14.11 is applicable. The general rule is that there should be no peaks for $I = 0$ or $\frac{1}{2}$, one peak for $I = 1$ or $\frac{3}{2}$, two peaks for $I = 2$ or $\frac{5}{2}$, and so on. What remains of ambiguity can be dispelled. If there are two or more peaks, let $H_1, H_2 \ldots$ denote their abscissae; then all the intervals $(H_{i+1} - H_i)$ are equal in every case, but H_1 is equal to the half or the whole of their common value according as I is an integer or is midway between two integers. If there

is one peak only, a decision between $I = 0$ and $I = \frac{3}{2}$ may be drawn from finer features of the curve. This is the "method of zero moments." For caesium it yielded to Cohen a curve with three peaks, for indium the four-peaked curve of Figure 14.13 to Millman, Rabi and Zacharias; the intervals attest the I-values $\frac{7}{2}$ and $\frac{9}{2}$ for Cs^{133} and In^{115} respectively. Several alkali-metal atoms have exhibited one-peaked curves.[1]

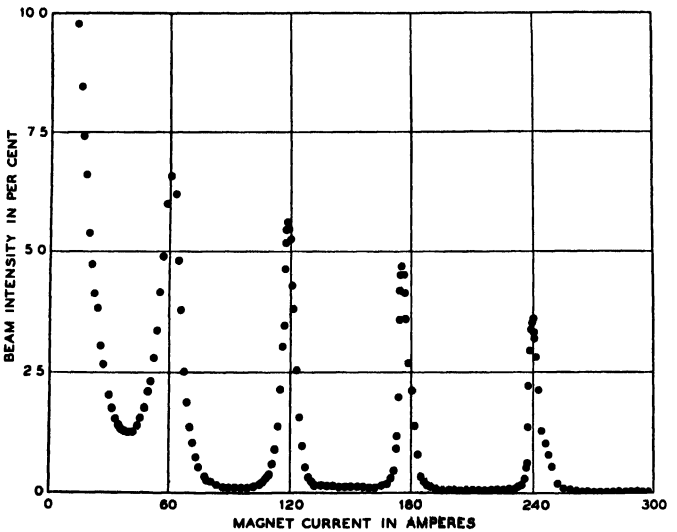


FIGURE 14.13—APPLICATION OF THE METHOD OF ZERO MOMENTS TO INDIUM, DEMONSTRATING 9/2 AS THE VALUE OF ITS NUCLEAR SPIN. (Millman, Rabi and Zacharias.)

The improvements in technique which made possible these results included the lengthening of the beam, to as much as 245 cm in the latest work; the sending of the beam through a preliminary inhomogeneous field, in which each one of the deflected pencils was broadened into a fan by virtue of the distribution-in-speed of the atoms, and a narrow portion of the

[1] Manley and Millman, **PR 51**, 19 (1937); Manley, Millman, Fox, **PR 49**, 867 (1936); Fox and Rabi, **PR 48**, 746 (1935); Millman, **PR 47**, 739 (1935); earlier literature cited in these.

broadened pencil was admitted through a slit to become the primary beam of the actual experiment; and the use of wires as detectors, the atoms being ionized by the wires in the manner mentioned on page 73.

Collation of nuclear spins

Estimates of nuclear spin have by now been made for upwards of a hundred different atom-types. This does *not* signify as many different elements, spin being a property which depends on mass-number M and thus may vary—and often does markedly vary—from one isotope to another of a single element. The student will divine that an element with two or more abundant isotopes will display a pattern of hyperfine structure composed of contributions from them all, which may be superposed perfectly but may on the other hand be entirely different. It seems rather wonderful that such patterns as those of mercury and lead should have been interpreted, and the contributions of seven isotopes of the one and four of the other distinguished.[1] Evaluation of spin for separated isotopes has been achieved for hydrogen only: the "alternating intensities" in the bands of H_2 are markedly different according as the molecules are H^1H^1 or H^2H^2.

From these values of nuclear spin a rule of great simplicity and importance has been derived. I quote it first in sections, because the data do not speak with equal clearness for all parts of it; and I preface it by saying that integer values of I are called "even spins," and values $\frac{1}{2}, \frac{3}{2}$ . . . are called "odd spins."

(*a*) *Nuclei of odd mass-number have odd spins.* This is attested by quantities of data obtained by all the methods.

(*b*) *Of the small category of four stable nuclei for which Z is*

[1] The hyperfine structure of various Hg lines including 5461, which had long been studied by spectroscopists who called it "fine structure" before the days of quantum theory, was first interpreted by Schüler and Keyston (**ZP 72**, 423; 1931). Boggs and Webb explained its variations with the mode of excitation (**PR 48**, 227 (1935) and literature there cited). For the case of lead see Rose and Stranathan (**PR 49**, 916; 1936) who carried forward earlier analyses (see literature there cited) and employed the hyperfine structure as an index of the relative abundances of the four principal isotopes.

odd and $M = 2Z$, *three have the "even" spin unity*, that of the fourth being unknown. Knowledge as to N^{14} is derived from alternating intensities, as to Li^6 from Rabi's technique, as to H^2 from both; B^{10} remains inviolate to all attempts.

(*c*) *The three nuclei* He^4, C^{12} *and* O^{16} *have zero spins*, as the total vanishing of alternate band-lines tells us.

(*d*) *Of other nuclei of even mass-number, all that is known is that they show no perceptible hyperfine structure.*[1] This is taken as meaning that they all have the even spin-value zero, but might possibly mean that because of smallness of the forces between nucleus and electron-shell the hyperfine levels are too close to each other to be distinguished.

Accepting the former of the two interpretations of section (*d*), we come to the presumptive General Rule:

The spin of any nucleus is "odd" or "even" according as the mass-number is odd or even.

This is one of the principal reasons for the doctrine that nuclei are assemblages of particles of two kinds only, protons and neutrons. The neutron is the corpuscle of mass-number 1 and charge 0 of which we shall learn much in Chapter XVII. There is no way of ascertaining the angular momentum of the isolated neutron. If however we assume that like the proton it is of spin $\frac{1}{2}$, and that any two spins of magnitude $\frac{1}{2}$ must orient themselves either parallel or anti-parallel to one another in the nuclear cosmos, then the General Rule follows immediately.

Let us ascend a little way along the sequence of nuclei in order of increasing mass. First come the proton of spin $\frac{1}{2}$ and the deuteron of spin unity, facts which indicate that the deuteron is composed of a proton and a neutron with parallel spins. Nuclei formed by transmutation are always too scant for these measurements, so we must leap over H^3 and He^3 and resume our progress at the alpha-particle He^4. Its spin

[1] Thus if an element is a mixture of odd and even isotopes, each of the latter produces just one unresolved component in the hyperfine structure of any line. These components may or may not coincide; if they do not, there are said to be "isotope shifts."

is zero, so that the protons and the neutrons in it must be anti-parallel two by two. At Li^6 we find again the spin unity, as though a deuteron had been attached to an alpha-particle; at Li^7 the spin is $\frac{3}{2}$, as though an extra neutron had set itself with spin parallel to this deuteron. Be^9 and B^{10} and B^{11} have been impervious to all the methods of assessing I, a great disappointment! C^{12} and O^{16} are both spinless, strengthening the notion that they may consist of three and four alpha-particles respectively; while N^{14} repeats the spin of the deuteron. As we continue on and on, higher and higher values of I make their appearance: $\frac{5}{2}$ first at $M = 35$, $\frac{7}{2}$ at 45 and $\frac{9}{2}$—the highest yet authenticated—first at 83. Yet even beyond mass-number 200, there are still examples of odd nuclei having the least possible "odd" spin-value of one-half.

Nuclear magnetic moments

In 1933 Stern himself, with the aid of his associates Frisch and Estermann, advanced the Gerlach-Stern technique to such a degree that with it they were able to measure the magnetic moments of hydrogen nuclei, even though these are but a thousandth as great as that of the spinning electron in the sodium atom, which at first it had been considered a triumph to detect! The ways in which Nature helped and hindered them are remarkably interesting, and illustrate many details of what we have lately been learning.

The technique cannot be tried at all upon a bare proton or deuteron, for the electrodynamic forces suffered by the moving charge in the magnetic field would be immensely greater than the force on which the experiment depends. If it is tried on a hydrogen atom the magnetic moment of the orbital spinning electron dominates the result, as we saw in Chapter XIII; indirectly one may yet evaluate the nuclear moment, as we shall presently see, but not directly. In trying it on H^1H^1 molecules we might anticipate another defeat, since each of them comprises two electrons. However in H^1H^1 molecules in the normal state, the two electrons are oriented anti-parallel

to one another: their magnetic moments like their spins are equal and oppositely directed, cancel one another and leave the scene free for the protons. Nature having helped to this extent we might expect her to overdo it, and orient the protons antiparallel to one another and so annul everything. So she does in some of the molecules but not in all; in the rest the protons are set parallel to one another, conveniently doubling the magnetic moment which it was Stern's problem to measure. In gaseous hydrogen which has had ample time to come to equilibrium at room-temperature, one-quarter of the molecules are of the former or "para-hydrogen" type, three-quarters are of the latter or "ortho-hydrogen" type. At lower temperatures the proportion is more in favor of para-hydrogen; the ratio appropriate to a lower temperature can be preserved for a while after the gas is swiftly brought to a higher. The specific heats of the gas vary appreciably with the ratio of "para" to "ortho" molecules; it was because of this that the existence of the two kinds was proved after having been predicted as a consequence of nuclear spin.[1]

"Para" molecules must accordingly go through the apparatus undeflected, but "ortho" molecules will be distributed among deflected and undeflected pencils in a manner capable of indicating (to a good enough experimenter) the value $2\mu_P$ of the doubled proton moment. With perfect resolving-power and ideal conditions, one would expect three of these pencils, the two protons in alignment having a net spin-value of unity. Ideal conditions are impossible because of the distribution-in-speed of the molecules which broadens the pencils, and the rotation of the molecules which imparts to them an extra magnetic moment. Both are minimized by maintaining at liquid-air temperature the box whence the beam of hydrogen proceeds (the name "oven" for the box now becoming rather ludicrous) but the unwanted magnetic moment due to rotation is never-

[1] This is another of the noteworthy episodes of contemporary physics (and chemistry) which cannot here be recited because of the omission of statistical theories from this book. See a series of articles by Bonhoeffer, Harteck, Eucken, Hiller, Clusius in *ZS. phys. Chem.* **B4**, 113 ff. (1929). The idea came first to Dennison (**PRS 115**, 483; 1927).

theless of the order of magnitude of the one which is sought. No such degree of resolution as that exemplified by Figure 13.4 is attainable, even with the refinements most lately introduced by Stern (slits 0.02 mm in width, fields 10 cm long and having inhomogeneity dH/dz amounting to 154,000 gauss/cm). The three maxima are however discernible in the latest curves.

When the experiment was first performed it gave as startling a result as any which has emerged of late to baffle the theorist. The expectation (page 451) had been that the proton moment would have the value, $eh/4\pi Mc$, baptized in advance "the nuclear magneton"; here, I recall, M stands for the mass of the proton. The world would seem simpler today, had this forecast been justified. It turned out however that the value is about two and one-half times as great! This unwanted result has stood the test of time, the latest experiment[1] from the school of Stern having given the result,

$$\mu_P = 2.46 \text{ nuclear magnetons } \pm 3\%$$

while the indirect method gives a value not far off. The value offered for the deuteron has not been raised to such a degree of exactitude, being given as between 0.5 and 1.0 nuclear magnetons.

What I just called "the indirect method" consists in ascertaining the energy-difference b between two levels of a hyperfine doublet (or multiplet), ascribing it to a purely magnetic interaction between the nucleus and the orbital electron-family, and embodying this theory in a formula. Such a formula was indicated on page 452, but that one is based upon assumptions too crude to be of any use. The quantal formula commonly used for hydrogen and alkali-metal atoms is the following,

$$b = (8\pi/3)\left(\frac{2I+1}{I}\right)\mu_N\mu_B\Psi^2(0) \qquad (14.15)$$

[1] Estermann, Simpson and Stern, **PR 52**, 535 (1937); prior literature there cited. This paper also comprises observations on beams of the molecules H^1H^2, which are divided into two deflected pencils: it is thought that the nuclei H^1 and H^2 are completely decoupled from one another. For the magnetic moment of the deuteron see Estermann and Stern, **PR 45**, 761 (1934).

μ_N standing here for the to-be-computed magnetic moment of the nucleus, μ_B for the Bohr magneton and $\Psi^2(o)$ for the squared amplitude of the Schroedinger wave-function at the position of the nucleus; this last is known exactly for hydrogen and approximately for the rest. The formula is due to Fermi, and has been much amended by Goudsmit and others in the hope of making it applicable to more complicated atoms. The value of b is derived sometimes from hyperfine structure observed with the spectroscope, sometimes from experiments done by Rabi's technique with values of unity or less for the variable called x in equation (14.14); one sees from that equation that the pattern of the pencils depends upon b. The latter method is still practicable when b is much too small to be discerned with the spectroscope. Such is the case with both of the isotopes of hydrogen, subjecting atomic (*not* molecular!) beams of which to the technique, Kellogg, Rabi and Zacharias[1] measured values of b from which by the foregoing formula they computed magnetic moments. Instead of these, I cite later values obtained by Rabi and his school, with still another ingenious method [2] depending on the fact (foretold by Rabi) that if a flying atomic magnet is passing through a uniform field in which it occupies one of its permitted orientations, and if superposed upon this uniform field there is a transverse oscillating field of small amplitude, then there is a high probability that the magnet will shift to another of its permitted orientations if and only if the frequency of the oscillating field is equal to the frequency of the Larmor precession which the magnet executes in the main field. In the experiments the uniform field is of strength so great that the nuclear magnet is completely decoupled from the orbital electron-family, and responds to the fields as though it were free. The values are:

$$\mu_P = 2.78 \pm 0.02, \quad \mu_D = 0.853 \pm 0.006$$

for proton and deuteron respectively, in nuclear magnetons.

Some forty values of nuclear moment have already been published, referring to nucleus-types sprinkled all along the

[1] PR 50, 472 (1936). [2] PR 53, 318 (1938).

list of elements from hydrogen ($M = 1, 2$) to bismuth ($M = 209$). The formulae are believed to be roughly but not perfectly correct, a view which is supported by the fact that they give somewhat different values of μ_N when applied to multiplets due to the interaction of a nucleus with different states of its orbital electron-family. (States of different electron-families may also be invoked, since hyperfine-structure occurs in the spectra not only of neutral but also of ionized atoms.) Even for hydrogen formula (14.15) must be wrong if the uncertainty of neither of the two values just given for μ_p has been underestimated. On this account I will say only that the accepted values range between 0.4 and 6 nuclear magnetons, with no discernible trend from one end to the other of the list. Even with these uncertainties, there are two inferences assured and important enough to be set down in culmination of this chapter.

First, *all the magnetic moments of nuclei are so very much smaller than that of the spinning electron as to forbid us from introducing electrons into our nuclear models* without so much changing them as to deprive them of their right to the name, or else imagining other particles to counteract them. This is one of the reasons for imagining nuclei as containing no particles but protons and neutrons.

Second, *in the composition of the deuteron the neutron figures as a particle of spin* $\frac{1}{2}$ *and of magnetic moment about* 2 *nuclear magnetons*, so linked with the proton as to be oriented alike as to spin but oppositely as to magnetic moment.

Now we have finished with the study of stable nuclei in their stable interrelations with the world about them, and turn to the study of the art and science of altering them.

CHAPTER XV

First Period of Transmutation

Near the beginning of this book the nucleus appeared before us as a particle: a thing possessed of mass and charge, and yet so small that it could adequately be conceived as a mere geometrical point—a mass-laden point surrounded by an inverse-square electrostatic field. Later on in these pages, it was invested with the further properties of angular momentum and magnetic moment. The image of a literal structureless point became hard to maintain, with all these qualities requiring to be assigned to it; yet even so, physicists might have preferred to sustain it rather than to engage in contriving nucleus-models, had they not been confronted with other facts. Comparison of the masses of the different kinds of nuclei (Chapter III) and comparison of the angular momenta (Chapter XIV) supplied reasons of the utmost force for supposing the nuclei—all but the lightest among them—to be composite. Already in these pages we, as a result of those comparisons, have begun to envisage each nucleus as an assemblage of particles of two varieties, the proton and the neutron.

Thus even if no one had ever transmuted or altered a nucleus, and even if nobody had observed any nucleus-type which disintegrates of itself, there would nevertheless be a large amount of knowledge of these bodies. This knowledge would by itself be amply sufficient to justify us in imagining the kernel of an atom as a microcosm or assemblage of protons and of neutrons, sufficient indeed to indicate the laws of these assemblages. And yet, today it is the lesser part of "nuclear physics." Of this the greater part consists of what is known about unstable and explosive nuclei which resolve themselves

into fragments, this being the field of radioactivity; and of what has been achieved in the disruption, fusion, and construction of nuclei, which is the art of transmutation and is the consummation of contemporary physics. As I said in an earlier passage, we learn to know the atom by taking it apart. So also are we now learning to know the atom-kernel, except that to speak of taking it apart is somewhat inexact: rather, we watch it falling apart of itself, or we bring two nuclei violently into contact with each other, so that the two composite systems merge and from their union two different systems emerge. This for the present is the most romantic, and ultimately is likely to remain by far the most influential, division of nuclear physics.

The art of transmutation is less than twenty years old, having been initiated no longer ago than 1919. In so brief a time, it has progressed so far that writers are able to systematize it, arranging their material without regard to the succession in which discoveries came. Some have already done so, and before very long such will probably be the universal practice; but the beginnings being still so near, the chronological order of telling the story will be adopted in this book for possibly the last time. The most serious departures from the order-in-time will occur when I speak of radioactivity, of which there have already been so many occasional mentions in these pages, but not yet a proper account. "Natural" radioactive nuclei were known to the number of some thirty-five, before the year 1910; but since the end of 1933, upwards of two hundred more have been made by transmutation, and though these are called "artificial" there is no deep reason for setting the two groups apart: I shall be speaking of them both when the chosen place is reached for treating of the latter.[1]

[1] These final four chapters would have been almost impossible to write but for the compendium of nuclear physics published in *Reviews of Modern Physics* (8 82, 9 69, 245; 1936–37) by Bethe, Bacher and Livingston, to which here once for all I make general reference, and express much gratitude. To it also I refer for its complete tabulation of the literature.

Early transmutations by means of alpha-particles

At the beginning of the art of transmutation, as at the beginning of the idea of the nuclear atom-model, stands (or rather, flies) the alpha-particle; and at the beginnings of both stands the figure of Rutherford, who used these particles to achieve in 1911 the theoretical, in 1919 the experimental feat.

The alpha-particle is potent in transmutation because it is itself a nucleus—because its charge is relatively low—and because its kinetic energy (as it springs from a natural radioactive substance) is colossally high. Were it not a nucleus, the presumption is that it could not transmute at all; for (as we shall see abundantly) all known transmutations, excepting for a few provoked by light, are results of interactions between nuclei which have come into contact with each other. I use deliberately so crude a word as "contact": it may indeed evoke a faulty picture, but understanding of the intra-nuclear world is not yet far enough advanced to permit anyone to decry it. It is certain at least that for transmutation, two nuclei must come extremely close together, within distances of the order of 10^{-12} cm or smaller—that so close an approach is hindered by the electrostatic force between the positive charges which both of them bear, unless one is a neutron—that for overriding this hindrance, either must be projected at the other with enormous kinetic energy—and that, other things being equal, the less highly the nuclei are charged the smaller the hindrance will be. It is not to be assumed that this electrostatic repulsion is the only obstacle, but at least we know that it is one.

Just how much of all this was realized by Rutherford before 1919 will never be known. It must not be imagined that he could anticipate success as a matter of course. If the speed of the alpha-particles at his disposal had been lower by forty per cent than it was, he would have observed nothing; if nuclei had been smaller by a couple of orders of magnitude than they are, he would have observed nothing; if he had been obliged to prove transmutation by detecting with chemical

means the new-made elements, he would have detected nothing. The margin was small, but with the technique of contemporary physics and with Rutherford to apply that technique, it was sufficient.

What Rutherford did was to project alpha-particles into gases or against very thin films of metal, and to look for scintillations on a fluorescent screen placed far beyond the reach of the alpha-particles themselves. Farther off yet, in fact, he placed the screen; he set it so far away that even the nuclei of hydrogen atoms, recoiling elastically from the impacts of the

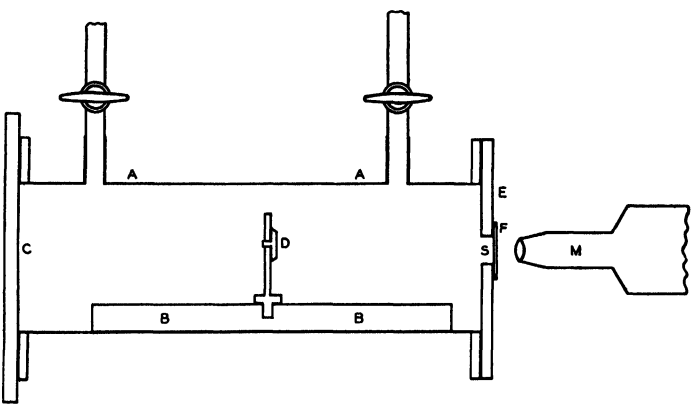


FIGURE 15.1—APPARATUS WITH WHICH RUTHERFORD DISCOVERED TRANSMUTATION. (*Philosophical Magazine.*)

alpha-particles, could not arrive at it. For a proton struck centrally by an alpha-particle, the one-to-four ratio of masses leads to a final speed 1.6 times as great as that of the impinging particle, and our general knowledge of range-vs-speed relations (briefly mentioned in Chapter VIII) indicates that the proton will run four times as far through air (or any gas) as the alpha-particle would have run had it not made the collision. In the apparatus sketched in Figure 15.1, where the source of alpha-particles is at D and the fluorescent screen at S, imagine that the tube is filled with air at normal temperature and pressure with a few hydrogen molecules intermingled. Most of the

alpha-particles will travel seven centimeters outward from D, losing their energy gradually (by ionizing the gas—Chapter VIII) over these seven centimeters so that they can go no farther. If, however, one of them makes a central impact against the nucleus of a hydrogen molecule very close to D, this proton will travel about thirty centimeters outward from D. Thus if air of normal temperature and pressure is bombarded by alpha-particles of seven-centimeter range, and scintillations are observed upon a screen beyond, the observer must reckon with the chance that they may be due to the nuclei of hydrogen contaminating the air, so long as the screen is less than about thirty cm away.

On the principle that the best way to deal with a possible source of trouble is to examine it minutely, Rutherford prepared for his attempt at transmutation by a study of the protons which recoil when alpha-particles are fired into gaseous hydrogen. His pupil Marsden had begun with such a study in 1914; resuming the research as soon as the war was ended, Rutherford counted the scintillations and plotted their number for what, in effect, were various thicknesses of matter interposed between screen and source. Usually in such work one does not draw the screen back steadily to increase its distance from the source; instead one leaves it fixed and varies the pressure of the gas, or else interposes a series of thin foils of aluminium or of some other metal or of mica, each of which has the same "stopping-power"—that is, slows down the particles, be they alphas or protons, to the same extent—as a known thickness of air. (For instance, a thickness of mica of weight 1.43 mg per cm^2 is equivalent in stopping-power to one cm of air at 15° C. and 760 mm Hg). In curves of the customary sort, number of scintillations is usually plotted the vertical axis, number of centimeters of air along the horizontal; but usually some other substance does duty for the air, and the thicknesses are translated into equivalent thickness of this standard gas at standard temperature and pressure[1] before the curve is drawn.

[1] 15° C. and 760 mm Hg, in Rutherford's convention; 0° C. and 760 mm Hg, in the convention used by certain other physicists, notably the Halle school.

Historic curves of this type appear in Figure 15.2. The bombarding alpha-particles proceeded from radium C′ and had a range of seven cm. Notice first the curve marked B: it was obtained with hydrogen (mixed with carbon dioxide) in the tube; it testifies that the scintillations did not cease until between the screen and source there was the equivalent (in gas and in mica) of thirty cm of air, and then they *did* cease. These were indeed the signs of the recoiling nuclei of hydrogen molecules conform-

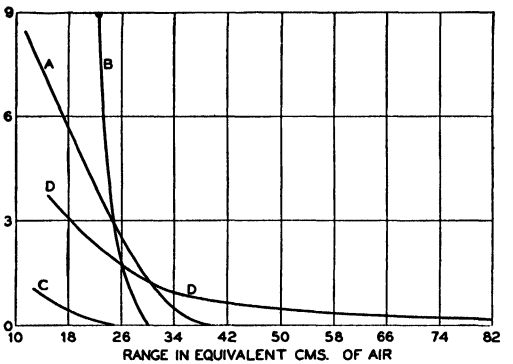


FIGURE 15.2—CURVES INDICATING THE RANGES OF PROTONS RECOILING FROM IMPACTS OF ALPHA-PARTICLES (A, B) AND OF PROTONS NEWBORN FROM TRANSMUTATIONS PROVOKED BY ALPHA-PARTICLES (C, D). (Rutherford; *Philosophical Magazine.*)

ing to the simple theory. The shape of the curve makes clear that many of the protons did not penetrate nearly so far as 30 cm; these were presumably those which were struck off-center and glancingly by the alpha-particles, and did not receive the maximum possible kinetic energy.

Curve C was obtained with oxygen; what there is of it was ascribed to commingled hydrogen; anyhow it does not reach so far as the critical abscissa 30. Curve A is more sensational: very definitely it extends beyond the critical abscissa, very definitely there are corpuscles more penetrative than either the primary alphas or such nuclei of stray hydrogen atoms as the alphas might have found to strike. This curve was obtained

with air. Since with pure oxygen there is no sign of such extraordinary corpuscles, nitrogen was held responsible for them, and so it has proved to be. The most astonishing of all the curves is D. It stretches far beyond the critical point, flashes appearing on the screen when even as much as 90 cm of air-equivalent lay between it and the substance which the alphas were striking, which was very thin aluminium foil. Aluminium under alpha-particle bombardment sends out particles three times as penetrating as the very fastest which could possibly occur among recoiling nuclei of such atoms of hydrogen as the metal might possibly contain by chance!

Rutherford guessed that these particles were protons, from the faintness of the scintillations (which were decidedly fainter than those of alpha-particles of equal range, and resembled those of recoiling protons). His guess has been confirmed in many ways: roughly from the aspect of the trails as rendered visible in Wilson chambers, much more sharply from the density of ionization along their paths as measured directly by apparatus which we shall presently consider (page 520), and from deflections in electric and magnetic fields: these last proved their charge to be positive, and gave for the charge-to-mass ratio a value agreeing with that for protons within the experimental error, which in Stetter's experiments [1] amounted to 5%. If anyone should wish to affirm that these particles differ in mass by as much as a couple of parts in a hundred from the hydrogen nucleus, it would probably be impossible to prove him wrong by these experiments; but if any such notion were to be adopted, the whole beautiful system of mass-interrelations among kernels which we shall later encounter would go badly askew; the integrity of this system is therefore the best individual piece of evidence that Rutherford was here observing protons.

Since in these interactions an alpha-particle impinges on a nucleus of nitrogen or aluminium and a proton emerges, we may express a part—the only part yet mentioned—of what is actually observed, by writing down the symbols of these nuclei

[1] PZ 42 741 (1927); reviewed in BSTJ, October 1931.

according to the rule of page 93. I recall that such a symbol is composed of the chemical symbol of the element, adjoined on the left below by its atomic number and on the right above by the mass-number of the isotope involved. *The symbols* were formerly used to denote the completed atoms, but *henceforth* they *shall denote the nuclei solely*; the reader ought not to be confused. *Impact of* $_2He^4$ *upon* $_7N^{14}$ *yields* $_1H^1$, *and impact of* $_2He^4$ *upon* $_{13}Al^{27}$ *yields* $_1H^1$.

In writing mass-numbers into this formulation, I have gone somewhat beyond what the experiments say outright; for though it is definitely known that the isotope 4 of helium and the isotope 1 of hydrogen are the ones here involved, there is nothing in the immediate data to show that N^{14} and Al^{27} are being transmuted, rather than other isotopes of nitrogen and of aluminium respectively. In these two cases it just happens that the doubt is practically negligible, since no other isotope of aluminium is known at all, and the only other isotope of nitrogen comprises only a few per thousand of all the nitrogen nuclei which there are. Of course one might have guessed at the start that all isotopes of an element would behave alike in respect of transmutation; but this is thoroughly false; of all the properties of the atoms, transmutability and radioactivity are *the* two which vary most dramatically from one isotope to another. To identify the right isotope is therefore very important and happily it is usually possible (though mistakes have been made in the past, and some may still be made); the ways will be mentioned later.

Now an assumption will be made for which the Rutherford experiments offer no proof, and yet it has been supported by so great a variety of evidence that it is more impregnable than many statements founded on observations directly. It will be assumed that *in the vast majority of cases, the transmutation results in just two nuclei: the proton (or whatever else is observed to emerge) and one other, the "residual nucleus" or "residue," which contains all the rest of the mass and the charge of the original nuclei.*

This assumption being made, we may immediately complete

the partial formulations at the end of the last paragraph but two, and write them in the guise of chemical equations—*equations of nuclear chemistry*:

$$_2He^4 + {}_7N^{14} = {}_1H^1 + {}_8O^{17}. \qquad (15.1)$$

$$_2He^4 + {}_{13}Al^{27} = {}_1H^1 + {}_{14}Si^{30}. \qquad (15.2)$$

The assumption (if correct) enables us to identify the element to which the residue belongs; and to determine as well the isotope to which it belongs, if the isotope of the struck nucleus on the left-hand side of the equation has been rightly identified (or reversely we can identify this last, if the residue reveals its own mass-number).

This grand assumption is directly verified for the reaction (15.1) by such expansion-chamber pictures as Figure 15.3, in which are seen the tracks of the impinging alpha-particle, of the emerging proton and of just one more particle, which is proved by the aspect, length and orientation of the track to be an oxygen nucleus. The expansion-chamber has also been used to verify it for a few other reactions which we shall presently notice (consult Figures 16.8, 17.5, 17.6), and incidentally has confirmed a couple of very exceptional cases in which there are three final nuclei instead of just two. Usually the transmutation is made by bombarding a solid target with

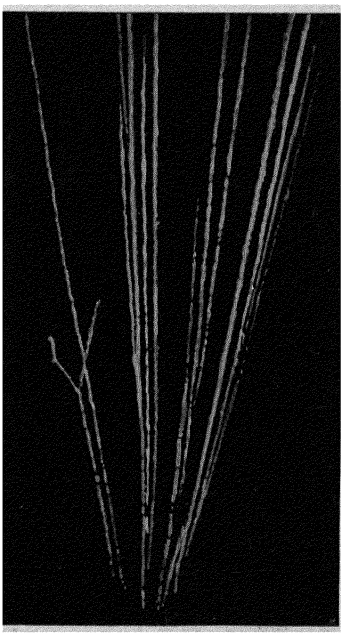


FIGURE 15.3—EXPANSION-CHAMBER PHOTOGRAPH OF THE REACTION OF TRANSMUTATION $He^4 + N^{14} \rightarrow H^1 + O^{17}$. (Blackett; *Proceedings of the Royal Society*.)

alpha-particles or other nuclei, and then the residues remain embedded in the solid; but even in such cases they may declare and identify themselves by their radioactivity. Even when the residue remains entirely undetectable, the mass-interrelations to which I have already alluded may speak very strongly for its continued and undivided existence.

These mass-interrelations having already been twice mentioned, it seems best to speak of them at once a little more explicitly, even though the proper place for discussing them is far ahead. In (15.1) as in (15.2) and all similar equations, the fourth mass-number is so chosen as to make a balance, the chemical equation being considered as an algebraic equation in the mass-numbers. The mass-number, however, does not represent the best available knowledge of the mass in question, which in the most favorable cases is known—by mass-spectrograph experiments, and in terms of the mass of ${}_8O^{16}$ taken as unity—to five or six significant figures. If into these equations we should insert those mass-values, what should we find?

Well, neither (15.1) nor (15.2) is the best available example, for in each there appear one or two nuclei for which the existing mass-spectrograph data are not first-rate. There are, however, some reactions (we shall presently consider a few) of which it can be said that the mass-spectroscopists have provided excellent mass-values for all the kernels concerned. When the equations are filled in with these mass-values, there emerges the momentarily-astounding fact that they do not balance!

The astonishment is only momentary, though; for actually there is a wonderful agreement here, and not a disagreement. In the mass-spectrograph the ions are moving so slowly, that the mass-values derived from it are not sensibly different from the rest-masses. In transmutation this is not usually true: the nuclei are often moving so rapidly that there is a sensible enhancement of their masses over the rest-values, according to the relativistic dependence-of-mass-on-speed (Chapter II). Another and usually more convenient way of saying the same thing is, that for a nucleus of rest-mass M_0 figuring in a reaction of transmutation the kinetic energy (call it K for the moment)

is likely to be so great that in the correct expression for the mass —*viz.* $(M_0 + K/c^2)$—the term K/c^2 is a small but appreciable fraction of M_0. Anticipating the conclusion of pages 532–533, it can be affirmed that *when the equations of nuclear chemistry are written as algebraic equations in the true masses of the nuclei, they balance.*

One may think this so obvious as hardly to be worth the verifying; and yet it is well deserving of as many tests as can be made. Each nuclear reaction which confirms it is a piece of evidence for the relativistic doctrine of the dependence of mass on speed and of the equivalence of mass and energy; and there can hardly be too many of these. Another point; in establishing this principle, it is taken for granted that the nuclei born out of transmutation are "normal" examples of their kind, *e.g.* that the nuclei designated by ${}_8O^{17}$ and ${}_{14}Si^{30}$ in (15.1) and (15.2) are identical with ordinary nuclei of oxygen 17 and silicon 30—identical not only in atomic number and in mass-number, but in exact mass. It turns out that this assumption is usually verified within the limits of experimental error; but conceivably it might never have been true. Transmutation might have led to newborn nuclei differing slightly from "normal" ones, and actually it sometimes does—an important fact, of which we could not be sure if we did not know so many cases in which the final nuclei are of the normal mass. Lastly, I mention that the principle is now regarded as so unshakably established, that the standard tables of nuclear masses are made up by drawing both upon the mass-spectrograph data and upon the equations of nuclear chemistry.

Now we see specific reasons for measuring as closely as possible the speeds or the kinetic energies of the nuclei which figure in a transmutation. The kinetic energies which are determinable are that of the bombarding or impinging nucleus, and that of at least one of the newborn kernels (the proton, in reactions 15.1 and 15.2). That of the other newborn kernel (the "residue") may be computed from these, the requirement of conservation of momentum being sufficient to fix it; it is not usually negligible! That of the remaining participant in the

reaction, the nucleus in the gaseous or solid target which received the impact, is, however, always negligible. These energies are essential for evaluating masses, and the task of ascertaining them is accordingly in the forefront of nuclear physics.

Equations like (15.1) or (15.2) may be considered complete as they stand, provided that the student will substitute for every such symbol as ${}_2He^4$ or ${}_1H^1$ the true mass of the flying particle in question with due regard to its speed. I hesitate though to urge this policy too strongly, for it is the custom to write such equations differently, with an explicit mention of the so-called "heat of reaction." Let Q_0 stand for the kinetic energy of the bombarding nucleus before the impact, Q_1 and Q_2 for the kinetic energies of the newborn nuclei at birth, and rewrite (15.1) thus:

$$ {}_2He^4 + Q_0 + {}_7N^{14} = {}_1H^1 + Q_1 + {}_8O^{17} + Q_2. \quad (15.3) $$

If we take such a symbol as ${}_2He^4$ as standing for the *rest*-mass in grammes of the particle in question and express Q_0 in ergs, then $({}_2He^4 + Q_0/c^2)$ is the *true* mass in grammes of the alpha-particle as it enters into the reaction; and, similarly, $({}_1H^1 + Q_1/c^2)$ and $({}_8O^{17} + Q_2/c^2)$ are the *true* masses in grammes of the departing proton and the residue as they separate from each other. The equation then affirms the conservation of mass.

Another way of describing the situation is to say that if the symbols ${}_2He^4$ and ${}_1H^1$ and ${}_8O^{17}$ represent rest-masses multiplied by c^2, then the equation as it stands affirms the conservation of energy in the reaction. However, it has become absurd to draw a sharp distinction between energy and mass. The most commendable procedure is to say that in (15.3) *all* the symbols are to stand indifferently for mass and energy, care being taken to express all of the quantities in the same units—in ergs or electron-volts, if it is energy which one wishes to stress; in terms of one-sixteenth of the mass of ${}_8O^{16}$ as unity, if it is mass which is to be emphasized. Such a symbol as ${}_2He^4$ now and henceforth stands for rest-mass or rest-energy, accord-

ing to the unit which one chooses to employ. The factor of transformation is such, that the measure of any one of the quantities in the equation is about a billion times as great in terms of electron-volts as in terms of the unit one-sixteenth-of-$_8O^{16}$ (the exact value is more nearly 931 Mev per mass-unit).

Let us write Q for $(Q_1 + Q_2 - Q_0)$, so that (15.3) becomes:

$$_2He^4 + {}_7N^{14} = {}_1H^1 + {}_8O^{17} + Q. \qquad (15.4)$$

By analogy with thermochemical equations, Q may be named the "heat of the reaction." It is the amount of energy transferred from the form of rest-energy to that of kinetic energy. According as it is positive or negative, the reaction is said to be "exothermic" or "endothermic." [1] Plenty of examples of both types are known, the absolute values of Q ranging as high as several millions of electron-volts. If the reaction is endothermic, it evidently cannot occur if the kinetic energy of the bombarding nucleus is less than $|Q|$. This is not, however, the same thing as saying that it must occur whenever Q_0 is greater than Q. The electrical repulsion between the striking and the struck nucleus, or some other inhibition not understood as yet, may prevent the reaction—or keep it from becoming frequent enough to be appreciable, which comes to the same thing—until Q_0 is considerably greater than $|Q|$. If the reaction is exothermic, there is nothing in the energy-relations to forbid it, however low the speed of the impinging nucleus may be; nevertheless it may not occur to any appreciable extent till Q_0 is raised to a high value. Thus a Q_0-value, which is a lower limit for the "onset of transmutation," is set by the Q-value, but the actual onset of a detectable effect is usually displaced to a much higher Q_0-value by other causes.

Notice how these equations imply that Q is independent of Q_0, so that if the kinetic energy of the impinging particle is

[1] Some people prefer to say "exoenergic" or "exoergic," and "endoenergic" or "endoergic" respectively. I avoid these terms for the reason that they do not suggest that it is *kinetic* energy which is augmented or diminished, whereas the termination "thermic" does give at least a faint suggestion of the circumstances owing to the relation between kinetic energy and heat.

augmented by a known amount an equal augmentation should appear in the sum of the kinetic energies of the newborn nuclei. Were this not so, the situation would be hard to interpret. It has, however, turned out to be so, in the rather few cases which have been rather carefully tested; though the custom now is to accept it without proof.

The (α, p) *reactions*

The heading of this section shows yet another notation: inside parentheses are placed, first a symbol for the impinging particle, then a comma, then a symbol for the simpler of the newborn nuclei. The usual symbols are: α for alpha-particle, p for proton, d for deuteron, n for neutron; to these are to be added ϕ or γ for photon, since corpuscles of light are often involved. When instead of a class of reactions a single member of the class is meant, the customary symbols for the other nuclei involved are placed before and after, thus: ${}_7N^{14}(\alpha, p){}_8O^{17}$ and ${}_{13}Al^{27}(\alpha, p){}_{14}Si^{30}$. Sometimes it is satisfactory to abridge these to $N(\alpha, p)O$ and $Al(\alpha, p)Si$, and even to leave the last of the symbols to the student's imagination.

Seventeen elements are known thus far to be subject to (α, p) reactions. They are B, N, F, Ne, Na, Mg, Al, Si, P, S, Cl, K, Ca, Sc; Ti, Cr, Ni. The first twelve are those which Rutherford transmuted, recognizing his success by observing the scintillations made by the emergent protons on fluorescent screens. All twelve of these transmutations have been verified by other physicists observing the protons by other schemes of detection. Only for one element (argon) is Rutherford's claim of success now held to be unfounded; and of the many elements with which he sought in vain for proton-scintillations, there is only one (calcium) for which anyone else [1] can boast of having detected emergent protons. This shows the method of the scintillations to be reliable when used by an expert—even though it is subjective and unregistrable, even though almost

[1] Pollard & Brasefield, **PR 51**, 8 (1937). Pollard discovered the reaction for scandium (**PR 53**, 679; 1938); I do not know whether Rutherford tested this element.

everybody now avoids it! As to the three elements listed after the semicolon, it is by a different way altogether that they have been proved transmutable. When the reaction $_{22}\mathrm{Ti}(\alpha, p)$ occurs the residual nucleus is radioactive, and by virtue of this property the experimenter can detect it and can prove that it is of the element $_{23}\mathrm{V}$; the proton, however, has not been detected. Corresponding statements can be made for the two cases of $_{24}\mathrm{Cr}$ and $_{28}\mathrm{Ni}$, the residual nuclei being provably of the elements $_{25}\mathrm{Mn}$ and $_{29}\mathrm{Cu}$ respectively. With calcium both the radioactive residue and the proton have been observed. Thus in every known (α, p) reaction, with the exception of that of calcium and one other,[1] one of the two newborn nuclei escapes observation: in twelve cases it is the residue, in three the proton.

Do not suppose that in listing seventeen elements subject to (α, p) transmutations, I have referred to only seventeen cases of this reaction! Some of these elements have two or three isotopes, of which it is almost or quite certain that more than one exhibits an (α, p) reaction; and each such case must be counted as distinct. Since the actual number is still obscure, it is not worth while here to give a tentative value. Do not suppose, either, that all of the elements not above listed are immune to transmutation initiated by alpha-particles! Many have already been transmuted in (α, n) reactions, a type to which a section will be devoted later. Success with most or all of the rest may be attained in due time, after alpha-particles of energy higher than any yet known shall have been made available.

This brings us to the topic of the yield or likelihood of transmutation.

Likelihood of transmutation

Like the study of ionization, that of transmutation has arrived at the stage in which the question "does it happen?"

[1] The already-mentioned reaction $\mathrm{N}(\alpha, p)\mathrm{O}$, for which the expansion-chamber photographs show the tracks of both proton and residue.

or "*when* does it happen?" for each specific process grows quickly into the question "*how much* does it happen?"

"Likelihood" and "cross-section" are defined for any specific reaction of transmutation just as they were defined on pp. 258–262 for any particular mode of ionization, with suitable changes of wording which the student can make for himself. Here as there, the ideal experiment consists in projecting a known number of bombarding particles of known and identical energy E against a slab of matter containing a known number of atoms per unit area, the slab being so thin that any particle which strikes an atom (here it is better to say: strikes a nucleus) in such a way as to produce the reaction in question strikes it with energy unabated. The newborn nuclei (usually the protons, when (α, p) reactions are being studied) are to be counted, and the resulting cross-section-value or σ-value is to be correlated with the E-value. Curves of σ plotted as function of E are known, by the way, as "yield-curves" or "transmutation-functions."

Cross-sections for transmutation turn out to be of the order of 10^{-25}cm^2 and downward, a great contrast with the values of the order 10^{-15}cm^2 which pertain to ionization! Owing to this contrast (which is perhaps the most vivid proof of the smallness of nuclei) the ideal experiment aforesaid is out of reach. To ensure that the number of transmutations shall be large enough to measure, one must use a slab so thick that among the very small proportion of alpha-particles which do make "lucky" hits, the great majority make these hits only after having ionized hundreds or thousands of atoms. Each ionization costs a part of the kinetic energy of the alpha-particle which performs it (some 33 electron-volts on the average, as we have seen in Chapter VIII). Thus if the particles impinge on the "near" side of the slab with uniform energy E, they arrive at the "far" side with lessened energy $(E - \Delta E)$, passing *en route* practically continuously through the intermediate energy-values; and ΔE amounts to many thousands of electron-volts. The measured σ-values thus pertain not to sharply-defined E-values, but to energy-intervals which seem extensive. Yet even these

intervals amount only to small percentages of the initial energies of the alpha-particles; and while this circumstance limits the hopes of detecting fine details of the yield-curves, experimenters on the whole are not yet so far advanced that it is already barring further progress.

Though I have been speaking above as if the kinetic energy of the impinging particle and that of a newborn nucleus were directly knowable, it is actually the *range* (in the sense of page 272) of each that is most often measured. Owing to this and other circumstances, the procedure of evaluating σ for processes of transmutation differs in many details from the corresponding one for ionization. When an (α, p) or an (α, n) reaction is being studied, the data for a yield-curve are usually obtained with a source providing alpha-particles of a uniform range R_{max} corresponding to some energy E_{max}; and the experimenter gets particles of ranges $R < R_{max}$ and hence of energy-values $E < E_{max}$ by interposing slabs or screens of various convenient materials between the source and the target. When a σ-value has been obtained for a given target, with a given source and a given interposed screen, the first step in reducing the datum is to translate the thickness of the screen into the equivalent thickness of air at standard temperature and pressure (page 272). Let x stand for this equivalent thickness (sometimes known by the ill-chosen name of "stopping-power"); then the range R of the particles impinging upon the near side of the target is given by $(R_{max} - x)$. Thus, one of the auxiliary problems of nuclear physics is to ascertain the stopping-powers or equivalent thicknesses of many substances in terms of air. To the extent to which this problem has been solved, all yield-curves may be plotted as curves of σ versus R, whatever the material used in the screens. Many curves to be found in the literature are of this sort, the further step of translating them into σ-vs-E curves not having been taken. For this step it is necessary to know the relation between E and R; and this circumstance accounts for the tremendous importance attached to the E-vs-R relation, and the great amount of experimental research which has been done upon it.

Many evaluations of σ have been made with "thick targets," meaning targets so thick that the alpha-particles "run themselves dead" and never emerge on the far side, and the energy-interval ΔE extends from E down to zero. With such an arrangement it seems at first glance that the yield-curve must be impossible of attainment, and the reported σ-values of very minor interest; but such need not be the case. Let the experiment be made once with impinging particles of range R, once with the range altered to $R + dR$. If the stopping-power of the target material is known, then the experimenter can compute the thickness dT thereof which corresponds to dR; and then, forming the difference dn between the numbers of transmutations counted in the two experiments, he may say that the excess of one over the other is the number of transmutations performed by particles of ranges between R and $R + dR$ in traversing a slab of thickness dT.

If dn and dR could really be made infinitesimal, then the datum would give a σ-value corresponding to R, by the relation:

$$dn = \sigma N dT = \sigma N(dT/dR)dR \qquad (15.5)$$

N standing for the number of atoms per-unit-volume of the target. Actually dR and dn cannot be made indefinitely small, and the use of the differential notation implies a precision greater than is attainable. Keeping this in mind, we may, however, keep the differential notation, in order to have a convenient way of describing two types of curve resulting from the data.

The first kind of curve is the directly-obtained curve of the number n of emergent protons or otherwise-observed transmutations, plotted against the range of the particles incident on the thick target: the n-vs-R relation, called the "integral curve" or "thick-target curve." The second is derived from the first by differentiating with respect to R; it is the (dn/dR)-vs-R relation, called the "differential curve" or sometimes the "thin-target curve." The yield-curve σ-vs-R is derivable from this last, if (dR/dT) is known. (This affords an incentive for determining the stopping-powers of all substances of which

transmutation may be observed, and not merely those of the substances which are convenient to use as screens.) Most of the so-called "thin-target" curves in the literature are differential curves obtained in this fashion, but a few have been made with targets in the form of films which are literally thin in the sense indicated at the beginning of this section.

It seems best to finish here with this topic of the determining of yield-curves, even at the cost of anticipating the historical order. When the bombarding particles are protons, deuterons or alpha-particles which have been accelerated in cyclotrons or high-voltage machines, their kinetic energy is usually measured in electron-volts and the measurement of range is circumvented.[1] Yield-curves then are plotted with energy in MEV as the abscissa; yet if they are to be compared with others obtained with natural alpha-particles, the E-vs-R relation must be known; or if the curve is to be obtained with a target thick in the foregoing sense, then it is necessary to know the E-vs-T relation, E now standing for the kinetic energy which the particles possess when they have penetrated through a thickness T measured inward from the surface of the target. Let R stand for the range corresponding to E; then the last requirement is satisfied if one knows both the E-vs-R and the R-vs-T relationships, both dE/dR and dR/dT. Now it becomes evident that *the* great auxiliary problem of nuclear physics is to know both of these relationships for as many substances as possible, with respect to the proton and deuteron as well as the alpha-particle; and indeed with respect to the other nuclei as well, for these may appear as newborn nuclei or be used in the future as bombarding particles.

When it is the radioactivity of the residual nucleus rather than the leaping-forth of the emergent particle which is detected, there is a neatly simple method of determining yield-curves with range for the abscissa. The target of the substance to be investigated is split into thin films, through which the

[1] Not invariably, since some experimenters using these devices prefer to measure the range of the particles which they produce rather than to estimate the energies which the apparatus is conferring upon these.

beam of the bombarding particles goes consecutively, arriving at each with a distinctive range which is calculable if the thicknesses of the preceding films and the R-vs-T relationship are known. After the exposure the films are removed to an electroscope, and the radioactivity of each is measured. All of them having been exposed to the identical number of bombarding particles, the relative radioactivities are sufficient for giving the relative ordinates of the yield-curve, even though (for lack of information of the total number of bombarding particles) the absolute values of these ordinates be undetermined.

General character of the yield-curves

It has already been implied—by the emphasis placed upon the high energies of alpha-particles, and by the reference to the electrostatic repulsion between nuclei—that transmutation is produced only when the energy of the bombarding particle, E, is great. One might expect a sudden onset at an E-value already high, followed by a steadily-rising σ-vs-E curve. It should be said at once that there is nothing remotely resembling a sudden onset, unless in the cases of endothermic reactions where owing to the energy-relations the process cannot occur until E is greater than $|Q|$ (page 483). For exothermic reactions the rise of the curve from the axis of abscissae seems to be smoothly gradual, and the least value of E at which transmutation is detected depends very much upon the efficacy of the scheme of detection. We shall later see that the gradual rise is taken as evidence for a quantal theory of nuclear phenomena. However, the yield-curve does not necessarily rise steadily with increasing E, as was supposed at first. It may have maxima and minima, suggesting what in other fields of physics is known as "resonance," a term which has come to be applied in this field also. The experiments (of Pose of Halle in 1930) which first revealed this feature were also the first to reveal a rather complicated situation, which on further analysis has shown that there may be several Q-values associated with a single reaction instead of only one. Since the resonance and the Q-values alike are

inferred from the distribution-in-range curves of emergent protons, we must now turn to the study of such curves.

Distribution-in-range and groups of emergent protons

It has already been stated that for knowledge of the energies of his bombarding particles, the experimenter must usually rely on measurements of their ranges. So also must he—and not just usually, but *always* [1]—measure the ranges of the emergent protons for knowledge of their energy-values. This is done by interposing, in the path of the protons between source and detector, strata of air of adjustable thickness or density, or strata of other substances (mica and aluminium, most often) for which the air-equivalents are known. The number of particles getting through the strata is plotted against the equivalent thickness thereof, and the result is what is called an "integral distribution-in-range curve."

Before viewing such curves the student should realize that even when there is only a single Q-value for a reaction, he must not expect to find protons of only a single energy and hence of only a single range. Such an ideal situation might indeed be approached, if available alpha-particle sources were much more generous than they are (and this is one of the reasons for developing strong artificial sources). Let θ stand for the angle between the directions of motion of the impinging alpha-particle and the emergent proton, in any single process of transmutation. In each experiment, θ should have very nearly the same value for every proton counted; therefore the experimenter should be able (*a*) to delimit his bombarding beam so that all of the bombarding alpha-particles would travel to the target along directions lying within a very small solid angle, and (*b*) to confine his observations to protons ejected in directions lying within another very small solid angle. The solid angles subtended by the target as seen from the source and by the

[1] This statement ceased to be valid in 1938, when S. K. Allison applied deflecting electric fields to the emergent particles of a (p, α) reaction (Allison & N. P. Smith, PR 53, 946). This new technique is of particular value for alpha-particles of energy-values inferior to some 3 Mev, for which the E-vs-R relation is not well known.

detector as seen from the target should both be very small. But this is impracticable for the present, because of the smallness of the number of transmutations producible per second with available sources of alpha-rays. The target usually subtends a large solid angle as seen from the source, and the detector or collector a large solid angle as seen from the target; and in many a distribution-in-range curve, each point by itself is obtained by counting protons of which the θ-values have a spread of fifty to sixty degrees.

What makes this spread of θ-values detrimental is the fact that the proton-energy depends upon θ, even when all the alpha-particles have the same E-value and the reaction exhibits only a single Q-value. It is only the *sum* of the kinetic energies E_p of the proton and E_r of the residue that under these conditions must be constant. On setting up the problem according to the rules of ordinary non-relativistic mechanics—assuming that momentum is conserved and that kinetic energy is augmented by Q when the initial pair of nuclei suddenly becomes the final pair—it is found that E_p depends upon θ as well as on E, Q, and the masses of alpha-particle and residue and proton. The details of the treatment are so like those for the case of elastic impact between nuclei which retain their identity (page 71) that they may safely be left to the student.[1]

We may now turn to some of the integral distribution-in-range curves exhibited in this book: for instance those of Pose's famous early work on the reaction $\mathrm{Al}(\alpha, p)$ shown in Figure 15.4, or those of Figures 16.9 and 16.11 [which refer to a (p, α) and a (d, α) reaction respectively so that the emergent newborn nuclei are alpha-particles]. They consist of horizontal lines known as "paliers" or "plateaus," connected by curving arcs. If these curving arcs were vertical straight lines, the curve would tell us that the emergent nuclei all have one or another of several discrete range-values, *viz.* the abscissae of the vertical lines. In most of the cases it is believed that these arcs would become vertical lines if only the spread of θ-values aforesaid

[1] The resulting formula is given by Livingston and Bethe, p. 277. They show that nothing worth while is gained by using relativistic mechanics in this problem.

could be reduced to zero. When this is true *the emergent nuclei are said to be distributed in "groups."* If there is only one Q-value for the reaction, there is only one group, and it is manifested by an integral distribution-in-range curve consisting of a single palier and a curving arc connecting this with the axis of abscissae. If there is more than one Q-value there are additional groups made manifest by additional curving arcs leading to additional plateaus.[1] These groups are a general attribute of (α, p) reactions.

After the range of a group has been ascertained from the curve and is translated into an energy-value E_p, the Q-value may be calculated from E_p by the formula just mentioned in a footnote. The accuracy of the estimated values is of course affected by the θ-spread which produces the curvature of the arcs. (It is usually the abscissa of the bottom of the curving arc which is quoted as "the range of the group," but this is not necessarily the wisest policy). As many as four or five Q-values have been ascertained in this manner for a single (α, p) reaction, *e.g.* the reactions $\mathrm{Al}(\alpha, p)$ and $\mathrm{F}(\alpha, p)$.

But how are we to reconcile this multiplicity of Q-values with the doctrine expounded on page 481, that the residue of a transmutation is a normal example of the nucleus-type having the requisite atomic number and mass-number? If any one of the Q-values is compatible with this conclusion, the others cannot be!

Let Q_0 stand for the Q-value calculated from such an equation as (14.4), by substitution into that equation of the rest-masses determined by mass-spectrograph for the normal nuclei of the proper mass-numbers and atomic numbers. If more than one group occurs in the reaction, it may be that the greatest of the Q-values is equal to Q_0. If so, it is possible to say that whenever a proton is ejected with one of the lesser Q-values, the residue is left in an "excited state." This name is

Groups may be shown more obviously by graphical differentiation of the integral curve, for the derivative curve ("differential distribution-in-range curve") then exhibits a peak for each; but this is usually held to be not worth the trouble of doing, and the student must accustom himself to the aspect of integral curves.

really significant only if it can be shown that the residue converts itself later into a normal nucleus, *e.g.* by emitting a photon of energy $(Q_0 - Q)$. Evidence for this conversion exists in some cases, and will be mentioned in due course. If all of the groups show Q-values smaller than Q_0, it is possible to say that the reaction always terminates with the residue in one or another excited state. This case is found in several reactions, too often indeed for comfort; with some of them physicists have deliberately sought for a previously-undetected proton-group with $Q = Q_0$, and have not found such a one. If one or more of the Q-values should be definitely greater than Q_0, it would be taken as a strong indication that one at least of the mass-values taken from mass-spectrograph data is wrong; failing this possibility of escape, one would probably have to admit two types of stable nucleus identical in mass-number and atomic number. There is as yet no indubitable instance of such a case. On the whole the situation in this field of research is still unsatisfactory, the experiments being so difficult and tedious that few people undertake them; and bad discrepancies still exist among the published Q-values. I will mention a few of the data for illustration, but only after taking up the so-closely-interwoven phenomenon of resonance.

Resonance

The important phenomenon of resonance, already mentioned, comes clearly into view only when one studies the behavior of individual proton-groups as E is varied. By inspection of the flock of curves in Figure 15.4 it will be seen that certain curving arcs and connecting paliers are present in some of the curves, absent in others. This is the observation which led to the discovery. Instead of following further the historical development, I will reverse it by postulating and stating the conclusion, and thence reason backward to the ways in which it has been tested.

We take it, then, as a fact that in some reactions a proton-group may be elicited by only such alpha-particles as have

energies between two values E_a and E_b (choosing the notation so that $E_a > E_b$). When this occurs it may be detected in the following ways:

(*a*) If the target is thick one may vary the "initial" energy E_0 with which the alpha-particles arrive at its face (this being done usually by varying the density of air in a narrow space

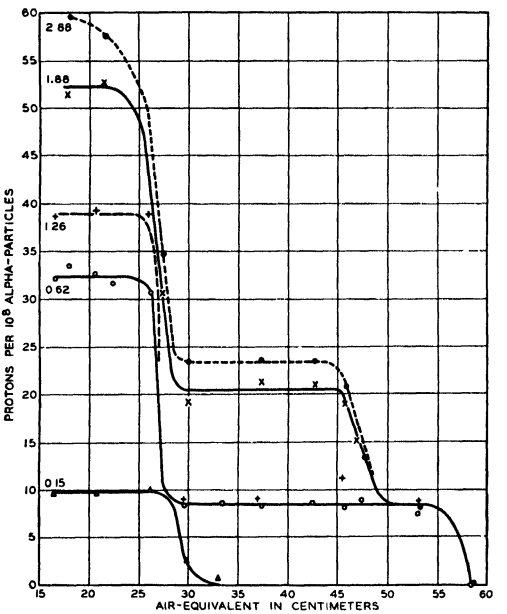


FIGURE 15.4—INTEGRAL DISTRIBUTION-IN-RANGE CURVES OF THE PROTONS FROM THE REACTION Al^{27} (α, p), TAKEN WITH VARIOUS VALUES OF KINETIC ENERGY OF IMPINGING PARTICLES AND SHOWING EVIDENCE OF RESONANCE. (Pose; *Zeitschrift für Physik.*)

between the target-face and the source) and plot the integral distribution-in-range curve for many different values of E_0. Suppose E_0 greater than E_a at the start. It will be found that as E_0 is lowered, the curving arc and the plateau to the left thereof which reveal the group remain almost unaltered until E_0 drops below a certain critical value, whereupon they vanish.

Some of Pose's curves are of this type; but those here reproduced as Figure 15.4 refer to the type which comes next.

(*b*) If targets thin enough to be traversed by alpha-particles are available, one may leave E_0 unchanged and vary the target-thickness t. The energies of the alpha-particles impinging on the nuclei then are comprised between E_0 and a least value E_1 which depends upon t. If integral curves are plotted for various thicknesses, it will be found that as t increases and E_1 decreases a curving arc and a plateau make their appearance at a critical value of E_1.

Figure 15.4 exhibits a family of curves obtained with five foils of aluminium exposed to alpha-rays of polonium with their full energy of 5.25 MEV and undiminished range of 3.72 cm (in air at 0° C. and 760 mm Hg). The lowest curve relates to the thinnest foil: its thickness corresponded to one mm of air, but for the reasons expounded above Pose had to set it so close to the source that it was traversed not only by alpha-particles moving nearly normally to its faces, but also by others moving so obliquely that their range was reduced by 1.5 mm. The nuclei in this foil were therefore bombarded by particles of all ranges between 3.72 and 3.57 cm, and this is what is meant by the number 0.15 marking the curve. The next-to-lowest curve relates to a foil of air-equivalent 0.62 cm for the most obliquely-moving of the alphas, and therefore to nuclei bombarded by particles of all ranges between 3.72 and 3.10 cm; and correspondingly for the rest.

The lowest curve informs us that so long as the alphas are moving so fast that they still have 3.57 cm of their range ahead of them, the protons which they eject are relatively slow (maximum range of 30 cm or thereabouts); but comparison with the next-to-lowest curve brings out the sensational fact that *when the alpha-particles are slowed down* to some critical speed corresponding to some range between 3.57 and 3.10, *they acquire the power of dislodging much faster protons*. The long palier and the curving arc coming down to the axis at 57.5 cm are the manifestation of a group excited only when E is less than an upper critical value E_a. Moreover, the fact that the curving

arc is duplicated in the upper curves is proof that the group is excited only when E is greater than a lower critical value E_b, for otherwise the arc would have been lengthened upward as the target was made thicker. The next-to-lowest and the middle curve are indistinguishable over the right-hand part of the graph, but in going onward to the second-highest curve one sees the vivid sign of a second group with a lower E_a-value than the first. Notice lastly how the most leftward of the curving arcs is present in all the curves, growing steadily in length as t is

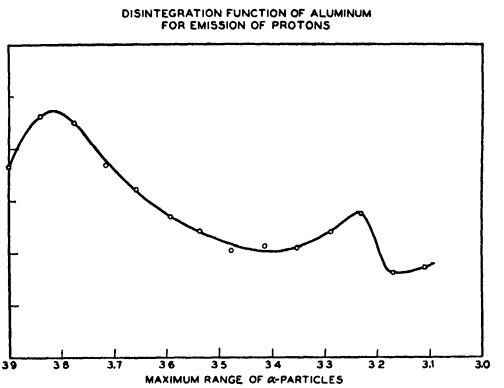


FIGURE 15.5—YIELD-CURVE OF THE REACTION Al^{27} (α, p) OBTAINED BY METHOD DESIGNATED BY (*c*) IN TEXT. (Chadwick and Constable; *Proceedings of the Royal Society.*)

increased and the alpha-particles are enabled to make more impacts in the foil. Here is the phenomenon of a group which particles of all energies within a wide interval are able to elicit—the opposite of the phenomenon of resonance.

(*c*) If the target is very thin the yield-curve for the total output of protons having ranges longer than some minimum may exhibit peaks. (The qualifying words "having ranges longer than some minimum" are necessary, for even if no deliberate attempt is made to exclude short-range protons, these will be undetectable if their ranges fall below a critical value dependent on the apparatus.) In Figure 15.5 appears such a

curve obtained (by Chadwick and Constable) with a film of aluminium of air-equivalent 0.8 mm only. Experiments of this type indicate resonance (when present) without disclosing Q-values. If the target is thick, the thin-target curve may "in principle" be deduced from it as indicated on page 488; experience has shown this a poor way of locating resonance.

(*d*) If the target is thick, the mere presence of distinct or even of somewhat distorted plateaus in the integral distribution-in-range curve of protons may be taken as a strong intimation of resonance: for if any group were made up of protons evoked by alpha-particles of so great a variety of speeds as exists in the thick target, then in spite of the fixed Q-value the distribution-in-range of the protons would probably be so broad as to spread out the curving arc to the extent of effacing the adjacent plateaus. Thus the topmost curve of Figure 15.4 practically proves resonance by itself; yet it does not permit the Q-values to be calculated, for one does not know from that curve by itself which energy-value of the alpha-particles is to be correlated with each group.

(*e*) If the target is very thin and the residue is radioactive the yield curve may be obtained in the way described on page 489; or else a single film may be exposed in successive experiments to alpha-particles which have passed through various thicknesses of matter on their way from the source, its radioactivity being measured after each exposure. With the latter scheme the target may be a mere coating laid down on an inert base. This method has given some of the best evidence of all; but obviously it does not indicate the Q-values, nor can it even prove the reaction in question to be an (α, p) reaction; this last must be proved in other ways.

Doubt of the phenomenon of resonance, which lingered a year or more after the first announcements, is now dispelled; but in respect of accuracy, concordance and completeness the data still leave much to be desired. Bad discrepancies still exist between the reports from different institutes, and there is no one experimenter so pre-eminent that his judgment is taken as law.

The group or groups for which σ rises steadily with increasing E (instead of passing through maxima and minima) ought not to be forgotten; but they have been rather neglected because of the interest taken in locating resonance-levels. The theory which we shall soon consider leads to the expectation that for any group taken by itself, the σ-vs-E curve consists of humps at low energies and of a steadily-rising-arc at higher energies. This assertion is sometimes supported by curves; it is also made in cases as to which one cannot make out whether it is derived from the theory only or from experiment as well. As for the Q-value for any such group, it can be ascertained only with difficulty if the experimenter is using a thick target; for then, the protons belonging to such a group will be elicited by alpha-particles of a wide interval of energies. The proper procedure in such a case is to correlate the fastest of the protons emerging in any chosen direction with the energy of alpha-particles just entering the target.

First notions of nuclear theory

It is probably high time for the student to be venturing some steps into the quagmire of current nuclear theory.

As a matter of fact, we have already ventured the first and safest of these steps; for the inverse-square electrostatic field verified by the scattering-experiments described in Chapter V is itself one of the attributes of the nucleus, though the most external and remote thereof. So also are the deviations from inverse-square character which, still according to the testimony of scattered alpha-particles and protons, the nuclear field exhibits at smaller values of r the distance from its center. We can indeed foresee that there should be deviations. If the nucleus contains separate positive particles, the fact that it does not instantly explode is to be interpreted by assuming that there are short-range attractive forces between these particles which, when these are sufficiently close together, are stronger than the electrostatic repulsions. This would be a pure and useless speculation, if it did not follow that an inward-flying

positive particle (such as a bombarding alpha-particle) should itself fall under the empire of such forces when it comes very close to the center ($r = 0$). If the notion of a force derivable from a potential is still acceptable at such small values of r, then the idea is expressed by a potential-curve such as that of Figure 15.6.

This curve constitutes by itself a large part of nuclear theory. No portion of it is to be regarded as known in advance, excepting the remoter part of the descending arc which sweeps off to the right. There, the ordinate is assumed to be equal to $2Ze^2/r$, in accordance with the information which scattered

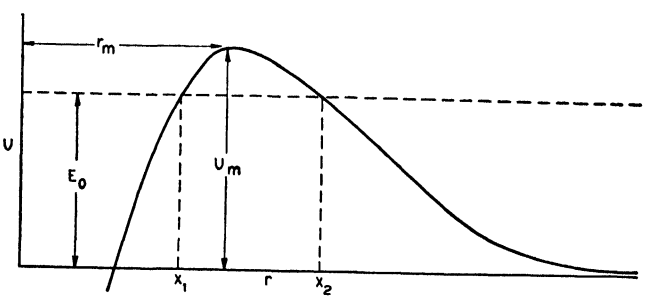


FIGURE 15.6—HYPOTHESIS AS TO THE TREND OF THE CIRCUMNUCLEAR FIELD.

alpha-particles give. Only there may the force upon a charged particle be taken as purely electrostatic; elsewhere it may *not*, so that care must be taken to distinguish the general sense of the word "potential" (here to be employed) from the narrower sense which is prevalent in electrical literature. The nature of the charged particle is important: when for instance we begin to consider oncoming protons instead of oncoming alpha-particles it may be seriously wrong to cut the ordinates of the potential-curve everywhere by half, as we could if the force were everywhere electric. The hump in the curve is called "the potential-barrier"; the ordinate and the abscissa of its top, U_m and r_m, are called "the height" and "the radius" of the barrier, respectively; the part to the left of the barrier is called

"the potential-well." Other names for the quantity r_m are "half-breadth of the potential-well" and "nuclear radius"—the latter being a usage which seems improper until one realizes that without a theory there is no way of defining any concept which may reasonably bear that name.

When the curve is used to interpret radioactivity and transmutation, the data apparently fix the magnitudes of the height and the radius of the barrier. There is, however, a good deal of latitude in the fixations, due partly to the fact that the values of U_m and r_m required by a given set of data depend on the details of the shape postulated for the hump. A shape which is often postulated involves a sharp corner to the hump, the curve being assumed to rise as r^{-1} as it comes in from the right, until suddenly it drops like a vertical wall. The simplest way to finish the curve is to terminate the vertical wall at some point below the axis of abscissae, and draw a horizontal line from that point to the axis of ordinates. The nucleus, or rather the nucleus-model, is then said to exhibit a "rectangular potential-well"—a comical name indeed, considering that it refers to a spherical region of space in which the potential is uniform. The distance downward to the horizontal line is called "the depth of the well." Its magnitude is immaterial for the present, only a lower limit thereto being fixed by the data of radioactivity and transmutation.

Imagine a particle coming straight toward the center of the nucleus. Let E_0 represent the value of its kinetic energy while it is infinitely far away, and also the value of its total energy. (The last clause is tantamount to saying that the zero from which we are to reckon energy corresponds to the state in which nucleus and particle are infinitely far apart; note the analogy with the theory of spectra!) Imagine now a horizontal line drawn in Figure 15.6 at the height E_0 above the axis. In a sense this line may be said to exhibit the history of the particle; for the length of the line drawn *downward* from it to the curve at any abscissa is equal to the kinetic energy which the particle has when at the distance from the nucleus represented by that abscissa.

The student will at once detect a marked contrast, according as E_0 is greater or less than U_m and the horizontal line passes above or below the crest of the hump. In the former case, the particle can pass into the potential-well; in the latter case it comes to a stop when it reaches the place corresponding to x_2. (If the particle is originally within the well it departs or remains according as E_0 is greater or less than U_m.) Using the word "probability" in a sense which is correct but may as yet seem rather foolish, I say that the probability of entry of the particle into the potential-well rises suddenly from 0 to 1 as E_0 passes upward through the critical value U_m.

The foregoing statement is rather too strong even by classical mechanics, for thus far I have disregarded particles which, as they start toward the nucleus from infinity are moving along lines directed toward the potential-well but not toward the exact center thereof. Taking these into account one sees that the probability-of-entry will not undergo a sudden change, but will instead rise gradually from 0 towards 1 as the value of E_0 rises from U_m. When quantum mechanics is invoked instead of classical, the gradualness becomes even more marked; for according to quantal theory, a particle may enter the potential-well even though E_0 be slightly smaller than U_m! This striking consequence follows from the fact that in quantal theory the particles are subordinated to the guiding waves. On solving the wave-equation for a wave-train coming from infinity and traversing such a field as that depicted in Figure 15.6, it is found that the wave-train continues into the potential-well whether or not E_0 is greater than U_m. True, it does not wholly pass: there is a beam reflected from the hump as well as a beam transmitted through the hump, and the ratio of reflected to incident amplitude is very much greater when E_0 is less than U_m than in the converse case, leaving much less for the transmitted amplitude. Nevertheless, the transmitted amplitude does not instantly vanish as E_0 descends below U_m; and therefore we must infer that particles for which E_0 is less than U_m have a probability of entry into the well which may indeed be extremely small but is not zero. (The status of the particles

while passing "through the hump" is one of the things which defy imagination).

We seem to have lost sight of transmutation altogether; but the connection resides in the fact that for most of the few years of the existence of nuclear theory, *it has been the custom to assume that transmutation occurs always when and only when a particle enters a potential-well.* Thus the foregoing paragraphs embody a theory of the dependence of likelihood-of-transmutation on energy-of-impinging-particle; a theory which can be made as explicit as one likes, by postulating a definite shape and scale for the potential-curve. The gradual rise of the yield-curves can thus be explained, and the theory even exhibits a feature to which one might be able to trace the phenomenon of resonance; for the ratio of transmitted amplitude to incident amplitude of waves of which I was just speaking, and accordingly the probability of entry of the particles, passes (as Gurney found) through minima and maxima as E_0 falls further and further below U_m. Yet the theory is certainly not adequate, nor can it be expected to be; for to assume that entry into the well is both a necessary and a sufficient condition for a transmutation of a given type is palpably too simple. A fact which shows this clearly is the coexistence of some of the (α, p) reactions which we have just been studying with (α, n) reactions which we shall study in the next chapter but one. There are nuclei from which either a proton or a neutron may spring when an alpha-particle enters; the yield-curves for both have been plotted, and the resonances do not coincide. Thus the probability of at least one of these modes of transmutation is assuredly not proportional to the probability of entry $P_e(E)$, and should be written rather as $f(E)P_e(E)$. Having been forced to admit this in one case, physicists will be more inclined to admit it in all, and to doubt whether yield-curves are exact or even close copies of probability-of-entry curves. Some are even prepared to discard the interpretation of resonance-peaks by Gurney's idea, preferring to ascribe the maxima in the yield-curves altogether to maxima in the function $f(E)$; this seems to me precipitate.

These last paragraphs may have impaired so much the student's esteem of this type of nuclear theory, that he wonders why I have given so much space to it. I must therefore say that the strongest evidence lies in the field of radioactivity. The model depicted by Figure 15.6 was in fact invented and the quantal theory applied to it—by Gurney and Condon, and independently by Gamow—in the hope of explaining the Geiger-Nuttall law of radioactivity (Chapter XVIII); and the hope was justified. Therefore in invoking for the study of transmutation this model and this theory, the physicist is employing ideas which have already shown themselves valuable. Moreover, these ideas have the historic glory of having led to a discovery. The notion of the penetrability of the potential-barrier to particles not having energy enough to "pass over the top" caused Gamow to encourage Cockcroft to try for transmutation by bombarding light elements with protons having relatively so little energy, that otherwise he might not have deemed the attempt worth while. The attempt was successful, and the second stage of the art of transmutation commenced in 1932.

CHAPTER XVI

Second Period of Transmutation

In all of the experiments described in the chapter preceding, the impinging particles provoking the transmutations were alpha-particles, and they had been provided by Nature with their kinetic energy, having just emerged from radioactive nuclei. In the experiments next to be considered, the impinging particles were deuterons or protons, and they had been provided with their kinetic energy by the artifice of man. These are two of the main distinctions between the first period of the art of transmutation, extending from Rutherford's pioneering achievement of 1919 until the Spring of 1932, and the second period which still continues. As nowadays the devices used for imparting energy to protons and to deuterons are being applied to helium nuclei also, the former distinction is fading. A third is also fading, that of the energy-magnitudes; for whereas the kinetic energies of the bombarding particles were of the order of several millions of electron-volts in the first period and only of the order of a few hundreds of thousands at the beginning of the second, the accelerating-engines of the laboratories are now beginning to impart kinetic energy equal and even superior to that of the alpha-particles from radioactive bodies. It seems likely that all of the data obtained with these so-called "natural" alpha-particles will soon be replaced by experiments with the intenser streams available from accelerating devices, whereupon the first period will be remembered only as an outstanding episode in the history of science. Yet a fourth distinction is important; it is to be made between the early years of the first period on the one hand, and all subsequent time on the other; the scintillation-method of the first days has long been superseded

by a diversity of others more delicate, more exact and more convenient by far. The science of transmutation will soon be based upon phenomena produced by the aid of machines and detected by the aid of devices, most of which were still uninvented and the rest only slightly developed when Rutherford began it.

Though in this book as a whole very little is said of apparatus, two of the accelerating-engines invented for nuclear physics are so novel and so remarkable that I will pause to describe them. It would however be unjust to single them out for description, without first emphasing that high-voltage devices of more traditional types have been and are very useful in nuclear research. Indeed when Cockcroft of Cambridge opened the second period of the art of transmutation by bombarding lithium with protons, he used a scheme in which two condensers (call them A_1 and B) were connected in parallel across a source of moderate voltage, and the charge imparted to B was then allowed to distribute itself over other condensers A_2 and A_3 connected in series with A_1; then by repeating the process several times he could obtain a potential difference (between the positive plate of A_1 and the negative plate of A_3) nearly three times as great as the voltage impresses upon the system, and he had reached 710,000 volts when he began on transmutation. Transformer-rectifier sets also have been improved to the point of producing one and even two million volts, and are used in several great laboratories (I mention Lauritsen's institute of Pasadena as an example). Mere attainment of high potential-differences is of course not enough; the apparatus must be able to maintain the high voltage during the flow of current constituted by bombarding nuclei as they fly from their source (generally a gas excited by a discharge) to the target, receiving their kinetic energy from the voltage on their way. Since worthwhile results have been obtained with currents of less than one millimicroampere (one millimicroampere, written as one mμA, is 10^{-9} ampere) this is not so stringent a condition as might be feared; but as time goes on, larger and larger currents are desired.

The two outstandingly novel devices invented for the acceleration of nuclei are the "cyclotron" of E. O. Lawrence and the "electrostatic generator" of R. J. Van de Graaff.

The cyclotron [1]

The cyclotron depends upon the principle that a charged particle, projected with kinetic energy K in a direction perpendicular to that of a uniform magnetic field H, describes in the plane at right angles to H a circular arc with an angular velocity independent of K. (The speed must be so low that the relativistic variation of mass with speed is negligibly small, else the principle is perceptibly false.) If in travelling along this arc the particle arrives at a point where the electrostatic potential abruptly changes by an amount ΔV in the proper sense for speeding it up, its kinetic energy increases by ΔV electron-volts, and it proceeds along an arc which is a prolongation of the first one but has a greater radius of curvature; *the radius of curvature goes up in the same ratio as the linear speed*, whence the foregoing statement about the angular velocity. Now imagine that the magnetic field is in the z-direction; that the charged particle—I will say "nucleus" henceforth—is moving in the xy-plane; and that the electrostatic potential has a constant value over each of the two halves of the xy-plane which are separated by the y-axis, but these two constant values differ by $\pm\Delta V$. The nucleus will cross the y-axis any number of times at equal intervals, and if the potential-difference between the two halves of the xy-plane remained the same in sign as well as magnitude, it would merely alternate between a higher and a lower value of K; but if the sign of the potential-difference were to be reversed between each two consecutive crossings, it would add $|\Delta V|$ electron-volts to its kinetic energy whenever it crossed, and K would ascend higher and ever higher. When one considers how swiftly the consecutive crossings must follow upon each other, it seems at first absurd to dream of reversing the

[1] Lawrence and Cooksey, **PR 50**, 1131 (1936); Livingston, **RSI 7**, 55 (1936).

potential-difference in the time-interval between each and the next; but the constancy of these time-intervals is the circumstance which saves the situation; for owing to that constancy, the periodically reversed potential-difference can be applied by so common and manageable a device as the vacuum-tube oscillator. The nucleus moves in the xy-plane, under the combined influence of the magnetic field normal to the plane

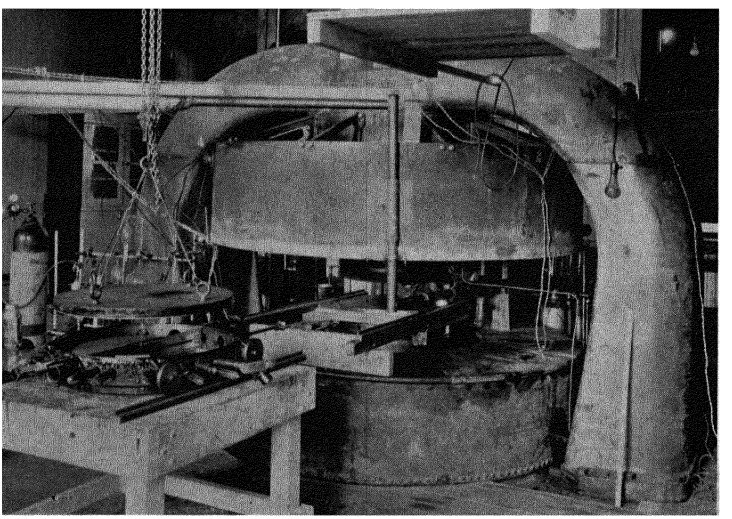


FIGURE 16.1—THE FIRST CYCLOTRON (BUILT BY LAWRENCE AT BERKELEY).

and an *alternating* potential-difference between the two halves of the plane; its path is composed of semicircles each broader than the one preceding, and is an approximation to a spiral continuously described with continuously rising speed.

All of the theory of the "ideal" cyclotron is comprised in equation (2.7) of page 39, which may be rewritten in the two forms:

$$\omega = He/mc, \quad K = \tfrac{1}{2}m(HeR/mc)^2, \qquad (16.1)$$

ω standing for the angular velocity of the nuclei which is 2π times the frequency of the oscillator, and K for the kinetic

energy which they possess while describing the semicircle of radius R.

The actual apparatus quite evidently cannot conform exactly to the idealized situation of the foregoing paragraphs! The nuclei revolve in a chamber which may be visualized as a drum which has been cut into two halves by a plane containing its axis of symmetry; the halves are slightly separated, the alternating voltage is applied between them. These drum-halves are called the "dees," after the letter of the alphabet of which the upper-case form recalls their half-moon shape. The

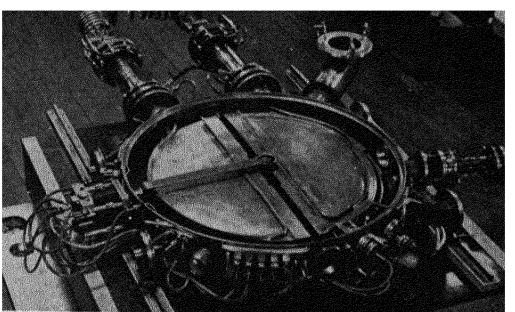


FIGURE 16.2—THE ACCELERATION-CHAMBER OF THE CYCLOTRON, OPENED TO SHOW THE DEES.

dees are enclosed in an air-tight drum-shaped case (shown coverless in Figure 16.2) so that the gas-density within them may be controlled. One would desire a perfect vacuum, but the to-be-accelerated nuclei must be liberated within the chamber itself; which accordingly is filled with hydrogen or deuterium or helium at the lowest density compatible with a discharge which will strip the orbital electrons from an adequate number of atoms. Usually the discharge is sustained by thermionic electrons from a filament set near the center of the chamber, but lately successful trials have been made with a high-frequency discharge pervading all the gas.

The magnetic field must be of considerable extent, since the radii of the semicircles grow rapidly with the energy of the

nuclei. This is definitely not one of the situations in which a high magnetic fieldstrength in a very small volume will do; the fieldstrengths occurring in cyclotrons are high, indeed, but not extraordinarily so, while the breadth of the field *is* extraordinary. The breadth of the magnet pole-pieces must, therefore, be great; and the first cyclotron (after a miniature made by Lawrence and Livingston for the initial trials of the idea) owed its existence to the availability of a gigantic magnet (Figure 16.1) which had been built for the generation of wireless waves by a method since superseded—a magnet with pole-faces twenty-seven inches and a half in diameter! Most but not all other existing cyclotrons have pole-faces as large or larger, and indeed the full breadth of the Berkeley cyclotron has not at this writing been utilized, the largest dees employed with it being twenty-three inches in width. Inhomogeneity of field in the interspace between the poles is usual, and is corrected or at any rate modified by pieces of magnetic metal called "shims," the proper placing of which within the interspace used to consume a good part of the labor involved in getting a cyclotron to going.

The behavior of the cyclotron is affected by the facts that the nuclei set free (from their orbital electrons) in the discharge begin their careers not all at one point, but anywhere in a broad region about the center of the dees; that they start off in all directions; and that they start at instants corresponding to all the phases of the alternating voltage, which, of course, varies continuously between its extreme values $+\Delta V$ and $-\Delta V$, instead of jumping cleanly from one to the other. Reflecting on these and other handicaps, one might well doubt whether the cyclotron could ever operate efficiently enough to be useful, and indeed it has often been characterized as an instrument which has worked far better than could reasonably have been expected—a pleasant surprise, in a world so full of devices which work less well than could reasonably be expected. These circumstances and others do lead to a broadish distribution-in-energy among the particles which have completed their spirals, and other instruments are being

preferred to the cyclotron for experiments in which a homogeneous beam of bombarding nuclei is of the first importance. It appears that not before last year (1937) was a serious attempt undertaken to work out the mathematical theory of what actually happens in the dees.[1]

Near to the cylindrical wall of one of the dees a supplementary field (electrostatic) seizes upon the nuclei which are

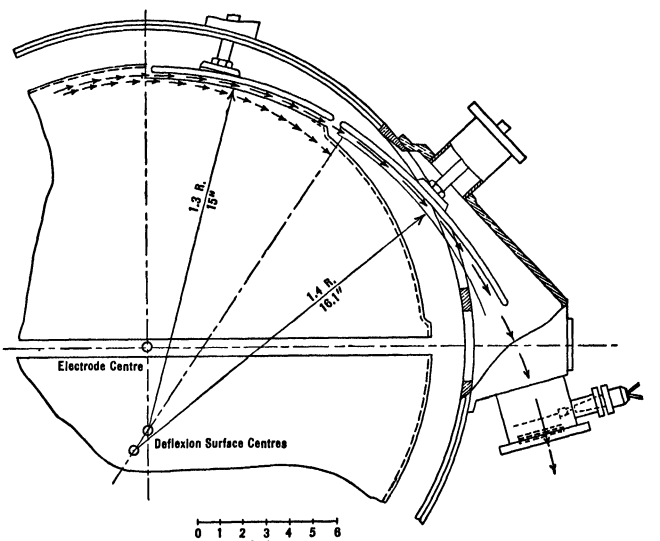


FIGURE 16.3—PATHS OF THE IONS IN THE CYCLOTRON JUST BEFORE AND DURING THE LAST LAP TO THE TARGET.

describing their last and greatest semicircle, and steers them off to the target. The target may be in a sort of recess in the wall of the dee, or in the enclosing chamber (Figure 16.3) or the nuclei may be directed to a "port" in the wall of the

[1] One circumstance which forbids unlimited advance is the relativistic dependence of mass on speed, which (as mentioned above) invalidates equations (16.1) and destroys the constancy of angular velocity on which the operation of the cyclotron depends; yet this may be compensated to some extent by making the field increase from the center to the edge of the region occupied by the dees. Literature: Rose, **PR 53**, 392, 715 (1938); R. R. Wilson, *ibid.* 408.

enclosing chamber, covered by a thin foil of metal through which they pass directly into the atmosphere. Their energy is either deduced from the range which they exhibit in air, or evaluated from the second of equations (16.1) with the radius of the dee substituted for R. (The latter and more obvious procedure is sometimes regarded as less reliable than the former, because of the chance that some of the emerging nuclei may have followed eccentric paths.) To give an idea of the energies and beam-intensities now attainable, I quote some data given by Lawrence and Cooksey near the end of 1936: twenty or twenty-five microamperes (μA) of deuterons of energy 5.8 MEV—fifty μA of deuterons of energy 4.3 MEV, in dees of smaller radius but greater height along the axis of symmetry—a tenth of a μA of helium nuclei of energy 11 MEV. As for protons, "the obtainable beams of high speed protons are usually equal in current to the deuterons, but less in energy on account of the required higher frequency [1] of the electrical oscillations." This is a good place to mention that the alpha-particles poured out by one whole gramme of radium per second in *all* directions amounts to a current-strength of only 0.011 μA. At the end of 1938 the records were 6.5, 11, 16 Mev for protons, deuterons, alpha-particles.

The electrostatic generator [2]

The "Van de Graaff apparatus" has the remarkable feature —rare indeed among the inventions of twentieth-century physics—that its fundamental principle could have been explained as readily to a natural philosopher of two hundred

[1] People often ask about the frequency and amplitude of the alternating voltage, but since these do not enter into the formula (16.1) for the energy achieved, they are often not recorded and not even measured. Lawrence and Cooksey speak of voltage-amplitudes "of the order of magnitude of 100,000" the nuclei then describing about 25 complete revolutions in the dees.

[2] R. J. van de Graaff, K. T. Compton and L. C. van Atta, **PR 43**, 149 (1933); van Atta *et al.*, *ibid.* **49**, 761 (1936); M. A. Tuve, L. R. Hafstad and O. Dahl, *ibid.* **48**, 315 (1935), **50**, 504 (1936); for apparatus in high-pressure tank, H. A. Barton, D. W. Mueller and van Atta, *ibid.* **42**, 901 (1932); R. G. Herb, D. B. Parkinson and D. W. Kerst, *ibid.* **51**; 75 (1937).

years ago as to a physicist of today. Only an endless belt of insulating substance, steadily moving because it is wrapped around two axles, one of which is being steadily turned (by a motor), and steadily carrying electric charge which is being

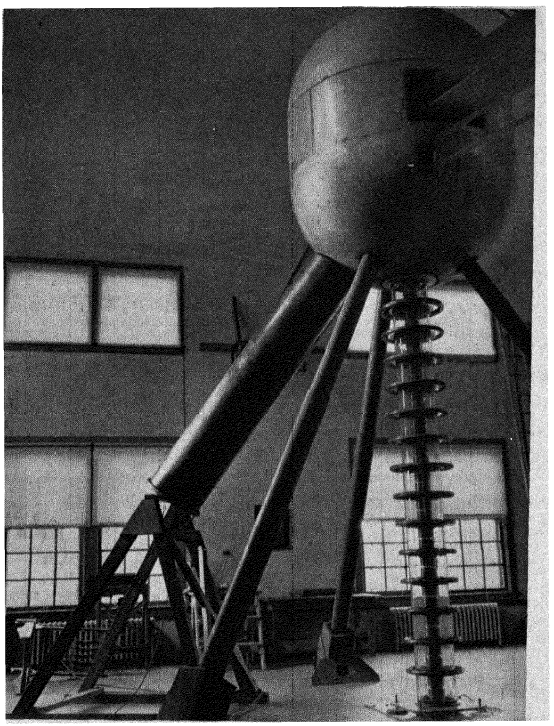


FIGURE 16.4—ELECTROSTATIC GENERATOR ACCORDING TO PRECEPTS OF VAN DE GRAAFF, BUILT AT THE CARNEGIE INSTITUTION OF WASHINGTON. (Tuve, Hafstad and Dahl.)

imparted continually to the belt at one place and withdrawn into a reservoir at another! Just before reaching the place where the charge is to be withdrawn, the belt enters through a hole into a big hollow envelope or "shell" of metal, spherical or oval in shape (Figure 16.4). This is what I just called the

reservoir: the charge leaves the belt and distributes itself over the outer surface of the shell. The transfers of charge from and to the belt are made in a way which Benjamin Franklin, if not his predecessors, would have recognized: very close to the surface of the running belt there are affixed two sets of needle-points or "corona-points"—one set is connected to a perpetually replenished source of electric charge, the other is connected to the shell—and owing to their sharpness the charge leaks across the gap, as it did between the clouds and Franklin's pointed lightning-rods of old.[1]

As the belt revolves, the charge and accordingly the potential of the shell rise steadily in absolute value; not, however, indefinitely. There must, of course, be a limit, and one might expect it to be reached when the power of the motor no longer suffices to drive the charge-laden belt into the shell against the repulsion of the charge which is already there. This imaginable limit is never even approached. The actual limit is set by the spark or corona discharge through which the charge escapes from the shell through the air when a certain critical potential is overpassed. The critical potential is greater, the less the curvature of the surface of the charged metal; it is on this account that the shell is made so large. The critical potential is also greater, the farther away the shell is placed from earthed objects such as ceilings and walls; on this account the apparatus is set up in large high-ceilinged halls, or even in the open air. Van de Graaff made two spherical shells fifteen feet in diameter, sitting atop of insulating hollow towers twenty-two feet high, inside which the belts were housed; charging one of them positively and the other negatively, he attained by the spring of 1936 to potential-differences of 5.1 million volts between the two. At the Paris Exposition of 1937 a similar

[1] The source of charge may be perpetually replenished by any convenient device (a transformer-kenotron set has sometimes been used); but the apparatus may actually function without any such source at all, the initially-transferred charge being produced apparently by friction, and the belt emerging from the shell with a charge of opposite sign (!) which provokes further transfers of charge in the correct direction. Perhaps the natural philosophers of two hundred years ago would understand these things better.

but somewhat less huge apparatus was daily producing voltages in excess of four million.

At this point, however, the experimenter meets with a new great difficulty, encountered in fact by all those who work with any instrument for the high acceleration of nuclei, with the sole but important exception of the cyclotron. In the cyclotron, the maximum potential-difference between any two parts of the apparatus is still only of the order of a few tens of thousands of volts, when the nuclei are invested with energies of millions of eV; but in the electrostatic generator (as in the Cockcroft scheme and the transformer-rectifier set) these two quantities, when measured in volts and electron-volts respectively, are equal. (This last statement is valid when the accelerated ions are protons and deuterons; *remember always that when helium nuclei are used, their double charge entails a doubling of the energy which is received from a given voltage*; but remember also that in overcoming electrostatic barriers such as those which surround bombarded nuclei, they are correspondingly handicapped.) If bombardment is to be carried out with nuclei of energy amounting to E electron-volts, the effort of the machine will generally be able to maintain the source (usually a gas excited by a discharge) at a potential of $+E$ volts relative to earth, while the target will be earthed.

The nuclei must fly from source to target through an evacuated tube or pipe, the wall of which is made of an insulating material. Between the ends of the tube must exist the voltage E. This means that the tube must be long in proportion, for much experience has shown that the ratio of E in volts to tube-length in centimeters cannot be raised beyond a few thousands without incurring danger of ruinous discharges. Even to do so well as this it is necessary to take elaborate steps to uniformize the potential-gradient along the wall, such as the insertion of channeled metal partitions at equal intervals along the hollow of the tube, these being connected through the wall to external metal rings or needle-points. The channeled partitions (illustrated in Figure 16.4) have the additional merit of

helping to focus the beam of nuclei, a service for which they must be carefully designed.

To make use of the voltage between positive and negative shells sitting on separate towers as in Van de Graaff's first design, it would be necessary to bridge the gap between the spheres by a very long tube of this sort supported in a horizontal position. The mechanical difficulties have been found to be very great; the first successful use of Van de Graaff's idea for transmutation was made with a vertical tube extending from a single shell toward the earth, as depicted in Figure 16.4. With this apparatus Tuve and his collaborators Hafstad and Dahl, of the Carnegie Institution in Washington, attained to voltages of 1,200,000 and hence to energies (of protons and deuterons) amounting to 1.2 MEV, with beam-strengths up to 15 mμA. Voltages below the uppermost attainable limit are obtained by facilitating the escape of charge from the shell through the air, this being done with the aid of "corona-points"—pointed spikes of metal arranged in a comb and earthed, which are placed closer to the shell the lower the voltage desired.

If the density of the atmosphere could be augmented at will, the critical potential of discharge which sets the limit to the attainable voltage could itself be considerably raised, even with a relatively small shell of relatively high curvature. This idea has led to the construction of electrostatic generators entirely enclosed in strong-walled tanks, first achieved by Barton, Mueller and Van Atta; a recent and excellent example is that of Herb and his collaborators Parkinson and Kerst at the University of Wisconsin. The shell, reduced to a mere cylindrical segment, is near the center of a tank 20 feet in length and $5\frac{1}{2}$ feet in diameter. From it the belt extends toward one end of the tank, and the evacuated tube—of corrugated porcelain, eleven feet long, with external corona-points at intervals of three inches connected to internal channeled partitions—extends to the other. Ingenious schemes are used for supporting the shell, the tube, and the other internal furniture of the tank, without introducing electrical leaks or local

potential-gradients of serious magnitude. When the air in the tank is at atmospheric pressure, the apparatus develops a voltage of half a million; but this comparatively modest (and yet quite useful) value is exalted to a full two million, when the pressure is raised to six or seven atmospheres.

Additional ways of observing transmutation

Of the ways of observing transmutation I have spoken already of three: the viewing of scintillations, never to be forgotten but now obsolete in research; the use of the expansion-chamber, still unrivalled in certain important respects and likely to be precious always; and the detection of radioactivity, available when one of the nuclei resulting from the transmutation happens to be radioactive. Describing again the typical transmutation as the coalescence of two nuclei followed by the separation of an "emerging particle" from a "residue," I recall that the first two of these are ways of detecting the emerging particle (though in rare cases the Wilson chamber shows the tracks of all the newborn nuclei) while the third is a way of detecting the residue. A fourth method, very like that of the expansion-chamber, consists in mixing the substance to be transmuted into a photographic emulsion, bombarding it and developing (in the photographic sense): the grains of the emulsion which were traversed by flying nuclei now appear sharply contrasted with the rest, forming visible tracks or trails. This has been developed and employed by Chadwick and Goldhaber and by Wilkins.

We now turn to the *electrical* ways of detecting the emerging particle.

In all of these, the flying newborn nucleus is allowed to traverse a gas contained in an "ionization-chamber"; it ionizes the gas, as it would in a Wilson chamber; but instead of immobilizing the ions by condensing water-vapor upon them, the experimenter leaves them free to drift to collecting-electrodes and even to make new ions in the gas. Thirty years ago it would have been impossible to detect by electrical means

the ionization directly produced by a single newborn nucleus (about a quarter of a million ion-pairs for an alpha-particle of energy 8 MEV, and much less for most emerging particles produced by transmutation). This problem was, however, destined to be solved in many ways, which I will group under four headings:

(E1) By arranging to have each particle touch off a brief but violent discharge—a sort of invisible spark—in the ionization-chamber. There is a strong electric field applied between the electrodes of the chamber, whereby the "primary" ions which the particle forms as it travels across the gas are enabled to produce vast numbers of extra or "secondary" ions; and these constitute a transient current large enough to be perceived in the external circuit. The idea was first put into practice by Rutherford and Geiger in 1908, and the instrument is often called a "Geiger counter." One of the electrodes must be either a fairly sharp point or a fairly thin wire, and there are many empirical rules about the size and shape of the chamber, the proportioning and conditioning and placing of the electrodes, the nature and purity and density of the gas, and the magnitude of the field. The name describes the use of the instrument well, since the vigor of the discharge does not depend on the nature or speed of the particle which touched it off, and therefore the counter serves best for counting particles of which the nature and speed are known or plausibly guessed (or irrelevant).

The methods to which we now turn give fuller information.

(E2) By modifying the foregoing scheme so that the number of secondary ions shall be proportional to the number of primary ions, and the measurement of their aggregate charge shall give at least a relative value of the ionizing-power of the traversing particle. This modification, which immensely expands the value of the method, is a relatively recent achievement of Geiger and Klemperer.

(E3) By developing an electrometer or electroscope so sensitive that it is able to detect and even to measure the total charge of a few thousands of ions. This was first achieved, or

at any rate first applied to transmutation, by Hoffmann of Halle and his associate Pose. The latter, in his observations of the protons of the Al(α, p) reaction which led to the discovery of resonance, succeeded in detecting flying nuclei which produced as few as three thousand ion-pairs. This sensitiveness is paid for by sacrificing all possibility of rapid work, for the electroscope takes a large fraction of a minute to complete its deflection and then recover its readiness to respond to the next particle. Pose in his experiments observed only some thirty protons to the hour.[1]

(E4) By applying external amplification to the feeble impulse which the primary ions due to a single particle produce in the external circuit, and which is imperceptible to an electroscope of normal and convenient quickness of response. This is done by developing the wonderful techniques of amplification which modern vacuum-tubes have made feasible, and like the two preceding schemes is an achievement of the last few years, having been advanced especially by Wynn-Williams of the Cavendish Laboratory and Dunning of Columbia. Some records made with Dunning's apparatus are here reproduced [2] as Figure 16.6; before discussing them, I pause to specify the most valuable of the powers which these last three methods have given to physicists.

These powers are those of determining the *nature* and the *energy* of the charged particle. To take the latter first: if the chamber is so deep and the gas so dense that the entering particle "runs itself dead" therein, the kinetic energy which this possessed at entry is (when measured in eV) about 33 times the number of ion-pairs produced, according to the extremely convenient theorem already so often cited. As to the nature of the flying nucleus, such measurements tell noth-

[1] It deserves to be commemorated that in their blank experiments, Hoffmann and Pose during one research observed deflections at the average rate of 1.22 per hour, but observed altogether 197 of them.

[2] I am greatly indebted to Dr. J. R. Dunning for these pictures, made at my request. He writes of (*b*): "The minimum amount of ionization here detectable is well under 1000 ions; probably it could be pushed down to 250 ions." Consecutive dots at the bottom of each record mark off the elapsing minutes.

ing. Suppose, however, that the chamber is so shallow and the gas so rarefied that in traversing it from side to side the nuclei lose only a relatively little amount of energy, preferably not more than a few tenths of one MEV. If now screens of various stopping-powers are placed in the path of the nuclei between the source and the chamber, and the ionization in the chamber is plotted as function of the air-equivalent of the screens, one gets the "Bragg curve" already mentioned in Chapter VIII. An example of this curve is shown in Figure 16.7. While the shape of the curve is the same for charged particles of all types, its ordinates in general—and the ordinate of the maximum in particular—depend upon the kind of particle. The curve of Figure 16.7 has the proper ordinates for an alpha-particle, and yet it was obtained with newborn nuclei (resulting from a transmutation which we shall presently study) which were not at the time known with certainty to be alpha-particles: the curve demonstrated their nature.

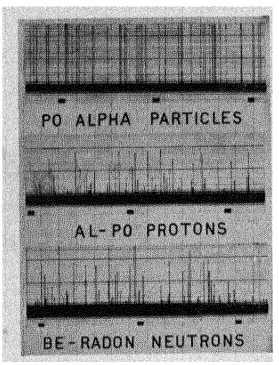


FIGURE 16.6—THREE RECORDS OF THE IONIZATION PRODUCED BY INDIVIDUAL PARTICLES IN A SHALLOW IONIZATION CHAMBER: (*a*) ALPHA-PARTICLES OF NEARLY UNIFORM SPEED; (*b*) PROTONS OF DIVERSE SPEEDS; (*c*) PARTICLES OF VARIOUS KINDS WHICH HAD BEEN SET INTO MOTION BY IMPACTS OF NEUTRONS. FOGGING ALONG THE BASE LINES MUCH DARKER HERE THAN IN THE ORIGINAL PRINTS. (Dunning.)

I digress to remark that the shape of the Bragg curve makes feasible a convenient way of obtaining something very like the differential distribution-in-range curve (Footnote 1, page 493) corresponding to an integral curve such as those of Figures 16.9 and 16.11. If the emergent protons (or flying nucleus of whatever kind is being studied) pass through screens of adjustable air-equivalent x and then across a shallow ionization-chamber, then whenever x is about equal to or slightly smaller than the abscissa of any of the curving arcs of the integral curve, the particles entering the chamber comprise

an exceptional number of (relatively) slow-moving ones for which the ionizing-power is relatively large. It follows that when the ionization in the chamber is plotted as function of x, a peak appears in the curve for each "group" in the sense of page 493. This is illustrated by the curve of Figure 18.5. The potency of this method is increased by having the particles traverse two shallow chambers in immediate succession, these being so connected that it is the difference between the currents flowing in them which affects the measuring-instrument.

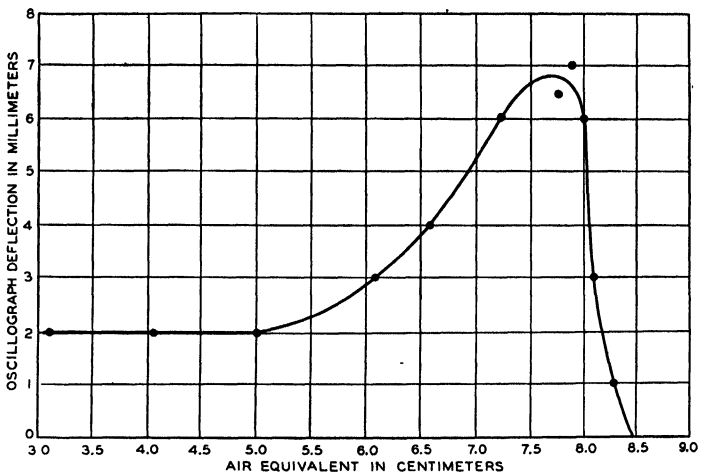


FIGURE 16.7—IONIZATION PRODUCED IN A SHALLOW CHAMBER BY NUCLEI (OF He) WHICH SPRANG FROM TRANSMUTATION AND PASSED THROUGH ABSORBING SCREENS OF THE INDICATED AIR-EQUIVALENT BEFORE REACHING THE CHAMBER. (Cockcroft and Walton.)

Now to return to the problem of identifying the particles: it is not always necessary for this purpose to trace the entire Bragg curve. The ordinates of this curve being about four times as high for alpha-particles as for protons, it is possible to fix (for any particular ionization-chamber and gas-content) a value I_0 such that if any traversing particle generates more than I_0 ion-pairs in the chamber it certainly cannot be a proton, and must be an alpha-particle if no yet-more-highly-

charged nuclei are present. With this in mind we turn to Figure 16.6, in which every vertical line (the image of a spot reflected from a turning galvanometer mirror) is due to an ionizing particle traversing a shallow chamber, and is proportional in length to the number of ions which the particle made in crossing.

The lines of great and nearly uniform length which appear in record (*a*) are due to alpha-particles from polonium; these all had nearly the same speed and were moving in nearly parallel lines when they entered the chamber, and it is clear that in crossing the gas they all made nearly the same number of ions; they left with most of their kinetic energy unspent.

The lines in record (*b*) are due to protons which sprang from the Al(α, p) reaction, their diversity in length revealing the diversity in speed among these last. The greatest amount of ionization which (in this particular chamber and gas-content) a proton could ever produce, with its most favorable speed, is indicated by the longest lines in (*b*). One sees that even these longest lines are definitely shorter than the lines in (*a*) due to the alpha-particles; the contrast in fact is much more striking than the pictures suggest, for the amplification was four-fold greater when (*b*) and (*c*) were made than when (*a*) was made.

Record (*c*) was made when neutrons (Chapter XVII) were traversing the chamber and a piece of paraffin beside it; not the neutrons, but the atom-nuclei which they strike and set into rapid motion, produced these lines. Lines longer than the longest in (*b*) were certainly not caused by protons: they were due to recoiling nuclei of the atoms of the gas which fills the chamber (air); the diversity among their lengths is ascribed to diversity of speed among the neutrons and to diversity in angle of recoil. Shorter lines were due in part to such nuclei, in part to hydrogen nuclei originally in the paraffin and knocked out of it into the chamber by neutron-impacts. Observations with expansion-chambers, and measurements made in the presence of deflecting fields, have confirmed these classifications of the particles.

The (p, α) reactions

The second-to-be-discovered of the types of transmutation was the (p, α) type. Five examples have now been authenticated; they pertain to four elements—lithium, beryllium, boron and fluorine—of which the first exhibits two of these reactions, one for each isotope. I pass over three of the five—$Li^6(p, \alpha)He^3$, $Be^9(p, \alpha)Li^6$ and $F^{19}(p, \alpha)O^{16}$—with a mere mention. Notice that in two of these three cases the residue is a nucleus of a familiar kind; not, however, in the first, for He^3 is obtained only by transmutation, though another reaction (page 552) produces it more richly than this one. To the other two I give subsections; but beforehand, a general remark is worth the making.

The student will have noticed that this (p, α) type of transmutation is the reverse of the first-discovered type. I say that the *types* are reverse to one another—not, however, the individual reactions. There is as yet no case in which a particular reaction of either class and its reverse have both been produced. For some of the known reactions a valid reason is easy to find: the residue (of the known reaction) is a kind of nucleus so rare in Nature that either it has not been found at all (such as He^3) or else it is so scanty that physicists are not able to get enough of it together for a satisfactory test (such as O^{17}, the residue of the reaction 15.1). In other cases another good reason is clear. If a reaction is exothermic, its reverse must be endothermic; if the former has a positive heat of reaction Q, the onset-energy for the latter cannot be less than Q. Now, of the five listed (p, α) reactions, all but one have Q-values positive and greater than 8 MEV, a value so near to the topmost energy of practically all natural alpha-particles that failure to detect their reverse reactions is not astonishing.

But how about reversing $Be^9(p, \alpha)Li^6$? There is no difficulty about finding plenty of Li^6 in Nature, and the Q-value of the (p, α) is only about 2 MEV; yet lithium has been bombarded with 7.5 MEV alpha-particles expressly in the hope of

performing the reaction $Li^6(\alpha, p)Be^9$, and protons have been sought in vain.[1] Moreover, if one looks over the list of the accepted Q-values for the known (α, p) reactions, many are seen to be negative: the reverse reactions must be exothermic, and yet have not been observed. Evidently it is much too naive to suppose that in transmutation, what the energy-relations do not forbid is necessarily permitted. There is nothing surprising about this; but the possibility of these processes being really irreversible is hard indeed to accept.

First example of a (p, α) reaction: Li^7 *plus* H^1 *becoming a pair of alpha-particles*

The transmutation described by this caption is second only to $N^{14}(\alpha, p)O^{17}$ in historic moment. It was the first to be performed with bombarding particles other than He^4 nuclei, and with bombarding particles which had derived their kinetic energy from a man-made engine. It was also the first for which the Q-measurements confirmed the relativistic dependence of mass on speed; it was and is one of the easiest to produce with low-voltage bombardment, one of the simplest and one of the most thoroughly studied.

The two newborn nuclei being of the same kind, there is no distinction to be drawn between emerging particle and residue; both are emerging particles in the sense of page 478. The fact that two emerge at once, in nearly opposite directions, was among the first to be sought and proved by Cockcroft and Walton after their discovery of the reaction. They had made the discovery with a fluorescent screen inside the evacuated tube in which the bombarding proton-stream impinged upon the target. On this screen they observed scintillations, about five per minute per μA of bombarding current, at the (then) amazingly low voltage of 125,000. Putting two screens on opposite sides of the target (a thin sheet of lithium) and assigning to each an observer who made a record of every scintillation on a moving tape, they found coincidences occurring a

[1] W. G. Shepherd, R. O. Haxby, E. L. Hill, PR 52, 674 (1937), and work at the Cavendish there cited.

great deal oftener than by sheer chance they should have. Magnificent expansion-chamber photographs were thereupon made by Walton and Dee, and these displayed "paired tracks" springing forth in nearly opposite directions from the knob that terminated the evacuated tube and enclosed the target (Figure 16.8). The proof that these newborn nuclei are alpha-particles has already been introduced into this book, for the

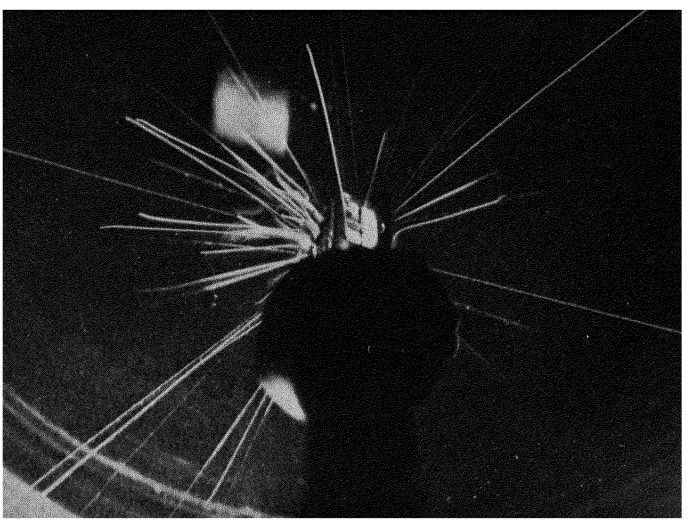


FIGURE 16.8—TRAILS OF PAIRED FRAGMENTS, He NUCLEI RESULTING FROM IMPACTS OF HYDROGEN ON LITHIUM: H^1 ON Li^7, H^2 ON Li^6. (Dee and Walton; *Proceedings of the Royal Society*.)

curve of Figure 16.7 (obtained by Cockcroft and Walton) relates to them.

Shortly after these discoveries, Oliphant, Kinsey and Rutherford with the aid of one of the electrical schemes of detection (the one listed *supra* as E4) got the integral distribution-in-range curve of the alpha-particles emerging within a small solid angle about a chosen direction. In this curve (Figure 16.9) the bends at the left are due to the He^4 and He^3 nuclei spring-

ing from the reaction $Li^6(p, \alpha)He^3$ already mentioned. At the right is a fine example of a steep ascending arc climbing from the axis of abscissa to a beautiful long plateau. Its resemblance to the ideal vertical step which should correspond to a group is much closer than is usually found with (α, p) reactions —largely because the strong bombarding current and high yield make it possible to reduce considerably what on page 492

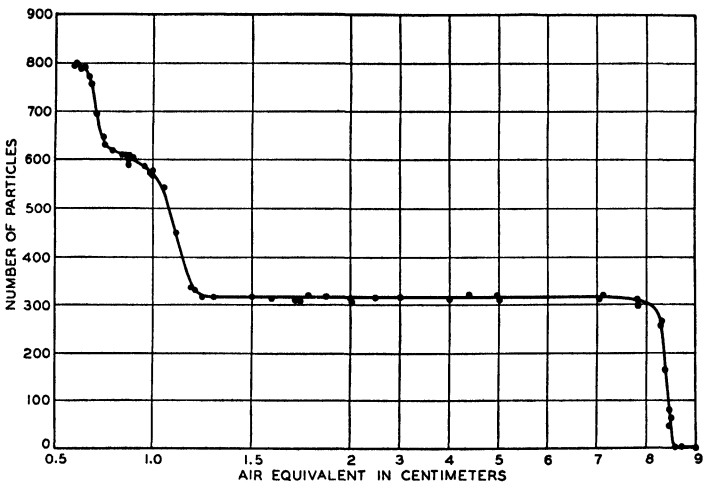


FIGURE 16.9—INTEGRAL DISTRIBUTION-IN-RANGE CURVE OF NEW-BORN NUCLEI RESULTING FROM REACTIONS OF TRANSMUTATION BETWEEN PROTONS AND LITHIUM. (Oliphant, Kinsey and Rutherford; *Proceedings of the Royal Society*.)

was called "the spread of θ-values." (For this curve the θ-spread was about 15° and the mean θ-value was 90°). This is the group—the only one—arising from the reaction which for consistency we should call $Li^7(p, \alpha)He^4$, though here the notation implies a nonexistent distinction between the two alpha-particles.

Deducing the kinetic energy of the individual particle from the range of the group, and taking account of the energy of the paired particle computed [1] as on page 492, one finds the

[1] It will have occurred to the student that by using the Wilson chamber and measuring the ranges of both members of a couple of "paired tracks," the observer

Q-value. It is extraordinarily high, of the order of 17 MEV. It is so high that the E-value of the protons is almost (not always quite) negligible by contrast, and the two alpha-particles fly off in directions almost (not quite) opposite, with energies which are almost (and may be quite) equal. The fact that these newborn nuclei agree not only in kind but also in energy (usually 8 to 9 MEV for either member of a pair) with the commonest "natural" alpha-particles is a queer but lucky coincidence; for the aspect of their scintillations on a screen and of their tracks in an expansion-chamber suggested their nature to Cockcroft and Walton before the curve of Figure 16.7 was available to clinch the proof.

To facilitate the statement which follows, I now write the equation of this reaction in the manner already exemplified by (15.4), *viz.*:

$$ {}_3\mathrm{Li}^7 + {}_1\mathrm{H}^1 = 2\,{}_2\mathrm{He}^4 + Q \qquad (16.2) $$

The Q-value deduced from the data of Oliphant, Kinsey and Rutherford embodied in Figure 16.9 (and substantiated by the Berkeley school) is 17.13 MEV. The sum of the rest-masses of Li^7 and H^1 is greater than the doubled rest-mass of He^4, according to measurements with mass-spectrographs, by 0.0185 in terms of the customary unit (rest-mass of ${}_8\mathrm{O}^{16}$) or 17.25 MEV. The agreement is admirable.[1] It may be regarded as proof of the Einstein mass-energy relation, as proof that the nuclei in question are normal nuclei of their several kinds, or best as fortifying both.

The yield-curve has been investigated over the remarkably long E-interval from 0.0236 to 1.125 MEV, though not

could circumvent this computation. This has been done, but the general preference is for adopting the same procedure as is customary when there is an unobservable residue. See in this connection A. Roberts *et al.*, **PR 49**, 783, 1936.

[1] I take the figures from Livingston and Bethe, who in turn took the data as to Q from Oliphant *et al.* and the data as to rest-masses from Aston, Bainbridge and Jordan, with slight modifications and compromises in their treatment thereof. It is disconcerting at first glance to learn that the agreement was practically as good in 1933 when the accepted values for the rest-masses of Li^7 and He^4 differed sharply from those now received. The then-existing error (if error it was) affected the ratios of these masses to that of ${}_8\mathrm{O}^{16}$ much more than it did the difference (Li^7–$2\mathrm{He}^4$).

throughout by the same experimenter.[1] It is interesting to note the diversity of the accelerating-engines: the cyclotron (M. C. Henderson) and the electrostatic generator in the atmosphere (Hafstad and Tuve) at the highest energies, the electrostatic generator in a high-pressure tank (Herb Parkinson and Kerst) from 0.1 to 0.4 MEV, the Cockcroft system (by physicists at the Cavendish) from 0.2 MEV downward, and ordinary transformers (Doolittle, Burhop, several German physicists) at the lowest energies. Unfortunately there is disagreement between the two authorities who worked in the

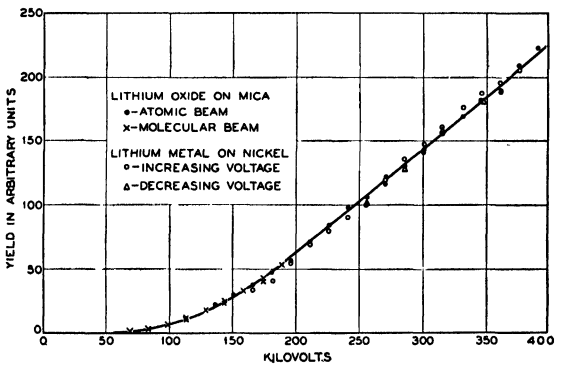


FIGURE 16.10—YIELD-CURVE OF THE REACTION OF TRANSMUTATION Li^7 (p, α) He^4 OBTAINED WITH THIN FILMS UP TO ENERGY 0.4 MEV OF THE IMPINGING PROTONS. (Herb Parkinson and Kerst; *Physical Review*.)

region beyond 0.4 MEV, while at the lowest energies the yield is so small that any measurement at all is a sort of *tour de force*. Between 0.1 and 0.4 MEV there are ample data of the Wisconsin school, obtained with five thin targets (thin films of lithium metal or lithium compounds laid down on solid bases) and a thick one. The thin-target curve bends gradually upward from the axis of abscissae, and then continues in a linear arc of remarkable straightness from about 0.15 to 0.4 MEV. Linear

[1] Hafstad and Tuve, **PR 48**, 314 (1935); Herb Parkinson and Kerst, *ibid.* **48**, 118 (see also Haworth and King, *ibid.* **51**, 59); Doolittle, *ibid.* **49**, 779 (1936); see Doolittle for earlier references.

arcs are uncommon in yield-curves, most of which (as published) are concave-upward throughout the energy-interval over which they extend, except when resonance-humps deform them. No sign of resonance appears in the curve of Figure 16.10.

In general it can be said that yield-curves obtained for the same element by different experimenters, or by the same experimenter in different experiments, are often alike in form but usually different in height; by which is meant, that to bring two of them into fairly good agreement it is usually necessary and often sufficient to multiply all the ordinates of one by a constant factor. Such discrepancies may be ascribed to one or more of a variety of things, such as errors in the estimate of the number of impinging particles, differences in the placing and aperture of the counters, and differences in the state of chemical combination of the transmutable elements in the target (particularly likely with an element so active as lithium). This last does *not* imply that the transmutability of an atom depends upon the kind of neighbors which it has, but only that the neighbors can slow down the oncoming particle before it arrives at the transmutable nucleus. Owing to all these things the form of a yield-curve is a great deal easier to get than a proper σ-value. For the present I give only a vague idea of the yield of the $Li^7(p, \alpha)He^4$ reaction by citing various values for the number of alpha-particles emitted [1] from a thick target per 10^9 incident protons (equivalent to 0.16 mμA sec): at 1 MEV, 80 according to Henderson; at 0.36 MEV, 63 according to Herb, Parkinson and Kerst; at 0.1 MEV, ca. .001 according to Oliphant and Rutherford; at 0.02 MEV, 10^{-5} according to Burhop. Trautenberg followed the reaction down as far as 0.013 MEV.

Ordinary lithium being a mixture of the isotopes Li^6 and Li^7, both are exposed to transmutation when it is bombarded by protons; how then can the physicist tell to which he should attribute any group of emerging particles? For the group with

[1] Number observed, multiplied by the ratio of 4π to the solid angle subtended by the detecting-device at the target. Divide by two to obtain the number of transmutations.

which we have just been concerned the answer has been given, and is clear. All particles of the group are known to be He nuclei, and they are known to occur in pairs; no reaction between Li^6 and H^1 could lead to such a result. If the pairing had not been proved, the physicist could base his belief on the fact that the beautiful numerical agreement just quoted for the Q-value—the "balancing" of equation (15.4)—is contingent on the rightness of the assumption that the reaction is $Li^7(p, \alpha)He^4$. For many a transmutation the opinion as to the isotope involved is based on just such a contingency. With lithium this now is doubly needless, for samples of practically pure Li^6 and others of practically pure Li^7 have been prepared,[1] and the alpha-particles of this group are obtained only by bombarding those of the latter isotope. As to the groups responsible for the bumps near the leftward end of the curve of Figure 16.9, they appear only when the Li^6 target is used, so verifying the ascription already made of them. It is much to be regretted that for most of the chemical elements the separation of the isotopes has not yet been done, nor is there any immediate prospect of its being done.

The reaction which we have just considered is not the only thing which can happen when protons impinge on Li^7 nuclei; there is also emission of photons, which by the testimony of the yield-curve is found to have no connection whatever with the formation of alpha-particles; it discloses an alternative reaction (page 503).

Second example of a (p, α) *reaction*: B^{11} *plus* H^1 *becoming alpha-particles* (*and* Be *nuclei*?)

Details of individual reactions must be omitted from now on except in remarkable cases, else half of this book would be nuclear physics; but that which happens[2] when protons

[1] Oliphant, Shire and Crowther, **PRS 146**, 122 (1934); L. H. Rumbaugh and L. R. Hafstad, **PR 50**, 681 (1936).

[2] Dee and Gilbert, **PRS 154**, 279 (1936); Cockcroft and Lewis, *ibid.* 246; Kirchner and Neuert, **PZ 34**, 897 (1933); J. H. Williams *et al.*, **PR 51**, 434 (1937); earlier literature cited by these.

impinge upon boron is very remarkable. Alpha-particles spring forth, as from lithium and other elements under such bombardment; but their integral distribution-in-range is not composed of steep ascending arcs between plateaus as heretofore—it is instead a continually bending curve from end to end. This signifies for the emerging particles a broad *continuum* of energies, as distinguished from a partition into groups. The only suggestion of a group is near the high-energy end, where Kirchner discovered (by measurements of tracks in the Wilson chamber) a small hump on the differential curve; this is believed to show a true group, that of the particles springing from the reaction $B^{11}(p, \alpha)Be^{8}$, with a Q-value of 8.60 MEV.

To its discoverers Cockcroft and Walton the continuum at once suggested (or perhaps confirmed a preconceived idea) that from the merging of B^{11} and H^{1} *three* alpha-particles spring forth. With three particles among which the available energy and momentum are to be distributed, one readily sees that a continuum is perfectly compatible with a unique Q-value. This, however, can no longer be learned by measuring the energy of alpha-particles going off in any one direction. It is necessary to use the expansion-chamber and to recognize, if not a full triad of correlated alpha-particles, at least a pair belonging to such a triad: and this turns out to be more difficult than to recognize paired tracks from the $Li^{7}(p, \alpha)He^{4}$ reaction. Some apparent triads published in the early days are now regarded as mere simulations, three unrelated particles having happened to spring out nearly at once; but in the latest work of Dee and Gilbert the danger of such mistakes is minimized. In their principal set of experiments they looked for pairs of tracks springing from sensibly the same point of the target, and for every pair they calculated [1] the direction and range which the third alpha-particle had to have if the pair was truly part of a triad; they calculated also the Q-value,

[1] The momentum-vector of the third particle must close the triangle formed by the momentum-vector P of the proton and the sum S of the momentum-vectors A_1 and A_2 of the two observed alpha-particles, if these last are truly two members of a triad. As P and A_1 and A_2 are all deducible from the data, A_3 is calculable; and since all four momenta are thus known, so are all four kinetic energies, and Q is calculable.

of which they knew in advance that to fit with the rest-masses of H^1, He^4 and B^{11} it should be about 9 MEV. Out of 79 pairs, there were 56 which gave Q-values so close to 9 MEV that Dee and Gilbert accepted them as "true pairs" belonging to triads; the rest they discarded. In 48 of those 56 cases, the direction or length (or both) of the track computed for the third alpha-particle was such that it could not have appeared distinctly on the photograph; in six of the other eight "the third track was observed within 3° of the calculated angle" and in two of these six it could be followed to its end. After making divers corrections and allowances for the uncertainties in the experiment, Dee and Gilbert arrived at 8.7 ± 0.2 MEV for the Q-value.

From the rest-masses the to-be-anticipated Q-value is figured as 8.72. Here again is a splendid agreement! but to pass on without saying more about it would not be fair, for it is connected with a remarkable bit of history. When the earliest estimates of Q from observations upon this reaction were made, they already were between 8 and 9 MEV, but the then-accepted rest-masses were requiring a value of 11 MEV or more. This was one of the several discrepancies which inspired Oliphant in Cambridge and Bethe at Cornell independently to take the drastic step of affirming that the accepted mass-values must be wrong, even though they had been determined with the best mass-spectrographs available.[1] It was sufficient to affirm that there must be something wrong with the accepted ratio of the masses of H^1 and O^{16}, since the custom had been to "determine" the mass of He^4 (as also those of Li^6 and Li^7) by measuring its ratio to the former of those, while the masses of all the atoms from Be^9 onward had been "determined" by measuring their ratios to the latter. This presumptuous idea was shortly to be ratified in the most convincing way. A re-measurement by Aston of the ratio of the rest-masses of H^1 and O^{16} gave a new value concordant with the one which Oliphant and Bethe had demanded; Bainbridge and Jordan confirmed it; and now

[1] Oliphant, Kempton and Rutherford, **PRS 150**, 241 (1935); Bethe, **PR 47**, 633 (1935).

the rest-masses and the Q-values form a unified and a coherent system, as I had occasion to emphasize in Chapter XV.

An interesting feature appears in the yield-curves. J. H. Williams and his colleagues at Minnesota counted the alpha-particles emerging with ranges greater than 4 cm; which is to say, chiefly if not altogether the alpha-particles of the aforesaid group ascribed to the reaction $B^{11}(p, \alpha)Be^8$. Varying E up to 0.275 MEV, they found a clear indication of resonance at 0.180 MEV. When alpha-particles of smaller ranges are counted together with these, the great majority of the counted particles proceed from the other reaction; and now the yield-curve is found to rise smoothly, being of the familiar concave-upward type.[1]

The nucleus Be^8 involved in one of the foregoing reactions is an absentee from the list of the nuclei known by direct observation; for it is known only as the unobservable residue in this and a couple of other reactions of transmutation. Should its rest-mass be demonstrably greater than double that of an alpha-particle, one might reasonably suppose it predestined to early disintegration according to the formula $Be^8 \rightarrow 2He^4$; well, the best available estimate of the Q-value for the just-mentioned group permits of this supposition. There is also some reason (Dee and Gilbert, *l.c. supra*) for imagining that when a triad of alpha-particles is emitted, one of the three springs out an inappreciably short time before the others, leaving these two to form what may be called an "excited nucleus of Be^8" which explodes almost instantly.

The (d, α) reactions[2]

While the Cambridge school was a few months in the lead over the Berkeley school in achieving transmutation by protons, the order was reversed when the deuteron became avail-

[1] J. H. Williams *et al.*, **PR 51**, 434 (1937); Neuert, **PZ 36**, 629 (1935). Williams obtained thick-target curves; that relating to particles of range greater than 4 cm was of a form which indicated a maximum in the thin-target curve (page 488).

[2] Lewis, Lawrence and Livingston, **PR 44**, 55 (1933); Oliphant, Kinsey and Rutherford, **PRS 141**, 722 (1933); Dee and Walton, *ibid.* 733; Cockcroft and W. B. Lewis, **PRS 154**, 246 and 261 (1936); for other references see Livingston and Bethe.

able in the spring of 1933. G. N. Lewis had already developed the art of enhancing the concentration of H^2 in hydrogen, and provided a sample of the gas composed largely of molecules of the formula H^1H^2 (also written as HD) instead of the familiar molecules of formula H^1H^1 (usually written as H_2). This was put into the cyclotron, which was already able to confer kinetic energies of 2 MeV on the ions; for each molecular ion of HD, the energy was necessarily divided in the ratio 2 : 1 between deuteron and proton, giving for the deuterons an

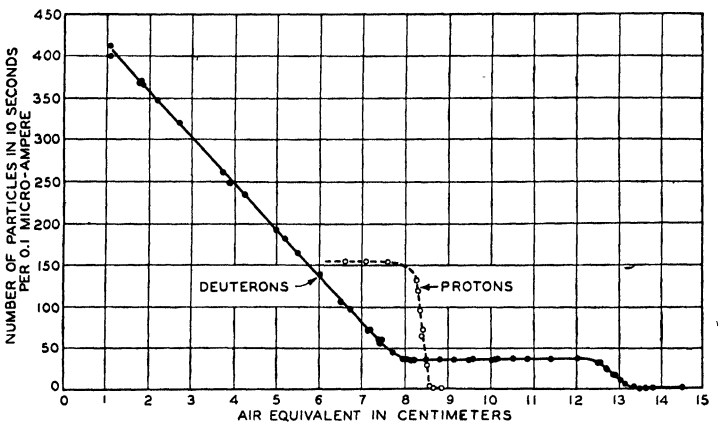


FIGURE 16.11—INTEGRAL DISTRIBUTION-IN-RANGE CURVE OF NEW-BORN NUCLEI RESULTING FROM REACTIONS OF TRANSMUTATION BETWEEN DEUTERONS AND LITHIUM. (Oliphant, Kinsey and Rutherford.)

available *E*-interval extending up to 1.33 MEV. With this at their disposition, Lawrence and Livingston bombarded lithium, and discovered and identified the alpha-particles springing from the reaction $Li^6(d, \alpha)He^4$; they bombarded several other elements, and alpha-particles came forth. Within four years, alpha-particles were detectably engendered from nine elements (aluminum being the heaviest) and attributed to sixteen different reactions; while of six more elements it could be said (not always with perfect confidence) that deuteron bombardment engenders radioactive residues, from the identification of which it follows that alpha-particles are driven out.

Some of these (d, α) reactions bear a striking resemblance to some of the (p, α) reactions which we have been considering. It may have occurred to the student to expect that when deuterons fall upon lithium, they treat the Li^6 nuclei in the same way as protons treat the nuclei of Li^7. Wrong as many obvious expectations turn out to be in nuclear physics, this one is correct. There *is* a reaction $Li^6(d, \alpha)He^4$ producing alpha-particles in couples, which in the expansion-chamber make splendid paired tracks (see Dee and Walton for examples). The rest-masses teach us not to expect the same Q-value as in the reaction $Li^7(p, \alpha)He^4$, but a decidedly greater one—22.1 MEV to be distributed between the alpha-particles of a pair. The data confirm this value admirably. The paired tracks are readily distinguishable by their greater lengths [1] from those resulting from the proton-reaction, and the kinetic energies of the individual particles (10 to 12 MEV) were unprecedented in human experience (exception being made for cosmic rays). Deuterons also react with the lighter isotope of boron as do protons with the heavier; for in the complicated "spectrum" of the alpha-particles elicited when boron is bombarded by deuterons, there is a continuum attributable to the reaction $B^{10} + H^2 \rightarrow 3He^4$ and a group ascribable to $B^{10}(d, \alpha)Be^8$. The rest-masses supply the Q-values 18.0 MEV and 17.9 MEV to be distributed between the three particles in the former reaction, the two in the latter; the latter value has been beautifully checked from the data, the former is at least not incompatible with them.[2]

[1] Incidentally it is well worth noting that the abundance of these long tracks may be taken as a convenient index of either the proportion of deuterons in a bombarding beam containing both deuterons and protons, or the proportion of lithium 6 in a bombarded film made from ordinary lithium by a process involving partial separation of the isotopes.

[2] Triads have not yet been observed as Dee and Gilbert observed them in the proton-reaction with boron 11, and one is driven to estimate the average kinetic energy of all the emitted alpha-particles, a procedure having two disadvantages: the slowest particles may escape detection, and the particles emitted in one direction may not have the same average energy as those emitted in all directions. There is also uncertainty as to the contribution of the two isotopes of boron; a separation of these would be a great service to nuclear physics.

Reverting to lithium: on the integral distribution-in-range curve of Figure 16.11 the short descending arc on the extreme right shows the group consisting of the paired alpha-particles aforesaid, while the dashed curve intersecting the main one shows the corresponding group from the reaction $Li^7(p, \alpha)He^4$. The long descending arc on the left is the sign of a continuum such as we now have learned to associate with the formation of three particles: the attribution is indeed to such a reaction,

$$_3Li^7 + {}_1H^2 \rightarrow {}_2He^4 + {}_2He^4 + {}_0n^1, \qquad (16.3)$$

one member of the triad being now a *neutron*. It is becoming impossible to proceed without formally introducing this particle, to which we now must turn.

CHAPTER XVII

Mostly Concerning Neutrons

Discovery of the neutron

The discovery of the neutron was carried from start to completion in three stages which followed quickly one on another in time but were effected far apart in space, in entirely independent institutes of three different nations. Such histories have not been rare in physics, yet this one is astonishing. It is natural to presume that when a physicist has made and given out the first advances in a field till then untrodden, those who wish to follow will have to spend so long a time in gathering resources like to his, in imitating his equipment and in learning his technique, that in the meantime he will go the rest of the way. In the field of transmutation this was peculiarly to be expected. Yet, when from the Reichsanstalt in the closing month of 1930 it was published that alpha-particles elicit penetrating rays from certain elements, the Institut du Radium and the Cavendish Laboratory were equipped and were alert. Their contributions came with intervals of months, not years; and the progress could hardly have been swifter, had all the work been ordered by a sole far-seeing mind.

Bothe and (H.) Becker were the authors of the work at the Reichsanstalt, with which the tale commences.[1] They

[1] It should perhaps commence with mention of the observations which F. P. Slater made in 1921 at Cambridge (England) of penetrating rays sent forth from tin and lead subjected to alpha-particle bombardment. The status of these observations is peculiar; they have never been confirmed or disproved, so far as I know; H. C. Webster of the same laboratory, writing in 1932, speaks of unpublished vain attempts to produce such rays with the alpha-particles from polonium, but does not distrust Slater's data; another vain attempt was made by Bothe and Becker. Slater used alpha-rays from radon, having energy about 3 per cent greater than that of the polonium rays which are now chiefly employed; the difference is slight, but possibly great enough to bring into play a phenomenon of "resonance."

bombarded various elements with the 5.3 MEV alpha-particles from polonium and used a Geiger point-counter for observing such rays as were able to come out from the bombarded substance through a barrier of two millimeters of zinc and brass. (A barrier such as this is amply sufficient for stopping alpha-particles, and such electrons as they may set free by ionizing atoms, and even practically the whole of the X-rays which may be emitted by atoms so ionized.) Lead and silver and calcium, and several lighter elements (N, C, O, Ne) were disappointing; there was no certainty of radiation. Magnesium and aluminum gave just a trace of a result; lithium and boron and fluorine notable amounts of rays able to penetrate to the counter and affect it; and beryllium a *relatively* tremendous amount, towering up over all the others on Bothe and Becker's chart.

It was at first assumed that the rays are electromagnetic. It was assumed, that is to say, that the primary rays proceeding from the source are of the nature of high-frequency light, that they consist of photons of great energy; and that the corpuscles which effect the ionization are electrons which these photons have detached from atoms and endowed with great velocities. No other theory was entertained by the early observers, German as well as French; the assumption seemed so inevitable that they did not even bother to state it explicitly, much less to justify it; it enters undefended into the titles of their papers, "Artificial Excitation of Nuclear Gamma-Rays" and the like. The tacit argument must have been, that of all the previously-known types of radiation (cosmic rays alone excepted) the gamma-rays approach nearest in penetrating power to these, and are therefore the natural analogue. The conclusion was largely right; it has been thoroughly proved that photons are abundant among these penetrating rays. Nevertheless there are also particles of quite another character, as we shall see.

The next work to be chronicled is that of Joliot and Irene Curie-Joliot, who undertook to continue Bothe's work with the three elements for which he had proved the rays to be strongest. To bombard these elements with alpha-particles they had

available *une source extrêmement forte*, 100 millicuries of polonium. In 1931 this spoke eloquently of the resources of the Institut du Radium, especially when compared with the 3 to 7 millicuries which were all that Bothe and Becker had. (In 1938 it looks trivial beside what cyclotrons can supply). They were thus able to detect the radiation from beryllium on the farther side of several cm of lead (later on, Thibaud and Dupré la Tour were to detect it beyond no fewer than 30 cm of even this ponderous metal) and to trace the curve of its intensity—which is to say, of the current in the ionization-chamber—as function of the thickness of intervening metal.

Nothing, so far, to suggest that the primary rays are other than electromagnetic! but in the first weeks of 1932 (the report appeared on 18th January) Curie and Joliot put various thin screens over the window through which the rays were entering the ionization chamber, and found that if these screens were of metal nothing sensational happened, but *if they were of paraffin, water or cellophane*—materials containing *hydrogen*—*the ionization-current went up instead of down.* This was not the first time in history that a screen had been observed to enhance the effect of what supposedly were gamma-rays, but in the previous authentic cases it was inferred—and rightly so—that the rays were expelling electrons from the substance of the screen.[1] Here the substances were distinguished not by abundance of electrons, but by abundance of hydrogen atoms in their structure; and Curie and Joliot conceived the idea that the primary rays were ejecting *protons* from the screen, which entered the chamber and in it ionized abundantly. This theory they fortified at once by applying magnetic fields, and finding that the ionization persisted (electrons issuing from the paraffin would have been twisted back, unless extremely fast); by interposing 0.2 mm of aluminum, and finding that the extra ionization ceased (electrons, if extremely fast, might have got through); and by taking cloud-chamber photographs, and observing tracks of the aspect of proton-tracks springing out of

[1] This is frequently observed in the study of cosmic rays

the paraffin and traversing the ionization-chamber partly or altogether.

At once it was suspected by Curie and Joliot that these protons were recoiling from elastic impacts of the high-energy photons which the primary rays were still supposed to be—that they had suffered, in fact, the very same sort of blow as electrons suffer in the well-known "Compton effect." So great, how-

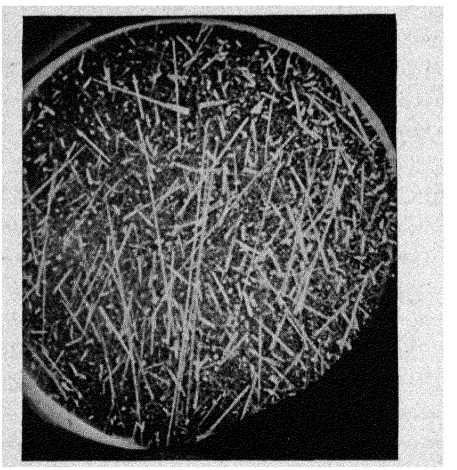


FIGURE 17.1—TRAILS OF PROTONS DRIVEN BY IMPACTS OF NEUTRONS OUT OF HYDROGEN MOLECULES IN A HYDROGEN-FILLED EXPANSION-CHAMBER. (Kurie.)

ever, was the energy of the protons (as evinced by their range) that photons of energy almost incredibly great had to be postulated; such would probably have a penetrating power even greater than that of the actual rays, and there were other objections more or less solidly founded on theory, which now it would be scarcely worth while to discuss. The French physicists were aware of these difficulties, and published them; but it was reserved for one of the Cavendish group to reject the idea altogether, and supplant it with the one which at present is accepted. Chadwick seized upon the revelations from the Institut du Radium with such alacrity that by February 27th he was reporting data obtained by counters and by cloud-

chambers—data which confirmed that the rays emitted from beryllium when bombarded by alpha-particles are able to confer great speeds not only upon protons, but on nuclei of other elements of low mass (a later list comprises Li, He, Be, B, C, N, O, A). Out of these data emerged the fact which spoke most clearly for his theory that the corpuscles which impel the protons and other nuclei are material particles of nearly the mass of a proton, instead of being corpuscles of light.

The argument is as follows. For simplicity let us consider solely the nuclei which are projected in directions pointing straight away from the source of the primary rays, and therefore must have suffered central impacts. Specially, let us take the cases of hydrogen and nitrogen nuclei thus projected. The ranges of these were promptly measured (of N by Feather, of H by various physicists) and their maximum speeds deduced by means of knowledge earlier acquired of the range-*vs.*-speed relations of charged particles. The values of speed accepted by Chadwick were $3.3 \cdot 10^9$ and $4.72 \cdot 10^8$, respectively. Now if the corpuscles which in central impacts give to these nuclei these speeds are photons, it is easy to compute by the Compton-effect equations the energy U of the photons; if the impinging corpuscles are material particles of mass M and speed v, it is easy to compute both v and M. It turns out that by the first procedure, one gets different values of U from the two cases (55 and 90 million electron-volts, respectively); by the second, one gets compatible values of M and v. With the first theory, then, one would have to say that nuclei of different kinds are struck by different photons. This is not absolutely incredible, as there *might* be a mixture of gamma-rays of different energies, and a greater likelihood of the higher-energy photons interacting with the more massive nuclei. But it seems less acceptable than the other theory, which permits one to postulate a single kind of corpuscle to explain the impacts against both kinds of nucleus. This corpuscle must be neutral, as it traverses layers of matter much thicker than a particle of charge e and the computed mass and speed could ever penetrate; it is therefore called the "neutron."

The value of M deduced from the foregoing data was given as 1.15 times the proton-mass, the possible error in the estimate of the speed of the recoiling nitrogen nuclei being such that Chadwick was able to say "it is legitimate to conclude that the mass of the neutron is very nearly the same as that of the proton." Later-to-be-discovered ways of ascertaining the neutron-mass were indeed destined to show that it differs from the proton-mass by less than one part in a thousand of either. Before introducing them I will speak further of the method of measuring neutron-energies already implied, and then of divers reactions from which neutrons are born.

Evaluation of neutron-energies

Taking it for granted that a neutron is a material particle of practically the same mass as a proton, we see at once that if a proton is initially free and stationary and a neutron strikes it with an elastic central impact, there occurs a total transfer of kinetic energy and momentum. (It is incorrect, of course, to assume that the protons which are nuclei of atoms in gaseous hydrogen or in hydrogen compounds are stationary and free; but the energy of *newborn* neutrons is so great by comparison, that the error in the assumption is negligible.) Thus if a physicist could manage to observe only the protons which are projected *straight ahead* by neutrons emanating from a source, he would know that every energy-value which he found represented among the protons was represented also among the neutrons from the source.

Actually, of course, it is an unattainable ideal which the foregoing sentence describes; there must be what I earlier called a "θ-spread," θ now standing for the angle between the directions of oncoming neutron and recoiling proton. This θ-spread has, however, been reduced to fairly innocuous proportions. In the great sequence of experiments at Pasadena by Bonner, Mott-Smith and Brubaker [1] which I shall often be

[1] T. W. Bonner and L. M. Mott-Smith, **PR 46**, 258 (1934); Bonner and H. M. Brubaker: **PR 47**, 910, **48**, 742, **49**, 19, **49**, 778, **50**, 308 (1935–36); Stephens and Bonner, **PR 52**, 527 (1937); Stephens, Bonner and Djanab, *ibid.* 1079; Bonner, *ibid.* 685.

quoting, the only protons of which account was taken were those of which the paths deviated by more than 8° (or sometimes 10°) from the line drawn from the center of the neutron-source to the point where the impact occurred. This statement does not give a full idea of the θ-spread, since the source was not a mere point, and θ-values as high as 16° had to be reckoned with; but even with so high a θ-value there is more than a 92% transfer of energy, and with $\theta = 8°$ there is a 98% transfer.

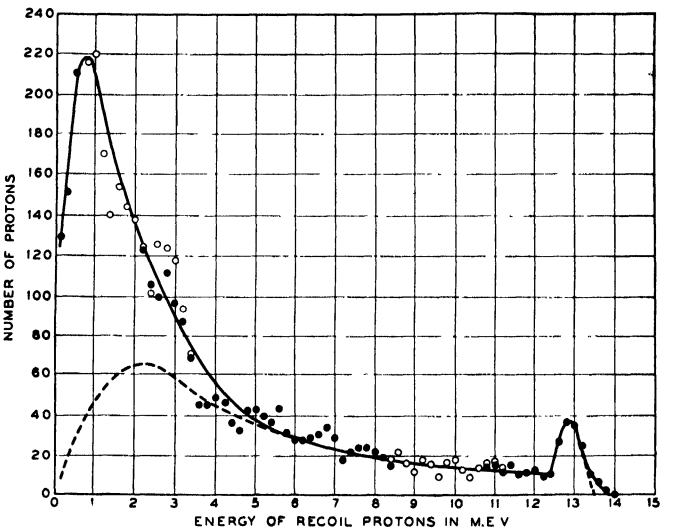


FIG. 17.2—DISTRIBUTION-IN-ENERGY OBSERVED FOR RECOILING PROTONS (SOLID CURVE) AND INFERRED FOR IMPINGING NEUTRONS (DASHED CURVE) BORN FROM REACTIONS OF TRANSMUTATION BETWEEN DEUTERONS AND LITHIUM. (Bonner and Brubaker; *Physical Review.*)

One may therefore expect the distribution-in-energy of the protons observed in such an experiment to be a fairly close copy of the distribution-in-energy of the neutrons. There is, however, a complication which must never be overlooked when distribution-curves are being sought. Even if θ were zero for every observed proton, the distribution-curve for the protons could not be an exact copy of that for the neutrons unless the

"cross-section for proton-neutron scattering" were independent of neutron-energy; now, there are powerful reasons for believing the contrary. Reserving these reasons till later, I point out in Figure 17.2 the full and the dashed curves, the former representing the observed distribution-in-energy of recoiling protons, the latter the inferred distribution-in-energy of the impinging neutrons: the discrepancy below 6 MEV is due to the allowance for the variation of cross-section.

The detecting-device of Bonner and Brubaker was an expansion-chamber filled with compressed methane; owing to the nature of the gas there was a notable number (four) of protons per molecule to receive the neutron-impacts, and owing to its high density the range of even rapid protons was relatively low. With a gas-pressure of 14.7 atmospheres (at the end of the expansion!) a proton of which the range in normal air would be as high as 126 cm had its track completely contained within the chamber; actually there were sometimes protons of 190-cm range, which had to be slowed down by screens interposed in their way. The neutron-sources were targets of various elements under bombardment by deuterons. In the experiments on the neutrons proceeding from lithium targets—or in another phrasing, neutrons from Li(d, n) reactions—19,600 stereoscopic pairs of photographs were made and 1550 proton-tracks were taken into account in forming the full curve of Figure 17.2.

More about the reaction $Li^7(d, 2\alpha)n$

The symbol $Li^7(d, 2\alpha)n$, though putting somewhat of a strain on the system of notation, should be recognized as referring to the reaction (15.3), of which at the end of Chapter XV I mentioned the newborn alpha-particles. The neutrons were first observed by Crane and his colleagues at Pasadena,[1] at about the same time as the alpha-particles were discovered in Cambridge. Of the spectrum exhibited in Figure 17.2, it is

[1] Crane, Lauritsen and Soltan, **PR 44**, 692 (1933); for further literature, see **PR 47**, 910 (1935).

believed that all of the curve except the hump at the right-hand end is due to neutrons born from this reaction. The average energy of the neutrons is 3.9 MEV; the average energy of the alpha-particles composing the continuum of Figure 16.11 is estimated (by Bonner and Brubaker from the Cavendish data) as 5.2 MEV. It can be said that these values are not incompatible with the Q-value (about 15 MEV) supplied by the rest-masses for the reaction; data justifying a more positive assertion would be welcome indeed. As for the hump, it is ascribed to neutrons springing from the reaction $Li^7(d, n)Be^8$. Here we meet again the nucleus Be^8, this reaction affording a second way of estimating its mass; the estimate agrees with the first one (page 533).

The (α, n) reactions

In their observations on the supposed gamma-rays evoked by impact of alpha-particles, Bothe and Becker were unwittingly discovering the (α, n) type of transmutation. It has already been introduced into this book (Chapter XV) as an alternative to another type: when alpha-particles are projected against aluminium (an element conveniently consisting of but a single isotope, so far as mass-analysis has yet been able to find) neutrons as well as protons are observed to spring out, and there is good reason for believing that the protons are ejected from some and the neutrons from other nuclei. There are other types of nucleus for which these alternative reactions are observed, and yet others which so far as our knowledge yet goes are subject only to (α, p) or only to (α, n) reactions when alpha-particles bombard them. Indeed, failure to observe the first of these reactions has been known to stimulate a successful search for the second.[1]

About a score of examples of the (α, n) reaction, distributed among some fourteen elements ranging up to 33 in atomic number, have now been recognized. The neutrons themselves have been detected in most of these cases, but in some the

[1] Pollard *et al.*, PR 51, 140 (1937).

reaction is inferred from a radioactive residue having the chemical properties befitting an element two units higher in atomic number than the element bombarded. Beryllium over all, then boron and fluorine, remain the outstanding examples of elements having isotopes subject to this form of transmutation.

With beryllium Dunning obtained a famous curve,[1] an integral-distribution-in-range curve for the protons projected more or less nearly straight ahead from a sheet of paraffin on

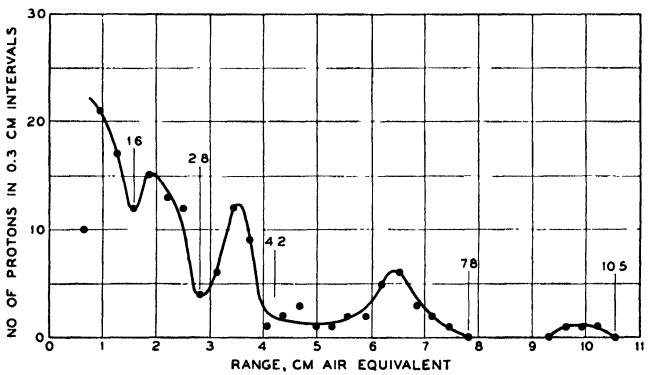


FIGURE 17.3—DISTRIBUTION-IN-RANGE OF PROTONS RECOILING FROM NEUTRONS BORN FROM THE REACTION OF TRANSMUTATION $F^{19}(\alpha, n)$; INDICATING NEUTRON-GROUPS AND EXCITED STATES OF Na^{22} NUCLEUS. (Bonner and Mott-Smith; *Physical Review.*)

which the neutrons were impinging. This curve exhibits plateaus connected by ascending arcs, betokening a multitude of groups among the protons and hence a corresponding multitude of groups among the neutrons. Since it was complicated by the fact that alpha-particles of three speeds were bombarding the beryllium at once, I show here in its stead the differential distribution-in-range curve (Figure 17.3) obtained by Bonner and Mott-Smith when they bombarded fluorine with the mono-kinetic alpha-particles from polonium. This curve

[1] PR 45, 586 (1934).

refers to the protons projected out of methane molecules in the expansion-chamber of which I spoke in an earlier passage. For the reason there mentioned we must not assume that the true curve for the neutrons is identical; but it is excessively unlikely that such humps as here appear (or the features of Dunning's curve) should be due merely to the variability of the cross-section for neutron-proton scattering.

For beryllium and for fluorine, then, there are several neutron-groups; and these being elements of one known isotope only, there is little hesitation in assuming that the reactions are $Be^{9}(\alpha, n)C^{12}$ and $F^{19}(\alpha, n)Na^{22}$ respectively and that the residual nuclei C^{12} and Na^{22} are liable to be left in any one of several different states. May we identify one of these as the normal state? In other words, does the Q-value computed from the kinetic energy of the fastest neutrons agree with that deduced from the rest-masses of the four nuclei involved in the reaction? There is no use in asking about Na^{22}, for this nucleus has never been observed in mass-analysis with the spectrograph, probably because it is unstable. With C^{12} this difficulty is absent, but there are two others: which value to choose for the neutron-mass, and which value to adopt for the kinetic energy of the alpha-particles? (I refer to the latter in a moment). Dunning from his data, and Bernardini and Bocciarelli from their more recent studies of the neutrons evoked from Be by the alpha-particles from polonium, conclude that C^{12} *is* formed in its normal state when the fastest neutrons are emitted; the several other states (three at least), formed when neutrons of other groups escape, are then to be classified as "excited." With boron, the neutrons from which also consist of a multitude of groups, there arises the further question as to which isotope is responsible for which groups. The occurrence of the reaction $B^{10}(\alpha, n)N^{13}$ is confirmed by detection of the radioactive residue N^{13}, but there is good reason for believing that some other reaction—probably $B^{11}(\alpha, n)N^{14}$—occurs with a much higher yield, since (according to observations by Bonner and Mott-Smith) the number of neutrons emitted from bombarded boron is ten or twenty times as great

as the number of newborn N^{13} nuclei revealed by their radioactivity.

Thus we have seen that the existence of groups—that is to say, the definiteness and multiplicity of Q-values—is a quality common to (α, p) and to (α, n) reactions. Another quality common to both is "resonance" in the sense of Chapter XV. Indeed the most vivid available proof of resonance is supplied by curves relating to (α, n) reactions, giving neutron-yield from *thin* targets as function of E the energy of the alpha-particles, and showing remarkable peaks: such is the dashed curve of

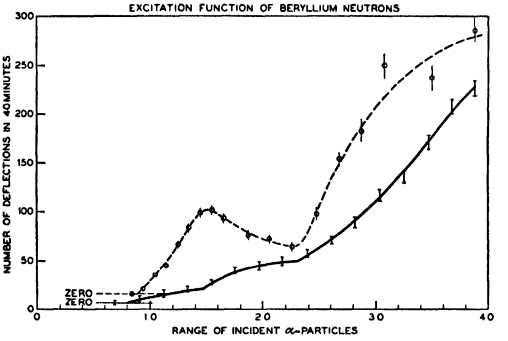


FIGURE 17.4—THIN-TARGET YIELD-CURVE (DASHED CURVE) AND THICK-TARGET YIELD-CURVE (SOLID CURVE) OF THE REACTION OR REACTIONS OF TRANSMUTATION Be (α, n). (Chadwick.)

Figure 17.4 (below which is a curve obtained with a *thick* target). This is the manner of proof of resonance labelled with (*c*) on page 497; the manner there labelled with (*d*) is exemplified by many curves which display such unmistakable peaks despite the fact that they were obtained with *thick* targets subjected to the alpha-particle bombardment. But when resonance is detected in these manners only, and there is more than one group, one does not know which E-values to correlate with which groups. This difficulty, with the consequent doubt as to Q-values, was mentioned on page 498 and has just recurred in the paragraph preceding this one.

I recall from page 503 that, according to the most recent

opinion, the resonance-levels of the reaction $Al^{27}(\alpha, p)Si^{30}$ are not coincident with those of the alternative reaction $Al^{27}(\alpha, n)P^{30}$—a fact full of import for nuclear theory.

The (ϕ, n) reactions and the mass of the neutron[1]

Now we consider a wonderful type of transmutation differing markedly from all the rest: there is but one initial nucleus, and the agent of the transmutation is a *photon* which employs its energy in detaching a neutron from that nucleus, vanishing altogether as it does so.

The discoverers of this process were Chadwick and Goldhaber in Cambridge, who exposed deuterium—in the form of gaseous heavy hydrogen filling their ionization-chamber—to gamma-rays of photon-energy 2.62 MEV (belonging to the gamma-ray spectrum of thorium C″) and observed the ionization due to individual protons emerging from the reaction,

$$\phi + {}_1H^2 \rightarrow {}_1H^1 + {}_0n^1. \qquad (17.1)$$

Shortly afterward they detected the emergent neutrons—not, however, by any method yet mentioned in this book, for these neutrons are too slow-moving; the appropriate methods will be described later on (page 563). The next-to-be-discovered reaction of this kind was found by Szilard and Chalmers, who detected the neutrons springing from the reaction,

$$\phi + {}_4Be^9 \rightarrow {}_4Be^8 + {}_0n^1. \qquad (17.2)$$

These two seem to be the only reactions which 2.62 MEV photons are energetic enough to produce. (The neutrons arising from them are known, by the way, as "photo-neutrons"). When the much more energetic photons engendered by the

[1] For reaction (17.1) see Chadwick and Goldhaber, *Nature* **134**, 237 (1934) and **PRS 151**, 479 (1935); Feather, *Nature* **136**, 468 (1935); Chadwick, Feather and Bretscher, **PRS 163**, 366 (1937); Richardson and Emo, **PR 53**, 234 (1938); Stetter and Jentzsche, **ZP 110**, 214 (1938); D. P. Mitchell, *et al.*, **PR 50**, 189 (1936); Bethe, **PR 53**, 313 (1938). For reaction (17.2) see Szilard and Chalmers, *Nature* **134**, 494 (1934), Chadwick and Goldhaber, **PRS** *l.c.*; Mitchell, *l.c.* For others see Bothe and Gentner, **NW 25**, *passim*; Chang, *et al.*, *Nature* **139**, 962 (1937).

impact of protons on lithium (page 630) were discovered, Bothe and Gentner at Heidelberg exposed a great many elements, sought for radioactive residues in the targets after the exposures, and found more than a dozen; these indicate as many reactions, of which it is almost certain (for the reason mentioned on page 590) that some and plausible that the rest are of the (ϕ, n) type. Chang, Goldhaber and Sagane added to the list a reaction believed to be $O^{16}(\varphi, n)O^{15}$. For the present these are interesting chiefly as showing how widespread is the liability of nuclei to being divested of neutrons by incident photons, when these last have energy enough.

The reaction (17.1) is of especial importance in that it affords what is the simplest and seemingly ought to be the exactest way of determining the neutron-mass. I rewrite (17.1) in the manner shown in (15.3), the quasi-chemical symbols standing for rest-masses, $h\nu$ for the photon-energy, K_p and K_n for the kinetic energies of proton and neutron respectively:

$$H^2 + h\nu = H^1 + n^1 + K_p + K_n, \qquad (17.3)$$

and putting Q for $(K_p + K_n - h\nu)$,

$$n^1 = (H^2 - H^1) - Q. \qquad (17.4)$$

So small is the momentum of the photon and so nearly equal are the rest-masses of proton and neutron, that K_n must be practically equal to the kinetic energy K_p of the proton, which is therefore all that need be ascertained. So small is K_p itself that the range of these protons is too small to be measured by the ordinary scheme of interposing screens of various thicknesses and finding out which thickness just suffices to stop them, while their tracks in ordinary expansion-chambers are almost too small to measure. This, however, was no impediment to the ascertaining of K_p, for Chadwick and Goldhaber went more directly to the point at issue, determining the number of ion-pairs produced in the ionization-chamber by each proton [1] and multiplying by 33 eV. (So also did Stetter and Jentzschke.) Chadwick, Feather and Bretscher then suc-

[1] They compared the kicks of the oscillograph (method E_4) due to the protons with those due to alpha-particles which were producing a known amount of ionization.

ceeded in measuring tracks in an expansion-chamber filled with a deuterium-containing gas ("heavy methane"), and from his measurements computed, by use of a particular range-vs-energy relation, a value 0.18 for K_p. Richardson and Emo used a cloud-chamber filled with deuterium, and deviated from all the rest by irradiating it with gamma-rays contingent upon a reaction of transmutation (page 629) having a photon-energy which Richardson had earlier evaluated as 3.00 ± 0.05 MEV.

The corresponding Q-values with their uncertainties, as given by their authors, are -2.14 (± 0.16), -2.25 (± 0.05 and -2.18 (± 0.07) in MEV. Also the reaction has been found to occur, feebly but unmistakably, when the photon-energy of the irradiating gamma-rays is only 2.20 MEV; Chadwick and Goldhaber detected the protons, D. P. Mitchell and his colleagues at Columbia the neutrons. On the whole it seems premature to estimate the neutron-mass from this reaction only; an expert weighting of the conclusions from many reactions is needed. If -2.20 MEV is correct for the Q-value, and the mass-spectrograph values for the rest-masses of H^1 and H^2 are right, then the neutron is a very little more massive than the proton: 1.0089 against 1.0081.

In the process (17.2) we meet the nucleus Be^8, for the third time in this book. No more than in the previous cases has it actually been observed in this one; and very lately word arrives that Glückauf and Paneth in Oxford have observed the formation of helium in beryllium exposed to gamma-rays, and have not only observed but measured it, comparing the amount of helium with the number of neutrons—and so far as can be inferred from the wording of the first disclosure, they have found nearly twice as many He nuclei as neutrons.[1] So, here again it appears that Be^8 explodes almost as soon as it is born. The yield of this reaction has been measured only for a couple of photon-energies, but is less for the higher of these than it is for the lower.[2]

[1] *Nature* **141**, 259 (1938).

[2] Fleischmann and Gentner, **ZP 100**, 440 (1937); Rotblat, *Nature* **139**, 963 (1937).

The deuteron-deuteron reactions [1]

The deuteron-deuteron reactions—often called D–D reactions for short—are the following two:

$$H^2 + H^2 \rightarrow H^1 + H^3, \qquad (17.5)$$

$$H^2 + H^2 \rightarrow n^1 + He^3. \qquad (17.6)$$

They are outstanding for several reasons: as being examples of alternative reactions—as being the only reactions yet known in which the initial two particles are identical—as having great yields at low bombarding energies, so that (17.6) in particular supplies more neutrons than any other available to anyone not having particles of more than 0.7 MEV at his behest—and as leading to the nuclei H^3 and He^3 which are presumably the simplest after the proton and the deuteron, and therefore are most welcome to the theorists.

The reaction (17.5) presented itself as soon as deuterons were first projected against targets in the cyclotron at Berkeley; for its newborn protons were observed springing forth with a characteristic range, whatever the target material. The first idea was that the deuterons were being broken in twain by their impacts against the heavier nuclei which they struck. Even in retrospect it seems a very plausible idea: yet this imagined process has never been proved to occur, and its apparent absence from the scene of Nature shows us vividly how far wrong we may go, if we take too literally the picture of a nucleus as a group of particles held together by links which are fracturable if struck with a blow sufficiently hard. It is now agreed that the first deuterons to arrive had settled down upon the targets, where presently they were struck by the later comers and the reaction (17.5) commenced. Paired particles from this reaction produce paired tracks such as are seen in Fig. 17.5.

This reaction being strongly exothermic, and the positive charges on the nuclei being as small as ever can be the case

[1] Oliphant, Harteck and Rutherford, **PRS 144**, 692 (1934); Dee and Gilbert, **PRS 149**, 200 (1935); Bonner and Brubaker, **PR 49**, 19 (1936); **149**, 406 (1934).

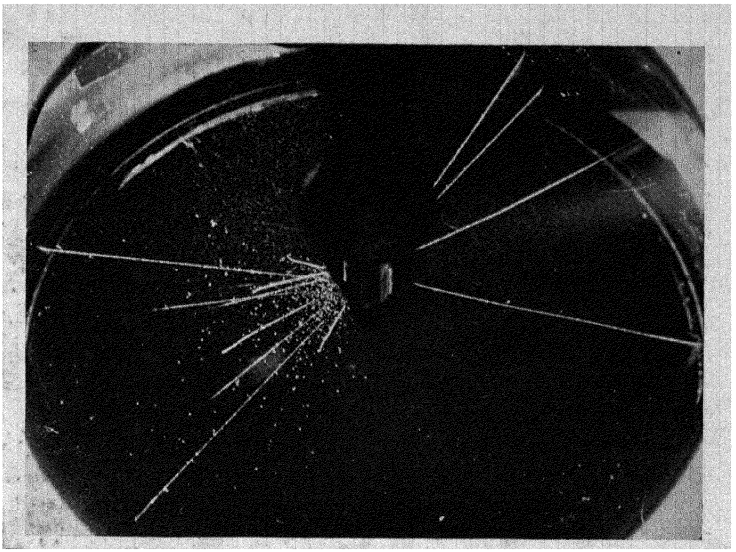


FIGURE 17.5—THE D–D REACTION RESULTING IN PAIRED H^1 AND H^3 NUCLEI. (Dee.)

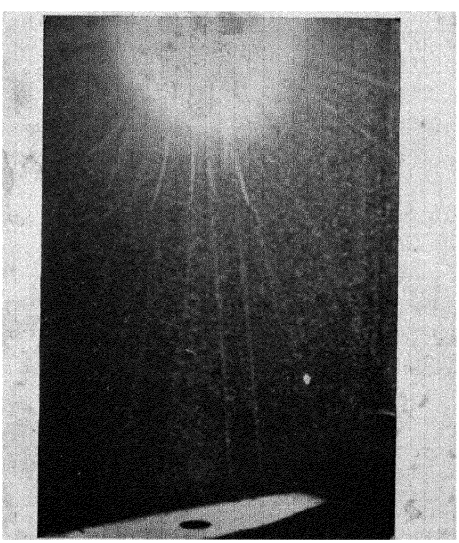


FIGURE 17.6—THE D–D REACTIONS: LONG TRACKS OF H^1 AND H^3 NUCLEI RESULTING FROM ONE, SHORT TRACKS OF He^3 NUCLEI RESULTING FROM THE OTHER. (Dee and Gilbert; *Proceedings of the Royal Society.*)

except with neutrons, one should not be surprised to find it detectably beginning at a low E-value; nevertheless it *is* surprising to learn that the protons have been detected (by Burhop) with as few as 8000 volts applied across the tube whence the impinging deuterons come! He estimated the yield at 10,000 volts as one per 10^{12} (it increases about a millionfold when the voltage is decupled). The paired tracks of the H^1 and H^3 nuclei, moving in almost exactly opposite directions, have been shown in beautiful photographs by Dee. The Q-value, based on measurements of the ranges of the emerging protons by Oliphant, Kempton and Rutherford, is extolled by Livingston and Bethe as the most accurately measured of all; it is 3.98 ± 0.02 MEV.

The alternative reaction (17.6) is also exothermic—not quite so much so, the Q-value being 3.18. The neutrons were discovered earlier than the charged nuclei which are paired with them—a rather odd fact, explicable because the range of the latter is low. To make visible the tracks of these He^3 nuclei, it was necessary for Dee and Gilbert to fill the expansion-chamber with gaseous deuterium and admit the bombarding deuterons into the chamber itself, so that not even the thinnest possible tube-wall should stop the newborn nuclei or detract from their energy before their entry into the expansible gas. The flare at the top of Figure 17.6 is the region within the chamber which the bombarding deuterons invade; the long tracks crossing the picture completely are those of the H^1 and H^3 nuclei born from reaction (17.5); the short tracks ending in the upper part of the picture are those of the He^3 nuclei, not discovered before the taking of the pictures of which this is one.

There appears to be only the one Q-value, for by the testimony of such pictures all of the He^3 nuclei are of a single group. The neutrons also are of a single group, the most vivid evidence for this statement being the form of their distribution-in-energy curve as obtained by Bonner and Brubaker; it has a single sharp peak, instead of the several shown by the corresponding curve (Figure 17.3) for the neutrons springing from a different

reaction. Of course the neutrons do not all have precisely the same energy, there being a spread due to the dependence of energy on direction (explained on page 492 for another type of transmutation) and on deuteron-speed (which enters in because of the unavoidable thickness of the target). Nevertheless, this reaction comes nearer to the ideal of supplying neutrons of uniform energy than does any other which is prolific, the others being defective from this viewpoint either because of continua as is (16.3), or because of multiplicity of groups as are the principal (α, n) reactions. The "D–D neutrons" emanating from a target in a direction inclined at about 90° to a beam of 0.5–MEV bombarding deuterons have energy of about 2.6 MEV.

There are interesting questions about the nuclei H^3 and He^3: do they exist already in Nature, or only when physicists have created them by transmutation? If the first alternative is not provable, can they be stable? Almost the last publication of Rutherford himself [1] was devoted to these questions, on the former of which a really vast amount of laborious research has been expended with sad results. The search for H^3 has been carried on almost ever since H^2 was discovered, and by the same procedure: treatment of huge quantities of water in such ways as to enhance the concentration of this heaviest of hydrogen isotopes—if it exists—to a high degree in the last tiny residue. Rutherford recounts that a Norwegian company electrolyzed 13,000 tons of water down to a residue of 11 cc. The test was made by introducing ionized vapor of this residual water into Aston's mass-spectrograph, and looking for appropriate ions. The choice of appropriate ions is not so easy as it may sound: an ion of mass 3 might be an ionized molecule of $H^1H^1H^1$ or H^1H^2, an ion of mass 6 might be any one of several things; even an ion of mass 5 might be H^2H^3 or $H^2H^2H^1$, but the resolving-power of Aston's apparatus was such that on his plates he could have distinguished between these two, and only the line due to the latter was visible! Seemingly positive results reported from other places are

[1] *Nature* **140**, 303 (1937); see also R. Sherr, *et al.*, **PR 54**, 388 (1938).

beclouded by the possibility that the ions observed might have been of the wrong character even though of the right mass-number to imply H^3 or He^3 or compounds thereof. Even with samples of gas in which H^3 and He^3 had been produced by the D–D reactions a few seconds or minutes before the mass analysis, there was no unassailable sign that they were still to be found.

Mass-analysis failing, the mass-values of H^3 and He^3 must be deduced from transmutations, preferably from reactions in which the rest-masses of all the other nuclei involved and the Q-value of the reaction have all been well measured. Reaction (17.5) is admirable from this viewpoint, and the mass of H^3 may be considered well-known. Reaction (16.6) is not so good because it involves the free neutron, as to the rest-mass of which there is the perplexity already mentioned; but for He^3 there is an additional check from the reaction $Li^6(p, \alpha)He^3$ earlier mentioned. It turns out that H^3 and He^3 have masses which are not distinguishably different; their common value (according to Livingston and Bethe) is 3.0171, with an uncertainty less than 2 units in the last place given. This value being definitely less than the mass of the deuteron augmented by that of either neutron or proton, neither H^3 nor He^3 should be of itself unstable; nor is there any evidence that they are. Physicists are left in the position of having evidence for neither the existence of these nuclei after the moment of their creation, nor the cessation of their existence thereafter.

The (d, n) reactions

Of the (d, n) type of transmutation we have already met with instances: that one of the deuteron-deuteron reactions which results in He^3 nuclei and neutrons of a single group, and the process $Li^7(d, n)Be^8$ which also gives neutrons of one group. The reaction (16.3) may also be included without too great a strain on the definition, and so may certain reactions in which (it is believed) a neutron and an alpha-particle are both evicted from a platinum nucleus by an entering deuteron. Whether

or not we count these in, the type is abundant even though there are but seven elements from which neutrons have actually been observed to spring under deuteron bombardment. The neutrons from these seven are classified as coming from eleven different reactions; and in addition there are more than a dozen cases of radioactive residues, created by deuteron bombardment and having the chemical properties appropriate to elements one unit higher in atomic number than those from which they are respectively made. Excepting the very special case of platinum mentioned above, the highest element yet proved susceptible to this sort of reaction is tin ($Z = 50$).

The studies of Bonner and Brubaker on the distribution-in-energy of neutrons were extended to those resulting from deuteron-bombardment of Li, Be, B and C. The case of beryllium is the prettiest, with only one isotope to consider, and the reaction $Be^9(d, n)B^{10}$ resulting in four groups of neutrons of which the fastest has just the right Q-value to agree with the rest-masses: in other words, there is admirable evidence for the formation of the B^{10} nucleus in either its normal or any one of three excited states. In the case of carbon there are two groups, believed to be evoked one from the preponderant isotope C^{12}, the other from its hundredfold less abundant comrade C^{13}. The former is found to have a negative Q-value (about -0.3 MEV) so the reaction is endothermic and therefore must not commence until the energy of the bombarding deuterons is raised past a threshold-value; Cockcroft and Lewis in fact found it setting in suddenly at about the right point. The case of boron is confused by uncertainty as to how to assign the features of the neutron-spectrum (a complicated one) between the two isotopes; this uncertainty has been happily lifted for lithium, of the case of which I have already spoken.

Other reactions resulting in free neutrons

The caption of this section is rather misleading, since at this writing nobody seems as yet to have detected any neutrons born out of any reactions excepting those of the types already

listed—(α, n), (d, n) and (ϕ, n). Yet when targets of various elements are bombarded by very energetic protons or by very energetic neutrons, radioactive residues are observed in a remarkable number of cases, scores upon scores.[1] For many of those engendered by protons, it has been shown that they share the chemical properties of the element one unit higher in atomic number than the one bombarded; for many of those engendered by neutrons, it has been shown that they retain the properties of the element whence they were made. These facts can be interpreted by assuming (p, n) reactions and $(n, 2n)$ reactions respectively: the latter symbol is a consistent but odd-looking way of expressing the idea that a neutron impinges on a nucleus and from it expels another neutron without itself being captured.

But *must* the facts be so interpreted? There are other possibilities which must be rigorously excluded: *e.g.* the "capture" or absorption of the incident proton or neutron into the nucleus, not attended by the emergence of any particle (unless a corpuscle of light). Sometimes a particular radioactive residue can be produced in a variety of ways, as I shall later be stressing (Chapter XVIII); then its mass-number and its atomic number are practically certain, and if it can be produced by proton-impact or neutron-impact against a nucleus of another known kind there is no doubt about the process. There is a strongly endothermic reaction which occurs when high-energy protons impinge upon oxygen; by themselves the observations upon it justify only the statement that it results in a radioactive isotope of fluorine, but comparison with other reactions proves it to be $O^{18}(p, n)F^{18}$. Some of the residues created by Bothe and Gentner with photon-bombardment of various elements (page 550) were later created by Chang, Goldhaber and Sagane with bombardment of the same elements by fast neutrons: if the

[1]All the evidence for (p, n) reactions comes from the Rochester school (Du Bridge *et al.*, **PR 51**, 775, 995 and later); they had 3.6 MEV protons available from their cyclotron. The idea of a $(n, 2n)$ process was invented by Heyn, *Nature* **138**, 723; **139**, 742; he observed many radioactive residues produced by fast neutrons, and a great many more were found by Pool *et al.* (**PR 52**, 239; 1937). Idea of $(n, 3n)$ reaction, see Pool *et al.* (**PR 52**, 41; 1937).

reactions are interpreted as (ϕ, n) and (n, $2n$) respectively, each interpretation strengthens the other. Some residues can be created by impact of fast neutrons but not by impact of slow ones, a quality which suggests the (n, $2n$) process rather than neutron-capture. Many cases are still doubtful, and I refrain even from enumerating the reactions now classed as (p, n) or (n, $2n$), except to say that there are dozens of each.

Now we turn from the reactions which release neutrons from nuclei, to those in which neutrons are absorbed or captured into nuclei.

The (*d*, *p*) *reactions*: *capture of neutrons out of deuterons*

A (*d*, *p*) reaction is one in which the neutron belonging to an impinging deuteron is annexed and detained by the nucleus which receives the impact. A proton goes free: whether it is the same proton as belonged to the deuteron is perhaps a meaningless question, and yet there is ground for supposing that in some of these reactions the deuteron breaks up as it approaches the potential-barrier of the "struck" nucleus, the neutron sliding on through while the proton is rejected before it even enters the potential-well (the Oppenheimer-Phillips theory). Anyone feeling these distinctions to be too subtle may confine his imaginings to what the symbol (*d*, *p*) expresses.

The (*d*, *p*) type appears to be the most abundant of all, no fewer than fifty cases being catalogued at this writing. As with the (*d*, *n*) type, most of these reactions are recognized from radioactive residues and not from energent protons. Protons have been observed emerging from nine bombarded elements, and are attributed to eleven reactions, of which the right to first place goes to $H^2(d, p)H^3$ which we have already studied. There are three elements (F, Na, Al) for which it can be said that both emergent protons and radioactive residues have been detected; as all three are elements of one known isotope apiece, there is little doubt that in each case the two nuclei resulting from one process have been observed. (More quantitative proofs or disproofs, such as might be obtained from comparisons of yield-

curves for protons and for residues, or from energy-relations, are wanting as yet.) Bismuth is the highest element for which the process is known to occur.

The (d, p) process entails what may be called a "promotion" of the struck nucleus into the next-following isotope of the element in question (by next-following I mean the isotope of which the mass-number is greater by unity). Of the first eight elements of the Periodic Table there are five which have at least two stable isotopes of which one is abundant while the other is next following to it. One might, therefore, reasonably hope to detect the reactions $Li^6(d, p)Li^7$, $B^{10}(d, p)B^{11}$, $C^{12}(d, p)C^{13}$, $N^{14}(d, p)N^{15}$ and $O^{16}(d, p)O^{17}$. Deuteron-bombardment does in fact elicit groups of protons from each of these elements, and for each of the first four there is a group having the correct Q-value to confirm the interpretation. For nitrogen this Q-value was determined (both at the Cavendish and at Berkeley) before the mass of N^{15} had been measured with the mass-spectrograph; on this measurement being made by Jordan and Bainbridge at Harvard, the agreement turned out to be excellent.[1] For oxygen the mass-spectrograph has not yet supplied the mass-value of O^{17} (I suppose because this isotope is so scanty), which, accordingly, is estimated by assuming the observed Q-value to correspond to the formation of O^{17} nuclei in their normal state. For all of these five elements but carbon there are other groups in addition to the one of greatest energy, and these are taken as indications of excited states of the residues Li^7, B^{11}, N^{15} and O^{17}.

This process of promotion may and often does occur when the next following isotope is not stable: I have already implied as much in telling of the numerous examples which result in radioactive residues. Li^7 and B^{11} can be promoted in this manner. Other striking instances are furnished by elements having one stable isotope only, such as Be and F and Al. The reaction $Al^{27}(d, p)Al^{28}$ exhibits no fewer than five groups: the highest in energy is assumed to correspond to the formation of

[1] Lawrence *et al.*, **PR 47**, 273 (1935); Cockcroft and Lewis, **PRS 154**, 261 (1936); Jordan and Bainbridge, **PR 50**, 98 (1936).

Al^{28} in its normal state (and the Q-value is used in computing the presumptive rest-mass of Al^{28}) while the others show that an unstable nucleus can vie with stable ones in the multitude of its excited states. A remarkable reaction is $Bi^{209}(d, p)Bi^{210}$, for the nucleus here designated as Bi^{210} is the very one which has been known for more than thirty years under the name of "radium E." It is one of the radioactive substances occurring in Nature as a member of the radium series (page 583) and when Livingood at Berkeley produced the reaction he was achieving the first artificial production of a natural radioactive body. I trench on the subject of Chapter XVIII so far as to say that Bi^{210} or RaE emits a negative electron, which leaves behind a residue (polonium, atomic number 82, mass-number 210) which is itself unstable and in due time—usually not for weeks—emits an alpha-particle. Livingood [1] observed not only the electrons but also the alpha-particles, these last becoming more numerous as time went on after the ending of the exposure (which was to a beam of 5.4 MEV deuterons amounting altogether to 13 microampere-hours).

Reactions involving capture of free neutrons

The reactions subsumed beneath this heading are of three varieties. Two are described as well as denoted by the symbols (n, α) and (n, p). The third is "pure" neutron-capture, not attended by the ejection of any material particle. Energy must, however, be released in any such reaction, and as it is released (in some cases provably, in others supposedly) in the form of photons, the third type is often denoted by (n, γ) or (n, ϕ).

Neutrons being chargeless, they are exempt from the great known hindrance which makes difficult the approach of charged particles to a nucleus. The potential-curve of Figure 15.6 has no bearing on them, and the potential-hill does not impede their entry into that central region of the nucleus which (with incoming *charged* particles in mind) nuclear physicists call the

[1] **PR 50**, 432 (1936).

potential-well. Exception being made for unknown hindrances, exothermic reactions may be expected to occur, however low the energy of the bombarding neutrons, and endothermic reactions to begin at the threshold-value of E imposed by the negative Q-value (page 490). It is, therefore, not too surprising to find that almost all of the reactions produced by free neutrons, apart from those known to be endothermic and those classified [1] as $(n, 2n)$, can happen when the neutrons have small kinetic energy (no greater perhaps than that of ordinary atoms in thermal agitation at room-temperature) and the word "bombardment" becomes an extravagant metaphor.

The earliest-to-be-discovered of these reactions were of the (n, α) type. Fast neutrons were sent into expansion-chambers, and occasionally at an expansion there appeared two tracks forming a V or a fork-without-a-handle, being the tracks of the two newborn nuclei (alpha-particle and residue) springing from such a reaction involving the nucleus of an atom of the gas. The directions of the two visible tracks are ascertainable, and from their lengths the observer may derive the energy-values of the newborn nuclei (if he has identified them correctly, and knows an adequate range-*vs*-energy relation; errors have arisen through confusing the $N^{14}(n, \alpha)B^{11}$ reaction with the $N^{14}(n, p)C^{14}$ reaction). If in addition the direction of the invisible path of the impinging neutron has been correctly guessed, there is knowledge enough to fix the Q-value and incidentally the neutron-energy. One may think that there should be no "if" about it, for surely the observer knows where the neutron-source is and can draw a line from it to the angle of the fork which is the scene of the impact! but the aptitude of neutrons for getting deflected is so great, that many such an impact is made by a particle flying from a quite unanticipated and unguessable direction. Some very odd results have been obtained, perhaps because of this defect or perhaps because the accuracy of the measurements on the visible tracks is not so great as was hoped.

[1] So classified, in many cases, precisely because they require high E-values of the neutrons (page 559).

The best-studied of all the (n, α) reactions occur with lithium and boron as the targets. With lithium there happens a process which undoubtedly is $Li^6(n, \alpha)H^3$; for the doubly charged alpha-particles and the singly charged H^3 nuclei can be distinguished (by ionization-chamber and external amplifier, method E_4 of page 519) and their ranges separately measured, and the Q-value agrees well with what is calculated from the masses. Taylor and Goldhaber used the technique (page 517) of mixing lithium into photographic emulsions, bombarding these and discerning the tracks of the newborn nuclei by virtue of developable grains along them. The tracks of the He^4 and H^3 nuclei born from a single encounter form not a V but a straight line, for this reaction is effected by slow neutrons, of momentum so small that the two newborn nuclei must spring off in directions practically opposite to one another. When boron is substituted for lithium in the emulsion, some of the tracks are isolated straight lines and some spring in threes from a common point. Those of the first sort are ascribed to $B^{10}(n, \alpha)Li^7$, the alpha-particle and the residue springing in opposite directions; those of the second sort are ascribed to an alternative reaction,

$$B^{10} + n \rightarrow He^4 + He^4 + H^3. \qquad (17.7)$$

Both reactions are confirmed by other methods, but all is not yet clear; the Q-value found for the former is too small, and one is obliged to postulate that $B^{10}(n, \alpha)Li^7$ always ends in an excited state of the Li^7 nucleus. Apart from their interpretations, these reactions of lithium and boron are of great value to nuclear physicists; for if an ionization-chamber is outfitted with either elements—being (say) walled inwardly with lithium, or filled with the gaseous compound boron trifluoride—it becomes one of the most sensitive devices available for detecting slow neutrons. Of the remaining (n, α) reactions I will mention only $C^{12}(n, \alpha)Be^9$ and $N^{14}(n, \alpha)B^{11}$, because they are reverse (in the sense of page 523) to two which have already been mentioned. The evidence for these is by no

means so neat as one would like to have it; but very probably here are two instances in which physicists are really able to do what with (α, p) and (p, α) reactions they have thus far failed to do—*viz.*, to produce both of two reactions reverse to one another.

It is from radioactive residues that most of the reactions (of all three kinds) involving capture of free neutrons are identified; for this was the method employed by Fermi and his colleagues at Rome when, in their first splendid onrush of discovery in 1934, they projected neutrons against a variety of targets and found in a few months more numerous transmutations than had been observed in all the other laboratories in all the years before. (Since the names of the principal members of Fermi's group occur continually and in groups of as many as five, I list them all at once: Fermi, Amaldi, d'Agostino, Pontecorvo, Rasetti, Segrè.) The Italians had an enviable opportunity, and they seized on it with promptitude and diligence not unworthy of its greatness. Barely three years had elapsed since the discovery of the neutron, and not three months since the demonstration in Paris that radioactive residues may be formed by transmutation and their atomic numbers may be learned. Before six months had passed they had applied this new agent of transmutation (used but little as yet, for expansion-chamber photographs of reactions due to neutrons had been few) and this new technique of detection to sixty elements, with success in forty cases out of the sixty! No comparable tempo had ever been known in nuclear physics, nor is it likely that such another will come.

Of the reactions then discovered, the chemical behavior of the residues suggested that some were of the already-known (n, α) type and some of the (n, p) type. Of others the residues were of the same chemical nature as the target elements from which they had been made, and these reactions were soon interpreted as cases of pure neutron-capture. Presently it was found that they are produced in greater abundance when the neutrons have been required to pass through water or paraffin on their way from source to target. This second discovery of

Fermi's school almost vies with the first in importance; for what they then discovered was an effect of *slow* neutrons, the bombarding particles having been slowed down by their many impacts with hydrogen nuclei as they made their way through the paraffin or the water. It strengthened the theory that the transmutations in question are cases of pure neutron-capture, and it pointed the way to the detection of many more examples of the type; and it opened up to physicists the whole subject of the properties of slow neutrons, a remarkable subject which will probably not be exhausted for many years to come.

The known reactions belonging to the (n, ϕ) class are now close upon a hundred, making this by far the richest class of all. All are known through radioactive residues, the ionization due to the recoiling of the newborn nuclei not having been detected. Perhaps this is not surprising, there being ample reason to believe that the transmutations often occur when the about-to-be-captured neutrons are moving so very slowly that the resulting nuclei have not enough kinetic energy to form even a single ion. If rapidly moving neutrons are really never thus captured, then the (n, ϕ) reactions produced by beams of fast neutrons must be due entirely to those particles which happen to get slowed down in the source or the target. The limitation is a serious one, since it debars the physicist from detecting such (n, ϕ) transmutations as result in stable isotopes, like some of the most interesting of the (d, p) reactions which we have lately considered. The prospect though is not entirely hopeless, as the photons might be detected, or the existence of an abnormally high "cross-section for interception" might be taken (and indeed has already been taken) as a sign of a (n, ϕ) reaction of high likelihood.

With this reference to interception, I pass over to other attributes, or to other ways of studying the attributes, of free neutrons; remarking in passing, that (n, ϕ) reactions may be used for detecting slow neutrons, a service which they fulfil less efficiently but sometimes much more conveniently than ionization-chambers containing boron or lithium.

Interception and scattering of fast neutrons

If a sheet of matter is put between a source and a detector of neutrons, the reading of the latter usually changes, and not necessarily in the downward sense. An increase in the reading does not, however, signify that the sheet has augmented the number of neutrons reaching the detector, nor does a decrease usually imply an equal percentage of decrease in that number. Detectors report only a fraction of the number of the neutrons which arrive at them, and this fraction depends very notably on the speed of the particles, so that if the interposed sheet of matter alters the speed it alters the reading even if it does nothing to the number attaining the detector. Allowance must be made for this variation of response with neutron-speed. This is in some cases so difficult (or impossible) that the course of the reasoning is inverted, and physicists try to deduce the law of this variation from the effect of interposed screens upon the reading.

The preceding is not the only difficulty met by those who try to design intelligible experiments of this kind. There are others, and all together they cause so much mystification that instead of describing at the start what is observed I will intimate what under ideal conditions might be expected, and then speak of some of the experiments which come nearest to showing it.

Let us imagine a beam of neutrons of uniform speed moving along parallel lines, and impinging on a thin sheet of matter of thickness t containing nuclei numbering N per-unit-volume. Let us first suppose that every neutron must either go through the sheet entirely unaffected, or else be captured. The fraction getting through will then be $exp(-N\sigma_c t)$, σ_c standing for a constant to be called "cross-section for capture of neutrons by the nucleus-type in question" or simply "capture cross-section" for short. The supposition is, however, too oversimplified to be of much (if any) use, as experience shows that neutrons are scattered in all directions from matter on which they impinge. Let us admit this scattering, and suppose that all of it is due to

elastic impacts. (The scattered neutrons will then be moving more slowly than the incident ones, and the question of the variation of detector-response with neutron-speed may arise, but probably not to a serious extent if the nuclei are more than ten or a dozen times as massive as the neutron). If t is large enough there can be confusion owing to neutrons making two or more impacts apiece before quitting the sheet, but usually t can be kept low enough to make it slight; we so assume it. Next, we have to consider the areas of beam, sheet and target. If only we could have a beam of infinitely small cross-section, a sheet just extending across the beam, and a detector extending across a cone of apical semi-angle θ with apex at the sheet, then the number of neutrons attaining the detector should vary thus with t:

$$I = I_0 exp[-N(\sigma_c + \sigma_s)t] = I_0 exp[-N\sigma_t t] \qquad (17.8)$$

Here σ_t, the quantity which could be measured by such an experiment, may be called "cross-section for interception." It is the sum of σ_c and σ_s, of which the latter may be called "cross-section for scattering," though it must depend on θ and a full description of it requires an allusion to θ.

To find how σ_t is divided between these two component terms, it is necessary to do other experiments or to have a plausible theory. If in some such way we could get the value of σ_s for any θ-value, and if we should then visualize neutrons and nuclei as elastic spheres of radii r_0 and r_m respectively, then we could calculate $\pi(r_0^2 + r_m^2)$. The value obtained for this last might turn out to depend on θ, in which case we should have been wrong to visualize in so naive a way. However, this is a subtler detail than we need worry about as yet, for the experiments fall so far short of the ideal just described that they give no clear testimony on this point.

The nearest approach to the ideal experiment is made with "fast" neutrons, a technical term signifying those which come straight from source to intercepting screen without being slowed down by passage through paraffin or water. Even with these one cannot come close enough to the ideal of a beam of particles

uniform in direction of motion, for neutron-beams cannot be focussed and can be collimated only with great difficulty (scattered neutrons being likely to fly all over the place); nor to that of a beam uniform in speed, most neutron-sources giving either continua or a plurality of groups (page 546; the particles from the D-D reaction (17.6) are the best). The values of σ_i exhibited in Figure 17.7 were obtained by Dunning.[1] The neutrons were highly diversified in speed (from 14.2 MEV

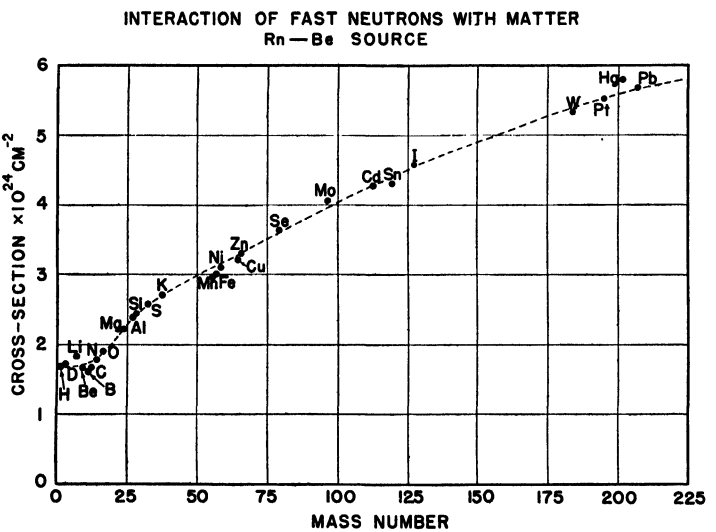


FIGURE 17.7—CROSS-SECTION FOR INTERCEPTION OF "FAST" NEUTRONS. DATA NOT ENTERED ON GRAPH: σ = 5.3, 5.8, 5.8 FOR Z = 79, 80, 82. (Data by Dunning.)

downward a long way, the average perhaps 4 MEV or lower) for they were derived from the $Be^9(\alpha, n)C^{12}$ reaction effected by alpha-particles of three groups (I mentioned the distribution-curve on page 546). The detector was an ionization-chamber containing air and having a sheet of paraffin against its left-hand wall. H nuclei in the paraffin, N and O nuclei in the air, were occasionally struck by some of the arriving neutrons and

1 PR 45, 586 (1934); 48, 274 (1935), and subsequent data kindly supplied for this figure by Dr. Dunning.

projected across the chamber; in it they produced spurts of ionization which were recorded, identified and counted as the lowest band of Figure 16.6 displayed to us in Chapter XVI, for this figure was made with this very apparatus. When t was varied the curve of detector-reading *vs.* t was not quite exponential. The deviations were ascribed to the departures of the experiment from the ideal aforesaid, and the values of σ_i computed by using what for neutrons are small values of t—actually 2 to 4 cm— and ignoring the deviations for these small thicknesses.

The curve of Figure 17.7 thus shows that for "fast" neutrons of a great diversity of speeds, σ_i is a smoothly-rising function of atomic number or nuclear charge. (No data are available for comparing different isotopes of one element.) The order of magnitude is that with which we are already familiar in nuclear physics.

Values of σ_i obtained with fast neutrons are usually considered to be values of σ_s, as though neutron-capture were negligible at such speeds (though this is dubious because of evidence that (n, ϕ) reactions occur when thin targets are bombarded by such neutrons). Accepting this, one may adopt the elastic-sphere model for the collisions resulting in scattering, and compute $\pi(r_0^2 + r^2{}_m)$. The process of computation and, accordingly, the result depend on whether one uses the classical or the quantal principles. Rabi and O. Clark, using the latter with Dunning's data, showed that these are roughly compatible with the picture of the neutron as an elastic sphere of constant radius and of nuclei as elastic spheres of radii proportional to the cube root of their mass-numbers M, the numerical values being:

$$r_0 = 1.2 \cdot 10^{-13}, \quad r_m = 3.8 \cdot 10^{-15} M^{1/3}. \qquad (17.9)$$

The phrase "elastic sphere" sounds a little less naive when translated into "spherical region of space surrounded by a narrow zone containing a strong repulsive field": as though for the neutron approaching a nucleus the potential-hill of Figure 15.6 were replaced by another (presumably non-electrical) hill

beginning much further in and rising with a much steeper slope.[1]

One expects to find in the literature a good deal of information about σ_s derived from observations on the scattered neutrons, but here is a surprising gap in the data of experimental physics. Dunning and Pegram made a few observations on the scattering from paraffin, water, carbon and lead, ancillary to their main experiments on interception.[2] There has indeed been a great deal of research on the distribution-in-angle of the protons which recoil from neutron-impacts, as they do in the expansion-chamber or the ionization-chamber when they serve as an index of incoming neutrons. This seems very relevant, as each recoiling proton indicates the direction and energy of a scattered neutron (if the impacts are elastic, which nobody doubts). But even if the total number of recoiling protons and, therefore, of scattered neutrons were measured, one could not determine σ_s because one does not know the number of neutrons in the incident beam.[3]

The problem of the distribution-in-angle of recoiling protons is therefore a different but an allied one. Its importance lies in its bearing on the elastic-sphere model (or other models) of nuclei and neutron. The distribution cannot be isotropic—not isotropic, that is to say, in the ordinary frame of reference, that in which the observer together with the protons awaiting the impacts in his expansion-chamber or ionization-chamber

[1] It could be an attractive field at the periphery of the potential-well, if one is willing to suppose that the well may be traversed by neutrons without obstruction.

[2] Dunning and Pegram, **PR 43**, 497 (1933); **45**, 599 (1934). One might also expect to measure σ_c by itself by surrounding the source with a sphere of the substance in question and surrounding the sphere with a detector on which all merely-scattered neutrons would land irrespective of their deflections. On trying this with detectors consisting of various metals subject to (n, ϕ) reactions, Collie and Griffiths (**PRS 150**, 434, 1936), found that with some of these the detector-response was actually increased by the presence of the sphere.

[3] This difficulty does not arise when σ_i is determined by using the same detector for measuring the neutron-beam with interposed sheets of matter of different thicknesses and with no interposed sheet at all (unless the presence of the intercepting matter alters the speed of the traversing neutrons and consequently the variation of detector-response with neutron-speed comes in). The problem of how to measure the number of neutrons in a beam is a serious one.

are practically at rest. Suppose, however, that the incident neutrons form a homogeneous beam, all having the same (vectorial) velocity $\mathbf{v}$ relative to the observer and his waiting protons. If the elastic-sphere model is adequate, the distribution-in-angle of the recoiling protons will be isotropic when translated into the frame of reference moving with velocity $\frac{1}{2}\mathbf{v}$ relative to the observer ("center-of-gravity frame," the allusion being to the center of gravity of any pair composed of impinging neutron and recoiling proton). If the incident neutrons do not all have the same velocity, knowledge of their distribution-in-velocity is needed for the test of the model, and the task is made much more difficult. The problem is an old one, as neutron-problems go; but the earlier researches were all impaired by this difficulty, and not till 1937 could physicists make use of the opportunity offered by the D-D reaction (17.6) with its plentiful supply of neutrons nearly uniform in speed. Several more investigations [1] have just been published, and three of them concur in confirming the elastic-sphere model by the test aforesaid, while one does not. The confirmation is to be expected, if the forces entering into play between nuclei and fast neutrons are confined to regions not much more extensive [2] than the elastic spheres specified by (17.9).

Interception and scattering of slow neutrons

The term "slow neutrons" has peculiar meanings in contemporary physics. It is applied primarily to neutrons which, after their birth from some reaction of transmutation, have passed through a notable thickness of water or paraffin before arriving at the detector. Fermi and his school observed that such neutrons enter into the (n, ϕ) or "pure neutron-capture"

[1] Bonner, **PR 52**, 685 (1937); Kruger *et al.*, 52, 678 (1937); Dee, **PRS 163**, 265 (1937); Lampson, Mueller and Barton, **PR 51**, 1021 (1937); earlier literature, see Bonner.

[2] Stehn and Share, **PR 52**, 48 (1937): "even for a simple potential-well of width $2 \cdot 10^{-13}$ cm, Bethe and Peierls showed that deviations from spherical symmetry [i.e., isotropy in the centre-of-gravity frame] would be negligible for neutrons with energies less than 20 Mev."

reactions much more abundantly than do those fresh from the source, a quality which suggests that they are relatively slow-moving. Now, water and paraffin both contain great numbers of protons; an elastic impact between a fast neutron and a nearly-stationary proton, if entailing a large deflection of the former, must also entail a large surrender of kinetic energy [2]; and, accordingly, the neutrons which enter a stratum of either substance (or any compound of hydrogen) are as liable to be slowed down as they are to be deflected. Indeed, if the water or paraffin is thick enough, the neutrons may come to constitute a "neutron-gas" pervading and in thermal equilibrium with the atoms of the solid body itself; a gas, however, of which the particles are not durable, since they are always liable to capture.

But how thick is "thick enough"? To answer this question properly, one would have to know the cross-sections σ_s and σ_c of hydrogen for neutrons of all speeds from the initial value characteristic of the source down practically to zero. As yet our knowledge of these cross-sections is very limited; but the value of σ_i for hydrogen for fast neutrons is great enough to indicate that when these last are shot into a block of paraffin several inches thick, a large proportion of them must emerge with energies not much or not at all greater than the thermal energy appropriate to the temperature of the block. Some, however, must escape before they are levelled down completely to this very modest energy-quota; and the attempts to detect, allow for, or obviate these "intermediate" neutrons have led to some of the most intricate experiments in neutron-physics or for that matter in the whole of physics. Leaving these aside for the moment, I say that such data as I next describe are believed to be germane chiefly to neutrons of thermal energy.

Interception of slow neutrons is measured like that of fast ones, by interposing sheets of matter between the detector and the source now shielded or enclosed by paraffin or water. To combat the difficulty of collimation physicists have designed

[2] If the elastic-sphere model is valid the average energy-loss per impact is $1/e$ times the energy at incidence (Wick, **PR 49**, 19, 1936).

what from their squatty shape are amusingly known as "howitzers," an example of which (from the Columbia school) is shown in Figure 17.8. In this howitzer the ultimate neutron-source, consisting of beryllium mixed with a natural radioactive substance which bombards it with alpha-particles and provokes the Be(α, n) reaction, is embedded in a block of paraffin. Into the block is scooped a hole extending to within 2 or 3 cm of the source. The only slow neutrons which can reach the detector are those which find their way into the channel, for the outer surface of the block is entirely covered with cadmium

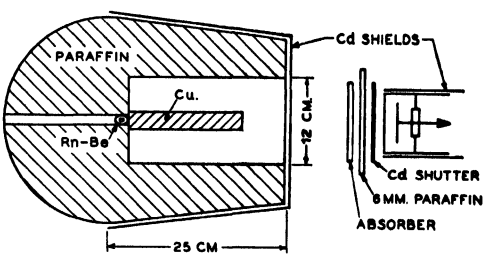


FIGURE 17.8—NEUTRON-HOWITZER DEVELOPED BY THE COLUMBIA SCHOOL. (After Fink.)

which they cannot traverse. A block of metal in the middle of the lumen of the channel stops some at least of the neutrons which by luck manage to fly from source toward detector without losing much of their energy in collisions. As to the detector, it must not depend for its functioning upon the ionization produced by nuclei recoiling from elastic impacts; these neutrons have not enough energy to cause appreciable recoils. The detector must depend upon the emission of alpha-particles resulting from the (n, α) reactions of lithium or boron (page 563), or else upon the emission of electrons from the radioactive residues resulting from (n, ϕ) reactions. For all of these the dependence of detector-response on neutron-speed is likely to be important.

The cross-sections σ_i for interception of slow neutrons were measured for no fewer than seventy-three elements in a single

investigation [1] by the Columbia school. (Since like the papers from Rome those from Columbia are often signed by a pleiad of names, I list the principal figures once for all: Dunning, Fink, Goldsmith, D. P. Mitchell, Pegram, and occasional visitors from Rome including Rasetti and Segre.) The neutrons, born out of the Be(α, n) reaction occurring in a capsule at the center of a solid sphere of paraffin of 6 cm radius, emerged through the surface of the sphere (there was no howitzer in these experiments) and travelled some 20 cm before impinging on the intercepting sheets of matter; just beyond these lay the detector, a lithium-lined ionization-chamber serving by virtue of the reaction $Li^6(n, \alpha)H^3$. The sheets or blocks of intercepting matter were sliced to thicknesses such that in all cases their presence reduced the reading of the detector by some 40%. Except for cadmium for which it was roughly tested, the exponential law (17.8) was taken for granted.

In the tabulation of the σ_i-values, the first astonishing feature is their diversity, and the second is the greatness of the greater among them. There is no systematic trend with mass-number or with atomic number, but quite the reverse! Thus for the three consecutive elements Li, Be and B, the σ_i-values are 45, 5 and 360 (all to be multiplied by 10^{-24}). One remembers that the reactions $Li^6(n, \alpha)$ and $B^{10}(n, \alpha)$ are of very high yield, and makes the guess that the first and the last of these σ_i-values exceed the middle one because σ_c is great. Elsewhere the contrasts are even more striking: for the five nearly consecutive elements of atomic numbers 46, 47, 48, 50, 51 (there is no datum for 49) the σ_i-values are 10, 55, 3300, 4 and 8. Again one may venture the guess that for cadmium ($Z = 48$) it is σ_c which is remarkably big, but here it is not substantiated by the formation of a radioactive residue, and one must assume (with some support from gamma-ray emission, page 565) a reaction of pure neutron-capture resulting in a *stable* isotope of cadmium.

Even if the high σ_i-values are really high σ_c-values, they are none the less remarkable. It appears that slow neutrons are

[1] **PR 48**, 265 (1935).

liable to capture if they pass within $\sqrt{3300/\pi}\cdot 10^{-12}$ or say $300\cdot 10^{-13}$ cm of the center of a cadmium nucleus, while for the elements of atomic numbers 62, 63 and 64 (σ_i = 4700, 3700, 30,000) the reach of the nuclear tentacles is even greater. Moreover in figuring the σ_i-value of cadmium, it has been assumed that all of the Cd isotopes are alike in this feature, a very unplausible assumption [1]! if one only among the eight known stable isotopes does the capturing, σ_i must be greater than $3300\cdot 10^{-24}$ by a considerable factor, the size of which must remain unknown until the proper isotope is recognized. This difficulty recurs for the elements 62, 63 and 64 (Sm, Eu and Gd) which are known to have at least seven, two and five stable isotopes apiece. But even if all the isotopes of an element were alike in this respect, the data would prove that for some nuclei at least, the forces whereby nuclei and neutrons act upon each other come into play at distances many times greater than the radius of the electrical potential-barrier which we have been calling the "nuclear radius."

Now it is possible to explain the important role of cadmium in experiments on slow neutrons, and in so doing to correct some implications of the last few paragraphs. Cadmium having so high a σ_i-value for slow neutrons and an ordinarily low one for fast, it is commonly used for shields to prevent the former from going where they are not wanted. What is more, it is commonly used as a filter. In most experiments on slow neutrons (*including those which I have just been quoting*!) two readings are taken for every set of experimental conditions (*e.g.* for every thickness of intercepting screen), one with and one without a cadmium shield somewhere in the path of the neutron beam; and it is the difference between these two which is given as the datum. This modifies the definition of " slow neutrons "; and the term has come to mean, in the customary practice "neutrons which having been slowed down in water or paraffin, are intercepted by a thin sheet of cadmium." These

[1] There is only one element of two or more isotopes—hydrogen—of which the σ_i-values for the individual nuclear species are known; they are given as 35 for H^1 and 4 for H^2.

are also called "C-neutrons," a better name because it connotes nothing beyond its definition. The σ_t-values which I have just been quoting pertain to C-neutrons.

Molecular hydrogen (in the liquid phase) intercepts a greater proportion of a beam of slow neutrons if the spins of the two nuclei in each molecule are parallel than if they are antiparallel.[1] The ratio of the σ_t-values is about 2 or about 4.5 for C-neutrons, according as these emerge from paraffin at room-temperature or at liquid-air temperature. This indicates that the forces between neutrons and protons are closely connected with the spins of these particles.

As with fast neutrons, one looks for data of scattering which might disclose the value of σ_s, and finds the literature [2] scant, yet fuller and more instructive than in the former case. It has been shown, with relative ease, that the σ_s-values for slow neutrons do not fluctuate from one element to the next in the extraordinary manner of the σ_t-values. Elements such as cadmium are by no means outstanding in scattering-power; cadmium indeed scatters slow neutrons much less than does copper, though it intercepts more than five hundred times as many (this was proved at Columbia by slipping metal cylinders over the source-and-detector-system). To assign a numerical value to a neatly definable cross-section is much harder. The technique of A.C.G. Mitchell and his colleagues at New York University consists of embedding the neutron-source in a solid paraffin cylinder, laying on top of the cylinder a foil of metal liable to a (n, ϕ) reaction with formation of a radioactive residue, and laying on top of the foil various thicknesses of the scattering elements. The amount of radioactive residue increases with the thickness of the superposed layers, and the rate of increase serves as an index of a cross-section which cannot be defined except by a detailed diagramming of the apparatus, but may roughly be called "cross-section for back-

[1] Brickwedde, Dunning, Hoge, Manley: **PR 54**, 266 (1938).

[2] Mitchell, Murphy and Whitaker, **PR 50**, 133 (1936), and papers there cited; Mitchell and Varney, **PR 52**, 282 (1937); Bradbury, *et al.*, **PR 52**, 1023 (1937); Dunning, **PR 48**, 278 (1937); Goldhaber and Briggs, **PRS 162**, 127 (1937).

ward scattering"; it ranges between 1 and 10 (times 10^{-24}) for about a dozen elements well sprinkled over the periodic table, and is singularly high for iron and nickel (11 and 18 respectively).

The belief that slow neutrons in general and C-neutrons in particular are largely of "thermal" speeds has been tested at Columbia [1] by a variant of the method of the gates (Chapter II). Four coaxial discs are identically made with 100 nearly equal sectors alternately of cadmium and of duralumin, metals which are opaque and nearly transparent to C-neutrons respectively. The first and fourth stand still, the second and third revolve at a common adjustable speed; and one sees that for neutrons travelling parallel (or not too obliquely) to the axis, the first and second discs constitute a gate of angular aperture varying rhythmically from 0 to π, the third and fourth another such a gate. In the experiments the neutron-sources were embedded deep in paraffin, and cadmium shutters were manipulated as aforesaid so that the eventual results refer to C-neutrons. Assuming these to have the Maxwellian distribution appropriate to the temperature of the paraffin, the experimenters computed the curve to be expected for number-of-neutrons-transmitted *vs.* angular-speed-of-discs, and compared it with the curve of detector-reading (the detector being a lithium-lined ionization chamber) *vs.* angular speed; the two curves agreed "fairly well." The comparison rested on the assumption that detector-response—that is to say, the likelihood of the reaction Li^6 $(n, \alpha)H^3$—is independent of neutron-speed. Supplanting this by the "$1/v$ law"—the assumption that detector-response varies inversely as neutron-speed—they found the agreement bettered.

This "$1/v$ law" is derived from a simple form of quantal theory, which may be paraphrased as being the notion that the likelihood of capture of a slow neutron by a nucleus is proportional to the length of time which the oncoming neutron spends in the immediate neighborhood of the nucleus. If this law is valid for a detector, and if in a given apparatus with paraffin-

[1] **PR 48**, 704 (1935); **49**, 103 (1936); **50**, 738 (1936).

surrounded neutron-source the number of neutrons emerging from the paraffin is independent of its temperature T_p, and if their distribution-in-energy is the thermal distribution appropriate to T_p—if all these things are true, then the detector placed near the paraffin should exhibit a reading proportional to $T_p^{-\frac{1}{2}}$. If it does not, one at least of the three assumptions must be wrong. Generally it does not.

The usual practice is to have the paraffin alternately at liquid-air temperature and at room-temperature. The ratio of the values of $T_p^{-\frac{1}{2}}$ is about 1.8 : 1; the ratio J of the detector-readings is nearly always definitely lower than this and definitely greater than unity. When one detector-type is substituted for another J is frequently altered, and this seems to prove that at any rate the first assumption is at fault; the $1/v$ law may be valid for some reactions of neutron-capture but cannot be valid for all. I mention in particular the researches of Moon and Tillman and their colleagues at the Imperial College [1] who subjected the neutrons to an initial slowing-down in paraffin at room-temperature before admitting them to the core of the apparatus, which was a hollow paraffin cylinder enclosed in a Dewar flask and enclosing in turn a cylinder of some metal liable to a (n, ϕ) reaction resulting in a radioactive residue. The ratio J depends on the thickness of the hollow cylinder; the values accepted by these physicists for fifteen elements range between 1.19 and 1.57; when by use of cadmium screens the J-values appropriate to C-neutrons were isolated, they were confined within the narrower range from 1.43 to 1.60.

The "$1/v$ law" being thus demolished for many examples of neutron-capture resulting in (n, ϕ) reactions, we might make a leap to the other extreme, and imagine that for each reaction there is a characteristic "resonance," *i.e.* that each can be performed only by neutrons of a narrow characteristic speed-range. Suppose a "detector" D liable to such a reaction, encased in a paraffin block full of slow neutrons; the radio-

[1] Moon and Tillman, **PRS 153**, 476 (1936); Michiels and Saxton, **PRS 162**, 441 (1937); see also v. Halban and Preiswerk, *Nature* **136**, 1027 (1935).

activity imparted to it in a given time would be decreased by separating it from the paraffin by an enclosing sheath of an element A; the proportion of decrease would depend both on A and on D, and would be abnormally great if A and D were one and the same element. This sort of thing has indeed been observed, first by Moon and Tillman; and already there have been listed many values of "absorption-coefficients" corresponding to various absorber-and-detector combinations, each being known as the coefficient for absorption, in some element A, of the "resonance-neutrons" of some element D. In some cases D exhibits two distinct radioactive residues, the signs of neutron-capture by two distinct isotopes; for these two (n, ϕ) processes there are found distinct absorption-coefficients in whatever element A may be used as absorber. If A is an element for which the "$1/v$ law" really does hold—as boron and lithium are still believed to be[1]—the ratio of the absorption-coefficients which it offers for any two groups of resonance-neutrons is the reciprocal of the ratio of the speeds of these last. The speed of any group of resonance-neutrons can then be expressed in terms of that of (for instance) the C-neutrons, and if it is correct to take this last as the mean thermal speed corresponding to the temperature of the paraffin, then the energy-values of the resonance-groups are established. Here we have the beginnings of an art of neutron-spectroscopy.[2]

[1] These, I recall, function as detectors by virtue not of (n, ϕ) but of (n, α) reactions. By an ingenious method devised at Columbia by Rasetti and others it has been shown that over a certain interval of neutron-speeds boron does conform to the $1/v$ law, while cadmium and silver do not (**PR 49**, 104, 777; 1936).

[2] Of the already big and rapidly-growing (and difficult) literature of this subject I select a few examples only: Moon and Tillman, *l.c.*; von Halban and Preiswerk, **JP 8**, 29 (1937); Goldsmith and Rasetti, **PR 50**, 328 (1936). The last-named give energy-values for various groups of resonance-neutrons, ranging from 0.026 to 84 E/V; their paper may also be consulted for the terminology, due largely to Amaldi and Fermi.

CHAPTER XVIII

Instability of Nuclei

A radioactive substance is one of which the atom-nuclei are unstable, being capable of surviving for an appreciable time, and then disintegrating or exploding. In such an explosion the nucleus breaks into two fragments, which spring apart with notable kinetic energy. One of these, the "emerging particle," is either an alpha-particle, a positive electron or a negative electron. The other, the "residue," is itself a nucleus, necessarily of different charge and often of different mass-number from the original one. If we denote the original kernel by the symbol (Z, M) comprising its atomic number and mass-number, then the residue is denoted by one of the symbols $(Z - 2, M - 4)$ or $(Z - 1, M)$ or $(Z + 1, M)$ according as the emerging particle is of the first or second or third of those types. Original kernel and residue are therefore isotopes of different elements, and radioactivity may be called spontaneous self-transmutation. The residue may be stable or radioactive; in the latter case it and the original kernel are said to belong to a "series." The durations of the individual nuclei of a radioactive substance are distributed according to the exponential law or law of chance, with a parameter called "half-period." The radioactivity of any such substance is described by stating its half-period, the distribution-in-energy of the emerging particles and the nature of these (in almost but not quite every case they will all be of one of the three aforesaid types). In addition there are often photons—"gamma-rays"—coming out of the residue, and X-rays and negative electrons that have their origin in the orbital electron-family and are emitted or engendered from this last because of the effect of the intra-nuclear process upon it: these are

incidental to the transmutation. Different isotopes of a single element may and usually do differ immensely in their radioactivity. Of radioactive nucleus-types some forty have been discovered in Nature, and since 1933 there have been created over two hundred more which are residues of such man-made reactions of transmutation as we have been considering in the three preceding chapters.

Except for the last clause and for the allusion to positive electrons, all of the content of the foregoing richly-packed paragraph is derived from roughly thirty years of ardent research on the "natural" radioactive substances, of which the first example was identified in 1897 by Henri Becquerel. It is phrased in the language of contemporary physics, but that language is itself to a great extent the fruit of research on radioactivity. The language, the ideas and the experiments which underlie it, are due in the main to Rutherford and Soddy.

Rutherford classified the rays which radioactive bodies emit, identified them and gave them "alpha," "beta" and "gamma" for names: the first and the third persist, the second is yielding to the practice of speaking of the rays as electrons with mention of their sign[1]. Later (but not much later) he conceived and proved the idea that a radioactive substance is an element engaged in transmuting itself into another. The theorem about the relation between a radioactive kernel and its residue was discovered and stated by Soddy, and independently by Fajans of Munich—not however in the form in which I expressed it above, for at that time (1913) the concepts of mass-number and atomic number were equally unknown. Fajans and Soddy said that the radioactive element transforms itself into the element two steps to the left or one step to the right along the Periodic Table, according as it emits an alpha-particle or a negative electron; and this was one of the facts on which the dual notion of nuclear charge and atomic number

[1] There has been lack of accord as to whether "β-rays" should be defined to include the electrons expelled from orbital electron-families as well as those shot out of nuclei, and a similar disaccord as to "γ-rays." At present there is disaccord as to whether to extend the former term to cover positive electrons.

was later to be based. It was also Soddy who invented the concept of "isotope" and the word itself, because of the existence of natural radioactive substances differing in their radioactivity but completely alike in all chemical properties; the

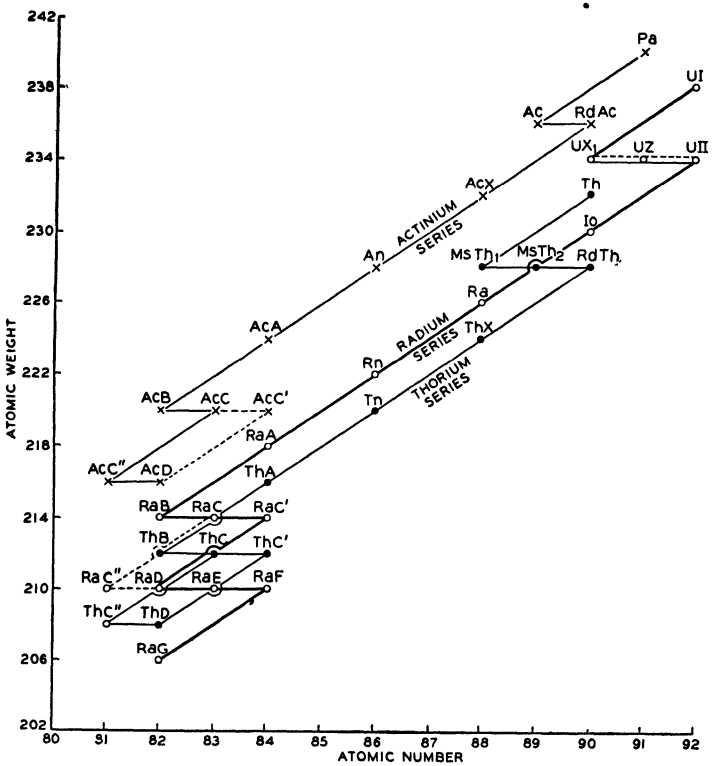


FIGURE 18.1—THE THREE SERIES COMPRISING MOST OF THE "NATURAL" RADIOACTIVE SUBSTANCES. THE ACTINIUM SERIES IS DISPLACED FOR CLARITY; THE ORDINATE OF AcD SHOULD BE 207. See page 594 for presumptive upward continuation of actinium series.

art of mass-analysis and the power of distinguishing the stable isotopes of an element, the ability to define and to measure mass-number, all came later. After the nuclear atom-model had been established by Rutherford, radioactivity was recog-

TABLE 18.1

Symbols, Names and Half-Periods of the Natural Radioactive Substances of the Long Series

The first column of this Table contains the usual symbols for the substances second, their atomic members; the third, their usual names; the fourth, their half periods as collated by Rasetti (*Elements of Nuclear Physics*; Prentice-Hall, 1936) excepting the values for UI and UII taken from Nier (**PR 53**, 922; 1938); the fifth, the nature of the particles which they emit at transmutation. In cases of branching, the less common of the two resulting substances is italicized. The annotation *est.* signifies that the half-periods in question are estimated by extrapolating the Geiger-Nuttall relation. The abbreviations *s*, *m*, *d*, *a* stand for second, minute, day, year. For the probable commencement of the actinium series, see page 594.

Uranium-Radium Series:					
UI	92	Uranium I	$4.56 \cdot 10^9 a$	α	
UX_1	90	Uranium X_1	24.5*d*	β	UX_2, β to UZ
UX_2	91	Uranium X_2	1.138*m*	β	
UZ	*91*	*Uranium Z*	*6.69h*	β	
UII	92	Uranium II	$2.68 \cdot 10^5$	α	
Io	90	Ionium	$7.43 \cdot 10^4 a$	α	
Ra	88	Radium	$1.55 \cdot 10^3 a$	α	
Rn	86	Radon, radium emanation	3.82*d*	α	
RaA	84	Radium A	3.05*m*	α	
RaB	82	Radium B	26.8*m*	β	
RaC	83	Radium C	19.7*m*	β	to RaC', α to RaC''
RaC'	84	Radium C'	$10^{-6} s$	α	
RaC''	*81*	*Radium C''*	*1.32m*	β	
RaD	82	Radium D	22*a*	β	
RaE	83	Radium E	5.0*d*	β	
RaF	84	Radium F, polonium	140*d*	α	
RaG	82	Radium G, radium lead apparently stable			
Thorium Series:					
Th	90	Thorium	$1.8 \cdot 10^{10} a$	α	
MsTh1	88	Mesothorium 1	7.6*a*	β	
MsTh2	89	Mesothorium 2	6.13*h*	β	
RdTh	90	Radiothorium	1.90*a*	α	
ThX	88	Thorium X	3.64*d*	α	
Tn	86	Thoron, thorium emanation	54.5*s*	α	
ThA	84	Thorium A	0.14*s*	α	
ThB	82	Thorium B	10.6*h*	β	
ThC	83	Thorium C	60.5*m*	β	to ThC', α to ThC''
ThC'	84	Thorium C'	$10^{-9} s$ (*est.*)	α	
ThC''	*81*	*Thorium C''*	*3.1m*	β	
ThD	82	Thorium D, thorium lead apparently stable			
Actinium Series:					
Pa	91	protactinium	$3.2 \cdot 10^4 a$	α	
Ac	89	actinium	13.5*a*	β	
RdAc	90	radioactinium	18.9*d*	α	
AcX	88	actinium X	11.2*d*	α	
An	86	actinon, actinium emanation	3.92*s*	α	
AcA	84	actinium A	.002*s*	α	
AcB	82	actinium B	35*m*	β	
AcC	83	actinium C	2.16*m*	α	to AcC'', β to AcC'
AcC'	*84*	*actinium C'*	.005*s*	α	
AcC''	81	actinium C''	4.76*m*	β	
AcD	82	actinium D, actinium lead apparently stable			

nized as a quality of the nucleus; and with this knowledge physicists ceased to wonder at their inability to modify or alter it by heat or chemical action or physical forces, a fact which in the early days was deemed mysterious and striking. Remember also that the nuclear atom-model and the art of transmutation were due to alpha-particles, and realize how greatly modern physics is dependent on the fact that about one in six among the nuclei available in Nature is unstable!

The discovery of one after another of the "natural" radioactive substances, to the eventual number of forty or thereabouts, makes a fascinating story; but already it lies so far in the past that here all but a few details must yield to later work. When Becquerel detected that uranium is radioactive, it was already known as an element; and it was destined later to be shown that several other well-known elements have radioactive isotopes. Twenty-five among the forty however belong to elements never known at all before the radioactivity of their isotopes betrayed them. Pierre and Marie Curie observed that certain uranium-containing ores were more intensely radioactive than their quota of uranium justified them in being; and from these, after long and arduous labors of analytical chemistry, the Curies succeeded in isolating what they called two new elements, and what we now must call isotopes of the till-then-entirely-unknown two elements 84 and 88. They gave the names "polonium" and "radium," and these we must now accept as belonging to the individual isotopes $M = 210$ and $M = 226$ of these elements, but *not* to the ensembles of isotopes sharing the atomic numbers 84 and 88 respectively. Likewise the names of nearly all the other natural radioactive substances belong to individual isotopes and not to elements, a dangerous source of confusion!

The radium extracted by the Curies amounted to only 2 parts in 10^8 by weight of the ores from which it was taken, and the polonium to an even smaller fraction. Yet they had detected its presence in the rocks before they had even begun with the concentration; and this was the first fact to show what has already been said so often in this book, that a sub-

stance which is radioactive can be detected far more easily and in far smaller quantities by its radioactivity, than it can be detected by any other property or than any non-radioactive substance can be detected at all. This was wonderful at the time, and even in retrospect it is impressive; for then there existed not a single one of the devices for discerning the passage of individual emerging particles, and even the other instruments were much less sensitive than they have since been made. Even to this day most of the natural radioactive isotopes have been detected only by their radioactivity, and there are fewer than half-a-dozen of which quantities have been assembled large enough to permit of using upon them either the chemical balance or the mass-spectrograph. The mass-numbers of nearly all of the radioactive nuclei (artificial as well as natural) remain beyond the reach of measurement; and yet we shall see that for most of them the indirect evidence is so powerful as to fix them with practically no doubt.

Advent of artificial radioactivity

The first who consciously made new radioactive nuclei by the art of transmutation were Frederic Joliot and Irene Curie, who in the early days of 1934 exposed aluminium, boron and magnesium to the 5.3-Mev alpha-particles emitted by polonium; left the samples under bombardment for a few minutes; then removed them to the neighborhood of a Geiger-Müller counter, and observed hundreds of counts per minute. These they found to be due to *positive* electrons; had this particle not been recognized a year and a half sooner among the cosmic rays, it would have been discovered together with these first examples of "artificial" or "induced" radioactivity. The number per minute was found to decline exponentially with lapse of time, in the manner of the emission of natural radioactive bodies.

Curie and Joliot supposed, and rightly, that the impinging alpha-particles were producing transmutations of which the "residues" (in the sense of page 478) were radioactive. The

nature of the residues was not however settled, for one could not be sure of the reactions. Boron was known at the time to be subject both to (α, p) and to (α, n) reactions, and the same was soon proved true of the other two targets. If the radioactive residue was that of the (α, p) reaction, it had to be a nucleus of C or Si if the target was B or Al; if of the (α, n) reaction, it had to be a nucleus of N or P in the two cases. Reactions otherwise unknown could not have been excluded.

Confronted with this situation, Joliot and Curie applied a technique old in the field of radioactivity but new in the field of transmutation. Relying on the presumption that the newborn nuclei would surround themselves with electrons and so become the kernels of completed atoms, they subjected their bombarded targets of B (or Al) to such chemical processes as would remove N (or P) commingled with the original element. *The radioactivity was itself removed and carried away.* I quote the words with which this technique was introduced to the world of science at large: "We have bombarded BN. By heating boron nitride with caustic soda, gaseous ammonia is produced. The activity separates from the boron and is carried away with the ammonia. This agrees very well with the hypothesis that the radioactive nucleus is in this case an isotope of nitrogen. When bombarded Al is dissolved in HCl, the activity is carried away with the hydrogen in the gaseous state, and can be collected in a tube. The chemical reaction must be the formation of PH^3 or SiH^4. The precipitation of the activity with zirconium phosphate in acid solution seems to indicate that the radioactive substance is an isotope of phosphorus." These accordingly are (α, n) reactions.

The student will remember how often, in the three preceding chapters, the nature of a reaction was settled by invoking evidence such as this. The Joliot-Curie pair had devised the first (and still the only) way of recognizing a product of transmutation from the chemical behavior of atoms having the newborn nuclei for their kernels—of identifying a new-made element in the way that elements generally have been identified from the beginning of modern chemistry. This is not the same

as saying that the methods of detecting transmutation listed in Chapter XV have been superseded, for the newborn nucleus is manifested through the ionizing-power of a fast-moving particle, not however the nucleus itself but an electron bursting out of it.

The problem of choosing the method appropriate to each particular case is chemical, and in many institutes of nuclear physics is confided to a professional chemist. I mention here only a general principle and a special case. The principle is, that the target must contain not only the nuclei concerned in the transmutation but some of the element (*i.e.* of the stable isotope or isotopes of the element) of which the residue is destined to be an unstable isotope. For instance, if the reaction in question is a (α, n) reaction of a nucleus (Z, M), then the target must initially contain not only the kernels (Z, M) but also some of the stable isotope or isotopes of the element $(Z - 2)$, or else must be dissolved and intimately mixed with this last; the subsequent chemical processes then remove the stable and the unstable isotopes of $(Z - 2)$ together. The lines just quoted from Joliot and Curie illustrate this point. The special case arises when the residue is of the same atomic number as the original kernel itself. If the target contains the original element uncompounded, attempts to remove the radioactivity are unavailing. Negative results of such attempts are usually taken as sufficient evidence that the reaction is of the (n, ϕ) or the $(n, 2n)$ type. Positive results of the same significance may be obtained in a way discovered by Szilard and Chalmers.[1] The target is made of a chemical compound of the element in question, and after bombardment is mixed with some of the uncompounded element; the subsequent chemical processes are such as to separate the latter from the former; some of the radioactivity "follows" the latter, indicating that when a kernel involved in a molecule either captures a neutron or has a neutron knocked out of it, it breaks away from its moorings and rebuilds its orbital electron-family in such a way as to form a free atom.

[1] *Nature* 134, 462 (1934).

Figures 18.2 and 18.3 exhibit the stable and the radioactive nuclei of mass-numbers from 1 to 23 inclusive, shown as circles and as stars respectively. From each star an arrow points upward or downward, to indicate that the unstable nucleus

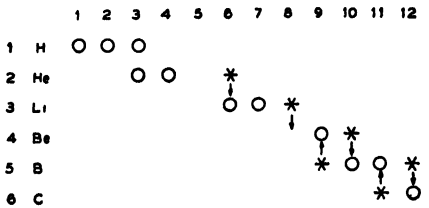


FIGURE 18.2—ISOTOPES OF THE FIRST SIX ELEMENTS. CIRCLES MARK STABLE NUCLEI, STARS UNSTABLE NUCLEI, ARROWS INDICATE THE TRANSMUTATIONS OF THE LATTER INTO THE FORMER; THE ARROW POINTS UPWARD OR DOWNWARD, ACCORDING AS THE NUCLEUS FROM WHICH IT POINTS EMITS A POSITIVE OR A NEGATIVE ELECTRON.

emits a negative or a positive electron. In every case but one, the arrow points directly to a circle; this signifies that the residue left behind by the departing electron is identical, in charge and in mass-number, with a stable type of nucleus already known. The residues are themselves apparently

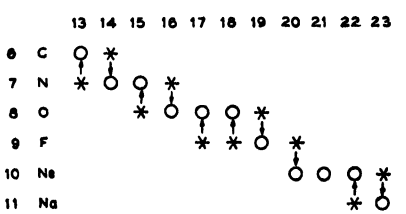


FIGURE 18.3—ISOTOPES OF THE ELEMENTS $Z = 6$ TO 11. SYMBOLISM AS IN FIGURE 18.2.

stable, *i.e.* they are not observed to be radioactive; this strengthens a principle to be mentioned in the next section, *viz.* that a given pair of values of Z and M is never shared by a stable and a radioactive nucleus. (Yet it must be conceded that in certain cases where there is no other indication of the

M-value of a radioactive isotope known to belong to a certain element, its star is deliberately located so that its arrow will point to a circle!) Were the arrows to point in the opposite directions, they would point to vacant spaces, presumably corresponding to pairs of values of Z and M for which stability is impossible. This suggests (to put it in a very animistic way) that the unstable nucleus "knows" whether stability is to be found by the emission of a positive or by that of a negative electron, and acts accordingly; there is an automatic tendency to go to the stable condition. The vacant space to which the arrow points from Li^8 is of Be^8, the mysterious nucleus of which so much has been said in previous chapters.

Turning back to Figure 18.1, and realizing that the place of each link between consecutive crosses, dots or circles corresponds to an arrow in the sense of Figure 18.2, we see that nearly all of the natural radioactive substances have radioactive residues. There is no violation here of the principle aforesaid, since there are no stable nuclei coinciding both in Z and in M with any of these. A small number of artificial radioactive nuclei, all of them high in mass-number, have radioactive residues; some of these form a chain (or chains?) leading through elements of higher nuclear charge than even uranium ($Z = 93, 94, 95, 96$), while others form the beginning of a chain comparable with the three exhibited in Figure 18.1.

Assignment of mass-numbers to radioactive nuclei

To assign the right atomic number to a radioactive substance is thus a problem of chemical analysis. To assign the proper mass-number is a problem to which there is no universal mode of approach, mass-analysis nearly always failing because of the scantiness of these bodies.

If the radioactive substance in question is man-made there may be no doubt whatever as to its *M*-value, provided that it is made by a reaction in which the bombarded target is an element of a single isotope. The phosphorus resulting from the reaction $Al(\alpha, n)P$ can only be P^{30}, since aluminium consists

entirely[1] of the one (stable) isotope Al^{27}. When however the bombarded target is an element of two or more isotopes, it is necessary to seek for auxiliary evidence to settle which of them is concerned in the reaction, and what accordingly is the M-value of the active residue. Often the physicist is enabled to make up his mind as to an M-value because what seems to be one and the same radioactive substance can be made in a variety of ways.

Illustrations will make this last point clear. Substances emitting negative electrons and having half-periods equal to about $2\frac{1}{2}$ minutes are produced by the following bombardments: (*a*) magnesium with alpha-particles; (*b*) silicon with neutrons; (*c*) phosphorus with neutrons; (*d*) aluminium with *slow* neutrons; (*e*) aluminium with deuterons.[2] Of these substances it has been shown in addition that (*c*) and (*e*) have the chemical properties of aluminium. The same inference is drawn as to (*d*) from the fact that slow neutrons produce it—that is to say, the rate of production of (*d*) is considerably enhanced by causing the neutrons to pass through water before they impinge on the target, which is *not* the case with (*b*) or (*c*). If all five substances are isotopes of aluminium, then the five reactions are necessarily $Mg(\alpha,p)$, $Si(n,p)$, $P(n,\alpha)$, $Al(n,\phi)$ and $Al(d,p)$. If all five are the same isotope of aluminium of mass-number (say) M, then the bombarded nuclei entering into these reactions are necessarily Mg^{M-3}, Si^{M}, P^{M+3}, Al^{M-1} and Al^{M-1}. If for M we put 28, then all of these last are to be found in the list of known and stable nuclei; not however if for M we put any other value than 28. Thus if the substances (*a*) and (*b*) are really identical with (*c*) and (*d*) and (*e*), the multiplicity of the known stable isotopes of Mg and Si respectively does not pre-

[1] Yet it is not possible quite to exclude the chance that an element may comprise an isotope too rare to be detected in mass-analysis, and yet abundant enough to yield, under bombardment, a perceptible amount of a radioactive residue.

[2] The references are: for (*a*), (*b*) and (*c*), Curie Joliot and Preiswerk, **JP 5**, 153 (1934) and **CR 198**, 2089 (1934), giving 2m 45 s (±20 s) for all three half-periods; for (*b*), (*c*) and (*d*), the Roman school, **PRS 146**, 483 (1934), and **149**, 522 (1935), giving 2.3 m for all three half-periods; for (*e*) McMillan and Lawrence, **PR 47**, 343 (1935), giving 2 m 36 s (±5 s), and Cork *et al.*, **PR 49**, 208 (1936), giving 2 m 36 s.

vent us from recognizing that (*a*) and (*b*) are both Al^{28} and that the reactions in question are $Mg^{25}(\alpha,p)Al^{28}$ and $Si^{28}(n,p)Al^{28}$ respectively. Another good example is afforded by the residues of five reactions occurring when sodium is bombarded by deuterons or by neutrons, aluminium by neutrons, magnesium by neutrons or by deuterons. All of these residues emit negative electrons and have half-periods of about 15 hours; the first and the third are known to have the chemical properties of sodium. If all of them are the same substance Na^M, the reactions are $Na^{M-1}(d,p)$, $Na^{M-1}(n,\phi)$, $Al^{M+3}(n,\alpha)$, $Mg^M(n,p)$ and $Mg^{M+2}(d,\alpha)$; if and only if M is 24, the nuclei Na^{M-1} and Mg^M and Mg^{M+2} and Al^{M+3} are all in the list of those which are known to exist and be stable. Such accordingly is the conclusion. Incidentally this presumably-single substance formed in five different ways is probably the most-mentioned of artificial radioactive bodies, because of its anticipated services in medicine. (It is often called "radio sodium" by a usage which is to be deprecated, inasmuch as an element may have more than one radioactive isotope; sodium itself has at least one other, believed to be Na^{22}).

Note the tacit assumption underlying all this reasoning: *to each pair of values of Z and M is supposed to correspond not more than one type of radioactive nucleus.* Though derived from no compelling theory, this seems to be *a rule agreeing nearly but not entirely with the truth*; exceptions are very few but strongly certified. The rule was evidently in the minds of Joliot and Curie when they published the first disclosure of artificial radioactivity. Having made—by the reaction $B^{10}(\alpha,n)$—what they rightly regarded as N^{13}, they foretold that it could also be made by projecting deuterons against carbon. The hint was taken immediately in the laboratories having the requisite apparatus, and in less than four months reports of success were published from three (Pasadena, Berkeley, and the Carnegie Institution of Washington).

Another slightly-different rule—*to any pair of values of Z and M there never correspond both a stable and a radioactive nucleus type*—is invoked very often in the study of transmu-

tation. When a physicist produces a radioactive body by bombarding a target, he begins his attempts at interpretation by rejecting all of the possible reactions which lead to residues identical in Z and M with known stable nucleus-types. When Curie and Joliot obtained their radioactive isotope of nitrogen by projecting alpha-particles against boron which is a mixture of B^{10} and B^{11}, they assumed that it sprang from the reaction $B^{10}(\alpha,n)N^{13}$ rather than from the reaction $B^{11}(\alpha,n)N^{14}$, on the presumption that the N^{14} resulting from the latter could be none other than the known stable isotope of nitrogen.

To this latter rule there seems to be no exception. To the former there seem to be a few, the very rarity of which has caused them to be studied with unusual care. For an example I take the case of bromine, attention to which was first directed by the fact that from bombardment of this element by slow neutrons there are born three radioactive substances, though it has only the two known stable isotopes Br^{79} and Br^{81}. Knowledge of this fact was supplemented soon by knowledge (due to Bothe and Gentner) that from bombardment of bromine by high-energy photons there also result three radioactive bodies. The first three are distinguished by different half-periods which I will call T_1, T_2, T_3; the second three by half-periods T', T'', T'''. It is found that within the uncertainty of the measurements, T'' is equal to T_1 (about 18 minutes) and T''' to T_2 (about $4\frac{1}{2}$ hours); and from these and many other facts it is inferred that there are two nucleus-types or "isomers" (such is the technical word) of Br^{80} distinguished by these half-periods, and that both can be formed either from Br^{79} by (n,ϕ) reactions or from Br^{81} by (ϕ,n) reactions. (Of the two other half-periods, T_3 is attributed to Br^{82} and T' to Br^{78}.) The other well-authenticated cases of isomers are those of In^{116} and Ag^{106}. The last is perhaps more remarkable than the first two, for both isomers of Br^{80} and both of In^{116} are emitters of negative electrons, but of Ag^{106} one type emits negative electrons and the other positive.[1]

[1] For the case of Br^{80} see Snell, **PR 52**, 1007 (1937); Bothe and Gentner, **ZP 106**, 236 (1937); Heyn, *Nature* **139**, 842 (1937); earlier literature cited by Snell. For

As to the mass-numbers of the natural radioactive bodies, the situation is very satisfactory for members of the thorium and uranium series, moderately so for the actinium series.

The *M*-values of all the members of the thorium series are fixed, first of all, by the mass-analysis of thorium which has revealed only a single isotope $_{90}Th^{232}$. Accepting this as the ancestor of the series, we readily see how to figure the *M*-value of every other nucleus marked on the zigzag line of Figure 18.1 beginning at Th. The final nucleus of the series must then be that isotope of lead which has the mass-number 208, and this is confirmed by the observation that the lead which is found in the earth in close association with thorium has generally an abnormally high quota of this isotope.[1] The confirmation is well worth having, since it practically removes the last lingering suspicions that some other isotope of thorium than Th^{232} *might* be the head of the series and that there *might* still be some members of the series between thorium and lead which had escaped detection.

The *M*-values of all the members of the uranium-radium series are fixed by several concordant facts. Mass-analysis of uranium shows one isotope preponderant by far (more than 99% of the whole) over all the rest; this one, $_{92}U^{238}$, is taken as the ancestor "uranium I" of the series, while another (U^{234}) lately discovered by Nier may well be [2] its descendant "ura-

In^{116} see Mitchell and Langer, **PR 53**, 505 (1938). For Ag^{106} see Pool, **PR 53**, 116 (1938).

[1] The subject of the isotopic constitution of lead is a long and important one. "Ordinary" lead—meaning, I take it, lead taken from minerals *not* rich in uranium or thorium—has three principal isotopes 206, 207, 208 (as well as four or five others constituting less than 2% of the whole) and a chemical atomic weight given as 207.2. The relative abundances of the three main isotopes were given by Aston as 27.76 : 20.20 : 49.55 (**PRS 140**, 535; 1933). Lead extracted from minerals rich in U or Th has an excessive proportion of 206 or 208 respectively—a fact which was proved by mass-analysis after the mass-spectrograph was invented, but had already been inferred from the abnormally low and abnormally high chemical atomic weights of these two varieties of lead. In the early days of the study of radioactivity this fact was one of the chief items of evidence for the general theory of radioactive bodies. According to Nier the composition of "ordinary" lead is itself far from constant (**PR 53**, 680; 1938).

[2] It has the correct mass-number, and its abundance agrees with that to be expected for uranium II in "radioactive equilibrium" (page 600).

nium II." Figuring the *M*-values of the other nuclei of the series from this presumption, we find 226 for radium, 222 for radon, 206 for the terminal nucleus which is an isotope of lead. The first two of these values are confirmed by measurements of the chemical atomic weight of Ra and the density of Rn, both of which have been isolated in amounts great enough to permit of weighings. The last is confirmed by the fact that the lead which is found in the rocks in close association with uranium has generally an abnormally high quota of the isotope 206 and a correspondingly low chemical atomic weight. This is convincing proof that the isotope 238 of U has been correctly identified as the ancestor and that none of its descendants has been overlooked.

The *M*-values of the actinium series were long among the mysteries of physics, not one of these substances having been isolated in quantity sufficient for mass-analysis or chemical weighing. There was a steadily-strengthening suspicion that the true ultimate ancestor of this series is an isotope of uranium, and its apparent head—protactinium, an isotope of atomic number 91—a descendant thereof; also a suspicion that its terminal nucleus, known to be an isotope of lead, is the isotope 207. About a tenth of a milligram of protactinium was finally isolated in the Kaiser-Wilhelm Institut, and upon its chemical atomic weight von Grosse made two measurements giving values between 230 and 231, and slightly favoring 231. The hyperfine structure of the optical spectrum of Pa is of a type distinctive of atoms having nuclei of odd mass-number (page 463). There exists an isotope 235 of uranium, scanty and yet abundant enough to be detectable; and there is some evidence that Pb^{207} is abnormally abundant in lead derived from minerals rich in uranium. Considering all these things it is very probable that the two ends of the presently-known actinium series as exhibited in Figure 18.1 should be located at ${}_{91}Pa^{231}$ and ${}_{82}Pb^{207}$, and that the zigzag line of that figure should take its departure from ${}_{92}U^{235}$. The linking U^{235} and Pa^{231} must occur through the emission of an alpha-particle and that of a negative electron, but which of these two processes

comes before the other is something that remains to be learned.[1]

The few remaining natural radioactive bodies do not belong to series, and each therefore constitutes a special case. A considerable amount of effort has been expended in trying to separate the isotopes of potassium adequately for finding out which is the radioactive one. In the mass-spectrograph the two desiderata of high resolving-power and copious stream of ions are at odds with one another, in the sense that in striving for either the experimenter recedes from the other; yet it now seems established by means of this instrument that the very rare isotope K^{40} (to which only one nucleus in 8000 of potassium belongs) is the source of the negative electrons emitted by this element. Smythe and Hemmendinger of Pasadena, to whom this conclusion is due, made a like investigation of rubidium, and traced its radioactivity to the nuclei of mass-number 87.[2] Meanwhile Hahn and his associates had come by a different route to the same conclusion; analyzing a mineral rich in rubidium, they found in it a quantity of strontium small and yet sufficient to be subjected to mass-analysis; Mattauch with his mass-spectrograph found it to consist of pure Sr^{87} (of "ordinary" strontium this isotope constitutes less than one-twelfth!).[3] As for samarium,[4] Dempster put a sample into his mass-spectrograph and used for his photographic plate an emulsion in which alpha-particles form tracks which are visible after development (Wilkins' method, page 517); after some months Wilkins examined the plate and found such tracks emanating from the "line" (similar to those of Figure 3.13) composed of the isotope Sm^{148}.

[1] See Western and Ruark, **PR 44**, 675 (1933) for earlier work; Aston and Rutherford, *Nature* **123**, 313 (1929) for evidence about Pb^{207}; von Grosse, **PRS 150**, 363 (1925), for chemical atomic weight of Pa; Aston, *Isotopes* and Nier, **PR 53**, 922 (1938) for isotopes of uranium; Rayton and Wilkins, **PR 51**, 818 (1937), for indication of a radioactive U^{235}.

[2] Smythe and Hemmendinger, **PR 51**, 178, 1052 (1937), and earlier literature there cited.

[3] Hahn, Strassman, Walling, Mattauch, **NW 25**, 189 (1937).

[4] Wilkins and Dempster, **PR 54**, 315 (1938).

The exponential law

The exponential law of radioactivity is so simple and so universal that there is really little to be said about it. I write it down *pro forma*:

$$N = N_0 \, exp\,(-\lambda t) = N_0 \, exp\,(-t/T), \qquad (18.1)$$

$$dN/dt = -\lambda N_0 \, exp\,(-\lambda t)$$
$$= -\lambda N_0(1 - \lambda t + \tfrac{1}{2}\lambda^2 t^2 - \ldots), \qquad (18.2)$$

and mention the well-known feature that if—as this formulation implies—there exist at $t = 0$ a number N_0 of radioactive nuclei of a single kind, then not only is their average future duration equal to T, but also at any future instant t the average future duration to which the $N_0 \, exp\,(-t/T)$ then-surviving nuclei may still look forward is equal to T. This is otherwise expressed by saying that the probable future duration of an unstable kernel is entirely independent of the length of time during which it has already existed.

The allusion to probability recalls that this exponential law or so-called "law of chance" has often been taken as suggesting that the number of particles in the nucleus is enormous. I quote a passage from Poincaré,[1] taking only the liberty of writing "nucleus" in two places where he wrote "atom." "If we reflect on the form of the exponential law, we see that it is a statistical law; we recognize the imprint of chance. In this case of radioactivity, the influence of chance is not due to haphazard encounters between atoms or other haphazard external agencies. The causes of the transmutation, I mean the immediate cause as well as the underlying one (*la cause occasionelle aussi bien que la cause profonde*) are to be found in the interior of the atom; for otherwise external circumstances would affect the value of the coefficient in the exponent. . . . The chance which governs these transmutations is therefore internal; that is to say, the nucleus of the radioactive substance is a world, and a world subject to chance. But, take note! to say 'chance'

[1] *Dernières Pensées*, pp. 204–205; he credits Debierne with the idea.

is the same as to say 'large numbers'—a world built of a small number of parts will obey laws which are more or less complicated, but not statistical. Hence the nucleus must be a complicated world. . . ." Poincaré's inference in these last two sentences was temporarily discarded after the advent of quantal theory, for the exponential law can be interpreted by supposing that in the unstable nucleus the to-be-emitted particle is wobbling around by itself in a central potential-well, and eventually escapes by penetrating a potential-barrier. This imagined process is the reverse of the process of entry of (say) an alpha-particle into a nucleus, described in Chapter XV. Nevertheless the theorists are now returning to the notion of a very complicated nucleus, in which the to-be-emitted particle may actually come into being as a result of a fortuitous concourse of circumstances, shortly before it is emitted.[1]

Measurement and values of half-periods

The exponential law is a single-constant formula, but the constant may be variously specified. The "decay constant" is λ of (18.1); the "mean life" is T or λ^{-1}; the "half-period" is about 0.7 T or accurately $(log_e 2)T$, being the lapse of time at the end of which just one-half of the initially existing kernels yet survive.

The standard way of evaluating these constants is, of course, to plot the activity of a sample of the substance on logarithmic cross-section paper against time, and determine the slope of the resulting straight line. "Activity" is a convenient general word for anything proportional to the number of emerging particles emitted per second, which is dN/dt of (18.2) and hence is proportional to $exp\ (-\lambda t)$. The plotting of such data often results in wonderfully fine examples (*e.g.*,

[1] Further and important evidence that transmutations are governed by the "laws of chance" is furnished by the variations or fluctuations (Schwankungen) of the numbers of alpha-particles emitted from a sample of any radioactive substance in consecutive equal time-intervals very short compared with the half-period of the substance. These are analogous to the fluctuations in thermionic emission called "Schroteffekt" (Chapter I). Cf. Kohlrausch, *Ergebnisse d. exakt. Naturwiss.* 5 (1926).

Figure 18.4) of linear relations which are perfect within the (often very low) uncertainty of the measurements. Sometimes when the sample is a recently-bombarded target, the data lie on a curve which straightens out into a line as t increases. It is then assumed that two or more radioactive bodies are

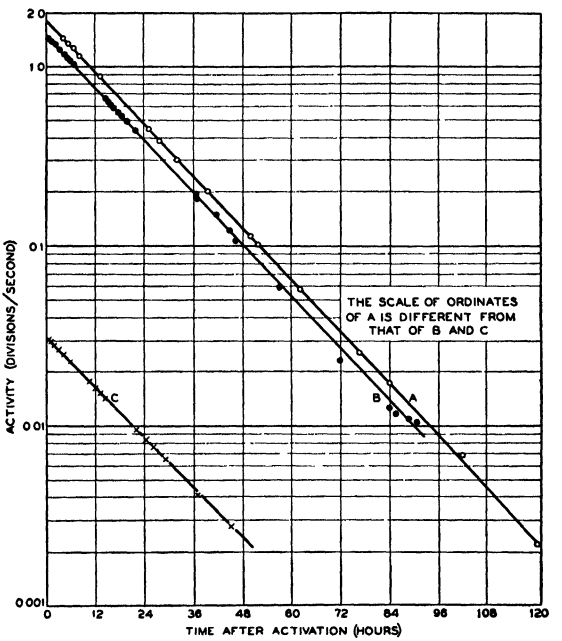


FIGURE 18.4—EXPONENTIAL DECAY-CURVES OF RADIOACTIVE BODIES OBTAINED FROM THREE REACTIONS (A, POTASSIUM BOMBARDED BY DEUTERONS; B, POTASSIUM EXPOSED TO SLOW NEUTRONS; C, SCANDIUM BOMBARDED BY FAST NEUTRONS) AND CONSIDERED ALL TO CONSIST OF THE SAME RADIOACTIVE SUBSTANCE K^{42}. (Hurst and Walke; *Physical Review.*)

present, the λ-value corresponding to the asymptotic line being that of the body of longest life. The activity A_1 due to this last is then computed for smaller values of t by backward extrapolation and deducted from the whole activity A, and the difference $(A - A_1)$ is plotted against t on logarithmic paper; the process may be repeated again and again. This way of

describing the procedure may suggest that it is arbitrary and void of physical meaning, but actually the radioactive substances thus distinguished can often be separated from each other, justifying the procedure fully.

It is not hard to see that for half-periods of more than a certain length, this "standard" method of determining λ becomes unusable; T-values of several years have actually been determined thus by people of patience, but who can stay long enough to trace a sufficient part of the curve for radium with its 1590-year half-period, not to speak of samarium with a million million years? For these two substances, and uranium and thorium also, there happens to be a method of the greatest simplicity in principle, available because one can isolate enough of the radioactive substance to weigh and one can also count the emerging particles (alpha-particles, in all these cases) which the weighed quantity emits per unit time. Thus the observer measures N_0 on the right and dN/dt on the left of equation (18.2), and calculates λ with supreme ease because the second and remaining terms in the series-expansion on the right are utterly negligible. A gramme of samarium, containing $4 \cdot 10^{+21}$ atoms, emits seventy-five alpha-particles per second, according to Hevesy [1]; this corresponds to a half-period of $1.2 \cdot 10^{12}$ years. (Notice that the calculation of half-period involves the assumption that all the isotopes are alike in their radioactivity, a most improbable assumption! This makes the value just quoted for Sm too high,[2] but the error is probably not grave for U and Th in each of which one isotope is predominant, and does not occur with radium which is prepared as a pure isotope.) The rate of emission of alpha-particles from a gramme of radium is one of the best-measured of all the constants of radioactivity, and supplies the definition of the unit

[1] Hevesy *et al.*, **ZP 83**, 43 (1933); see also Libby Peterson and Latimer, **PR 48**, 571 (1935), who applied the method to Sm and Nd, and Mühlhoff, **AP** (5) **7**, 205 (1931), who applied it to K and Rb; in these cases the number of emitted particles was determined not absolutely, but by comparison ultimately with another radioactive substance for which the number was taken as known.

[2] From the observation of Wilkins just cited it follows that the true value is about 14% as great.

called the "curie"; the now-accepted value is $3.71 \cdot 10^{10}$ alpha-particles per second.

As most of the natural radioactive substances belong to one or another of the three long series which we shall later consider (they may be viewed in Figure 18.1), a rock or a preparation which contains any at all is likely to contain several, growing one out of another and decaying one into another. The "ancestors" or first members of the series are by far the longest-lived of all (otherwise they and their descendants would no longer be here) and this entails an important consequence. The rocks containing uranium (or thorium) are so old that the ancestor has had plenty of time to get into what is termed "radioactive equilibrium" with its descendants; by which is meant, that the quantity N of any one of the radioactive members of the series has ceased to vary with time because per unit time it is gaining just as many nuclei by self-transmutation of the preceding member as it is losing by self-transmutation of itself. It is easy to see that in radioactive equilibrium the products λN for all the members of a series have the same value; people have determined certain λ-values by availing themselves of this fact.

When the radioactive equilibrium is broken by removing some of the members of the series by chemical processes, the subsequent dependence of activity on time may be very complex for a while, and difficult to analyze; yet some of the natural radioactive bodies have been recognized and their λ-values estimated by analysis of such curves (so have the λ-values of uranium I and thorium and radium, the results being concordant with those obtained by counting the particles emitted by measured amounts of these bodies). It is easy to see that if at $t = 0$ the ith member is completely freed of all of its descendants and if furthermore λ_i is much smaller than λ_{i+1}, then as t increases the amount and the activity of the $(i + 1)$th member increase from zero, being proportional to $(1 - exp(-\lambda_{i+1}))$. So long as t is small compared to λ_{i+1}^{-1}, the activity of this $(i + 1)$th member is proportional to t. This last statement is not very important by itself, but becomes so when combined

with the statement that the activity of the $(i + 2)$th member increases at first as t^2. It follows that if the $(i + 1)$th member escapes detection because the particles which it emits are too slow-moving, the $(i + 2)$th member reveals itself as such because of obeying this quadratic law. In this manner radium was shown by Soddy to be the "grandchild" and not the "child" of uranium II; the alpha-particles proceeding from the intervening substance (ionium) were later observed. If in such a case as we are now imagining this undetectable member of the series were to be isolated by chemical processes from its ancestors and its posterity alike, the observer would have a piece of matter which was inactive when prepared and became radioactive as time went on. Cases of this type have occurred; and a good deal has thus been learned about these undetectable (formerly called "rayless") bodies; actinium and mesothorium I are to this day known in no other way.

The half-periods of the natural radioactive substances are listed in Table 18.1. One sees that the diversity in order-of-magnitude is great! Among the two hundred and more artificial radioactive isotopes the diversity is very much less, but there are compelling reasons why this must seem to be true even if complete knowledge would not confirm it. Since only four years have elapsed since artificial radioactivity was discovered, there simply has not been time to follow any such curve as that of Figure 18.4 for several years: the three-year half-period of Na^{22} is the longest yet to be measured. Not enough of any such isotope to be weighable has yet been made, so that the method of counting the particles emitted per unit time by a weighed quantity is unavailable. Worse yet, equal amounts (*i.e.*, equal numbers of nuclei) of different radioactive substances have activities which are smaller, the greater the half-periods; this lessens the chance of even detecting the longer-lived bodies which may result from transmutation. At the other extreme, half-periods of the order of one second have been evaluated by using either the rim of a turning wheel or the bob of a pendulum to carry a target swiftly from a place where it is under bombardment to another place where it is

beside a detector: He^6 and Li^8 (formed by neutron-bombardment of Be and Li) afford examples, and there is an indication that B^{12} has a half-period of the order of 0.02 second. The still lower values of actinium A and radium C′ in the Table of the natural radioactive substances were measured by *tours de force*, depending on the fact that these nuclei are born from explosions of the preceding nuclei of actinon and radium C, and consequently have appreciable velocities of recoil. In the former case, the wheel-device is used and the wheel-rim receives its load of actinium A by passing through gaseous actinon; in the latter, the newborn nuclei of radium C′ fly along an evacuated tube with a calculable speed, and the activity of the beam is measured at various points along its path.

From the viewpoint of the theorist the most important feature of the half-periods is their relationship with the energies of the emerging particles. Before taking this up we must learn the main points of what is known about these energies.

Distribution-in-energy of alpha-particles and the associated Q-values

Formerly it was commonly believed that all of the alpha-particles proceeding from a single radioactive substance have initially the same kinetic energy, or—in the language of Chapter XV—that they belong to a single group. Putting Q for the sum of this kinetic energy and that of the recoiling residue (the latter to be computed from the former with the assumption of conservation of momentum), we should then have for every reaction of α-emission an equation of the type:

$$ {}_{Z}\mathrm{R}^{M} = {}_{Z-2}\mathrm{R}^{M-4} + {}_{2}\mathrm{He}^{4} + Q \qquad (18.3) $$

this being the analogue of the equation (15.4) for processes of artificial transmutation. As in (15.4), the quasi-chemical symbols stand for rest-masses of nuclei. R is used as a general symbol for a nucleus of the mass-number and atomic number specified by superscript and subscript respectively (it is hard to

find a suitable capital letter not already bespoken for some particular element).

If there is truly only one group of alpha-particles and hence only one Q-value, the natural inference is the same as in similar cases of artificial transmutation: the residual nuclei ${}_{Z-2}R^{M-4}$ are taken to have all the same mass. This pleasantly simple conclusion is still acceptable for some reactions of α-emission, but by no means for all. Of eleven such reactions it is now known that there are two or more α-particle groups, hence two or more Q-values. By the reasoning of Chapter XV one would deduce that in every such case the residual nucleus has two or more mass-values, one perhaps corresponding to its normal state and the rest to excited states. Before rushing to this conclusion we should, however, look at the data more closely.

The cases in question fall into two classes, one of which comprises the "three C′-products"—RaC′, ThC′, AcC′. The phenomenon here is, that while the great majority of the alpha-particles all have the same kinetic energy there are a very few "long-range" particles which make scintillations on screens held far beyond the uttermost distance reachable by the main group. Their distribution-in-range was thoroughly studied at the Cavendish: first by the method described on page 521, in which the detector beyond the absorbing screens consists of a pair of ionization-chambers so set up that the particles traverse them in succession and the difference between the amounts of ionization in the two determines the detector-reading, which accordingly shows a maximum for each group. A curve obtained in this way appears in Figure 18.5. Later the distribution-in-energy of these particles was determined by using a transverse magnetic field of adjustable strength to bend their paths through a semicircular arc to an ionization-chamber located in a fixed position, as ions are deflected in the elementary type of mass-spectrograph shown in Figure 3.10; here the field had to be made immensely strong by the building of a specially-designed annular magnet. The analysis of the long-range particles of RaC′ disclosed twelve groups with Q-values all greater than

that of the main group; but the abundance of this last was more than 10^4 times as great as those of all the subsidiary twelve added together.

The α-emitters of the other class produce no long-range particles. For each the α-particles were formerly supposed to belong to a single group, but "increase of resolving power" made it clear that actually there are several groups having Q-

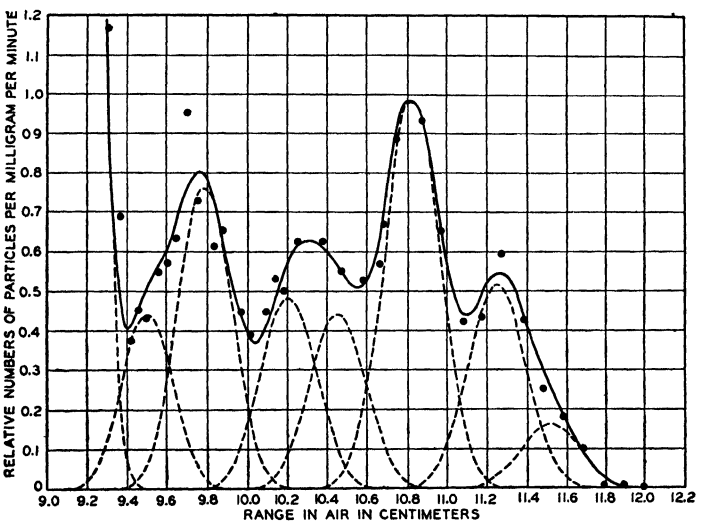


FIGURE 18.5—PART OF THE DISTRIBUTION-IN-RANGE CURVE OF THE LONG-RANGE ALPHA-PARTICLES FROM RaC'. (Rutherford, Ward and Lewis; *Proceedings of the Royal Society*.)

values relatively close together and abundances not very unequal. This increase in resolving-power was attained by deflecting the alpha-particles through semicircular arcs by the aid of very powerful magnets: at first the great magnet at Bellevue near Paris, employed by Rosenblum who with it discovered this so-called "fine-structure of the α-ray spectrum"; later the annular magnet of the Cavendish already mentioned. (Note that part of the advantage of this scheme resides in the fact that the α-particles travel through a good vacuum on their

way from source to detector, and hence do not undergo the many encounters with atoms which impress a wide distribution-in-energy on a beam of particles initially all of the same energy.) Photographic plates were used as detectors by Rosenblum, so that the visual resemblance to an optical spectrum is close. By 1936 Rosenblum had detected 15 groups of RdAc, 12 of AcX, 4 of An and 2 of AcC; 2 of RdTh and 6 of ThC; 2 of Ra and 2 of RaC. According to Lewis and Wynn-Williams there are (or were, in 1932) at least five cases in which the distribution-in-range curve obtained with the method of page 603 shows a single symmetrical peak suggesting only one group: AcA, Tn and ThA, Rn and RaA. Several α-emitters remain to be investigated to this end.

Now as to the theory. Inspecting (18.3), one sees that a Q-value is a value of a certain linear combination of three rest-masses, and a reaction of α-emission exhibiting several Q-values is disclosing several different values of this combination. If we repeat the procedure applied in Chapter XV to cases of artificial transmutation, we assume that ${}_zR^M$ and ${}_2He^4$ are always the same and that different Q-values imply different values of ${}_{z-2}R^{M-4}$: that is to say, we assume that the initial nuclei are always in the same state, the residual nuclei sometimes in one and sometimes in another of several different states. That assumption was acceptable in the cases of Chapter XV, where the bombarded nuclei were stable and had had all the time in the world to arrive at their normal states before the bombardment commenced. It is still currently made for the processes of α-emission involving fine structure. Yet the initial nuclei have themselves not been in existence for an indefinite time; and in particular the nuclei RaC′ and ThC′ and AcC′ have such fantastically small values of half-period as 10^{-6} and 10^{-11} and 10^{-2} second. It is therefore quite admissible that they should themselves be in excited states up to and at the moment of α-emission, and this in fact is the supposition currently held of the C′-products about the nuclei which eject the long-range particles. If this is correct, then when equation (18.3) is written for any one of the three C′

products the residual nucleus $_{Z-2}R^{M-4}$ is always in the same state, and different values of Q correspond to different states of the initial nucleus.

Interrelation of energy and half-period for emission of alpha-particles: the Geiger-Nuttall relation

One of the broad vague rules of radioactivity is, that the longer the half-period of an α-emitter the slower-moving are the α-particles which it emits. This is a rule over which one is likely to stumble at first, for it involves two time-rates—the rate at which α-particles come out of a great assemblage of nuclei, and the rate of speed with which the individual particle comes out of the individual nucleus. When the student can keep these two rates clearly distinct, he may say that the more hurriedly the α-particles leave in the one sense, the more hurriedly also they leave in the other.

Even in this vague form, and even when one compares α-emitters belonging to one and the same series, the rule is not without exception: comparison of AcX and RdAc shows that the former exhibits both the shorter half-life and the slower alpha-particles, and comparison of ionium and uranium II reveals—according to very recent data [1]—the same situation. Nevertheless, there is value in expressing it in the sharper though inexact form of the "Geiger-Nuttall relation" —or rather the Geiger-Nuttall relations, for actually there are three of these, one for each of the three great series of natural radio-active bodies. A Geiger-Nuttall relation between decay-constant λ of the α-emitters and range R of their alpha-particles is of the form:

$$\log \lambda = A + B \log R \qquad (18.4)$$

If it were exactly correct for (say) the uranium-radium series, the constants A and B could be so chosen as to define a line

[1] Taking the previously-established values for the ranges characteristic of Io and U II and the half-period of the former, and computing the half-period of the latter from the relative abundances of the isotopes 238 and 234 of uranium as lately found by Nier, radioactive equilibrium among them being assumed.

passing exactly through the several data-points of Figure 18.6 (notice that this is a "log-log" plot, logarithm of decay-constant along the vertical axis and logarithm of range along the horizontal). Figure 18.6 shows that it is not exactly correct, yet close enough to be interesting and important. The like can be said of the thorium series, and of the actinium series also if actinium X is disregarded.

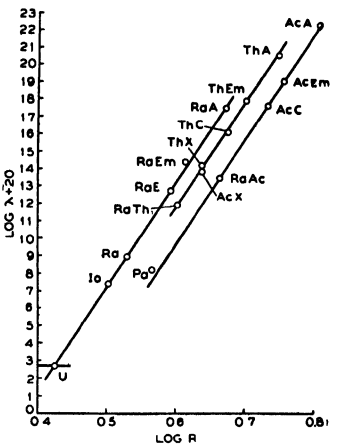


FIGURE 18.6—THE GEIGER-NUTTALL RELATIONS FOR THE THREE GREAT SERIES OF NATURAL RADIOACTIVE SUBSTANCES.

The constant B of (18.4) is given by Hevesy and Paneth as 53.9 for all three series, signifying that the decay-constant varies about as the fifty-fourth power of the range! I know of no other relation between physical variables in which so high a power occurs. As an illustration, I recall that the ranges of the alpha-particles from uranium I and from polonium differ by less than 50%, but the two half-periods stand in the ratio of 10^{10} to one. Different values must be assigned to A to fit the three series; Hevesy and Paneth give it as -37.7 for the uranium-radium sequence.

The range-values used in the Geiger-Nuttall relations refer to the main group, in the cases of the C′-products; while for α-emitters exhibiting fine-structure they correspond (I take it) to mean range-values obtained with apparatus of resolving-power too low to separate the groups. Owing to the relations among range, α-particle speed, α-particle energy, and Q-value, it is probable that the logarithm of any one of these last three comes as close to being a linear function of $\log \lambda$, as does $\log R$ itself. Still other formulae for the interrelation have been proposed.

People have used the Geiger-Nuttall relations for esti-

mating the λ-values of substances for which R had already been measured, but not as yet the half-period. Such estimates have now been superseded by more direct knowledge, except for thorium C′; for this the extrapolation of the line of Figure 18.6 indicates a half-period of 10^{-9} second, while the experimenters can only say that it is too short to be measured.

The exponential law and the facts embodied in the Geiger–Nuttall relation furnish what was the earliest and is still the strongest support of the nucleus-model of Figure 15.6 with its potential-barrier penetrable to charged particles according to the precepts of quantal theory. Just as (according to the quantal doctrine) an alpha-particle approaching from outside has a possibility of entering the potential-well even though its energy E_0 be lower than the top of the barrier, so does an alpha-particle initially within the well with such an energy-value have a possibility of escape whenever in its presumptive to-and-fro motions in the well it is bumping against the enclosing barrier. Interpreting E_0 as the kinetic energy which the particle will have if it makes its escape and gets completely out of the circumnuclear field—to wit, the kinetic energy which it will have when it reaches the observer's measuring-apparatus—we foresee that the probability of passage through the barrier will decrease very rapidly as E_0 diminishes below U_m of Figure 15.6. The half-period will depend on this probability and on the number of bumps made per unit time by the α-particle against the barrier. The quantitative theory seems difficult, and is even more difficult than it seems.[1] The resulting expression, depending as it does on the precise shape and scale assumed for the curve of Figure 15.6—depending, that is to say, on quantities largely unascertainable which to a great extent can be adjusted to fit any data—seems to disappoint all hopes of a convincing test.

One essential test can however be applied, which the

1 Gurney and Condon, *Nature* **122**, 439 (1928), **PR 33**, 127 (1929); Gamow, *Atomic Nuclei* (Oxford Univ. Press), Chapter V. The values of r_m computed by this theory must be regarded as lower limits if, as Bethe suggests (**PR 50**, 677; 1936) the α-particle exists as such only occasionally in the potential-well.

theory must pass; and pass it does. Let the shape and the scale (*i.e.* the values of r_m and U_m) be so chosen that for some particular α-emitter, RaA for instance, the theoretical expression gives the correct half-period when for E_0 is substituted into it the observed energy of the α-particles. Now let precisely the same values of r_m and U_m be postulated for some other α-emitter, RaC′ for instance, and let the half-period corresponding to the observed energy of these α-particles be computed. One should not expect a perfect agreement, since the two nuclei are not identical; but we should be disconcerted by a sharp disagreement, since both nuclei belong to elements of which the nuclear charges differ at most by only a few per cent and the nuclear masses by little more. A very great disagreement would in fact be gravely injurious to the theory. Making the test, Gurney and Condon found, however, that there is no grave disagreement: the theory survives the test.

A very similar test was applied with greater minuteness by Gamow. For each element he assumed a potential-hill having a vertical rise on the inward side, and on the outward side a curved slope conforming exactly to the function $(Z - 2)e^2/r$, where Z stands for the atomic number of the element before the alpha-particle quits it and consequently $(Z - 2)e$ stands for the charge of the residual nucleus. In other words, he postulated a classical inverse-square electrostatic field from infinity inward to a distance r_m from the centre of the nucleus, and at r_m a discontinuous potential-fall. This is a potential-distribution distinguished by a single disposable constant, to wit, r_m; for U_m itself is determined by r_0 and Z.

Gamow proceeded to compute what value must be assigned to r_m in order to achieve agreement between theory and experiment for each of the twenty-three alpha-emitters. Had the values proved very different for the various alpha-emitters, it would have spoken ill for the theory; but all the values were comprised between 8.9 and 5.7 times 10^{-13} cm. The order of magnitude is satisfactory; the differences between the several values are by no means disagreeably great; and there

are even signs of a systematic upward trend of the values of r_m with the atomic numbers of the nuclei. Quite apart from these numerical agreements, the theory offers an instant explanation of a fact that long mystified Rutherford: the fact that α-particles emerging from nuclei of uranium have kinetic energy distinctly smaller than they would have, had they travelled outward through the whole of the region in which the existence of an inverse-square field is confirmed by the experiments described in Chapter V.

Distribution-in-energy of positive and negative electrons emitted from nuclei (β-emission)

A strange and mysterious quality is displayed by each of the many processes of nuclear self-transmutation (nine times as numerous as those of α-emission!) in which from the unstable nucleus there springs an electron of either sign. There is no single group, nor a finite number of separate groups; *there is only a smooth continuum. The electrons emitted from radioactive nuclei of any one type have a continuous distribution-in-energy.* What makes this mysterious is, of course, the implication of a continuum of Q-values, which (if accepted) in its turn implies a continuum of permitted energy-values and permitted states for either the initial nucleus or the final nucleus (or both) in the reaction of the type,

$$ {}_{Z}\mathrm{R}^{M} = {}_{Z\pm1}\mathrm{R}^{M} + {}_{\mp1}e + Q \qquad (18.5) $$

The notion of a continuum of permitted states is unprecedented and unwelcome. Moreover, if the initial nucleus is the residue of a process of α-emission or of artificial transmutation, there is good reason (as we have just observed) for believing that it must always be in the same state or in one or another of a few discrete states; and if the final nucleus is either stable or liable to α-emission, the same presumption holds. There remains a possibility that the final nucleus might at the moment of formation be in any one of the states of a continuum, and

convert itself almost at once into a normal nucleus by emitting a photon; but no gamma-radiation such as this idea would require has as yet been observed. Most theorists of the present time have adopted the dubious idea that whenever a nucleus emits an electron it also emits an extra particle both chargeless and imperceptible, and equation (18.5) is incomplete because of taking no account of this particle.

Facts that lead to so singular a conclusion must be firmly certified, and evidently these are. As to the emission of *positive* electrons, there is no doubt: the distribution-curves

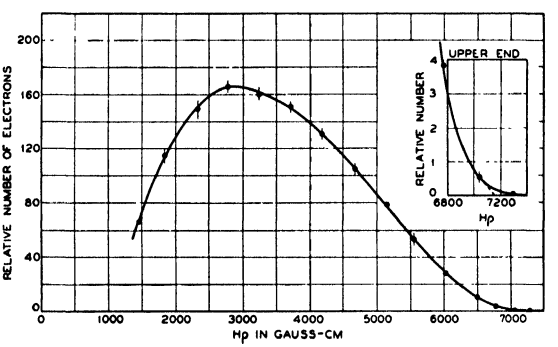


FIGURE 18.7—DISTRIBUTION-IN-MOMENTUM OF THE (NEGATIVE) ELECTRONS EMITTED BY THE ARTIFICIAL RADIOACTIVE SUBSTANCE P^{32}. (Lyman; *Physical Review.*)

are extensive, single-peaked and smoothly-rounded, like the one exhibited in Figure 18.7. As to the much-earlier-discovered emission of negative electrons, doubt subsisted for a long time. The distribution-curves of these are likely to be serrated, with many sharp peaks rising from a smoothly-rounded base. On a photographic plate which the electrons reach after passing through a deflecting field the peaks appear as lines—the "lines of the β-ray spectrum"—with a continuous background between. We have already encountered these lines in Chapter VIII, and the meaning thereof. They are due to electrons which did not come out of the nuclei, and, therefore, have no right to figure in such an equation as (18.5). These (negative)

electrons were ejected from the orbital electron-families; the nuclei were ionizing and depopulating their own planetary systems, by a direct transfer of energy out of themselves. This is the process which we visualized (in probably too crude a fashion) as the ionization of the atom by a photon proceeding from the nucleus thereof; and to which we shall soon return, to use it as an index of the emission of gamma-rays from kernels. For this interpretation of the β-ray peaks the evidence (Chapter VIII) is so strong that even though no one would claim that every line in every β-ray spectrum has been properly correlated, yet it would be a hardy physicist who at present would attribute any such peak to electrons fresh from the nucleus. For these true electrons of transmutation, then, no place remains excepting the continuous spectrum. To these negative indications are added the positive items of evidence, that certain emitters of negative electrons (a good example is P^{32} to which Figure 18.7 refers, and another is RaE) exhibit only a continuum, appearing on the photographic plate as a single broad hazy band without superposed lines; and that the number of electrons appearing per-unit-time in a continuum due to a known quantity of a known β-emitter is about the same as that of the number of kernels disintegrating during that time, while a far smaller number appears in the peaks.[1]

(The foregoing paragraph falls far short of portraying the confusion which for years resulted from the fact that negative electrons bear no infallible earmark of their origin, whether from nucleus or from orbital family. Even the emission of α-particles from α-emitters is often attended by efflux of electrons, now believed to come from the orbits; and one wonders not a little how that part of the displacement-law which applies to β-emitters could have been divined as long ago as it was. As to positive electrons, our present atom-model excludes them

[1] The measurement is made upon a β-emitter which is in radioactive equilibrium (page 600) with a known quantity of some α-emitter. Such measurements have been difficult and not precise; they have indicated a total number of electrons 10% or so greater than the number of disintegrating nuclei, and the excess has been ascribed to orbital electrons not recognized as such. See Gurney (**PRS 109**, 540; 1925) for RaB and RaC, and Emeleus (**PCPS 22**, 400; 1924) and Riehl (**ZP 46**, 478; 1928) for RaE.

from the orbital family and thus seems to fix the nucleus as their place of origin. Yet for electrons of both signs, there is always the possibility that they are created in pairs in the neighborhood of the nucleus.)

The distribution-in-energy of these electrons of the continuum has been studied of late with ever-increasing attention. In some experiments they are deflected by an adjustable magnetic field to a fixed ionization-chamber or Geiger counter, while sometimes the observer uses a cloud-chamber in a fixed magnetic field, and subjects himself to the labor of measuring the curvatures of many hundreds or even of thousands of photographed tracks. (Incidentally, it is often the distribution-in-momentum of the electrons which is plotted and published, for reasons of convenience which should be obvious.) The cloud-chamber is esteemed in particular because the electrons need not pass through any solid matter, and the slower ones therefore are less likely to be missed or to have their initial energies incorrectly assessed. Already it has given strong indications that in the earlier-published curves for certain β-emitters, the portion to the left of the maximum is unduly high and the maximum itself too conspicuous and perhaps too far to the right. Perhaps then it is wise to reserve judgment as to these parts of the curves, and to accept as reliable only the declining arc to the right and the end or end-point thereof.

In using the word "end-point" I am writing as though it were settled that the distribution-curves come down at a slant into the axis of abscissae instead of trailing off asymptotically to that axis. This has been a question at issue for some years, recalling the question of a quarter-century ago as to whether the distribution-curves of photoelectrons from metals come down sharply into the axis and so indicate a true finite upper energy-limit. Some of the distribution-curves of the nuclear electrons have indeed been traced with especial care toward the high-energy end, and have been found to slant sharply toward the horizontal axis. The same may be said of certain integral distribution-in-range curves, obtained by the method described in connection with α-rays on page 491; for although

the concept of range is not so well defined for electrons as for α-particles, yet an integral curve descending sharply to the axis is a reliable token of a distribution-in-energy having a sharp upper limit. Moreover, if there were no upper limit to the electron-energy, there could be no finite Q-value; and the mere fact that in a reaction of β-emission both the initial nucleus and the residue are of definite nucleus-types is sufficient to show that Q cannot be excessively large. This fact indeed may be used to show that the measured end-points are deeply significant, as I now proceed to tell.

Suppose for a moment that for a reaction of β-emission it were possible to ascertain by mass-analysis the rest-masses of both the initial nucleus and the residue, ${}_zR^M$ and ${}_{z\pm1}R^M$ of (18.5). Then either Q would be fixed by (18.5), or else that equation would have to be fundamentally wrong. Well, it happens that mass-analysis has not yet been applied to any two nuclei thus related, and nevertheless there are cases in which we have equivalent data. As an example,[1] I take the radioactivity of nitrogen 13, a reaction of β-emission of which the electrons are positive: ${}_zR^M$ is then the radioactive nucleus ${}_7N^{13}$, and ${}_{z-1}R^M$ is ${}_6C^{13}$. The mass of the latter has (lately) been measured with the mass-spectrograph, not however that of the former. Yet the former is itself the residue of the reaction $C^{12}(d, n)N^{13}$, and therefore its rest-mass is a linear combination $(C^{12} + H^2 - n^1 - Q)$ of the Q-value of this last reaction and the rest-masses of deuteron and neutron and C^{12} nucleus. All these four have been determined, though with different degrees of experimental uncertainty (which, of course, are summed up in estimating the uncertainty of the combination). Computing this combination from what he judged the best available data, and inserting it into (18.5) together with the value supplied by the mass-spectrograph for C^{13}, Cockcroft calculated 1.08 ± 0.10 Mev for the Q-value of the β-emission. Now comes the comparison: the end-point of the distribution-

[1] The examples in this paragraph and the following one are taken from Cockcroft, **PRS 161**, 450 (1937). I leave out most of the mass-values and Q-values employed in the computations, as many are still liable to minor changes.

curve for the positive electrons of N^{13} corresponds to a Q-value of 1.25, with a margin of uncertainty expressed as +0.1, and is therefore concordant with the inference from the masses!

Even had the rest-mass of C^{13} not yet been well determined with the mass-spectrograph, we should not have been helpless. When deuterons impinge on carbon, two alternative reactions occur; one is $C^{12}(d, n)N^{13}$ which we have just been considering, the other is $C^{12}(d, p)C^{13}$. It would be correct to say that the latter reaction permitted physicists to compute the rest-mass of C^{13} from its Q-value and the rest-masses of C^{12} and deuteron and proton; but this gives a wrong impression of what must actually be done to predict the Q-value of the β-emission of N^{13}. The better way of viewing the situation is, that a transition from C^{12} to C^{13} occurs alternatively by two routes: (*a*) by impact of a deuteron, followed instantly by emission of a proton; (*b*) by impact of a deuteron, followed instantly by emission of a neutron and later by emission of a positive electron. Whichever route is chosen, the same amount of mass must be rejected. By route (*a*), this amount is the rest-mass of the proton plus the Q-value (in mass-units, of course) of the (d, p) reaction; by route (*b*), it is the sum of the rest-masses of neutron and positive electron, plus the sum of the Q-values of the (d, n) reaction and the β-emission. The argument thus leads to an equation for the Q-value of the β-emission, solving which Cockcroft again got 1.08 ± 0.10 Mev, again to be compared with 1.25 Mev. Alternative reactions likewise occur when B^{10} is bombarded by deuterons, O^{16} by deuterons and Al^{27} by alpha-particles; and they lead to Q-values for the disintegration of the β-emitters C^{11}, F^{17} and P^{30}, which agree with the end-points of the continua of these.

Thus it is established that if *all* of the electrons emerging from the nuclei of a β-emitter had the same energy as do the fastest among them, not only would the reaction have a definite Q-value but this value would be concordant with the mass-difference between initial nucleus and residue. Yet the vast majority of the emitted electrons have less energy than this seemingly-necessary amount; what then?

Concept of the neutrino

The difficulty to which we have thus been brought is circumvented by contemporary theorists with the assumption that whenever an electron departs from a nucleus there also departs another particle, chargeless and as yet undetected. This particle is called the "neutrino" (some call it either "neutrino" or "antineutrino," according as the electron is negative or positive). It is supposed that the mass or energy of the neutrino is in each case just equal to the mass or energy which the electron ought to take away but doesn't. To put the same idea more formally: equation (18.5) is supposed to be after all a properly-balanced equation, in which however Q is not merely the sum of the kinetic energies K_e and K_p of ejected electron and recoiling residual nucleus, but is the sum of these and a third term ν:

$$ {}_{Z}\mathrm{R}^{M} = {}_{Z\pm1}\mathrm{R}^{M} + {}_{\mp1}e + K_p + K_e + \nu \qquad (18.6) $$

ν standing for the mass or energy of the presumptive neutrino and being always so chosen that the sum of the last three terms on the right is the same whatever the value of K_e. We have just seen the reasons for supposing that ν is zero when K_e has its highest observed value.

Many attempts have been and are being made to detect the neutrino, but as yet without success. One might hope that it would make elastic impacts against electrons and these then manifest themselves by their ionizing-power, but no such recoiling electrons have yet been identified. A famous experiment, though planned before the advent of the neutrino-theory, gave a result which has an immediate bearing upon it. Let a β-emitting substance be surrounded completely by matter thick enough to absorb all of the electrons springing from the nuclei, and let the whole be used as a calorimeter. It turns out that the amount of heat developed in a given time in this calorimeter is equal to the number of nuclei which disintegrate during that time, multiplied *not* by the Q-value *but* by average

kinetic energy of the electrons.[1] Therefore the neutrinos do not give up an appreciable part of their energy while traversing a thickness of matter amply sufficient to take away the whole kinetic energy of their associated electrons.

If there is no neutrino, the departing electron and the recoiling nucleus in a process of β-emission must spring in opposite directions, the latter with a kinetic energy calculable from that of the former. Could these predictions be shown to be at variance with the facts, the case for the neutrino would be greatly strengthened. In the expansion-chamber, the track of the recoiling kernel is unfortunately not long enough to show a well-marked direction, for it should and evidently does consist of a few droplets only. The number of these droplets should however be an index of the kinetic energy of the residue; and very lately Halpern and Crane[2] have counted it for a few examples of the emission of (negative) electrons from Cl^{38}, and conclude in favor of the neutrino.

Disappointing as is the scantiness of the positive evidence for the neutrino, the case for imagining this particle is much stronger than it implies. Not only the difference in energy between initial nucleus and residue, but also the difference in angular momentum, invite us to accept it. It has been mentioned here (Chapter XIV) that the spins of nuclei are "odd" or "even" according as the mass-number is an odd or an even integer—a principle founded on tests of so many stable kernels and on so much theory, that its extension to unstable nuclei seems almost if not quite unquestionable. Now when a process of β-emission occurs, the mass-number of the residue is the

[1] Ellis and Wooster, **PRS 117**, 109 (1927); Meitner and Orthmann, **ZP** 60, 143 (1930). The experiments were performed upon RaE, because of the absence of nuclear gamma-rays and their attendant complications. The average kinetic energy of the electrons (that of the nuclei is small enough to be neglected) is only one-third of the Q-value; the rate of generation of heat agrees with the value computed from the former. It was thought at the time that the electrons of the continuum might be giving up, in the immediate vicinity of the atoms of their birth, energy enough to account for the differences between their observed energy-values and the endpoint of the distribution-curve.

[2] **PR 53**, 789 (1938).

same as that of the initial nucleus, and yet the departing electron carries off with itself its own spin of $\frac{1}{2}$, the abstraction of which from the initial kernel ought to have changed the spin of this last from even to odd or from odd to even, according as it was even or odd to start with! Theorists circumvent this difficulty by assigning to the neutrino also a spin of $\frac{1}{2}$, which saves the situation.[1]

Another situation in which the neutrino is invoked is that presented by the existence of stable isobars belonging to adjacent elements. If two nuclei differ in Z by one unit, and in mass by an amount Δm much smaller than the mass-unit (so that they have the same mass-number) but larger than the rest-mass of the electron, then it is suggested by contemporary theory that the more massive of the two ought to convert itself spontaneously into the less massive by emitting an electron of appropriate sign and appropriate kinetic energy. Yet there are cases of "stable isobaric pairs" exhibiting a difference of one unit in Z; not so many such pairs as used to be thought, and not nearly so many as there are of stable isobaric pairs showing a difference of two units in Z, but still several well-certified cases. It may be that in every such case Δm *is* less than the rest-mass of the electron, but this does not seem likely; and if together with every electron a neutrino must be emitted, then the permissible mass-difference is augmented by the rest-mass of this extra particle. If these Δm-values are ever measured with adequate accuracy, they may serve for fixing the neutrino-mass.[2]

[1] The argument remains substantially the same if the electron, or the neutrino, or both, may have orbital in addition to intrinsic angular momentum—presuming that orbital angular momentum is always "even" for intra-nuclear motions, as already we know it to be for the extra-nuclear electrons. In the following section we shall see that the spin of the nucleus is supposed to change from one to another "even" value, or from one to another "odd" value, in certain cases of β-emission.

[2] See Bainbridge and Jordan, **PR 50**, 292 (1936) and literature there cited. The pairs confirmed by them were at mass-numbers 113 (Cd and In), 115 (In and Sn), 123 (Te and Sb); to these may be added 187 (Re and Os; Nier, **PR 52**, 885 1937). The simple theory quoted here is modified if the possibility of capture of an extra-nuclear electron by the nucleus exists (Yukawa and Sakata, *Proc. Phys. Math. Soc. Japan* **17**, 467, 1935; Sizoo, *Physica*, **4**, 467, 1937).

Mention must here be made of the so-called Fermi and Konopinski-Uhlenbeck theories,[1] even though their origins lie beyond the scope of this book. Based on certain assumptions as to the neutrino, they lead to formulae for the distribution-in-energy of the electrons of β-emission. Each formula has the proper qualitative character, for each describes a single-peaked continuum with an endpoint; each involves the energy-value at this endpoint (U_e, say) as a disposable constant. There are many β-emitters for which the central region of the actual distribution-curve can be well fitted by the latter or "K-U" formula, with suitable choice of U_e; but when this is done it is found that the high-energy end of the actual curve is *not* well fitted, and in particular the actual endpoint is at an energy-value ten or fifteen per cent lower than the chosen value of U_e. Since it is the actual endpoint and not the "K-U upper limit" which is corroborated by the data quoted in the foregoing section, this is a definite fault of the theories, which however are to be thanked for stimulating many of the recent careful studies of continua.

Suppose that the residual nucleus ${}_{Z\pm 1}\mathrm{R}^{M}$ resulting from a β-emission has one or more excited states in addition to the normal state: may the β-emission not lead sometimes to one and sometimes to another of these states (as happens with α-emission) and the observed continuum be a superposition of two or more continua with different endpoints? This is a distressing possibility—distressing because there is no way of distinguishing these presumptive component continua, so long as we lack a thoroughly adequate theory. It is in fact a distressing certainty, for there is good evidence (supplied by gamma-rays and also by alpha-rays[2]) that the residues of many β-emissions do appear in excited states. Deviations of actual distribution-curves from the K-U formula are frequently taken as signs of such a situation.

[1] Fermi, **ZP 88**, 161 (1934); Konopinski and Uhlenbeck, **PR 48**, 7 (1935); see also Peierls, **PRS 161**, 452 (1937).

[2] We have already considered (page 603) the evidence given by the long-range alpha-particles of the C′-products; it bears on the excited states of the C-products. We shall consider a little of the evidence given by gamma-rays in a subsequent section.

Interrelation of energy and half-period for emission of electrons

With β-emission as with α-emission, there is a broad vague rule that the longer the half-period is, the slower-moving the emitted particles are likely to be. In its most careful formulation this is known as the "Sargent relation," after B. W. Sargent who plotted the disintegration-constants λ and the maximum energies U_m (those at the endpoints of the continua) against one another on a log-log scale for a dozen of the natural β-emitters. (Note incidentally that all of these are of high mass-number.) Seven of the points lay close to one slightly bending curve, four close to another, the twelfth by itself. Along each of the two curves lay points belonging to each of the three series, a sharp contrast with the Geiger-Nuttall relations governing the α-emitters. Trying out linear relations,

$$\log \lambda = A + B \log U_m \qquad (18.7)$$

Sargent found values of B between 3 and 7 to be the most nearly adequate.

The Konopinski-Uhlenbeck theory leads (approximately, at least) to the conclusions that all of the points should lie on the same straight line, if all processes of β-emission entail the same difference Δi between the spin of the initial kernel and the spin of the residue; whereas if different integer values are permitted to Δi in Nature, there should be parallel straight lines, of which the uppermost (*viz.*, the one with the greatest log λ for given log U_m) will correspond to $\Delta i = 0$ and the next to $\Delta i = 1$ and so forth. It is therefore conjectured that in Sargent's set of points, the more abundant group belongs to processes in which the nuclear spin remains the same when electron and neutrino leap out, and the other group to processes in which it changes by one unit. Latterly A. C. G. Mitchell has collected the data and made a similar plot for about a score of artificial β-emitters, the electrons being negative for most but positive for some.[1]

[1] It should be said that Mitchell plots the "K-U upper limits" (page 619) instead of the actual endpoints, and that the situation is complicated and confused for substances of which the residues seem to have excited states (see previous section). References: Sargent, **PRS 139**, 659 (1933); Mitchell, **PR 52**, 1 (1937); for theory see Bethe and Bacher, **RMP 8**, 189 ff. (1936).

The points are found to cluster about three parallel straight lines, interpreted as corresponding to $\Delta i = 0, 1, 2$ respectively. Their slope, the constant B of (18.7), has a value of about 6, which is concordant with the theory for not-too-massive nuclei. Values of Δi greater than zero are called "forbidden" but the prohibition cannot be very imperative, to judge from the number of cases which exhibit them. The theory also declares that the phenomena vary with the nuclear mass in such a way that it would be unreasonable to plot Sargent's data and Mitchell's on the same graph; and makes the anomaly presented by the natural β-emitters potassium and rubidium (their half-periods are amazingly long and their electrons rather slow-moving) less difficult to accept by suggesting that for these two processes Δi is 3 or greater.

Process alternative to emission of positive electrons

Suppose a nucleus capable of ejecting a positive electron. It would achieve the same alteration of charge, and the same lack of alteration of mass-number, either by doing this or by absorbing a negative electron out of its own orbital family. This suggests that the latter process may occur as an alternative to the former, but the suggestion was not taken seriously until it emerged from the quantal theories already mentioned under the names of Fermi and Konopinski-Uhlenbeck. Efforts were then made to verify it by looking for K-radiation coming from radioactive bodies which emit positive electrons; for if a nucleus were to capture one of the K-electrons of its own entourage, the event might be revealed by K-rays consequent on the refilling of the vacancy so created in the K-shell (page 429). The first attempt—by Jacobsen on Sc^{43}—led to a negative result; the second—by Alvarez on a block of titanium which had served as target to deuteron-bombardment—to the detection of a strong K-radiation. Several other cases have been reported in the few intervening months, but in these the K-rays are associated with the emission of *negative* electrons. This queer result is interpreted by supposing that (*a*) the positive

electrons escape detection because they are emitted much more rarely than K-electrons are captured, and (*b*) the negative electrons come from the orbits, being expelled either by Auger effect (page 434) supplanting some of the K-rays or else by nuclear gamma-rays. Many oddities in the recorded phenomena of β-emission might be explained if we could postulate the capture of K-electrons occurring *ad libitum* as an alternative to the emission of positive electrons. To say more at this moment would probably be premature.[1]

Gamma-rays and excited states of nuclei

In these final four chapters it has often been said that nuclei may have excited states, and it has often been implied and sometimes stated that they may pass from one to another of these or from one of these to the normal state by emission of photons. These photons are the "gamma-rays."

The gamma-rays first to be discovered were of course those which spring from natural radioactive bodies. It is interesting to recall that they made themselves known before the alpha and the beta-rays, which could not penetrate the wrappings with which Becquerel endued the photographic plates which he placed beside his uranium salts and of which he observed the fogging. The discovery of gamma-rays from artificial radioactive substances followed closely upon the making of these. Shortly before had come the discovery of the gamma-rays from stable nuclei, emitted just after these kernels are born out of transmutation. The penetrating rays observed by Bothe and Becker in 1930, and subsequently proved to consist in part of neutrons (as I told at the beginning of Chapter XVII) comprised in addition a strong admixture of photons.

The spectroscopy of gamma-rays did not begin till after

[1] Experiments: Jacobsen, *Nature* **139**, 879 (1937); Alvarez, **PR 52**, 134 (1937), **54**, 486 (1938); Williams and Pickup, *Nature* **141**, 199 (1938); Oldenberg, **PR 53**, 35 (1938); Ridenour *et al.*, **PR 53**, 770 (1938). For theory see Bethe and Bacher, **RMP 8**, 196 (1936), and literature cited by Pool and Campbell, **PR 53**, 272 (1938). In the early days of transmutation, the apparent inability of nuclei to capture their orbital electrons caused much wonderment.

that of X-rays. In a strict sense it has not even yet progressed very far, since gamma-rays are mostly too short-waved to be diffrangible by crystals; though some remarkable feats of diffraction have been achieved in France by Thibaud and Frilley the latter of whom produced a line-spectrum (of the radiations from RaC) comprising a line of wavelength 0.16X and photon-energy 0.773 Mev.[1] Usually the distribution-in-energy of gamma-ray photons is inferred as is that of neutrons: from the distribution of charged particles to which they have transferred their energy in whole or in part. These representatives of the photons are however *electrons* instead of protons (here, I recall, is one of the chief criteria for distinguishing photons from neutrons in practice).

In the study of *natural* radioactive bodies, the electrons which disclose the γ-spectrum are usually those which spring from the orbital families of those very atoms of which the nuclei are disintegrating: they are said to be released by "internal conversion" of the photons, a process which we considered in Chapter XIV and shall presently review. Sometimes in the study of natural radioactive nuclei, and always in the study of kernels made by transmutation, the photons which emerge from the atoms of their birth are allowed to fall upon targets and the electrons springing from these are observed. The inferences from the resulting "electron-spectrum" are complicated by the fact that three processes may be involved: those which we have studied in Chapters VI and VIII under the names of "photoelectric effect," "Compton effect," and "pair-production." Which of the three is predominant will depend, as said in Chapter VIII, on the energy of the photons; with increasing energy and lessening wave-length, the second gains over the first and then the third over the second; but for a given wave-length the observer can alter to some extent the ratios among the three by properly choosing his target-substance.

During the 1920's attention was directed to gamma-rays of photon-energy 3 Mev or less (often much less), and they were

[1] Thibaud, *Annales de Physique* **5**, 73 (1926); Frilley, *ibid.*, **11**, 483 (1929).

analyzed chiefly by means of internal-conversion electrons from the atoms of their birth or by means of photoelectrons from target atoms. The source would consist of the natural radioactive substance in question, often mixed or covered with the element serving as target. Magnetic field was used to show the energy of the electrons by curving their paths. The apparatus was of the type already so often mentioned in this book, the electrons falling either upon a photographic plate and making a picture like an optical line-spectrum while the field was held constant, or else into an ionization-chamber or a counter and giving a curve with peaks when the field was varied. The peaks or lines[1] are indications of what, by analogy with the usage of Chapter XV, I will call "electron-groups."

The peaks or lines are those which we have already twice encountered in this book: lately, when they had to be set apart from the continuum due to the electrons coming out of the nuclei, and earlier in Chapter VIII, where they were used to prove the classification of the orbital electrons of atoms into electron-shells. In the notation of that chapter, each shows a group of electrons of energy $(h\nu - W)$, with $h\nu$ standing for the photon-energy of some particular γ-line and W for the extraction-energy of some particular electron-shell. There we regarded $h\nu$ as known independently, as it could be if X-rays were used; here we regard W as known, and must infer the value of $h\nu$. But how shall the physicist choose the right W-value for a given electron-group, in view of the facts that often there are many gamma-ray lines from a nucleus and always there are many electron-shells in massive atoms?

Occasionally the answer to this question is simple, as in the case of radium, for which there are three electron-groups only and a common $h\nu$-value (0.189 Mev) is computed for all three by giving to W the values appropriate to the K, L_I and M_I

[1] They are usually much sharper in the spectrum of internal-conversion electrons than in that of the electrons expelled from targets; this is (partly, at least) because the radioactive substance may be spread out in a film much thinner than the thinnest target which is thick enough to yield an observable spectrum. Most of the inferences as to gamma-rays are accordingly drawn from spectra of the internal-conversion type, but those of the other sort have been much studied by Thibaud in particular.

shells; and in that of RaD, for which there are five groups and a common $h\nu$-value (0.0472 Mev) is computed for all by giving to W the values appropriate to the L_I, L_{II}, L_{III}, M_I and N_I shells.[1] Most cases are more complicated. The analyst works over the data until he succeeds in setting up a hypothetical gamma-ray spectrum in which not too small a fraction of the electron-groups is explained by the action of not too great a number of γ-frequencies upon various shells, and reversely there is no too-obtrusive case of an electron-group being predicted from his hypothetical γ-spectrum and failing to appear. Studying RdAc, Meitner and Hahn located 49 electron-groups, and could attribute 33 of them to the action of one or another of 10 types of nuclear photons on one or another of 9 shells; while for AcX they explained 12 out of 21 groups by postulating 5 photon-energies. Fifteen γ-ray lines are listed for RaC by the Cavendish school.

But which is the element from which the photoelectrons come, and of which, accordingly, the W-values must be taken? If the radioactive substance R is surrounded by a sheath of some element E so thick that no electrons from R can get through it, then as to the electrons springing from the outer surface of E the answer is obvious. Suppose now the opposite extreme, in which the source consists at first entirely of R, and is uncovered. The daughter-substance S must, of course, be growing as R decays, but for a while its quantity will be very small compared with the quantity of R as yet untransformed. We might then expect electron-groups coming from R only. However, the test reveals that only the electron-groups from S appear, or anyhow they are by far the strongest.

The test was made repeatedly with a good deal of care, in the years around 1925 when there was controversy over the to-be-expected result.[2] To see the principle, imagine it to be possible to identify pairs of electron-groups produced by single

[1] Numerical values taken from Rutherford Chadwick and Ellis, *Radiations from Radioactive Substances*, Chapter XII (a good reference for the subject of these and the succeeding paragraphs).

[2] Meitner, **ZP 34**, 807 (1925); Ellis, Rutherford, Wooster, Black, **PCPS 22** (1925).

gamma-ray lines operating on two electron-shells, say K and L_I for definiteness. The difference ΔU between the kinetic energies of the groups of such a pair must be practically equal to the difference $\Delta W[=W(K) - W(L_I)]$ between the extraction-energies of the shells; now, ΔW *depends on atomic number.* Meitner applied this principle, often identifying not merely pairs but larger sets of groups due to single gamma-ray lines operating on several different shells. She identified 10 sets of groups for RdAc and 5 for AcX, ascribed to as many gamma-ray lines. The atomic numbers of these two α-emitting substances are 90 and 88 respectively, but the ΔU-values agree with the ΔW-values 88 and 86 respectively which are appropriate to the residues of the α-emission. Suppose that instead one were to allow some of the gamma-rays to fall on another element, platinum for instance, and were to receive the resulting electron-groups on the same plate as those from the radioactive substance. One might then identify pairs of groups produced by single gamma-ray lines operating on the K-shells of the atoms of the source and on the K-shells of platinum atoms respectively, and equate their energy-differences to $[W_{Pt}(K) - W(K)]$, solving the equations then to get the K extraction-energy $W(K)$ of the atoms of the source. Doing this with RaB and RaC, Ellis and Wooster found that the values of $W(K)$ are compatible not with the respective atomic numbers 82 and 83 of these two β-emitters, but with the respective atomic numbers 83 and 84 of their residues. Other variants of the test applied to RaB and ThB led to like results.

The outcome of the test has shown two things: first, that the atoms in which the photon-emission occurs are by far the most likely to eject the electrons; and, second, that the photon-emission occurs *after* the moment of transmutation, when the emitting nucleus is already the residue left behind by the α-emission or the β-emission, already possessed of its new charge and already having organized its own orbital electron-family. Restating the former consequence in the language of Chapter XIV: "internal conversion" is far more likely than extraction of an electron from an atom by a photon coming

from outside. (There is reason for preferring not to speak of photon-emission as occurring when internal conversion takes place, but for our present purpose it may be disregarded.) As for the latter consequence, it is likely nowadays to seem self-evident, which it is not. Even when accustomed to it, one is deplorably liable to forget that gamma-rays belong, not to the substance of which by the present usage they bear the name, but to the daughter-substance thereof.

A gamma-ray spectrum having been mapped, the next step is to construct a system of levels or stationary states of the kernel in question, such that the photon-energies shall be the energy-differences between the states. In general, one would wish to invent the smallest possible number of states needed for interpreting the lines observed, with due regard to the combination-principle which is often found roughly to hold. Apart from whatever support it may give to this principle, nuclear theory is still too rudimentary to offer any aid in making up a system of states or in confirming one when made. There are, however, cases in which an alpha-ray spectrum may be invoked in confirmation. The prettiest of these is furnished by the disintegration of radium, for which the recently-quoted fact about the gamma-ray spectrum implies two states of the daughter nucleus Rn differing in energy by 0.189 Mev. If the α-emission described by the equation,

$$_{88}\mathrm{Ra}^{226} = {}_{86}\mathrm{Em}^{222} + {}_{2}\mathrm{He}^{4} + Q \qquad (18.8)$$

(Em being the symbol of the element 86 of which Rn is the isotope 222) should leave the residue sometimes in the one and sometimes in the other of these states, the α-spectrum should be composed of two groups differing in energy by 0.189 Mev. (More precisely, this should be the difference between the Q-values, but the correction is trivial.) When this inference was first drawn the α-particles from Ra were still supposed to consist of a single group, but the spectrum was formed by Rosenblum upon a photographic plate, and after long exposures a faint companion to the chief line was discovered.[1] The

[1] S. Rosenblum, CR 195, 317 (1932); E. Stahel, CR 195, 317 (1932).

energy-difference was found to be 0.185 Mev, with a sufficient latitude to be concordant with the energy of the photons; and rough estimates of the intensities of the γ-line and the new α-group sustained the obvious forecast that in number-of-particles-per-second they should be equal. Note how this strengthens the presumption (page 605) that fine-structure of an α-spectrum is due to excited states of the *residual* nucleus.

At the opposite extreme in complexity lies the γ-spectrum of RaC, presumed to be associated with the excited states of the daughter-nucleus RaC′. We have already (page 603) considered the spectrum of the "long-range α-particles" of RaC′ which are believed to be associated with excited states of the initial nucleus, to wit, of RaC′ itself. The same system of excited states should then give rise to both spectra; can this be substantiated? It is no small task to make the comparison between two spectra so rich (at least 13 and 15 lines in the α-spectrum and the γ-spectrum respectively) and especially when many of the γ-lines can only be inferred from still a third spectrum, that of the β-rays. Nevertheless, this was done by Rutherford, Lewis and Bowden.[1] To give a mere notion of the status in which the problem was left, I quote the following lines: "When we consider in broad terms the data which have been presented, there can be no doubt that there is a high correlation between the alpha-particle levels which have been observed and the emission of gamma-rays. In more important cases the numerical agreement is well within the experimental error of measurement, while the relation between the intensity of the alpha-ray groups and the gamma-rays associated with them is of the right order of magnitude to be expected on general theoretical grounds. In other cases the agreement is very uncertain, and more definite information on the gamma-rays is required to make the deductions trustworthy. It is unfortunate that we have been unable to detect the alpha-particle groups corresponding to certain

[1] **PRS 142**, 347 (1933); also Lewis and Bowden, **PRS 145**, 235 (1934), for the corresponding case in the thorium series.

postulates [*i.e.*, postulated from the classification of the gamma-ray lines] . . ."

We turn now to the gamma-rays from artificial radioactive substances, the manner of studying which is strikingly different. Internal conversion ceases to be observed, supposedly because of a general law according to which it should be rare in atoms of low atomic number. The photons fall upon targets set up inside Wilson chambers pervaded by magnetic fields, and the tracks of the ejected electrons are observed and their curvatures measured, the observer persisting until he has material enough for plotting a distribution-curve or "electron-spectrum." Both photoelectric effect and Compton effect are usually in operation, and the observer must make an estimate of their relative contributions to the electron-spectrum: the estimate is based on prior knowledge or theory applied to the actual target-substance and the actual range of photon-energies, or on the profile of the peaks in this spectrum. The outstanding work on the artificial β-emitters is that of J. R. Richardson,[1] who alone or with Kurie has studied Na^{24} Mg^{27} and A^{41} which emit negative electrons and N^{13} V^{48} and Cu^{64} which emit positives. For each of these excepting Na^{24} the electron-spectrum exhibits a single notable peak, implying a single unmistakable γ-ray line. Sodium 24 is richer, there evidently being three γ-lines of which the photon-energies are given as 3.00 ($\pm$0.05) Mev, 2.04 and 1.01 Mev. In addition, each of the three which emit positive electrons emit also photons of the proper energy to be interpreted as having been formed in pairs when the positive electrons combine with negatives and convert themselves into light (page 174). The distribution-in-momentum of the electrons emerging from V^{48} is shown in Figure 18.8.

Transmutation by impact is frequently attended by the emission of gamma-rays. While many such cases are already known it cannot be said that our knowledge is broad, for most of the research has been devoted to a few examples for which

[1] J. R. Richardson and F. N. D. Kurie, **PR 50**, 999 (1936); Richardson, **PR 53**, 124, 610 (1938).

the photon-energies are immense. These comprise the impacts of protons on lithium, boron, and fluorine: photons of 16 or 17 Mev are born from the first-named two, and the last provides a γ-line of 6 Mev. They have been studied by means of the cloud-chamber in the manner just mentioned.[1] At these

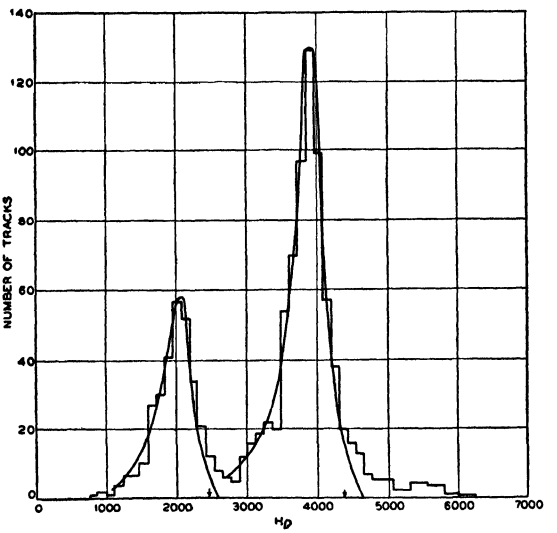


FIGURE 18.8—A GAMMA-RAY SPECTRUM (THAT OF V^{48}, RADIOACTIVE WITH EMISSION OF POSITIVE ELECTRONS) INDICATED FROM THE DISTRIBUTION-IN-MOMENTUM OF "COMPTON" ELECTRONS RECOILING FROM ELASTIC IMPACTS OF THE PHOTONS. TALLER PEAK ASCRIBED TO RADIATION FROM THE NUCLEI OF V^{48}, THE OTHER TO THE PHOTONS FORMED BY MERGING OF POSITIVE WITH NEGATIVE ELECTRONS. (J. R. Richardson; *Physical Review.*)

high energies the photoelectric effect has vanished from the scene, but the Compton effect has a new rival in the phenomenon of pair-production. The photographs often show splendid examples of pairs; the experimenter measures the curvature of each of the coupled tracks, adds the thence-calculated kinetic energies, adds the one Mev corresponding to the rest-masses

[1] Delsasso, Fowler, Gaerttner, Lauritsen, **PR 51**, 527 (1937), **53**, 628 (1938); Gaerttner and Crane, **PR 52**, 282 (1937); McMillan, **PR 46**, 868 (1934), this last being an investigation by measuring absorption-curves.

of the two electrons, considers the sum to be the photon-energy; this he does for a great enough number of pairs to furnish a distribution curve. Usually he also measures the curvatures of the unpaired tracks of negative electrons, and forms a second distribution-curve. This latter he may interpret by assuming that all of these unpaired tracks are due to Compton recoil; or he may make the better assumption that some are members of pairs of which the other members escape detection, and attempt an estimate of the number of these. The inferences as to the γ-spectrum drawn from the two distributions have not always been fully concordant, which shows that something remains to be understood. The situation seems best in the case of boron bombarded by protons, for which both of the curves (as reported by Fowler, Gaerttner and Lauritsen) imply three γ-lines at 16.6 ($\pm$0.6), 11.8 ($\pm$0.5) and 4.3 ($\pm$0.3) Mev.

The reactions in question are thought to be examples of proton-capture ("p-ϕ reactions") in which the protons merge with nuclei of Li^7 and B^{11} and F^{19} respectively, producing excited kernels of Be^8 and C^{12} and Ne^{20} respectively. The photons are supposed to be emitted from these last as they go to or towards their normal states, which Ne^{20} and C^{12} presumably attain, while Be^8 (the "undiscoverable" nucleus which I have already had several occasions to mention) is obliged to resolve itself into a pair of slow-moving alpha-particles. As to this last, we have already found (page 527) how to estimate the energy set free when a proton and a Li^7 nucleus become a pair of nearly-motionless He^4 kernels: it amounts to 17.25 Mev, to which in this experiment 0.44 Mev must be added for the kinetic energy of the impinging protons. This value agrees with the best information as to the highest-energy photons from the presumptive Li (p, ϕ) reaction. Similarly, for protons impinging on boron: the rest-masses of H^1 and B^{11} and C^{12} as most recently determined show that 16.0 ($\pm$0.1) Mev must be released if the first two unite and become the third. Of the three-line spectrum mentioned just above, the highest line conforms to this prediction.

The lower-energy lines reported for the reaction $B^{11}(p,\phi)C^{12}$ —note how the sum of their energies agrees with that of the highest line!—suggests that the nucleus has an excited state which may be at either 4.3 Mev or 11.8 Mev above the normal, and that when going from the 16-Mev state to the normal it chooses sometimes to stop there momentarily and sometimes to go direct. This item of information refers, I recall, to the C^{12} kernel. Its level-system should be well mapped out in the course of time, for there are several different reactions of which C^{12} is the residue, and all of them have evidence to offer. There is, for instance, the reaction $Be^{9}(\alpha, n)C^{12}$ already so much discussed; its γ-spectrum has been studied by Bothe and his colleagues, who enclosed a mixture of beryllium and polonium in a sheath of lead and by magnetic deflection obtained the distribution-curve of the electrons (Compton-recoil electrons mainly) springing from the outer surface of the sheath. Groups of neutrons have been observed from this reaction, and each is supposed to give the energy-value of a state of the C^{12} kernel; so also the groups of neutrons from the reaction $B^{11}(d,n)C^{12}$ and of alpha-particles from $N^{14}(d,\alpha)C^{12}$. All of these diverse phenomena concur in indicating an excited state of the C^{12} nucleus about 4.3 Mev over the normal.[1] There must be other states, and possibly some day the codified knowledge about the level-systems of this and other kernels may rival that already assembled about their electron-families.

But is the 16-Mev state of the C^{12} nucleus really a definite one? There is another possibility; to show which, I put H^{1} and B^{11} for the rest-energies of the nuclei so styled, C^{12} for the rest-energy of the carbon-12 kernel in its *normal* state, K_1 for the initial kinetic energy of the proton, K_2 for the calculable kinetic energy which the C^{12} kernel must have at formation in order to conserve momentum. The energy of excitation is then $(H^{1} + B^{11} + K_1 - C^{12} - K_2)$. If the proton-capture could occur for any value of K_1 over a continuous range, this expression would have a continuum of values, and there would be no

[1] Data collated by Fowler *et al.* (PR 53, 628, 1938); for work from Bothe's school, see Bothe, ZP 100, 272 and Maier-Leibnitz, ZP 101, 478 (1936).

definite excited state. Conversely, if there is an excited state, the proton-capture must be possible only at a definite K_1 value; there must be resonance in the production of γ-rays, in the same sense as earlier I spoke of resonance in the production of protons from (α, p) reactions. Now for all the three (p, ϕ) reactions with which we have been concerned, experiment gives a beautifully clear-cut answer. There *is* resonance, and

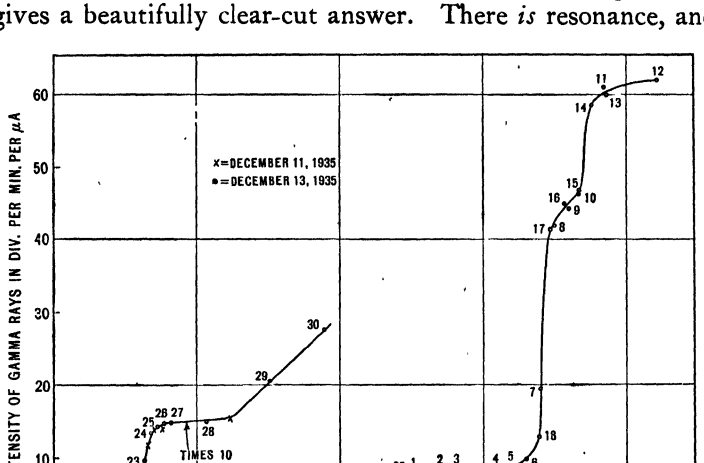


FIGURE 18.9—YIELD-CURVE FOR THE PHOTONS RESULTING FROM IMPACTS OF PROTONS ON A THICK TARGET CONTAINING FLUORINE. (Hafstad Heydenburg and Tuve; *Physical Review.*)

it is remarkably sharp, as evinced by the yield-curve of Figure 18.10 which refers to the bombardment of fluorine by protons. This is an integral yield-curve obtained with a thick target (page 488) so that the sharp resonances are manifested by steep rises at 0.440 Mev. The resonance of the reaction Li(p, ϕ) at 0.440 Mev is so well marked that it is used for calibrating high-voltage devices.[1]

[1] Hafstad, Heydenburg and Tuve, **PR 50**, 504 (1936); Bothe and Gentner, **ZP 104**, 685 (1936). Three resonances are indicated for the reaction F(p, ϕ) and two or more for B(p, ϕ), suggesting an equal number of excited states for the residues.

And so, quite suddenly, we come to the end of that portion of contemporary physics which it has been possible to put between the covers of this book. The situation seems to call for an exposition of nuclear theory to terminate the work. I have however preferred to give the available space to nuclear data instead, they being mostly exempt from doubt while the theory is still difficult, variable and unsure. That incredibly tiny object at the heart of the atom, which in most of the phenomena of chemistry and physics figures only as a point endowed with mass and charge to hold the orbital electrons in their bonds—that object now develops into a strange new world. Narrow though it be, it is a world into which the physicist can pry. It is, however, not a simple world. Even the lightest and presumably simplest of all nuclei has presented us already with a mystery, in the magnitude of its magnetic moment which impairs the hoped analogy of proton and electron. Of the more massive, those which are stable have impeded us, by the magnitudes of their magnetic moments and their spins, from building models for them out of protons and electrons as formerly was tried. Yet of those which are unstable, some absorb electrons and are thus converted into stable ones, and many emit electrons; and the manner of the electron-emission offers mysteries too, those for the palliation of which the neutrino was conceived.

The tendency at the moment is, to imagine the nucleus as a system of protons and neutrons (to which some would add particles not surely recognized as yet in the extranuclear world). As to the forces which these exert on one another in the narrow room of the nuclear system, the theorist has thus far had much latitude of choice, since in extranuclear space such particles rarely approach one another so closely as to expose his assumptions to test (but experiments on proton-proton scattering and proton-neutron scattering are beginning now to limit him). He also has and uses the right to point out that the problems of three and more bodies have never been fully solved by mathematicians, except by grace of such simplifying factors as exist in the solar system and among the orbital electrons, but in the

nuclei do not exist. It is supposed that the protons and the neutrons are roaming about in the narrow region called the nucleus, being endowed with kinetic energy as well as with potential; and that there are configurations of the system determined by something like quantum-conditions, corresponding to the normal and excited states which experiment reveals. It is supposed that emerging photons and electrons are born at the instant of departure, while alpha-particles may be formed in the nuclear world by temporary aggregation of protons and of neutrons, either to fall apart in time or to fare forth into the outer world. Of late it has been believed (largely by Bohr's suggestion) that particles projected into the kernel from without will share their energy of motion with those already there, and that in the course of time a large part of the energy so shared may be gathered into the possession of another of the nuclear particles, which thereupon will fly out and by its leaving consummate a transmutation. Some have gone so far as to liken the nucleus to a droplet of liquid, making comparison thus with that one of the three phases of matter which has been the hardest to interpret by atomic theory. No symbol of the times could be more apt. The chemistry and physics of the past have culminated now as a translation of a chaos of phenomena into a cosmos of elementary wave-guided particles; not a complete translation of course, very far indeed from being complete, but of notable extent. They who have brought that achievement to its present height now find themselves in view of a second chaos of phenomena, and the task of building a cosmos confronting them anew.

SUBJECT INDEX

Absorption-edges, 256–258
Absorption-spectra: *see* spectra
Alpha-emitters, 583, 589–610
Alpha-particles, alpha-rays:
  charge, 30
  charge-to-mass ratio, 64
  mass, 100
  spin, 465
  identification as He nuclei, 63–64
  emission from radioactive bodies, 583, 580–610
  product of transmutation, 523–536
  scattering by nuclei, 122–138
  agent of transmutation, 473–479
  long-range, 603
  scintillations due to, 63
Alternating intensities (band-spectra), 455
Alternative reactions (transmutation), 503, 530
Angular momentum:
  electron-family of atom, 311, 374, 379, 396–407
  electron (spin), 403
  nuclei, 138, 451–466, 617
  neutrino, 618
  ratio to magnetic moment, *see* magnetic moment.
Atomic masses, 90, 92, 93
Atomic number, 120, 133
Atomic weights:
  chemical, 16, 94–95
  from isotope-abundances, 94–95
  ordering of elements, 108
Auger effect, 434
Avogadro number, 15
Azimuthal quantum-number, 309

Balmer series, 298, 335
Band-spectra, 439–450
Barytron; *see* mesotron
Beta-emitters, 583, 588–589, 610–622
Beta-ray continuous spectra, 610–622
Binding-energy, 248
Bohr magneton, 394, 460
Bohr theory of atom, 303, 329–339, 421
Bragg method, 202
Bragg curve, 271, 520
Breadth of spectrum lines, 187
Bremsstrahlung, 274
Broglie (de) waves, 163

C neutrons, 576
Canal rays, 74
Cloud-chamber, 65
Collisions of second kind, 290–291
Composition of angular momenta, 396
Compton effect, 169–174
Correlation, rules of, 164
Cosmic-ray particles, 58–61, 275–276
Cosmic-ray showers, 275–276
Counters, 518
"Crystal wavelengths" (of X-rays), 28
Crystals, 10–16, and *see* diffraction
Curie (unit), 600
Cyclotron, 40, 507

D-D reactions, 552
Decay-curves (radioactivity), 596–602
Decoupling (of angular momenta), 407–410, 456–469
Degeneracy, 338, 371
Deuterium, 93, 335
Deuteron:
  mass, 100
  spin, 465
  magnetic moment, 466–470
  agent of transmutation, 533–536, 552–561
Deuteron-deuteron reactions, 552
Diffraction:
  at a plane grating, 2–10
  of X-rays and electrons by crystals, 12–15, 190–226
  of atoms by crystals, 226

Displacement-law, 116
Dissociation of molecules, 244
Doublet term-systems, 317, 396, 415–421, 430
Doublets in X-ray spectra, 430–434

Eigenfunctions, 347
Eigenvalues, 347
Elastic impacts:
  alpha-particles and nuclei, 70
  photons and electrons, 169
  electrons and atoms, 277
Electron (charge *e*), 20–33
Electron, heavy: *see* mesotron.
Electron, negative:
  charge, 31
  charge-to-mass ratio, 34–56
  spin, 403
  magnetic moment, 403
  dependence of mass on speed, 54–58
  speed by method of gates, 42–45
  liberation from matter, 45–46
  genesis from radioactive bodies, 583, 588–589, 610–622
  capture by nuclei, 621
Electron, positive:
  appearance in cosmic rays, 58–61, 174–179,
  genesis by pair-formation, *ibid.*
  genesis from radioactive bodies, 588–589, 610–622
Electron-pairs:
  formation from photons, 59, 174–179
  merger to form photons, 178
Electrostatic generator, 512
Elements:
  table of the, 105
  periodic properties, 102–107
  non-periodic properties, 108–111
Emerging particle (transmutation), 478
Excitation of atoms:
  by electron-impact, 281–288
  by light, 288–290
  by excited atoms, 290–291
Excitation of light, 285–289, 323–326, 622–633
Excited states:
  of atoms, 284–296, 326
  of nuclei, 493, 547, 622–633
Expansion-chamber, 65
Exponential decay, 596

Faraday (constant), 18
Faraday's laws of electrolysis, 18–19
Fermat's principle, 157
Fine structure:
  optical spectra, 313–319, 395–407, 415–421
  alpha-ray spectra, 604–606
Fine structure constant alpha, 420
Fluctuations:
  in alpha-particle emission, 597
  in electron-emission, 31–32
Fluorescence yield, 437–439
Forbidden lines, 307

***g*-factor**, 388
g-formula, 402
Gamma-rays:
  genesis from radioactive bodies and transmutation, 622–633
  as agent of transmutation, 549
  production of electron-pairs by, 174–179, 273
  absorption and scattering, 174, 254, 273–274
Gas-kinetic cross-sections, 263
Gates, method of the, 42–45
Geiger counter, 518
Geiger-Nuttall relation, 606
Gerlach-Stern technique, 390–395
  refined by Rabi, 457–464
Group-speed, 145, 179
Groups (in transmutation), 493–494, 546–548
Gyromagnetic effects, 386–390

Half-periods, 596–602, 583
Heat of reaction (transmutation), 483
Huyghens construction, 5, 152
Hydrogen atom:
  Bohr model, 329–339, 421
  Wave-mechanical theory, 363–377, 421–427
Hyperfine structure, 115, 451–457, 594

Indefiniteness, principle of, 181
Inelastic impacts:
  electrons and atoms, 281
  photons and atoms, 289
Integral distribution-in-range curve, 491
Internal conversion, 434, 622, 626
Internuclear distances (molecules), 443
Ionization:
  by electrons, 230–245, 258–272
  by very fast charged particles, 266–272
  by positive ions, 246
  by excited atoms, 246
  by solid metals, 73, 247
  by light, X-rays, gamma-rays, 248–258, 272–273
  of molecules, 242
  of excited atoms, 293
  multiple, 240
  thermal, 247
  total, 265
  primary, 268
  likelihood of, 258–273
Ionizing-potential, 231–237, 285, 304
Ions, positive, 62–90
  mass-analysis, 74–90
  properties, 117
  spectra, 319
Isobaric pairs, 618
Isoelectronic spectra, 317, 415–421
Isomers, 589
Isotopes:
  defined, 82, 582
  stable isotopes tabulated, 92–93
  stable and unstable ($Z = 1$ to 11), 587
  discussion, 80–100
Isotope effect (band-spectra), 448–450

jj coupling, 407

Larmor precession, 380
Lattice, 11, 191, 223–224
Laue patterns, 202, 209
Least action, 157
Least time (principle of), 157
Levels, 251, 308
Line-series, 298, 311
Line-spectra, 297–328
Lines, breadth of spectrum, 187
Linear amplifier, 519
Linear oscillator (quantal theory), 361
Loschmidt number, 15
Lyman series, 299

Magnetic moments:
  atoms, 394
  spinning electron, 394, 403
  nuclei, 466–470
Magnetic moment and angular momentum, ratio of:
  electron-orbits, 379
  spinning electron, 389
  atoms, 395–407
Magneton:
  Bohr, 460
  nuclear, 468
Mass-analysis, 80–90, 237–245
Mass-spectroscopy, 80–90, 237–245
Mass-number:
  defined, 93
  of radioactive nuclei, 589–595
Mean free paths (electrons in gas), 279
Mesotron, 272, 276
Millerian indices, 219
Molecules:
  dissociation of, 244
  spectra of, 439–450
Moseley's law, 432
Multiplets, multiplicities, 317–319, 395–407

Neutrino, 616–619
Neutron, 98, 465, 470, 537–579
  discovery, 537–542
  mass, 551
  spin, 470
  magnetic moment, 470
  as agent of transmutation, 557–565
  energy-measurement, 542
  capture, 561–565
  interception and scattering, 566–579
  slow, thermal, C neutrons, 571–579
  photoneutrons, 549
Neutron-howitzer, 573
Nuclear atom-model: 119, 122–138, etc.
Nuclear magneton, 468
Nuclear chemistry, equations of, 479–484

Nuclei:
tabulation of stable, 92–93
partial tabulations of radioactive, 582, 583, 588
masses, 98–100, 482, 493, 532
spins, 451–470
magnetic moments, 466–470
inverse-square field, 135–136
model of Gurney, Condon and Gamow, 499–504, 608–610
disintegration of, *see* transmutation
instability of, *see* radioactivity

Oil-drop experiment, 20
Orbits in Bohr atom-model, 331–339
magnetic moment and angular momentum, 379
Ortho-hydrogen, 467

Packing fraction, 98
Pair-production, 174–179, 274–276
Parabola-method, 75–79
Para-hydrogen, 467
Paschen-Back effect, 407–410
Pauli principle, 415
Periodic table, 101–121, 410–415
shown, 105
Perturbations, 371
Phase-speed, 143, 179
Photoelectric effect:
external (of metals), 166
ionization of gases by X-rays, 248–258, 272–274
compound, 434
Photons; *see* X-rays, gamma-rays, photoelectric effect, ionization, spectra
Photoneutrons, 549
Planck's constant h, 164, 168, 275, 301
Polarization in Zeeman effect, 382, 410
Positron; *see* electron, positive
Potential-barrier:
at surface of metals, 167
of nucleus, 499–504, 608–610
Powder method, 203, 209
Precession, Larmor, 380
Pressure of light, 181
Proton:
mass, 100, 336

Proton:
spin, 451–466
magnetic moment, 466–470
as agent of transmutation, 523–533, 558
as product of transmutation, 473–499, 559–561
projection by alpha-particles, 474
scattering with protons, 137
scattering with neutrons, 570

Q-value: *see* heat of reaction.
Quantum-numbers:
j, 316; k, 309; l, 365–367, 373–377; m, 383–386, 399; n, 316, 365, 415–421; s, 415.

Radiation-potential, 232, 235, 296
Radioactivity, 114, 489, 580–633
artificial, discovery of, 585
tabulations, 582, 583
theory, 608–610
Raman effect, 289–290
Ramsauer effect, 280
Range, 272, 487, 491–499
Refraction by wave and corpuscle theories, 151
Refraction of X-rays, 51
Relativistic mechanics, 52–54,
test by e/m of electrons, 54–56
tests by transmutation, 56–57, 481–504
Residue (transmutation), 478
Resonance (in transmutation), 494–493, 548
Resonance-neutrons, 579
Resonance-potentials, 284, 304
Resonance-radiation, 288, 457
Rest-mass, 53
Reverse reactions (transmutation), 523, 563
Rotation of molecules, 439–450
Russell-Saunders coupling, 403–407
Rydberg constant R, 298, 306, 332, 420

Sargent relation, 620
Scattering:
alpha-particles with nuclei, 122–138
protons with protons, 137

Scattering:
protons with alpha-particles, 474–476
protons with neutrons, 570
neutrons with nuclei, 566–579
X-rays with matter, 169–174
Schroedinger wave-equations, 359–361
Scintillations, 63
Selection-principles, 306, 384, 409, 431
Self-reversed lines, 328
Series:
in optical spectra, 298, 311
in X-ray spectra, 425–434
of natural radioactive bodies, 582, 583
Shells (of electrons in atoms), 251, 410–415
Showers, 275–276
Single-line spectra, 323–325
Singlet term-systems, 317, 404–405
Slow neutrons, 571–579
Space-charge, 52, 233
Space-quantization:
theory, 385
verification, 392
Spectra:
band (optical), 439–450
line (optical), 297–328
single-line, 323–325
X-ray, 252, 256–257, 274, 427–434
alpha-particle, 602–610
beta-ray continuous, 610–622
gamma-ray, 620–633
mass, 80–90, 237–245
electronic, 250, 611
Speed of light in refractive medium, 155
Spin: *see* angular momentum.
Stark effect, 339
Stationary states, 284 *ff.*, 308
Statistical weight, 373
Stern-Gerlach technique, 390–395
refined by Rabi, 457–464
Stokes' law, 23
Stopping-power, 475, 487
Surface potential-barrier (metals), 167

Terms (spectrum), 302–304
primed and unprimed, 322
Transitions, 302
Transmutation, 471–633
apparatus for observing, 63, 65, 489, 517–522
Transmutation, reaction-types of:
(α, p), 473–499; (α, n), 545–549, 586; (d, α), 533–536; (d, n), 556–557; (d, p), 559–561; (n, α), 559–564; $(n, 2n)$, 558; (n, ϕ), 561–565, 587; (p, α), 523–533; (p, n), 558; (p, ϕ), 631; (ϕ, n), 549–551, 587; $(D\text{-}D)$, 552–556
Triplet term-systems, 317, 404–405

Uncertainty, principle of, 181
Unit-cell, 15, 223
Units in spectroscopy, 305

Van de Graaff generator, 512
Vibrations:
of molecules, 439–450
in wave-motion, 139–143
of stretched string, 342
of tensed membrane, 348
of ball of fluid, 352

Wave-equations, 359–361
Wave-mechanics, 340–377
Wavelets, 3–8
Waves, correlation of, with corpuscles, 139–189
Wilson chamber, 65

X-ray refraction, 51
X-ray spectra:
absorption, 256–258
continuous, 274, 433
doublets in, 432–434
emission, 252, 425–434
selection-principle, 431

Yield-curves of transmutation, 485–491, 527–529, 633

Zeeman effect, 339, 381–407, 455–457
Zero moments, method of, 462

AUTHOR INDEX

Agostino, d', 564
Akesson, 284
Akiyama, 70
Allison, 173, 437, 491
Almy, 450
Alvarez, 621, 622
Amaldi, 564, 579
Ampere, 38, 379
Anderson, C. D., 58, 59, 177, 272, 276
Angstrom, 204, 263, 305
Arago, 155, 156
Archimedes, 23
Aston, 83, 86, 95, 98, 99, 255, 527, 532, 555
Auger, 70, 129, 131, 133, 134, 135, 436, 437, 438, 439
Avogadro, 17, 95

Babcock, 449
Bacher, 472, 620, 622
Back, 409, 457, 458
Backlin, 17, 27
Bainbridge, 86, 87, 88, 89, 90, 94, 98, 527, 532, 560, 618
Balmer, 298, 312, 313, 314, 333, 335, 422, 424
Barkla, 437
Barnett, 387, 388, 389
Barton, 512, 516, 571
Bates, 134, 387, 389
Baxter, 105
Bearden, 17, 28, 29, 51
Becker, 537, 538, 539, 545, 622
Becquerel, 581, 584, 622
Bergmann, 308, 312
Bernardini, 547
Bestelmeyer, 55
Bethe, 90, 91, 92, 269, 472, 492, 527, 532, 533, 554, 556, 571, 608, 620, 622
Bhabha, 276
Bichowsky, 422
Bieler, 133, 136
Birge, 18, 27, 31, 449
Black, 625
Blackburn, 447
Blackett, 58, 59, 70, 71, 72, 129, 131, 134, 135, 136, 137, 138, 175, 177, 479
Bleakney, 83, 84, 85, 88, 94, 239, 240, 243, 244
Bocciarelli, 547
Boggs, 464
Bohr, 107, 301, 302, 303, 304, 310, 312, 329, 331, 332, 333, 335, 336, 337, 339, 340, 341, 365, 375, 377, 394, 403, 416, 419, 420, 422, 423, 460, 469, 635.
Bollman, 29, 275
Bonhoeffer, 467
Bonner, 542, 543, 544, 545, 546, 547, 552, 554, 557, 571
Bothe, 107, 176, 537, 538, 539, 545, 549, 550, 558, 592, 622, 632, 633
Bowden, 628
Bowen, 420, 421
Bozorth, 206
Bradbury, 44, 576
Bragg, W. H., 202, 271, 520, 521
Bragg, W. L., 213, 215, 219, 220, 221
Brasefield, 484
Breguet, 154
Breit, 459, 460
Bretscher, 549, 550
Brickwedde, 335, 576
Bridgman, 160
Briggs, 576
Brillouin, L., 146
Brode, 270, 272
Broglie, L. de, 8, 163, 191, 340
Broglie, M. de, 252, 253, 256
Brosi, 449
Brubaker, 542, 543, 544 545, 552, 554, 557
Bucherer, 55
Burhop, 437, 528, 529

Campbell, 622
Cario, 290, 291
Carlson, 276
Chadwick, 58, 129, 131, 132, 133, 134, 137, 138, 177, 497, 498, 517, 541, 542, 548, 549, 550, 625
Chaffee, 44, 49
Chalmers, 549, 587
Champion, 137, 138
Chang, 549, 550, 558
Chao, 179
Chattock, 387
Clark, 569
Clusius, 467
Cockcroft, 504, 506, 515, 521, 524, 525, 527, 530, 531, 533, 557, 560, 614, 615
Cohen, 462
Collie, 570
Compton, A. H., 14, 171, 172, 173, 437
Compton, K. T., 232, 233, 285, 290, 512
Condon, 243, 244, 504, 608, 609
Conrad, 79
Constable, 497, 498
Cooksey, 507, 512
Cork, 109, 590
Corson, 270, 272
Coster, 428
Coulomb, 136
Crane, 177, 274, 544, 617, 629
Crookes, 46, 63, 68
Crowther, B. M., 530
Crowther, J. G., 139
Curie, I., 58, 60, 176, 538, 539, 540, 585, 586, 587, 590, 591, 592.
Curie, Marie, 584
Curie, Pierre, 584
Curtiss, 298
Czerny, 443

Dahl, 512, 513, 516
Dallas, 437
Danysz, 249, 250, 251, 252, 255
Darwin, 134
Dauvillier, 110
Davis, 236
Davisson, 165, 191, 202, 203, 208, 209, 210
Debierne, 596
Debye, 170, 171, 172, 173, 203
Dee, 525, 530, 531, 532, 533, 535, 552, 553, 554, 571
Delsasso, 629
Dempster, 79, 83, 84, 85, 86, 91, 92, 99, 242, 255, 595
Dennison, 467
Dewar, 578
Dirac, 175
Djanab, 542
Doolittle, 528
Doppler, 625
Dunning, 519, 546, 547, 568, 569, 570, 574, 576
DuBridge, 558
Duffendack, 290
DuMond, 18, 29, 173, 275
Dupre la Tour, 539
Dunnington, 49, 50

Einsporn, 287
Einstein, 53, 174, 301, 387, 527
Eisl, 265, 266
Eldridge, 284
Ellett, 457
Ellis, 617, 625, 626
Emeleus, 612
Emo, 549, 551
Epstein, 339
Estermann, 227, 228, 390, 466, 468
Eucken, 467
Ewald, 190

Fajans, 116, 117, 120, 581
Faraday, 18, 19, 48, 64, 65, 89, 192, 202, 209
Feather, 541, 549, 550
Fermat, 157, 158, 159, 161
Fermi, 469, 564, 565, 571, 579, 619, 621
Fink, 573, 574
Fizeau, 154, 155, 156
Fleischmann, 551
Fleming, 27
Fletcher, 27
Foote, 298, 323
Foucault, 154, 155, 156
Fourier, 9, 144, 189, 250, 345
Fowler, 629, 630, 631, 632

Fox, 463
Franck, 282, 287, 290, 291
Franklin, 514
Fraunhofer, 2, 9
Fresnel, 8
Friedrich, 191
v. Friesen, 226
Frilley, 623
Frisch, 228, 466
Fry, T. C., 32
Fuchtbauer, 326

Gaerttner, 629, 630, 631
Galileo, 20
Gamow, 504, 608, 609
Geiger, 31, 130, 132, 133, 134, 504, 518, 583, 606, 607, 608, 620
Gentner, 549, 550, 551, 558, 592, 633
Gerlach, 54, 390
Germer, 165, 191, 202, 203, 208, 209, 210, 211
Gerthsen, 137
Giauque, 449
Gibbs, J. W., 156
Gibbs, R. C., 336, 421, 426
Giddings, 294
Gilbert, 530, 531, 532, 533, 535, 552, 553, 554
Glückauf, 551
Gnan, 226
Goldhaber, 517, 563, 576
Goldsmith, 574, 579
Goucher, 233, 234, 235, 236
Goudsmit, 422, 423, 457, 469
Grace, 426
Gray, 179
Griffiths, 570
Grotrian, 431
Gurney, 266, 503, 504, 608, 612
Guye, C. E., 55

Haas, de, 387
Hafstad, 137, 512, 513, 516, 528, 530, 633
Hahn, 595, 625
Halban v., 578
Halpern, 617
Harkins, 449
Harrison, 300
Harteck, 467, 552
Haworth, 528
Haxby, 524
Heisenberg, 183, 185, 423
Heitler, 276
Hemmendinger, 595
Henderson, 528, 529
Herb, 512, 516, 528, 529
Hertz, 234, 281, 282, 283, 324, 325
Herzberg, 417
Heusler, 389
Hevesy, 599, 607
Heyden, 426
Heydenburg, 137, 633
Heyn, 558, 592
Hill, 524
Hiller, 467
Hoffmann, 519
Hoge, 576
Houston, 426
Hughes, 88
Hull, A. W., 203
Hull, G. F., 181
Hull, W. W., 32
Hupka, 55
Hurst, 598
Huyghens, 3, 4, 8, 152, 158, 161, 162

Immelman, 176
Imes, 449
Irwin, 450

Jacobsen, 621, 622
Jenkins, 450
Jentzsche, 549, 550
Johnson, J. B., 32
Johnson, T. H., 227, 228
Johnston, 449
Joliot, 58, 60, 176, 178, 538, 539, 540, 585, 586, 587, 590, 591, 592
Jordan, 423, 527, 532, 560, 618

Kaufmann, 55
Keesom, 262
Kellogg, 392, 459, 469
Kempton, 532, 554
Kent, 426

Kerr, 292
Kerst, 512, 516, 529
Keyston, 464
King, A. S., 449, 528
Kingdon, 234
Kinsey, 525, 526, 527, 533, 534
Kirchner, 44, 49, 50, 226, 530, 531
Klarmann, 176
Klemperer, 518
Knipping, 191, 205
Kohlrausch, 597
Konopinski, 619, 620, 621
Kopfermann, 426
Kossel, 428
Kratzer, 449
Kruger, 571
Kurie, 540, 629

Lampson, 571
Landauer, 79
Langer, 593
Langmuir, 234, 279
Larmor, 29, 380, 381, 382, 388, 469
Latimer, 599
Laue, 190, 191, 202, 205, 206, 208, 209, 215, 216, 219
Lauritsen, 177, 506, 544, 629, 630, 631
Lawrence, 50, 507, 508, 510, 512, 533, 534, 560, 590
Lebedew, 181
Ledoux-Lebard, 110
Lees, 70, 72
Lenard, 232
Leprince-Ringuet, 176
Lewis, G. N., 534
Lewis, L. C., 393, 604, 605
Lewis, W. B., 530, 533, 557, 560, 628
Libby, 599
Livingood, 561
Livingston, 90, 472, 492, 510, 527, 533, 534, 554, 556
Locher, 272, 436
Loeb, 44
Loomis, 449
Loughridge, 270
Lozier, 244, 245
Lyman, E. M., 611
Lyman, T., 299, 312

Maier-Leibnitz, 632
Manley, 463, 576
Marsden, 130, 132, 133, 134, 475
Massey, 437
Mattauch, 595
Maupertuis, 159
McKellar, 450
McMillan, 31, 178, 590, 629
Meggers, 315, 323
Meissner, 327, 394
Meitner, 58, 60, 617, 625, 626
Mendelejeff, 111
Meyer, 227
Michelson, 142, 154, 155, 156
Michiels, 578
Millikan, 22, 24, 25, 26, 27, 28, 29, 168, 333, 420, 421
Millman, 463
Mitchell, A. C. G., 576, 593, 620, 621
Mitchell, D. P., 549, 551, 574, 576
Mohler, 232, 285, 294, 298, 323
Moon, 578
Mooy, W. H., 262
Moseley, 109, 111, 112, 432
Mott, 137, 138
Mott-Smith, 542, 546, 547
Mueller, 512, 516, 571
Mühlhoff, 599
Mulliken, 455
Murphy, E. J., 576
Murphy, G. M., 335

Naude, 449
Neddermeyer, 177, 276
Neuert, 530, 533
Neumann, 55
Newton, 36, 53, 152, 158
Nichols, 181
Nier, 84, 91, 92, 94, 244, 583, 593, 595, 606, 618
Nuttall, 504, 583, 606, 607, 608, 620

Occhialini, 58, 59, 177
Oersted, 38
Oldenberg, 622
Oliphant, 525, 526, 527, 529, 530, 532, 533, 534, 552, 554

Oppenheimer, 274, 276, 559
Orthmann, 617

Paneth, 551, 607
Parkinson, 512, 516, 528, 529
Paschen, 306, 409, 458
Pauli, 415, 451
Pegram, 570, 574
Peierls, 571, 619
Perrin, F., 70, 129, 131, 133, 134, 135
Perry, 49
Peterson, 599
Philipp, 58, 60
Phillips, 559
Pickup, 622
Planck, 164, 301, 361, 362
Plesset, 274
Poincaré, 596, 597
Pollard, 484, 545
Pontecorvo, 564
Pool, 558, 593, 622
Pose, 490, 492, 495, 496, 519
Preiswerk, 578, 590
Prout, 95, 96, 100

Rabi, 390, 392, 458, 459, 460, 462, 463, 465, 469, 569
Raman, 289, 290
Ramsauer, 280
Rankine, 262
Rasetti, 564, 574, 579, 583
Rayleigh, 112, 156
Rayton, 595
Rechou, 109
Regener, 31
Richardson, 178, 549, 551, 629, 630
Richter, 423
Richtmyer, 257
Ridenour, 622
Riehl, 612
Ring, 297
Roberts, 527
Robinson, 25, 27, 64, 89, 255
Rogers, 134
Rojansky, 88
Rose, 464, 511
Rosenblum, 604, 605, 627
Ross, 171
Rotblat, 551
Rouse, 294
Royds, 63, 64, 65
Ruark, 595
Rumbaugh, 530
Russell, A. S., 116, 120
Russell, H. N., 404, 405, 407, 408
Rutherford, 31, 63, 64, 65, 89, 122, 129, 130, 132, 133, 134, 136, 255, 473, 474, 475, 476, 477, 478, 484, 505, 506, 518, 525, 526, 527, 529, 532, 533, 534, 552, 554, 555, 581, 582, 595, 604, 610, 625, 628

Sagane, 550, 558
Sakata, 618
Salisbury, 10
Salles, 26
Sampson, 85, 94
Sargent, 620, 621
Saunders, 404, 405, 407, 408
Saxton, 578
Scheffers, 394
Scherrer, 174, 203
Schopper, E., 31
Schottky, 32
Schroedinger, 163, 340, 356, 359, 361, 365, 372, 469
Schüler, 464
Schuster, 156
Segre, 564, 574
Shane, 336, 426
Share, 571
Shaw, 49, 50
Shepherd, 524
Sherr, R., 555
Shire, 530
Shrader, 31
Siegbahn, 109, 428
Simons, 176
Simpson, 468
Sizoo, 618
Skramstad, 270
Slater, F. P., 537
Slater, J. C., 422
Smekal, 290
Smith, L. G., 243, 244
Smith, N. P., 491

Smith, P. T., 240, 241, 243
Smyth, 238, 239, 595
Snell, 592
Soddy, 82, 116, 117, 118, 120, 581, 582, 601
Soltan, 544
Sommerfeld, 104, 146, 337, 339, 377, 422, 423, 424, 434, 445
Spedding, 336, 426
Spees, 55
Sponer, 443
Stahel, 627
Stark, 339
Stehn, 571
Stephens, 542
Stephenson, 437, 438, 439
Stern, 227, 228, 390, 391, 392, 466, 467, 468
Stetter, 477, 549, 550
Stevenson, 109
Stokes, 23, 24
Stoner, 415
Stoney, 32
Stranathan, 464
Strassman, 595
Strauss, 299
Sucksmith, 387, 389
Szilard, 549, 587

Tarrant, 179
Tate, 84, 240, 241, 243
Taylor, 563
Terroux, 270
Thibaud, 58, 61, 178, 539, 623, 624
Thomas, 419, 422
Thomson, G. P., 78, 79, 165, 191, 203, 213
Thomson, J. J., 48, 74, 77, 79, 80, 82, 237, 270
Tillman, 578
Townsend, 26
Trautenberg, 529
Tuve, 137, 512, 513, 516, 528, 633
Tyndall, 44

Uhlenbeck, 422, 423, 619, 620, 621
Urey, 422

van Atta, 512, 514, 516
Van de Graaff, 44, 507, 512, 516
Van Vleck, 450
Varney, 234, 576
Vaughan, 240, 243
Vincent, 32
Visser, de, 323
von Grosse, 595

Walke, 598
Walling, 595
Walton, 521, 524, 525, 527, 531, 533, 535
Ward, 604, 605
Webb, 464
Webster, 429, 433, 537
Wendt, 79
Wentzel, 434
Western, 595
Wheeler, 51
Whitaker, 576
White, 421, 425
Wynn-Williams, 519, 605
Whytlaw-Gray, 95
Wick, 572
Wien, 291, 292, 293
Wilkins, 517, 595, 599
Williams, E. J., 269, 270, 622
Williams, J. H., 530, 533
Williams, N. H., 32
Williams, R. C., 336, 426, 427
Wilson, C. T. R., 65, 66, 67, 68, 69, 71, 122, 132, 272
Wilson, H. A., 26, 27
Wilson, R. R., 511, 516, 517, 526, 531
Wood, 326
Wooldridge, 450
Wooster, 617, 625, 626

Young, 8
Yukawa, 618

Zabel, 227, 228
Zahn, 55
Zacharias, 392, 459, 463, 469
Zeeman, 339, 381, 457
Zuber, 176