

Elementary Notions of Quantum Mechanics

KARL K. DARROW, *Bell Telephone Laboratories**

TABLE OF CONTENTS

BASIS OF ALL CORPUSCULAR THEORIES	25
BASIS OF ALL UNDULATORY THEORIES	26
SPEEDS IN UNDULATORY THEORY	28
BEGINNINGS OF THE FUSION OF CORPUSCULAR AND UNDULATORY THEORIES	30
TWO THEORIES MUTUALLY ADJUSTED TO ENABLE BOTH TO COPE WITH DEFLECTION AND REFRACTION	31
PRINCIPLE KNOWN AS "LEAST TIME" FOR THE WAVES, AS "LEAST ACTION" FOR THE CORPUSCLES	33
LEVEL SURFACES OF THE ACTION CONSIDERED AS WAVE FRONTS WITHOUT PERIODICITY	34
BEGINNING OF WAVE MECHANICS: A PERIODIC FUNCTION IS INTRODUCED WHICH FOR ITS WAVE FRONTS TAKES THE LEVEL SURFACES OF THE ACTION	36
FUNDAMENTAL INTERRELATIONS OF CORPUSCLES AND WAVES	36
PHASE SPEED, GROUP SPEED AND CORPUSCLE SPEED OF ELECTRICITY OF MATTER	39
PRINCIPLE OF INDEFINITENESS	39
REFLECTION AND TRANSMISSION OF CORPUSCLES AND WAVES	42
WAVE EQUATIONS	44
SCHROEDINGER'S WAVE EQUATION, ITS PROPER VALUES AND THE STATIONARY STATES OF ATOMS	47
LINEAR HARMONIC OSCILLATOR TREATED BY THE METHOD OF SCHROEDINGER	48
IMPLICATIONS OF THE SYMBOL Ψ	49
FREQUENCIES OF LIGHT EMITTED BY ATOMS COMPARED WITH FREQUENCIES INHERENT IN THE ATOM MODEL	51
VIBRATION FREQUENCIES OF A CLASSICAL ATOM MODEL; FOURIER ANALYSIS	52
TERM FREQUENCIES AND EMISSION FREQUENCIES OF A QUANTUM-MECHANICAL ATOM MODEL; MATRIX ALGEBRA	54
MORE ABOUT MATRIX ALGEBRA	56
DESIGN OF CLASSICAL AND OF QUANTUM-MECHANICAL ATOM MODELS	57
QUANTUM CONDITION IN MATRIX ALGEBRA, AND ITS CONSEQUENCES	58
CONCEPT AND EMPLOYMENT OF OPERATORS IN QUANTUM MECHANICS	61
RADIATION FROM ATOMS INTERPRETED BY CLASSICAL AND BY QUANTUM-MECHANICAL THEORY	65

QUANTUM mechanics is the latest stage of the ancient and stubborn endeavor of men to build a satisfying image of the world out of particles in a void. Also it is the present stage of

another and newer endeavor, the effort to picture the world as a flux of waves in a medium filling space. These seem antagonistic, and indeed were always regarded as rivals until the present century began, for the good and sufficient reason that in earlier times the theories were so formulated that they could not both be used of the same thing at once—they were truly incompatible. That was of little consequence during the nineteenth century, since all the known phenomena of matter were interpretable by the first, all those of light by the second of the images. In those days, it seemed as though the

* This article is substantially the same as one which was composed for publication in Germany, and has been brought out as a book by the Verlag S. Hirzel. I am much indebted to the criticism and advice of my mathematical colleague Mr. L. A. MacColl, and have profited by some of the additions and rearrangements made for the German version by the translator, Dr. E. Rabinowitsch. I have also taken heed of many good suggestions made by an anonymous critic while the manuscript was in the hands of a publisher.

kingdom of nature were sharply partitioned between the rivals, light owning the dominion of the undulatory theory, matter being atomic; as soon as electricity was recognized to be a substance, it was annexed to the domain of the corpuscular theory. But then came the twentieth century, and discovery after discovery broke through the supposed partition, until men were forced to admit that it did not exist at all. Certain phenomena of light demanded that corpuscles be invented, certain phenomena of electricity and matter required that waves be conceived. There was no recourse except to alter both the images until it should be possible to fuse them into one. Quantum mechanics proposes alterations and a fusion; it obliges the particles to submit to the guidance of waves, it humbles the waves to the rôle of pilots to the particles; and in prescribing laws for one, it dictates the behavior of both.

Such a programme makes the impression of being thoroughly revolutionary; and many among the workers in quantum mechanics have helped to confirm that impression, by writing or speaking of the downfall, the overthrow, or the repudiation of classical theories. No one, of course, would deny that the new system is full of radical ideas, nor that its history is a succession of breaches with the past. Yet, the changes which some of the earlier doctrines have suffered are no more remarkable—I should say that they are *less* remarkable—than the changeless survival of others. The Constitution of theoretical physics which our forefathers handed down to us has been drastically amended in several of its sections, but others are intact, and these are not the least important. If one wishes to say that quantum mechanics came about through revolution, one should always add that never has there been a revolution more gradual, more cautious, more tenacious of all the virtues of the old regime.

Indeed, as one reads onward in the literature of quantum mechanics, one is, turn by turn, surprised by the radicalism of some single new idea, and by the immense conservatism of the scientific mind, which can almost never be persuaded to relinquish the successes of an older theory, however avidly it may seize on the advantages of a newer. The nineteenth-century

form of the kinetic theory of gases with its rigid elastic spherules, the electromagnetic theory of light and its vibrating charges and its spreading waves, the Rutherford atom model with its central nucleus and its revolving electrons, the earlier mutations of the quantum-theory—all of them led to striking numerical agreements between experience and theory; and of these, almost all have been saved, hardly any have been renounced. Can we say that the “classical” mechanics of Newton and Lagrange and Hamilton is out of date, so long as atom-models are still constructed by imagining some assemblage of charged and massive particles and writing down the classical expression for its energy as function of their positions and their speeds? True, after writing this expression down we do strange things with it which Hamilton probably never dreamt of, and the model slips out of sight in the metamorphoses of our equations; but in the end we compute certain things which we call amplitudes, and then to get the intensity of the light which streams from the atom we multiply these amplitudes by the very factor, by which in the electromagnetic theory we multiply the amplitude of the vibrations of a charge in order to get the strength of the light which it emits! Maxwell and Lorentz have not been superseded, so long as this is so.

All this has bearing on the knotty question: how shall the beginner attack the quantum mechanics? Shall he follow in the footsteps of the pioneers, first being told of the earlier theories, and then of the several ways in which they have been altered? Or shall he be confronted at once with the present theory as though it had sprung into being complete and perfect, owing nothing to the past?

Well! if quantum mechanics were perfect, much could be said for the latter course. Already there is a tendency among expositors to take it, and to ignore the steps by which the theory has been carried to its present state:

*“He then unto the ladder turns his back,
Looks in the clouds, scorning the base degrees
By which he did ascend. . . .”*

But the present theory is not perfect. Some of its successes are properly those of earlier theories, not its own; some of its predictions are unveri-

fied; no one has yet discovered how to make it cope with certain problems. This strengthens the case for the opposite course—for treating the present theory as a stage in the evolution of thought, derived from the foregoing stages by making modifications in these which are no greater than is strictly needful. It is certainly the easier course for those who have already been broadly trained in general physics, of which the scope of quantum mechanics covers but a smallish part. Furthermore, one can scarcely believe that if quantum mechanics shall one day be perfected, it will isolate itself so completely from yesterday's theories that the interconnections will not be worth retracing.

In this article, therefore, the evolutionary mode of treatment shall be adopted instead of the revolutionary.

BASIS OF ALL CORPUSCULAR THEORIES

How far back should we go into the atomic theories, into the wave theories of earlier days, in order to provide the necessary basis?

To Democritus and to Lucretius we need pay only the tribute of a few respectful words. They spoke and fought for the idea of atoms in a void, Lucretius with splendid eloquence; perhaps without them the idea would never have been started for us to develop; but they had of course no notion whatever of the modern scientist's demand for a quantitative, a numerical agreement between experiment and theory. Of Dalton and the other chemists of the early nineteenth century we need say hardly more. Their atoms possessed no qualities except weight, and hooks to interlock with the hooks of other atoms; they could not give the weight of any atom in ounces or grammes, but only the ratios of weights of atoms of different kinds (frequently with an error, as the discovery of isotopes has proved). With the kinetic theory of gases came actual atom-models—rigid elastic spheres and (for diatomic molecules) rigid elastic ellipsoids—which began to fulfil the physicist's aspirations. Then came an idea of enormous importance: the notion of the atom as a mechanical structure whereof the parts could be made to vibrate about their common centre, which Helmholtz and others invented to explain why light goes more slowly

when it passes through matter. The present-day theory contains that idea, but could not itself be born till after the discovery that electricity is a substance composed of atoms, subtler even than those of matter.¹ When this discovery was made, soon after the turn of the century, modern atomic theory began. Our situation is distinguished in two principal ways from that of the atom-builders who preceded us. We live after that discovery, not before it; and unlike every earlier generation, *we have seen our atoms*.

This statement is perhaps a little too strong, but certainly not by much. I do not say that we have seen atoms in the way in which we see a pebble or a ball when it hurtles through the air; but I say that we have beheld them in the same way as meteors or skyrockets are seen. The reader may now say: "We have not seen meteors, but the trails of incandescent matter that they leave behind them in the night sky; we have not seen skyrockets, but the trains of flaming sparks which they shed in their flight; we have not seen atoms, but the rows of water-droplets which condense upon the ions which a flying atom forms along its path through moistened air—we have not seen these particles, but only their tracks." No doubt! But the track of a corpuscle may be just as clinching evidence for the existence of that corpuscle, as would be its image formed on the retina of a human eye or its feel as it rolled around in the palm of a human hand. Indeed, if the corpuscle in question is too small to be seen or felt, how else can we define it than by saying that it can follow a path and form a track?

We can readily observe the paths of charged atoms—the elementary corpuscles of electricity and of matter—through field-free space and through electric and magnetic fields. The straightness of the paths in field-free regions, their curvatures in fields, show us that the atoms conform to the laws of classical mechanics. They respond to the fields in just the same way as large and tangible pieces of charged matter,

¹ I use "electricity" hereafter in referring to beams or aggregations of electrons, and "matter" in referring to beams or aggregations of atoms (ionized or not). There are strong objections to this usage, but it seems to me on the whole less objectionable than to speak of "waves of an electron" or "electron waves" or "atom waves."

Coulomb's pithballs for example. The curvatures may be measured, and their values give us the ratio of charge to mass for the electron (the atom of negative electricity) and for many kinds of charged atoms of matter. The measurements are often made by methods in which the path is not made visible, but the principles are the same. We then appeal to Millikan's famous experiments with oil-drops, which supply us with the value of the charge which figures in all these charge-to-mass ratios. Knowing this, we can compute the mass of the electron and the masses of all other kinds of atoms. There is one other observable and measurable property of the atom—measured, it is true, by experiments on streams of atoms, not upon individuals; it is the magnetic moment.

Thus, when we wish to form a corpuscular image of negative electricity, we must conceive it not as atoms of arbitrary charge and arbitrary mass and arbitrary behavior, but as particles of definite mass (about 9×10^{-28} gramme) and definite charge (about 4.77×10^{-10} e.s.u.), which behave in the classical way as they move through such electric and magnetic fields as we can produce in the laboratory. We must conceive of sodium (for instance) as particles of mass about equal to 3.8×10^{-28} gramme, and magnetic moment of the order of 10^{-20} c.g.s. unit. Our freedom is thus very seriously restricted; not for us the soaring independence of the philosophers of the past, who were free to assign whatever values pleased them to the sizes and the masses and the numbers of their atoms, secure against all test! But what we lose in freedom, we gain in assurance. An atomic theory in which so much is fixed by first-hand knowledge cannot go wholly astray.

Yet by a strange coincidence this generation, the first in history for which the atomic theory of matter and electricity has been transformed into a fact, has been obliged to introduce an undulatory theory of both. In the study of light it has had the same experience, reversed: it has been required to invent a corpuscular image, but it has not been permitted to abandon the undulatory image which was handed down to it from the past. Twice it has been obliged to unite a corpuscular with an atomic theory, without

destroying either. This is the problem which now engages us.

BASIS OF ALL UNDULATORY THEORIES

We must begin by analyzing the nature of a wave theory, such for instance as the supremely successful traditional theory of light founded by Young and Fresnel, and developed in the first quarter of the nineteenth century. This is more difficult than the analysis of a corpuscular theory, at least for Anglo-Saxons; the English physicist Crowther has wittily suggested that this is because the ball has been for ages the favorite plaything of the Anglo-Saxon races.

The very name calls up a picture of a procession of water waves along the surface of a pool, crest following on trough and trough on crest one after another in an indefinite sequence. Of all its features, the most striking is this visible regular alternation of hills and valleys; the next most striking is the constant speed with which the sequence glides along. These constitute "wave motion" to the eye. A little observation shows us that the water does not move onward with the waves: a floating leaf will rise and fall as the waves pass under it, and have no tendency to follow them. By hydrodynamical theory one may determine how the water itself does move; the motion is not very simple, and perhaps it is better to think of transverse waves on a very long stretched cord. Here the waves move along the cord, a succession of valleys and hills like ripples on water; but each element of the string oscillates sidewise. 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:

$$\begin{aligned}\zeta &= A \cos (nt - mx) \\ &= A \cos 2\pi(\nu t - \mu x) \\ &= A \cos 2\pi(t/T - x/\lambda) \\ &= A \cos (2\pi/\lambda)(ct - x),\end{aligned}\tag{1}$$

wherein the names and meanings of the constants are the following: λ , the wave-length or crest-to-crest distance; μ , the wave number; ν , the frequency; T , the period; c , the speed with which the waves glide along, hereafter to be called the phase speed; and A , the amplitude.

Strictly there should be an additive constant inside each of the brackets, but we may imagine the zero of time so chosen as to annul it.

Instead of a single string running along the axis of x , let us 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 Eq. (1). We now have 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 simultaneous equation for H_y , the magnetic field strength; but for the present purpose we need not remember 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; therefore we do not perceive nor measure the frequency ν . We do not see any waves, hills, valleys, crests, nor troughs; therefore we do not perceive nor measure the wave-length λ . 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*, though different from it in any material medium. Apart from polarization (of which I shall say little) and momentum (to be discussed later on) there are only two qualities of light which we can measure. One of these is interpreted as being proportional to square-of-amplitude (A^2); the other is a speed which may be, but in general is not, equal to the phase speed. The former is the intensity of the light, which in a truly plane-parallel monochromatic beam is the same everywhere and always. Thus, of all the quantities appearing in the expression (1) which describes 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 value when we

superpose a pair of wave trains of equal wave-length but not coinciding in direction. Say that we have two wave trains such as that described by (1), excepting that instead of travelling along the x -direction, they travel along two directions in the xy -plane inclined respectively at angles $+\theta$ and $-\theta$ to the x -direction, and thus intersect each other at angle 2θ . Adding the displacements together, we get:

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

Now we have a wave motion of which the amplitude varies from place to place, being in fact a sinusoidal function of y , the square of its magnitude being given by the equation:

$$(\text{Amplitude})^2 = 4A^2 \cos^2 [(2\pi \sin \theta / \lambda)y]. \quad (3)$$

If we should send two beams of plane-parallel light across one another's paths at an angle 2θ , and should then observe that the intensity of the light varied with place according to Eq. (3)—if we should observe that the intensity varied as the square of the cosine of some constant multiple of y —then the wave model would justify itself.

This is precisely what happens. The experiment which I have mentioned is a typical experiment on what is called the "interference of light." The intensity of the light is found to vary as $\cos^2 (ky)$; here k stands for a constant which has different values for light beams of different colors, and can be determined by measuring (for example) the distance between two successive planes of zero light-intensity, the so-called "width of a fringe." When k is determined and θ is known, the wave-length λ can be calculated; that is to say, we can assign the value which the crest-to-crest distances in the wave models of the two beams of light must have, in order to produce an interference-pattern of the scale which is observed.

Such is the standard and ultimate way of ascertaining wave-lengths. Two beams of light, or more than two, are caused to overlap. There is then a peculiar and characteristic intensity-distribution in space, which is known as an "interference pattern" or a "pattern of station-

ary waves." It conforms with the wave model, provided that we identify light intensity with square-of-amplitude, and provided that we choose the proper value of λ . What is called "measuring wave-length" consists in making the proper choice.

The commonest way of doing the experiment is to send a single plane-parallel beam against a screen with a series of equal and equally-spaced slits, or a mirror with a series of identical and equally-spaced grooves. The beam is then divided into a number of narrow pencils, each emerging from one of the slits or reflected from one of the grooves; these spread out and overlap, and produce the remarkably sharp and distinctive kind of interference pattern which is especially known as a "diffraction pattern." This consists of beams diverging at fixed angles from the direction of the original beam. The wave model is competent to interpret this pattern, provided again that the proper value of λ be chosen. If a beam of a certain kind of light (x-rays) is projected against a mass of powdered metal, there appears a diffraction pattern consisting of cones, having the direction of the original beam for their common axis; this also is amenable to the wave model, and permits a determination of wave-length.

We have now seen that light waves are never observed (in any ordinary sense of this verb), and that the so-called "wave-length" is a property of the wave model which is evaluated by making observations on an interference pattern.

Next we must consider some of the speeds associated with wave trains; for the significances of some of these speeds, the comparative lack of significance of others, involve some of the subtlest points of wave theory.

SPEEDS IN UNDULATORY THEORY

This is by definition the speed at which the wave crests of the undulatory model advance. But we never perceive the wave crests, and therefore cannot measure their speed. The model described by Eq. (1) refers to an unlimited beam of uniform light. In such a beam, nothing visibly moves. All we can do is to 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 experiments upon the speed of light, by any of

the three methods (the toothed-wheel method, the rotating-mirror scheme, the recently-developed method in which Kerr cells are used).

However, the speed with which a segment of a wave train gets from place to place is *not necessarily* the speed of the wave crests. This is a fact of experience: one sees it exemplified in watching trains of ripples on water, where the individual wave crests are visible.

To get at the reason for this fact, we must first realize that a train of waves of finite length is never a perfect sine-wave such as Eq. (1) describes. Even though every single wave of the train should be a perfect sine-curve, even though the train should be broken off with absolute sharpness at both of its ends—even in this apparently ideal case, it could not be regarded as a perfect sinusoidal train. For it is of the essence of a perfect sinusoidal wave train, that it goes on forever in both directions. A segment of a wave train must be regarded as a resultant of superposed sine-waves of the type described by Eq. (1), having various amplitudes and various wave-lengths, which can be determined (if we know the exact shape of the train) by the method of Fourier analysis.

The onward gliding of a segment, therefore, is the result of the onward gliding of the individual infinite wave trains of different wave-lengths, into which it must be resolved. Now if all of these have the same wave speed,—if phase speed is independent of wave-length—the segment will travel along undistorted, with the same speed as that of the wave crests. But if phase speed is not independent of wave-length, the segment will travel at a rate which will differ at least from some, and may differ immensely from all of the phase speeds of its components. Worse yet, it will change its form as it goes along, and in the course of time it may even 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 may not speak of its speed at all, unless we hold in mind that we are speaking of something which is essentially somewhat indefinite, and grows more and more indefinite the longer time goes on.

This is a most important point about wave motion and wave models. Unluckily, the study

of light gives us too scant an opportunity to become familiar with it; for light *in vacuo* exemplifies the simple special case in which phase speed is independent of wave-length, and therefore segments of wave trains—or “pulses,” to use a common word—travel with the phase speed and travel undistorted. This is not however true of light in transparent material media (witness Michelson’s experiment on the speed of light in carbon disulphide), nor of water waves, nor of radio waves in ionized air, nor of electrical waves along cables (in both of which last two cases, the distortion of pulses is a well-known fact of experience); and it is not true of the waves which are introduced in quantum mechanics as partial models for electricity and matter.

There is a special case which figures largely in quantum mechanics: we will work it out at once. Say we have a pair of infinite sinusoidal wave trains, equal in amplitude but of slightly different wave-lengths. Superposing them, we obtain the familiar pattern of an endless sequence of beats (Fig. 1). The maximum or centre 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:

$$\lambda/D = \Delta\lambda/\lambda. \quad (4)$$

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 (Fig. 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

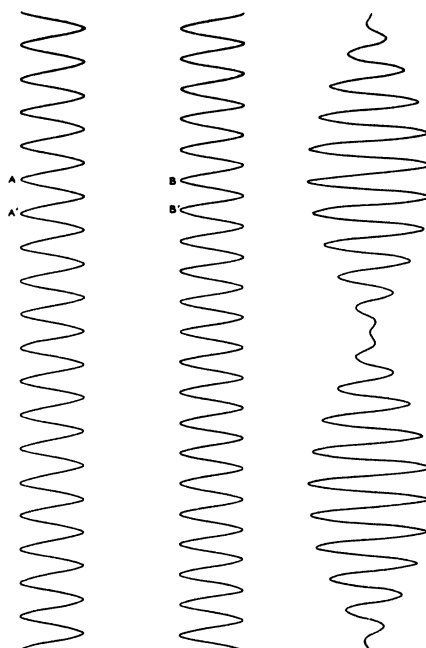


FIG. 1.

relatively to the first; for definiteness 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 B' of the long-wave train coincides with the crest A' 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}

We now deduce the formula for the actual value of the speed of the beats. Denote by v and

^{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. Sommerfeld and Brillouin, *Ann. d. Physik* **44**, 177–202, 203–240 (1914).

$v + \Delta v$ 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 observer. Relative to a crest of the former wave train, the speed of the latter train is Δv , that of the beats is $(g - v)$. 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:

$$(g - v) / \Delta v = -\lambda / \Delta\lambda. \quad (5)$$

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;

$$g = v - \lambda(dv/d\lambda). \quad (6)$$

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

$$g = -\lambda^2(d\nu/d\lambda) = d\nu/d(1/\lambda) = d\nu/d\mu. \quad (7)$$

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

We now return to the development of the main argument.

I have recalled to the reader 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 v 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.

The reader perhaps is wondering why I have spent so much space in recalling what has long been known about light. The reason is this: the wave theory of matter and of electricity has 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 of negative electricity such as we are in the habit of visualizing as a stream of electrons all going in the same direction with the same speed. Such a beam, as everyone knows, can be approximated by setting up a vacuum tube with a hot filament, a cylinder charged to a potential higher than that of the filament, and some metal diaphragms with holes in line with one another and with a hole in the cylinder. 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 an isolated piece of the beam is measurable, but the phase speed is not to be

² This statement, like certain others in the article, requires modification if radio waves are taken into account. The continuity between these and what is commonly 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.

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 carbon bisulphide.

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, as I said, we have seen them. 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 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

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.

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.

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? 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 $z = 0$.

Now imagine a stream of electrons of momentum p , approaching the plane along a direction inclined at θ to the normal—*LM* in Fig. 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

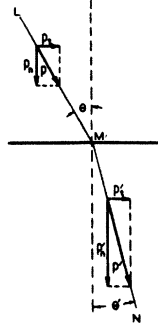


FIG. 2.

is no component of field strength parallel to the plane. Fig. 2 is almost self-explanatory; as one readily sees,

$$\sin \theta / \sin \theta' = p' / p. \quad (8)$$

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:

$$\sin \theta / \sin \theta' = p' / p = \text{quantity independent of } \theta. \quad (9)$$

But this is the law of optical refraction, which, as every physicist remembers, is compatible with the wave theory of light. We have found 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.

Fig. 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

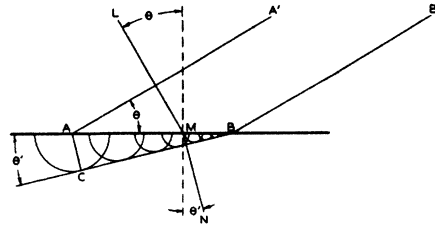


FIG. 3.

instant $t=T$, of the part which has not reached the surface; CB is the trace, at $t=T$, of the part which has traversed the surface. Denote by v and v' the speeds of the wave front above and below the surface. Evidently the perpendicular distance from AA' to BB' is vT ; 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 v' . This argument gives $v'T$ for the length of the line AC , and we readily derive:

$$\sin \theta / \sin \theta' = v / v' = \text{quantity independent of } \theta. \quad (10)$$

We are trying to represent a beam of electricity both as a stream of corpuscles and as a train of waves, which must travel the same course, and must therefore bend alike as they pass across a potential discontinuity. By comparing Eqs. (8) and (10) we see what we must assume in order to achieve this: to wit, the relation:

$$v / v' = p' / p. \quad (11)$$

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:*

$$v = \text{const.} / p. \quad (12)$$

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. Eq. (12) disposes of all chance of assuming that wave speed and electron speed

are the same. We can make the wave train travel along the same course as the electron stream, but we cannot make a wave crest and a particle stay together.

It takes only one more step to arrive at the relation between momentum and wave-length; but despite the fact that wave-length is much nearer to observable things than is phase speed, there are several points which it is better to treat before that step is made.

The condition expressed by Eq. (12) suffices to make corpuscle stream and wave train keep together, not merely at a sharp deflection such as would occur at a potential discontinuity, but also throughout a curved path such as the beam describes in an electric or a magnetic field. It is equally potent for the case of a beam of ponderable matter passing through a gravitational field. In all of these cases the path of the beam may be traced either by supposing it made up of corpuscles whereof the momentum changes because of the force which the field exerts upon them, or by imagining it as a wave train of which the phase speed varies inversely as the momentum computed for the corpuscles.

THE PRINCIPLE KNOWN AS "LEAST TIME" FOR THE WAVES, AS "LEAST ACTION" FOR THE CORPUSCLES

Select a couple of points on the path of the beam, or, to be more precise, on that of one of the corpuscles composing the beam; for instance, *L* and *N* of Fig. 2. Go over to the wave model; the path in question (*LMN* of Fig. 2) now has a new interpretation: it is the curve to which a wave front remains perpendicular, in going from *L* to *N* (Fig. 3). One could now draw an infinity 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 onward from that point to *N*. Each of these would be a conceivable path for a corpuscle going from *L* to *N*; also, a wave front could be conceived as going from *L* to *N* in such a way as always to remain perpendicular to any one of these. None of these, however, is the actual path. The actual path is outstanding among all of these by the feature, that the time spent by the wave front in going from *L* to *N*

along it is less than the time which would be required to follow any other. (In making this statement, it is of course assumed that the wave front would always travel with speed *v* when above the refracting surface, and always with speed *v'* when below it.)

This is Fermat's "principle of least time" which he proved for the case illustrated in Figs. 2 and 3, where every path is straight except for a sharp bend or corner occurring where it traverses a single refracting plane. It can be generalized to the case of a continuously-curving beam traversing a field of force, though then it may be strictly true only when the comparison is made between the actual path and others which differ therefrom by infinitely little. Also it is possible to contrive cases in which the time is a maximum rather than a minimum; so that on the whole, the principle is best entitled "principle of stationary time," and written thus:

$$\delta \int v^{-1} ds = 0, \quad (13)$$

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

This principle is, of course, not true for the corpuscles; the actual path (say *LMN* of Fig. 2) is not in general the one by which they could go in the shortest time from *L* to *N*. This difference enhances the prestige of the waves, which is a good thing, as it diminishes the advantage which the corpuscles have over them in our minds by virtue of being observable. 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 believed in "economy of time" as a fundamental principle of nature, and postulated—this was as long ago as the middle of the seventeenth century—that light should spend the shortest possible time in going from the start to the finish of its path. By postulating this, one reverses the trend of the argument of the foregoing pages, and comes to the conclusion that in such a case as that of Figs. 2 and 3, the speed in the upper medium must stand to that in the lower as $\sin \theta$ to $\sin \theta'$ —the ratio later prescribed for the wave-front speeds by Huyghens.

There is however a corresponding principle for the corpuscles. Evidently it is to be obtained by substitution from (12) into (13); we get:

$$\delta \int p ds = 0. \quad (14)$$

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. To express the idea in a less rigorous but more customary fashion: the line integral (14), taken along the actual path of the particle between any two points thereof, is smaller than the corresponding integral for any neighboring path joining those points.

This is the "principle of least action." 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 electricity or matter as a stream of corpuscles associated with a train of waves, we may make use of both.

There is a more familiar notation for the principle of least action, derived from (14) by simple substitutions based on the classical mechanics (the relativistic formulae will be given later on). Put u for the speed of the corpuscles and m for their mass, therefore mu for their momentum p and $(\frac{1}{2})mu^2$ for their kinetic energy T . We readily deduce:

$$p ds = p(ds/dt)dt = p u dt = 2T dt \quad (15)$$

and the principle of least action takes the form:

$$\delta \int 2T dt = 0. \quad (16)$$

The integral which appears in (14) and (16) alternatively written as $\int p ds$ and $\int 2T dt$, goes by the name of "action." I will denote it by A . It is a function of the two points on the path of the corpuscle between which the integral is taken. If we consider one of the points, say $P_0(x_0, y_0, z_0)$, as fixed, it is a function of the coordinates (x, y, z) of the other point P : one

might call it the "action" which the corpuscle has accumulated on its way from P_0 , by the time it arrives at P . It seems decidedly abstract; we are not nearly so well accustomed to it, as to such functions as energy and momentum; but quantum mechanics confers upon it, or rather upon a function closely allied to it, new prestige and importance.

LEVEL SURFACES OF THE ACTION CONSIDERED AS WAVE FRONTS WITHOUT PERIODICITY

The important feature of action, for our present purpose, is due to the fact that we may write

$$p ds = p_x dx + p_y dy + p_z dz \quad (17)$$

for $p ds$ is the scalar product of two vectors, one being the momentum whereof the components are p_x, p_y, p_z , and the other the line element whereof the components are dx, dy, dz . Hence:

$$\partial A / \partial x = p_x, \quad \partial A / \partial y = p_y, \quad \partial A / \partial z = p_z \quad (18)$$

the space derivatives of the action A , with respect to the coordinates of P , are the components of the momentum which the corpuscle has when it arrives at P . This quality is shared by the function W which I will now define:

$$W = \int (2T - E) dt = A - Et. \quad (19)$$

Here E stands for the energy of the corpuscle—the total energy, kinetic and potential together, which remains the same all along the path, so that its time integral along any section of the path is simply the product of its value by the time which the corpuscle spends in tracing that section.

The function W has the following properties:

$$\begin{aligned} \partial W / \partial x &= p_x, & \partial W / \partial y &= p_y, \\ \partial W / \partial z &= p_z, & \partial W / \partial t &= -E. \end{aligned} \quad (20)$$

Consider the function W at any particular instant of time. It has a system of "level surfaces," or surfaces over each of which its value is constant. These surfaces are perpendicular to the paths of the corpuscles at the points where they intersect it. Now let t vary. As it varies, the level surfaces wander through space. Each small section of each surface moves parallel to its own normal.

From Eqs. (20) it can readily be shown—indeed, it is almost obvious—that the speed with which these level surfaces travel is given by the equation,

$$\text{Speed of level surfaces} = E/p. \quad (21)$$

(I have given the argument in detail in another place,⁴ and therefore do not repeat it here.)

Now this is Eq. (12), with a specific value assigned to the constant which in that equation appears as an indeterminate. To recall the argument: in order that wave train and corpuscles may follow the same path, the wave fronts must move with a speed equal to $1/p$ multiplied by a constant. The value of the constant does not matter; we may make it equal to the energy of the corpuscles, if we will; and if we make this choice, the wave fronts of the train become the level surfaces of the function W .

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 Fig. 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 Fig. 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.

⁴ *Introduction to Wave-Mechanics*, Bell Sys. Tech. J. 6, 653-701 (October, 1927); translated into German as *Elementare Einführung in die Wellenmechanik* (Hirzel, 2nd Ed., 1932); hereinafter referred to as *Wave-Mechanics*.

BEGINNING OF WAVE MECHANICS: A PERIODIC
FUNCTION IS INTRODUCED WHICH FOR ITS
WAVE FRONTS TAKES THE LEVEL SUR-
FACES 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.⁴ 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 v is inversely proportional to the corpuscle mo-

⁴ A reader of the manuscript has objected that "monochromatic light" is by definition light of a single wave-length, so that this sentence is a definition (or a truism) and the previous sentence a theorem, instead of *vice versa*. He is probably right about the usage, but if he is, the usage is regrettable. "Monochromatic" can and should be defined without reference to theory, and as Newton might have defined it from his experience, had the word existed in his time: as referring to the kind of light which is *not* spread out or changed in appearance when passing through a prism.

mentum p according to Eq. (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 and λ for the frequency and wave-length of the waves, we get:

$$v = \nu\lambda = E/p. \quad (22)$$

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

$$\lambda = (E/\nu)/p = \text{const.}/p. \quad (23)$$

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 every reader of this article: that it is to be identified with Planck's constant first introduced in the theory of black-body radiation ($h = 6.55 \times 10^{-27} \text{ g} \cdot \text{cm}^2 \cdot \text{sec.}^{-1}$).

FUNDAMENTAL INTERRELATIONS OF CORPUSCLES
AND WAVES

So we come to two of the fundamental relations of quantum mechanics:

$$E = h\nu, \quad (24)$$

$$\lambda = h/p, \quad (25)$$

presumptively valid for light and electricity and matter alike, which are offered for the test of experiment.

As earlier pointed out, these waves are not observable, and neither are the qualities of the wave model which are called phase speed and frequency, but the wave-length may be deduced from observations on diffraction patterns. Thus for free electricity (electrons) and electrically-neutral matter (atoms), Eq. (25) is of the two the more amenable to test. Beams of electrons (or of atoms) must be caused to overlap one another; the region of overlapping must be explored for the peculiar distribution-in-intensity which is called a diffraction pattern; if found, it must be proved identical with that which would be expected for some particular value of λ , and this value is then to be compared with that demanded by Eq. (25).

The experiments made with this aim form already a great and illustrious group, though the earliest of them all—that of Davisson and Germer—was done so recently as the winter of 1926–27. Davisson and Germer measured the diffraction pattern of the beams of negative electricity which emerge from a single crystal of metal onto which an electron stream is projected. G. P. Thomson shortly afterward made the corresponding experiment with electrons of much higher kinetic energy and thin films of metal composed of many small crystals. Hundreds of analogous experiments have been performed with many values of electron energy and electron momentum, many kinds of crystals, and a variety of ways of mapping the diffraction pattern. Rupp reports an observation of a pattern imposed by a ruled optical grating upon a beam of electrons which it intercepted and reflected. O. Stern and T. H. Johnson independently have observed the diffraction of ordinary matter: that is to say, the pattern produced by beams emerging from the surface of a single crystal, against which a stream of gas such as hydrogen or helium is directed. There have also been strong though not perfectly clear indications of the diffraction of proton streams.

The general outcome of all these experiments is the full verification of Eq. (25).

The testing of Eq. (24) is a task of a different sort altogether, and one might deem it impossible in principle, in view of what was earlier said about the non-observability of frequency and phase speed. In truth, it has never been made for

electricity nor for matter. Corpuscle energy and wave-train frequency have been proved proportional to one another, only for light—only because of certain interrelations between light and electricity, or light and matter⁶—and only because of the unique coincidence of phase speed with the measurable speed of a segment of a beam, which characterizes light *in vacuo*.

To take one only of the proofs, the earliest and likewise the most accurate: when a beam of light of wave-length λ (not too large) falls on a metal from a vacuum, it expels electrons, which have a distribution in speed extending up to a maximum speed. This maximum speed u_m (like the distribution in speed), is independent of the intensity of the light, whereas the total number of electrons expelled per unit time is proportional to the intensity of the light. Moreover the kinetic energy $K_m (= \frac{1}{2} m u_m^2)$ corresponding to the maximum speed is a linear function of the reciprocal of the wave-length of the light:

$$K_m = A_1/\lambda + A_2, \quad (26)$$

where A_1 and A_2 stand for constants of which the second depends upon the metal, but the first is a universal constant. All these facts are well interpreted by assuming that the energy of the light is concentrated upon corpuscles each having a quantity equal to A_1/λ , and by further assuming that each electron which is expelled has taken over the entire energy of such a corpuscle. (The constant A_2 is then interpreted as the amount of kinetic energy which is converted into potential energy as the emerging electron surmounts a potential barrier at the surface of the metal.) Since λ is inversely proportional to ν , it is thus being assumed that corpuscle-energy is proportional to wave-train frequency, as in Eq. (24); and the facts sustain this idea. Further, we should have:

$$A_1/\lambda = h\nu = hv/\lambda; \quad A_1 = hv. \quad (27)$$

Here v stands for the phase speed of light *in vacuo*, which is unobservable; but profiting by the lucky coincidence aforesaid we put c , the measured speed of light *in vacuo*, for v :

$$h = A_1/c \quad (28)$$

⁶ Observations on the black-body spectrum form perhaps an exception to this statement.

and indeed it is found from the experiments that the quotient of A_1 by c has the same value as the experiments on electron-diffraction give for the constant denoted by h in Eq. (25).

Thus, the Eqs. (24) and (25) have both been checked, but—strange to relate—not for the same entities; one has been tested for light, the other for electricity. There is however a beautiful phenomenon which has been well interpreted by assuming both of the two for light, and therefore is a test of the two assumptions interlocked. It is, of course, the Compton effect. A beam of x-rays (high frequency light) of known wavelength is projected into a stratum of matter; electrons emerge from the matter with high speeds, from which their energy and momentum can be determined; there also emerge “scattered x-rays” of greater wave-length than the incident beam. All the phenomena are wonderfully well explained by assuming that elastic impacts occur between corpuscles of light and corpuscles of electricity—photons and free electrons; where by “elastic impact” is implied, that energy and momentum are both conserved by the pair of particles participating in the impact, without any loss or transfer to any other object; and where the energy and momentum of the photons are computed by Eqs. (24) and (25) (and the relation $\nu\lambda=c$) from the wave-lengths of the x-rays. The emerging electrons are those which are recoiling from such impacts, the scattered x-rays are the rebounding photons of lessened energy and momentum.

I quote the equations of the process, employing relativistic instead of ordinary classical mechanics.⁷ The symbols ν , λ refer to the incident x-rays (photons before impact); ν' , λ' to the scattered x-rays (photons after impact); u stands for the speed of an electron after impact (before the impacts the electrons have speeds which are relatively, often negligibly small; we here neglect them); ϕ and θ stand 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 is given by the expression,

$$E = mc^2 = [m_0/(1-\beta^2)^{1/2}]c^2, \quad \beta = u/c \quad (25A)$$

where m_0 stands for the rest mass of the electron. (We need not bother to include potential energy.) Thus:

$$h\nu = h\nu' + [m_0/(1-\beta^2)^{1/2}]c^2. \quad (25B)$$

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:

$$h\nu/c = (h\nu'/c) \cos \theta + [m_0u/(1-\beta^2)^{1/2}] \cos \phi, \quad (25C)$$

$$0 = (h\nu'/c) \sin \theta + [m_0u/(1-\beta^2)^{1/2}] \sin \phi. \quad (25D)$$

Eliminating ϕ and u 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:

$$\lambda' - \lambda = (h/mc)(1 - \cos \theta). \quad (25E)$$

This is a prediction of the wave-length which should belong to the x-rays scattered at angle θ . It is verified by experiment. Eliminating θ and ν' between Eqs. (25B), (25C) and (25D), one gets the corresponding relation between u , ν , and ϕ . This may be translated into a prediction of the energy which should belong to electrons recoiling in the direction ϕ ; it too is verified by experiment. Finally, there is evidence that when an electron recoils in a direction ϕ , a photon rebounds simultaneously in the direction θ which is computed by eliminating u and ν' between Eqs. (25B), (25C) and (25D), and substituting the given value of ϕ into the equation thus derived.

Such are the most conspicuous examples of the evidence sustaining Eqs. (24) and (25). Henceforth these two relations shall be taken for granted in this article.

⁷ As many physicists must have found out for themselves, the attempt to make the equations look simpler and be more easily solvable by using ordinary mechanics has precisely the opposite effect from that desired. The theory of the Compton effect affords one of the cases where the wisest policy is also the easiest.

PHASE SPEED, GROUP SPEED AND CORPUSCLE
SPEED OF ELECTRICITY AND MATTER

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 Eq. (24) 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 $(c^2/u + U/p)$. 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.⁸

Returning to Eq. (7) and substituting from Eqs. (24) and (25), we find:

$$g = -\lambda^2(d\nu/d\lambda) = dE/dp. \quad (29)$$

By the classical mechanics of material particles, we have:

$$p = mu, \quad E = \frac{1}{2}mu^2 + U, \quad (30)$$

from which we derive at once the result

$$g = u.$$

The speed of the corpuscles in a **beam** of matter or electricity—atoms or electrons—is **equal** to the group speed of the waves.

The same theorem is obtained from relativistic mechanics, the required formulae being these:

$$\begin{aligned} E &= m_0c^2/(1-\beta^2)^{\frac{1}{2}} + U, & p &= m_0u/(1-\beta^2)^{\frac{1}{2}}, \\ E &= c(m_0^2c^2 + p^2)^{\frac{1}{2}}, & (\beta &= u/c). \end{aligned} \quad (31)$$

⁸ One often sees the phase speed equated to c^2/u , 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. (12), which is **not** the case at all.

PRINCIPLE OF INDEFINITENESS

Consider again the beam of corpuscles of uniform momentum, and the accompanying sinusoidal wave train, with which the argument started. At the risk of seeming absurdly repetitious, I say once more 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 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 wave-length, 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 wave-lengths 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 wave-lengths, 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."⁹ 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:

Turn back to Fig. 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 Eq. (4):

$$D = \lambda^2 / \Delta\lambda, \quad (32)$$

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 wave-lengths. 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:

$$\begin{aligned} D \cdot \Delta p &= \Delta x \Delta p = (\lambda^2 / \Delta\lambda) \Delta(h/\lambda) \\ &= (\lambda^2 / \Delta\lambda) (h \Delta\lambda / \lambda^2) = h, \end{aligned} \quad (33)$$

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

⁹ 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.

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.¹⁰

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 Eq. (33). By making certain approxi-

mations, a relation resembling Eq. (33) 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:

$$p_x/p_z = \tan \theta. \quad (34)$$

The angle θ conforms to the equation:

$$\sin \theta = \lambda/D, \quad (35)$$

which is proved in all textbooks of optics.¹¹

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 fundamental interrelation of Eq. (25) between wave-length and momentum of the undiffracted beam, we get:

$$p_x/p_z = \lambda/D = h/Dp_z. \quad (36)$$

The difference between the values of p_x 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_x , we obtain:

$$D\Delta p_x = \Delta z\Delta p_z = 2h. \quad (37)$$

This seems to differ from Eq. (33) 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 permissible statement is, that the

¹⁰ In earlier allusions to diffraction patterns, I spoke for convenience as though they occur only when separate beams overlap. The diffraction pattern of a single slit may be explained in similar fashion if one wishes, though the convenience of the procedure is doubtful in this case. One imagines the wave front divided into strips parallel to the axis of the slit but much narrower than D , and envisages the overlapping of these strips as they simultaneously advance and expand.

¹¹ If we divide the wave front in the slit into strips as suggested in the previous footnote, 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.

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.¹²

Part of what has gone before has been devoted to phenomena which wave model and corpuscle model can be made to fit equally well: refraction at a surface, bending of a beam as it traverses a field of force. Part however has been devoted to a group of phenomena—those of diffraction—which only the wave model naturally fits; when both the pictures are employed, the corpuscles must be subordinated to the waves, in ways which have already been suggested. There is yet another group of phenomena for which the wave model is superior, and the corpuscles must be constrained to adapt themselves accordingly; these we will now take up.

REFLECTION AND TRANSMISSION OF CORPUSCLES AND WAVES

When a beam of light passes from vacuum (say) into glass, the passage is not attended solely by refraction, but also by reflection of a portion of the beam. The reflection is not irregular nor random, but is determined by well-known laws. It is difficult, if not impossible, to produce any reasonable explanation of these by a simple corpuscular model. Newton confronted this difficulty, and emphasized it by many clear and striking arguments. I should like to quote a few of his phrases, out of place though seventeenth-century words may seem to be in an article on quantum mechanics: "The cause of reflection is not the impinging of light on the solid or impervious parts of bodies, as is commonly believed. . . . First, in the passage of light out of glass into air there is a reflection as strong as in its passage out of air into glass, . . . and it seems not probable that air should have more strongly reflecting parts than glass. But if that should possibly be supposed, yet it will avail nothing; for the reflection is as strong or stronger when the air is drawn away from the glass (suppose by the air-pump invented by Otto Gueriet¹³ . . .) as when it is adjacent to it.

Secondly, if light in its passage out of glass into air be incident more obliquely than at an angle of 40° or 41° it is wholly reflected, if less obliquely it is in great measure transmitted. Now it is not to be imagined that light at one degree of obliquity should meet with pores enough in the air to transmit the greater part of it, and at another degree of obliquity should meet with nothing but parts to reflect it wholly. . . ." There is also the dependence of reflecting-power on color, which [Newton continues] makes it possible to set up an experiment in which a beam of red light is partially transmitted on arriving from glass at a boundary between glass and air, while a parallel beam of blue light is totally reflected; and "why should the blue wholly impinge on reflecting parts, so as to be all reflected, and yet the red find pores enough to be in a great measure transmitted?"

After these and many other cogent illustrations, Newton concludes: "This problem is scarce otherwise to be solved, than by saying that the reflection of a ray is effected not by a single point of the reflecting body, but by some power of the body which is evenly diffused all over its surface. . . ." This is an admirable statement of a portion of the wave theory. To it we should add, that this "power" evenly spread over the surface of the body acts upon a wave front evenly spread over the cross section of the beam. Instead of imagining corpuscles of light bouncing back from atoms of the matter beyond the boundary, it is expedient to imagine a homogeneous wave front, or better yet a regular sequence of wave fronts, reflected from a homogeneous surface coinciding with the boundary of the reflecting body. Having computed the reflection in this manner, we can later introduce the corpuscles, making them distribute themselves in the manner prescribed by the squared amplitude of the waves.

To go over to the case of beams of electricity: we have seen that the analogue to a refracting surface is a surface dividing two regions of different electrostatic potential. This would be a potential discontinuity. Actually, it is not necessary to assume a change of potential infinitely sudden, but merely one which is comprised in a distance small compared with the wave-length of the waves. Suppose however (for

¹² I have given a couple of additional illustrations in *Rev. Sci. Inst.* 4, 188-192 (1933).

¹³ Newton's spelling!

convenience) that there is a discontinuity at the plane $x=0$, the potential being lower by V electrostatic units on the right than it is on the left. Imagine a stream of electrons of uniform momentum p_x and kinetic energy U advancing along the axis of x , from the left; and keep also in mind the concomitant picture of a beam of plane waves of wave-length h/p_x , their wave fronts parallel to the plane $x=0$.

The unmodified corpuscle model leads to a simple definite prediction. Either U is greater than eV , in which case the electrons will proceed across the discontinuity, and continue their courses on the right-hand side, though with diminished kinetic energy ($U-eV$). Or else U is smaller than eV , in which case the electrons will not be able to cross the discontinuity, and it is natural (though not inevitable) to assume that they will be reflected back to the left with reversed momentum and unchanged kinetic energy. We then have either 100 percent transmission or 100 percent reflection, and a definite rule for deciding which occurs. The incident beam will never be divided into a transmitted part and a reflected part.

Were this to be the case, we should have to infer that the accompanying waves are always either totally transmitted or totally reflected. However, our general experience with waves—water waves, sound waves, light—suggests that this is not a plausible conclusion. Total transmission of the otherwise-known types of wave motion never occurs across a boundary. Total reflection, it must be conceded, does occur, but not (to my knowledge) with normal incidence. Division of the incident beam into a reflected part and a transmitted part is the usual rule. Were this rule to extend to the waves associated with electricity, we should be obliged—in carrying out the policy of subordinating corpuscles to waves—to infer that even when U is less than eV , some particles will cross the potential discontinuity; even when U is greater than eV , some will fail to cross it.

The appeal to direct experiment cannot be made in this case, as it was in discussing diffraction. It is beyond the powers of experimental physicists to make a potential drop so sharp and sudden, that it is confined to a space much smaller than the wave-length of an ordinary

electron beam (this being about 10^{-7} cm for the slowest electrons of which it is possible to form a reasonably nearly uniform stream). However, many phenomena of metals—such as the thermionic effect and the photoelectric effect—are interpreted by supposing that in the vicinity of the surface of any metal, the potential changes considerably over distances of the order of 10^{-7} or 10^{-8} cm, and through this zone of rapidly-varying potential all the electrons emerging from the metal must succeed (or fail) in escaping. Moreover there is a well-supported theory that the nucleus of any atom is surrounded by a region in which the potential changes considerably over distances of the order of only 10^{-12} cm; and across this region, the charged particles proceeding from the nucleus must make their escape (if the atom is radioactive) or else the impinging charged particles must make their entry (if the nucleus is transmuted by impact of alpha-particles or protons). These cases are capable of affording indirect tests of the wave theory of reflection.

The theory must however be formulated more precisely before such tests can be applied. In general, the feature of a wave theory which controls the laws of reflection and transmission is its set of boundary conditions. These may be chosen because they seem plausible, but ultimately it is the agreement of the observed with the predicted laws which justifies the chosen set. Thus, in the electromagnetic theory of light, it is postulated that when light-waves pass from one medium to another with a different dielectric constant, the following quantities suffer no discontinuous change at the boundary: *viz.*, the component of electric displacement perpendicular to the boundary surface, the component of electric field strength parallel to this surface, all the components of magnetic field strength. These conditions result in hypothetical laws of reflection which are found to agree with the empirical laws: for instance, with the observed relation between angle of incidence and percentage of light reflected.

The waves of electricity and matter must likewise be provided with boundary conditions, which will describe their passage across abrupt potential drops or potential rises, and will be tested ultimately by the service which they

render in interpreting experiment. At this point I will give only a few of the consequences of those which have been adopted, leaving the conditions themselves unstated (they will be stated along with the wave equations).

In the case already specified—uniform potential (V_0 , say) from $-\infty$ to $x=0$; uniform but lower potential, $V_0 - V$, from $x=0$ to $+\infty$; a potential discontinuity of V electrostatic units at $x=0$; a beam of electrons of kinetic energy U coming along the x -direction from the left—the prediction is partly the same, partly different from that of the pure corpuscle theory. If $U < eV$ there is total reflection, as formerly predicted; if $U > eV$ there is not total transmission, but a division of the beam into a transmitted fraction and a reflected fraction. Now modify the case by supposing the potential equal to V_0 from $-\infty$ to $x=0$; equal to $V_0 - V$ from $x=0$ to $x=a$; and again equal to V_0 from $x=a$ to $x=+\infty$; so that the shape of the potential curve suggests a trough of depth V and breadth a excavated in a horizontal plain. It now follows from the boundary conditions that there is *not* total reflection when $U < eV$. Part of the train of waves passes through the potential trough and continues its course towards positive infinity, despite the fact that the potential-drop on the “near side” of the trough—at $x=0$ —is so great that all of the electrons should be stopped by it. Subordinating the corpuscles to the waves, we are now obliged to declare that electrons of kinetic energy U can traverse a sudden potential-drop greater than U/e , if farther along there is a compensating (or only a partially compensating) potential-rise. Not every electron of a homogeneous beam can achieve this, but only a certain fraction of the total number; in other words, there is a certain *probability* of traversal or transmission. This probability is the transmission coefficient of the trough for the waves, the ratio of their squared amplitudes to right and to left of the trough; it is a function of V and a , and of U (or λ) and it is determined by the boundary conditions.

The reader will perhaps be more readily reconciled to this singular conclusion, if I remind him that a similar quality of light has been known for many decades. Let a beam of light be caused to enter a glass prism in such a

direction, that when it arrives at the far side of the prism it is totally reflected; then if a second piece of glass is brought up against that side, leaving a very thin but yet appreciable air film between the two, the totality of the reflection is suspended and a portion of the light goes inward through the second glass. This is, of course, only a striking special case of the well-known rule (enounced by Newton) that when a beam of light is sent against a transparent film or stratum between two media (which may or may not be the same, but must be different from the substance of the film) the fraction of light transmitted depends upon the thickness of the stratum and the wave-length of the beam.

WAVE EQUATIONS

We have now passed the midpoint of this article, and yet there has been no allusion thus far to a wave equation. This must be very annoying to mathematicians, for it is their custom to define a “wave theory” by specifying the differential equation to which they expect the phenomena to conform. In favor of their policy it can be argued that many of the quantitative tests of the theory, by which it is eventually to be justified, are possible only when the wave equation and the boundary conditions are assigned. Against their policy it may be said that all of their equations have developed out of the eighteenth-century efforts to find descriptions for vibrations of strings and waves on the sea, by way of the nineteenth-century triumphs in describing vibrations and wave patterns of sound in air and elastic distortions in solids. Some people would be glad to blot out all these memories, but to some they are essential to make the theories vivid. The trend toward perfect abstraction, which for several years has been dominant in quantum mechanics, is welcome to the ones; but the others crave to retain, for as long as possible, as many as possible of the links with the past. For the sake of the class of these last (to which, as the reader must realize already, I belong) I have retarded as far as possible the advent of the wave equations.

The plural seems odd in the foregoing sentence, but it is desirable, for there are several so-called wave equations of differing aspects employed in various problems. One feels that all of them

should be special cases (or accidental imitations) of a single universal wave equation, and much contemporary exposition is phrased as if this were certainly true; but the proof as yet is lacking.

The simplest of all wave equations is the one-dimensional example which is this:

$$\partial^2 f / \partial t^2 = v^2 (\partial^2 f / \partial x^2). \quad (38)$$

This is employed in such cases as that of transverse waves running along a wire stretched along the x -axis, and sound waves travelling through air with plane wave fronts perpendicular to the x -direction. In these cases f stands for transverse displacement of the string and condensation of the air, respectively.

Any expression of the type appearing in Eq. (1) is a solution, provided that the product of the constants ν and λ is equal to v :

$$f = A \cos 2\pi(\nu t - x/\lambda + \alpha); \quad \nu\lambda = v. \quad (39a)$$

Thus any sine-wave of whatever frequency is compatible with Eq. (38), but to each frequency there corresponds a prescribed wave-length, since the phase speed is fixed by the factor on the right of Eq. (38).

Another type of solution is obtained by changing the minus sign into a plus sign in Eq. (39a):

$$f = A \cos 2\pi(\nu t + x/\lambda + \alpha); \quad \nu\lambda = v. \quad (39b)$$

Another sort of solution is obtained by assuming f to be the product of a function of x only and a function of t only, and proceeding in the manner known as "integration by separation of variables." Putting $\Psi(x)\phi(t)$ for f in Eq. (38), we get:

$$v^2 \Psi''(x) / \Psi(x) = \phi''(t) / \phi(t). \quad (40)$$

The left-hand member is independent of t , the right-hand member independent of x . Each therefore must be equal to a constant, to the same constant, which I denote by $-C^2$ (employing an intrinsically negative constant, in order that the solutions shall be trigonometric functions). This gives a pair of differential equations for Ψ and ϕ respectively:

$$\phi''(t) = -C^2 \phi(t), \quad \Psi''(x) = -v^{-2} C^2 \Psi(x) \quad (41)$$

whereof the solutions are:

$$\phi(t) = A \cos 2\pi(\nu t + \alpha);$$

$$\Psi(x) = B \cos 2\pi(x/\lambda + \beta) \quad (41a)$$

$$f = \text{const.} \cos 2\pi(x/\lambda + \beta) \cos 2\pi(\nu t + \alpha) \quad (42)$$

and ν , λ , C and v are all interconnected by the relations

$$C^2 = (2\pi\nu)^2 = (2\pi v/\lambda)^2; \quad \nu\lambda = v, \quad (43)$$

so that ν and λ signify frequency and wave-length, as in the other types of solution.

The reader will surely recognize that Eq. (39a) describes a sine-wave progressing in the positive x -direction, Eq. (39b) a similar wave advancing in the negative direction, and Eq. (42) a stationary distribution of amplitude or "stationary-wave pattern" which is evidently of the class of interference patterns and diffraction patterns of which I spoke on earlier pages. As a matter of fact, the right-hand member of Eq. (42) is obtained by adding expressions such as (39a) and (39b); it is the interference pattern of two wave trains of equal wave-length moving in opposite directions; it is essentially the right-hand member of Eq. (2), the angle there called θ being given the value $\pi/2$.

Thus far it has been implied that the string or the air is of infinite extent in the x -direction. The chief importance of Eq. (42) lies in the fact that it is appropriate to a string of finite length or a stratum of air of finite breadth, with boundary-conditions at its ends. Say for instance that we have a string of length L clamped at both ends, so that f must perpetually be zero at (say) $x=0$ and $x=L$. These conditions permit such motions as are described by Eq. (42), when certain specific values of λ are chosen: these are the values which are quotients of $2L$ by the integer numbers, *viz.*, $2L, 2L/2, 2L/3$, and so on. These are the "resonance wave-lengths" of the bounded string (or air-chamber), and the corresponding values of ν/λ are the "resonance frequencies."

We have already employed expressions like those in Eqs. (39a) and (39b) to describe homogeneous beams of matter and electricity travelling through field-free space, and expressions like that in Eq. (42) to describe diffraction patterns. Since therefore the solutions of the wave equation (38) are of value in interpreting electricity and matter, it is natural

to suppose that these may be fully governed by an equation like (38). What more can be deduced from such an equation?

At the very outset there is a contrast with the cases of waves in air and vibrations of strings. With a string of definite tension and density, or with air of definite pressure and temperature, v is a fixed quantity; and sine-wave trains of any frequency are compatible with Eq. (38)—that is to say, they are physically possible—the wave-length always being equal to the quotient of the frequency by the fixed value of v . This is expressed by saying that phase speed is independent of frequency and wave-length. Such is not the case with the wave trains representing matter and electricity, for which the phase speed varies with wave-length in the manner prescribed by Eqs. (22) to (25). We cannot properly express this by saying that in Eq. (38), v must be regarded as a variable; for we are not (as yet) supposing that v varies with x or t . We are obliged to say that when the wave equation (38) is applied to beams of matter or electricity in field free space, any value of v is permissible, but to each value of v there corresponds only one solution of the type (39a), with a specific value of frequency.

Remembering the expression (from Eq. (22)) for phase speed of waves associated with a beam of corpuscles of energy E and momentum p , we may write:

$$\partial^2 f / \partial t^2 = (E/p)^2 (\partial^2 f / \partial x^2); \quad v = E/h, \quad \lambda = h/p, \quad (44)$$

the equation on the left being meant as the tentative wave equation for such a beam, while the others are to remind us that in the solutions we are not free to choose any pair of values of v and λ —not even any pair of which the product is (E/p) —but must use the particular pair E/h and h/p . Thus, the only solution of the form (39a) is this:

$$f = A \cos (2\pi/h)(Et - px + a). \quad (45)$$

Also the only solution of the form (42) is this:

$$\begin{aligned} f &= \Psi(x)\phi(t) \\ &= D \cos (2\pi/h)(px + b) \cos (2\pi/h)(Et + a). \end{aligned} \quad (46)$$

Here $\Psi(x)$ is the solution of a differential equation which I repeat from Eq. (41):

$$\begin{aligned} \partial^2 \Psi / \partial x^2 + v^{-2} C^2 \Psi &= \partial^2 \Psi / \partial x^2 \\ &+ (p/E)^2 (2\pi E/h)^2 \Psi \\ &= \partial^2 \Psi / \partial x^2 + (2\pi p/h)^2 \Psi = 0 \end{aligned} \quad (47)$$

in the integration of which it has been assumed that p is constant.

Thus far nothing of importance has been added to our concepts by these equations. The important developments begin when we go over to the case of a beam traversing a space which is not field free, so that the momentum of the corpuscles changes from point to point. Assume the potential gradient to be parallel to the x -direction, so that the beam will not be deflected and we can continue to use one-dimensional equations.

Now E and v are still constant throughout the beam, but p has become a function of x . We continue to accept the wave equation (44), but for the time being we neglect progressive waves, confining our attention to solutions of the type $\Psi(x)\phi(t)$ —solutions in which the variables are separated, and which describe stationary wave patterns. The function $\phi(t)$ is still of the form assigned to it in Eq. (46), since E is constant, and the differential equation in Eq. (41) which defines it has not been altered; it is still sinusoidal and of frequency E/h . The function $\Psi(x)$ has changed, for it is a solution of the differential equation (47) in which p has ceased to be a constant.

The next step consists in replacing p by its expression in terms of E , the total energy, and V , the potential energy¹⁴ of the corpuscles; in nonrelativistic mechanics this is $[2m(E - V)]^{1/2}$ (see Eqs. 30).

So we come to *the first of Schrodinger's wave equations*:

$$\partial^2 \Psi / \partial x^2 + (8\pi^2 m/h^2)(E - V)\Psi = 0. \quad (48)$$

Here E is a constant, V a function of x depending on the field of force which the beam is traversing.

¹⁴ This is a change from the previous meaning of V , which was electrostatic potential.

SCHROEDINGER'S WAVE EQUATION, ITS PROPER
VALUES AND THE STATIONARY STATES
OF ATOMS

The striking feature of the differential equation (48) is, that for certain particular values of E its solution (or solutions) have certain distinctive qualities. For instance, it happens with certain types of function $V(x)$ that there is a discrete set of values of E , such that if any one of them be substituted into Eq. (48), there is a solution which is everywhere single-valued, bounded, continuous, and has a continuous first derivative; while if any value of E not belonging to the set is placed in Eq. (48), every solution increases without limit as some (finite or infinite) value of x is approached. With other types of function V , there is a continuous range of values of E extending over a certain interval, and outside of this interval there is a discrete set of values, such that the solution of Eq. (48) possesses the foregoing qualities if E lies in the interval aforesaid or belongs to the discrete set, and otherwise does not. With yet other types there is no discrete set; thus with $V = \text{constant}$ any value of E yields a solution of the stated type, except values less than V which are (or, till quite lately, have always been considered) physically meaningless.

Thus if we conceive certain kinds of fields of force, and corpuscles moving in them attended by such wave trains as have been described, we are obliged to admit certain distinctive energy values and frequency values. These are distinctive in the sense, that stationary wave patterns can arise in which the amplitude (which is the function $\Psi(x)$) is afflicted neither by discontinuities in itself, nor by discontinuities in its slope, nor by becoming infinite at any point of the field. They are known as "proper values" or "eigenvalues."

It was Schroedinger's idea that these are to be identified with the energy values of the stationary states of atoms and molecules. For any type of molecule or atom a model is to be constructed, consisting of corpuscles in a field (for instance, of electrons in an electrostatic field) such that when the potential-energy function corresponding to the field is inserted into some such equation as (48), the proper values of E shall coincide with

the observed energy values of the stationary states.

The initial triumphs of this idea were impressive and convincing. Thus, the stationary states of the hydrogen atom (as deduced from its spectrum) have energy values which conform approximately with the proper values of the *three-dimensional Schroedinger equation* which follows:

$$\left(\frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial^2 \Psi}{\partial z^2}\right) + \frac{8\pi^2 m}{h^2}(E - V) = 0;$$

$$V = -e^2/(x^2 + y^2 + z^2)^{\frac{1}{2}}. \quad (49)$$

It is evident that this choice of the function V corresponds to the picture of an electron in the inverse-square electrostatic field of a stationary point charge $+e$ at the origin of coordinates—in other words, to the customary model of the hydrogen atom. Not only is the approximation very close; it can be converted into exact agreement (within the narrow limits of experimental uncertainty) by improving the theory in three ways: (a) by taking into account the motion of the nucleus which bears the charge $+e$; (b) by substituting for Eq. (49) the wave equation based on the relativistic, as distinguished from the classical, relation between E and p ; (c) by introducing the conception of "electron-spin," i.e., conceiving the electron and the nucleus as point charges possessed of magnetic moment. Further, there is impressive agreement with the predictions of this theory about the influence of electric and magnetic fields on the stationary states.

It would be an unjustifiable use of our limited space to make a further list of the successes of Schroedinger's wave equation. It would also be unjustifiable to reproduce the methods of determining the proper values and the corresponding solutions or "proper functions" of the Schroedinger equation, for the several cases—that is to say, for the several choices of the function V —which have proved both mathematically manageable and physically valuable. Some of the simplest are given in my article *Wave-Mechanics*, and the others may be found in any of several treatises; I refer in particular to E. C. Kemble's contribution to this Journal.¹⁵ I

¹⁵ E. C. Kemble, *Rev. Mod. Phys.* 1, 157 (1929).

will quote a single example, for the sake of having something definite to look at; it shall be the case $V = \text{const. } x^2$, that of the "linear harmonic oscillator."

THE LINEAR HARMONIC OSCILLATOR TREATED BY THE METHOD OF SCHROEDINGER

Let the field of force be such that a corpuscle (of mass m) experiences a force driving it towards the plane $x=0$ and proportional to its distance from that plane. Denote the force by $-k^2x$, the potential-energy function by $\frac{1}{2}k^2x^2$. As everyone knows, the corpuscle makes sinusoidal oscillations with a frequency $\nu_0 = k/2\pi m^{\frac{1}{2}}$, or rather it would make such oscillations, if it were not subordinated to the attendant waves. Eq. (48) becomes:

$$\partial^2\Psi/\partial x^2 + (8\pi^2m/h^2)(E - 2\pi^2m\nu_0^2x^2)\Psi = 0 \quad (50)$$

and by a simple change of variable [$q = x \times 2\pi(m\nu_0/h)^{\frac{1}{2}}$]:

$$\partial^2\Psi/\partial q^2 + (C - q^2)\Psi = 0; \quad C = 2E/h\nu_0. \quad (51)$$

This is solved by putting for q a series of the following type:

$$\Psi = e^{-\frac{1}{2}q^2} \sum_{n=0}^{\infty} a_n q^n. \quad (52)$$

Substituting this into Eq. (51); grouping all of the terms involving the same power of x ; and equating each of the groups separately to zero, we arrive at this set of equations:

$$a_{n+\frac{1}{2}}/a_n = (2n+1-C)/(n+1)(n+2); \quad n = 0, 1, 2, 3 \dots \quad (53)$$

Evidently, if C is an odd integer, the chain of coefficients in Eq. (52) comes to an abrupt end at that value of n which is equal to $\frac{1}{2}(C+1)$, and the solution of Eq. (51) consists of a finite number of terms; but if C has any other value than that of an odd integer, the series goes on forever. In the latter case, Ψ is infinite at $x = \infty$; in the former case, Ψ is finite at $x = \infty$ and everywhere else, it is said to be "bounded."

This illustrates the general statements of the foregoing pages. The distinctive values of E corresponding to the odd-integer values of C are the odd-integer multiples of $\frac{1}{2}h\nu_0$:

$$E_s = (s + \frac{1}{2})h\nu_0; \quad s = 0, 1, 2 \dots \quad (54)$$

These are the "proper values" of the wave equation with the specified potential-energy function, the "permitted energy values" of the model. The corresponding proper functions are obtained from Eq. (52); they are commonly written:

$$\Psi_s(x) = \exp(-2\pi^2m\nu_0x^2/h)H_s(2\pi x(m\nu_0/h)^{\frac{1}{2}}), \quad (55)$$

the symbol H_s standing for the so-called "Hermite polynomial of order s ," which is the summation appearing in Eq. (52), with the coefficients adjusted according to the rule stated in Eq. (53), and ending at the term for which $n = s$.

This "case of the linear harmonic oscillator" occupies a peculiar position in quantum mechanics. It possesses the one great advantage of extreme simplicity, to which is due its prominent position in all elementary expositions. *En revanche* it suffers from various disadvantages. The very simplicity of this oscillator makes it untypical. It is one-dimensional, whereas the atom models which have been most successful in accounting for the properties of matter are three-dimensional. Classical theory and quantum theory coincide in prescribing that it should emit the frequency ν_0 defined before Eq. (50), whereas the success and prestige of the new mechanics is derived chiefly from atom models (that of hydrogen, for instance) for which classical theory predicts one set of frequencies and quantum theory another, and it is the latter set which is verified by experiment. Moreover it is well known that atom-models and molecule models are constructed from electrified particles which attract or repel one another with inverse-square forces, and the student may ask of what use it is to discuss a particle attracted by a force which varies directly as its distance from some fixed point. To this question the answer is, that in (for instance) a diatomic molecule such as H_2 or Cl_2 , the two nuclei and the several electrons are apparently so arranged that each of the nuclei has a position of equilibrium, and when either of them is displaced by a little from its equilibrium position the resultant of the various forces acting upon it is to first approximation a force of precisely that character. This statement is however true *only* to first approximation; in general, the function $V(x)$ comprises not only a term $\frac{1}{2}k^2x^2$, but others

with higher powers of x ; and these entail new frequencies, by which the quantum theory can be tested. This is the case of the "anharmonic" oscillator, of which the case of the harmonic oscillator is a limiting subcase, of relatively minor interest.

In applying quantum mechanics to atoms, the art of choosing the wave equation is of the highest importance. I say "art" intentionally, since the universal and fundamental principle—if such a one there be—is not yet ascertained. One of the most generally-acceptable of the prescriptions will be given in the closing section of this article. It must suffice at this point to mention (what the reader may already have noticed) that the prescription embodied in (48) cannot be applied as it stands to all of even the very simplest cases with which quantum theory purports to deal. Take for instance the "rigid rotator," familiar and valuable in the theory of molecules. In its simplest form it consists of two rigid spheres maintained at a fixed distance apart, and revolving around an axis perpendicular to the line of centres and intersecting this line at its midpoint. The accepted wave equation for this case¹⁶ is the following:

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} + \frac{\partial^2 \Psi}{\partial z^2} + \frac{8\pi^2 EA}{h^2} \Psi = \nabla^2 \Psi + \frac{8\pi^2 EA}{h^2} \Psi = 0. \quad (56)$$

Here A stands not for the mass but for the moment-of-inertia (about the axis in question) of the system. (V has vanished, because the potential energy of the system is constant and may be put equal to zero.)

Though in Eq. (56) the combination of derivatives symbolized by $\nabla^2 \Psi$ (one of the customary symbols for this combination; another is $\Delta \Psi$) is written out in Cartesian coordinates, it is expedient to employ spherical coordinates in dealing with the rigid rotator, and also with the electron revolving in a central field of force. These are illustrations of the rule that a particular choice of coordinates may be of great practical importance; in the foregoing cases the wave equation when expressed in Cartesian coordinates is

cumbrous and difficult to solve, but when it is translated into spherical coordinates the requisite procedure is fairly simple.¹⁷

Finally I must not omit to say that accurate wave equations must be based on relativistic instead of classical mechanics, and must be adjusted to allow for the spins of electrons and nuclei.

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

¹⁶ *Wave-Mechanics*, pp. 691–692.

¹⁷ *Wave-Mechanics*, pp. 682–686, 691–692.

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 re-phrased.)

(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 (41) in such forms as (41a); 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,

$$\begin{aligned} A \exp (int) &= (R+iS) \exp (int) \\ &= (R+iS)(\cos nt+i \sin nt) \quad (60) \end{aligned}$$

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 (38), 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:

$$\Psi\Psi^* \propto N. \quad (61)$$

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 in 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. (54) 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 will suffice for the present.¹⁸

We come now to one of the greatest of the problems confronting quantum mechanics.

FREQUENCIES OF LIGHT EMITTED BY ATOMS COMPARED WITH FREQUENCIES INHERENT IN THE ATOM MODEL

Atoms and molecules are capable of emitting and absorbing light of various characteristic wave-lengths. Many of these wave-lengths have been ascertained from diffraction-patterns and have been translated into frequencies, in the customary way which was earlier described. It is almost inevitable to suppose that these are the vibration frequencies of the parts of the atom relatively to each other, much as the frequencies of the sound waves which radiate from a piano through the air are those of the vibrations of the parts of the piano. Indeed this notion is so deeply embedded in our experience that we can hardly conceive of any other, and it is involved in all of our dynamical theories. It seems as though the frequencies of waves simply *must* be the same as those of their sources, and an atom simply *must*

¹⁸ 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. (61).

be a structure having all of the vibrations which can be detected in the light which it emits.

The analogy with sounding-bodies and waves of sound seems indeed to be established by the fact that Schroedinger's wave equation, when applied to an atom model in the manner which has been described, determines a set of vibration frequencies. These are the frequencies of the stationary waves which form the patterns heretofore described, which are associated each with one of the proper functions Ψ and the corresponding energy value E . To the various permitted energy values $E_1, E_2, \dots, E_n, \dots$ correspond permitted frequencies, the values of which (by Eq. (24) which underlies all the subsequent theory) are $E_1/h, \dots, E_n/h, \dots$. The whole of the spectrum of the atom would then be expected to consist of this sequence. Nevertheless it is the amazing fact that *the spectrum of an atom actually consists of the difference frequencies or beat frequencies $(E_r - E_s)/h$.*

VIBRATION FREQUENCIES OF A CLASSICAL ATOM MODEL; FOURIER ANALYSIS

It is not necessary to make so definite a statement as the foregoing, in order to bring out the fact that spectra are not explicable by classical ideas. Any vibrating system conforming to classical mechanics, and possessing a limited number of degrees of freedom,¹⁹ is strictly limited in its vibration-frequencies. Its motion may comprise as many different fundamental frequencies as it has degrees of freedom, but it may have no others excepting integer multiples and linear combinations of integer multiples of these fundamentals. Say that it has three degrees of freedom (such is the case of a corpuscle in a constant field of force); denote by ν_1, ν_2, ν_3 the corresponding fundamental frequencies; then *all* of its frequencies are obtained by giving various integer values to a_1, a_2, a_3 in the expression $(a_1\nu_1 + a_2\nu_2 + a_3\nu_3)$. Spectra generally, and in particular the beautifully simple spectrum of atomic hydrogen, are entirely different from this.

The term "vibration frequency" in the foregoing lines may suggest too narrow a conception. The argument above is not confined to a particle

¹⁹ This statement is confined to mechanical systems not afflicted by frictional forces and damping.

moving back and forth on a line (which would be a system of one degree of freedom, all of the frequencies of which would be harmonic overtones of a single fundamental). It is not even restricted to the case of a corpuscle revolving in a fixed orbit, such as the ellipse characteristic of a particle of constant mass in an inverse square central field. In most kinds of fields, the orbit of a corpuscle consists of a succession of loops no two of which exactly coincide, though the path as a whole is confined within a limited region of space ("conditionally periodic motion"). To these cases also the foregoing theorem refers. On resolving the motion into its component frequencies, all of these are found to be linear combinations of integer multiples of the fundamental frequencies; there are two fundamentals if the field possesses central symmetry, otherwise there are three.

The classical operation of "resolving a motion into its component frequencies" is a very important one. What has just been said does not in the least imply that a particle moving back and forth on a line (to take the simplest case of all) must necessarily have a sinusoidal motion with a single frequency chosen from among the permitted list. In general the function $x(t)$, representing its distance from a fixed point on the line (say the midpoint of its to-and-fro motion) is not of the form $C \cos(\nu t + \gamma)$; but it may be expressed as a sum of such terms in which the frequencies are those of the permitted list aforesaid—to wit, the fundamental and its integer multiples—and the amplitudes and phase angles are suitably chosen:

$$x(t) = \sum C_s \cos(2\pi s \nu_0 t + \gamma_s); \quad s = 0, 1, 2, 3, \dots \quad (62)$$

This is the principle of Fourier analysis, on the technique of which I will dwell for a few lines, as it is helpful in understanding the process of quantum-mechanical analysis toward which we are advancing.

Instead of Eq. (62), I write its equivalent in the complex-variable notation:

$$x(t) = \sum_{-\infty}^{\infty} C_s \exp(2\pi i s \nu_0 t); \quad C_s = C_{-s}^*. \quad (63)$$

The introduction of negative values of s and

hence of negative frequencies is something of a hindrance to visualization, but not to the deriving of physical results. The condition $C_s = C_{-s}^*$ makes the right-hand member of Eq. (63) real, as the left-hand member is assumed to be.

Choose some integer r ; multiply both sides of Eq. (63) by $\exp(-2\pi i r \nu_0 t)$; integrate over an entire period of the fundamental, e.g., from $t=0$ to $t=T=1/\nu_0$, or over a finite number of periods. We have:

$$\int_0^T x(t) \exp[-2\pi i r \nu_0 t] dt = \sum_{s=-\infty}^{\infty} C_s \int_0^T \exp[2\pi i (s-r) \nu_0 t] dt. \quad (64)$$

Now the whole practicability and value of Fourier analysis is due to the fact that all the terms on the right of Eq. (64) are equal separately to zero, save only the one for which $s=r$. This is due to a peculiar property of the system of functions $\exp(2\pi i r \nu_0 t)$, for which write:

$$f_s(t) = \exp(iskt), \quad s=0, 1, 2, 3, \dots \quad (65)$$

This peculiar property is expressed by the equation:

$$\int_0^T f_r^*(t) f_s(t) dt = \begin{cases} 0 & \text{if } r \neq s \\ \neq 0 & \text{if } r = s. \end{cases} \quad (66)$$

A system of functions which possesses this property is said to be "orthogonal in the interval from O to T ." The set of functions (65) is therefore an orthogonal system, but by no means the only one. *The proper functions of any wave equation, such as the functions (55), constitute an orthogonal system.*

Owing to this property of orthogonality, we deduce from Eq. (64):

$$C_s = (1/T) \int_0^T x(t) \exp(-2\pi i s \nu_0 t) dt. \quad (67)$$

The coefficients in the right-hand member of Eq. (63), which is the "Fourier expansion" of $x(t)$, are thus determined. The orthogonality of the system of functions $\exp(2\pi i s \nu_0 t)$ makes it easy to expand the function $x(t)$ in terms of that system.

Evidently the function $x(t)$ is completely specified by writing down the fundamental frequency ν_0 and the values of the coefficients of its Fourier expansion. Of these last it suffices to quote those with positive (or zero) value of the index s , since the others are the conjugates of those.

$x(t)$ specified by $(\nu_0; C_0, C_1, C_2, \dots)$

$$\text{or } (\nu_0; C_s). \quad (68)$$

Each of the coefficients C_s is in general a complex quantity, specifying both the amplitude and the phase of the corresponding "component vibration" of frequency $s\nu_0$, as explained above in connection with Eq. (60). If we are concerned only with the squares of the amplitudes, it suffices to write down ν_0 and the quantities $C_s C_s^*$.

For dx/dt it seems natural to put:

$$\dot{x}(t) \text{ specified by } (\nu_0; 2\pi i s \nu_0 C_s) \quad (69)$$

and to specify second and higher time-derivatives in the corresponding way. This procedure is not always valid, but the cases in which it is valid form an acceptable analogy to the case towards which we are moving.

For x^2 the result is almost equally simple. On multiplying the right-hand member of Eq. (63) into itself, we get a series of terms containing the same exponentials as figure in the expansion of $x(t)$, and readily find:

$$x^2(t) = \sum_r D_r \exp(2\pi i s \nu_0 t); \quad D_s = \sum_r C_r C_{s-r} \quad (70)$$

and consequently:

$$x^2(t) \text{ specified by } (\nu_0; \sum_r C_r C_{s-r}). \quad (71)$$

This simplicity is due to the fact that the sum of any two frequencies of the set $s\nu_0$ is itself a member of the set. From this fact it further follows that any positive-integer power of $x(t)$ has the same frequencies in its Fourier expansion as does $x(t)$ itself. Moreover, it follows that if there are two functions $p(t)$ and $q(t)$ having the same fundamental frequency ν_0 and therefore the same overtone-frequencies $s\nu_0$, then any function of the form $p^m q^n - m$ and n standing for positive integers

(including zero)—has the same set of frequencies. So also does any finite or convergent infinite sum of terms of the form $p^m q^n$; and therefore, so does any function of p and q developable as a power-series in these variables—and this is a very wide class of functions indeed.

I summarize these latest statements as follows. Let p and q stand for two functions of t which are specified respectively by $(\nu_0; p_s)$ and $(\nu_0; q_s)$. Let f stand for any function of p and q which either is of the form $p^m q^n$, or is a finite or convergent infinite sum of terms of this form (even these restrictions are not necessary). Then f is specified by $(\nu_0; f_s)$, the coefficients f_s being calculable from the coefficients p_s and q_s .

The notation of the foregoing paragraph is chosen to harmonize with that which is customary in quantum mechanics, where q in general stands for a coordinate and p for a momentum, and the coefficient of any term in the expansion of a function is denoted by the symbol of that function with a subscript to indicate the term.

To acquaint the reader with the art of using these specifications or "representations" of functions in dynamical reasoning, I give the procedure for the simplest of all cases, that of the linear harmonic oscillator. Starting from the well-known equation:

$$m(d^2q/dt^2) = -k^2q \quad (72)$$

we put for $-k^2q$ its representation $(\nu_0; -k^2q_s)$ and for $m(d^2q/dt^2)$ its representation $(\nu_0; -4\pi^2ms^2\nu_0^2q_s)$. Corresponding coefficients are to be equated:

$$4\pi^2ms^2\nu_0^2q_s = k^2q_s. \quad (73)$$

It is evident that to solve this system of equations we must put $q_s = 0$ for every value of s excepting one. *The motion is therefore sinusoidal, of a single frequency.* Choosing $s = 1$ for the term of which the coefficient is not to vanish, we get $k^2 = 4\pi^2m\nu_0^2$, a familiar result. The value of q_1 remains indeterminate, any amplitude being compatible with Eq. (72). If on the right of Eq. (72) we put q^2 , or any positive-integer power of q except the first, or any sum of such powers of q , the values of q_s no longer vanish and some at least of the other frequencies come in.

In Eq. (63) at the beginning of this section, the Fourier expansion of the function $x(t)$ was written

down with plus signs connecting its terms. Later the plus signs quietly slipped out of the picture, leaving behind the ensemble of sinusoidal vibrations, denoted by their coefficients C_s or q_s . It was possible indeed to solve the dynamical equation (72) of the harmonic oscillator without bringing back the plus signs, though they could have been restored at any time. It has been good practice to dispense with them, for now that we are going over to quantum mechanics, we are about to meet with ensembles of sinusoidal vibrations from which the plus signs are altogether omitted. The vibrations are supposed to exist together, and yet it is worse than useless to add their expressions.

TERM FREQUENCIES AND EMISSION FREQUENCIES OF A QUANTUM-MECHANICAL ATOM MODEL; MATRIX ALGEBRA

We think of an ensemble of sinusoidal vibrations $q_s \exp(2\pi i\nu_s t)$, in which the frequencies ν_s form such a set as occurs in an actual spectrum. That is to say: these frequencies are not integer multiples $s\nu_0$ of a common fundamental ν_0 , they are the differences between the members of a set of "term frequencies" $W_1/h, W_2/h, W_3/h, \dots$, which themselves are in general not integer multiples of a common divisor. No plus signs are to connect the members of this ensemble, and no summation sign is to be placed before them.

The notation just suggested is inadequate. Each member of the ensemble should be marked not by a single symbol s , but by the two indices, say r and s , of the terms of which the difference gives the frequency. The typical member is to be written:

$$q_{rs} \exp(2\pi i\nu_{rs} t); \quad \nu_{rs} = (W_r - W_s)/h. \quad (74)$$

Negative frequencies (in the cases where $W_r < W_s$) and zero frequencies (in the cases where $r = s$) are admitted; as before, they are hindrances to visualization, but helpful in arriving at physical results. As with the Fourier expansion, q_{rs} is put equal to q_{rs}^* , and this is justified in the event. As a symbol for the ensemble as a whole, patterned after the symbol $(\nu_0; q_s)$ for the Fourier expansion of $q(t)$, I choose temporarily:

$$\text{Ensemble } q: (W_1, W_2, \dots; q_{rs}) \quad (75)$$

the system of values W (the energy values of the stationary states) implying their difference frequencies $(W_r - W_s)/h$.

The statements which follow are best regarded as qualities imposed on these ensembles by postulate, to be justified eventually by the usefulness of quantum mechanics in interpreting the phenomena of nature.

For the ensemble dq/dt we put:

$$\text{Ensemble } \dot{q}: (W_1, W_2, \dots; 2\pi i\nu_{rs}q_{rs}). \quad (76)$$

This is analogous to Eq. (69), and follows directly from the later-to-be-given rules for adding and subtracting ensembles. The rule for second and higher time derivatives follows immediately.

As soon as we begin to think of multiplying the ensemble q by itself, the striking peculiarity of quantum mechanics appears. If we had connected the various members (74) by plus signs, and had then multiplied the two summations together in the usual way, we should have got a procession of exponentials with frequencies $(\nu_{rs} + \nu_{r's'})$. The frequencies of the product would have been the sums of the frequencies of q , taken two by two. In general, these sum frequencies are *not* included among the frequencies of q . If we call the ensemble thus obtained the "square of q " (as we have a perfect right to do) we are obliged to say that q^2 has a different set of frequencies from q .

There are however some among the frequencies $(\nu_{rs} + \nu_{r's'})$ which are included among the frequencies or, let me say, in the "spectrum" of q . These are the ones for which s and r' are the same; because of these, by virtue of the second of Eqs. (74):

$$\begin{aligned} h(\nu_{rs} + \nu_{r's'}) &= h(\nu_{rs} + \nu_{ss'}) = (W_r - W_s) \\ &+ (W_s - W_{s'}) = W_r - W_{s'} = h\nu_{rs'}. \end{aligned} \quad (77)$$

If therefore in multiplying the ensemble by itself we leave out of the product all members having frequencies other than these—if we multiply each member (rs) of q not by itself and all the rest, but only by those members ($r's'$) for which $r' = s$ —then the resulting ensemble will have the same frequencies as q , and no others. This is what shall be called "the ensemble q^2 ."

The same thing happens when we multiply two ensembles q and p having the same spectrum; what by all precedent would be called the

"product" does not have the same and only the same frequencies as each of its factors. We however are going voluntarily to omit most of the terms which are obtained by the usual method of computation, saving only those which have the spectrum-frequencies of q and p . These which are left shall be called "the ensemble qp ."

The procedure is clear when one adopts yet another symbol for ensembles of the sort with which we are dealing:

$$\text{Ensemble } q: \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} & \cdots \\ q_{21} & q_{22} & q_{23} & q_{24} & \cdots \\ q_{31} & q_{32} & q_{33} & q_{34} & \cdots \\ q_{41} & q_{42} & q_{43} & q_{44} & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix} \quad (78)$$

the exponential factors being omitted for convenience, though strictly each member q_{rs} is to be thought of as $q_{rs} \exp(i\nu_{rs}t)$. Let \mathbf{q} and \mathbf{p} stand for two ensembles having the same set of constants W_s and frequencies ν_{rs} . (Two ensembles are never added nor multiplied together unless this is true of them.) Boldface type shall henceforth be used for ensembles considered as a whole, ordinary italics (with subscripts) for the individual members of such ensembles.

What shall be called "the product \mathbf{qp} " is the ensemble constructed according to the following rule of calculation:

$$(\mathbf{qp})_{ik} = \sum_j q_{ij} p_{jk}. \quad (79)$$

The member of \mathbf{qp} which stands at the crossing of the i th row and the k th column is obtained by multiplying each member of the i th row of \mathbf{q} (counting along the row from the left) by the corresponding member of the k th column of \mathbf{p} (counting down the column from the top), and adding these products all together. The frequency ν_{ik} of this member of \mathbf{qp} is the same as that of the similarly-situated members of \mathbf{q} and of \mathbf{p} . What shall be called "the square of \mathbf{q} " is obtained by putting \mathbf{q} for \mathbf{p} in Eq. (79). What shall be called "the product \mathbf{pq} " is the ensemble constructed according to the following rule of calculation:

$$(\mathbf{pq})_{ik} = \sum_j p_{ij} q_{jk}, \quad (80)$$

its member (ik) being obtained by multiplying

the members of the i th row of \mathbf{p} by the corresponding ones of the k th column of \mathbf{q} and adding these products all together.

Finally, the "sum ($\mathbf{p}+\mathbf{q}$)" shall be obtained by adding corresponding members of the two ensembles, and the "difference ($\mathbf{p}-\mathbf{q}$)" by subtracting the members of \mathbf{q} from the corresponding ones of \mathbf{p} . The product of \mathbf{q} by a scalar C shall be the ensemble of which each member is C times the corresponding member of \mathbf{q} .

Ensembles of the aspect of Eq. (78), subjected to the rules of calculation which have just been stated, are called "matrices." The rules themselves are those of "matrix algebra," a branch of mathematics invented and developed long before quantum mechanics was ever dreamt of. It was ready and waiting for Heisenberg and the other inventors of quantum mechanics, as the Greek theorems concerning conic sections had been ready and waiting for Kepler and Newton and the subsequent builders of celestial mechanics. It justifies the following brief statement of one of the principles of quantum mechanics:

In respect of the emission and absorption of light, atoms and molecules behave like systems of vibrations which conform to the rules of matrix algebra instead of the rules of Fourier expansions.

MORE ABOUT MATRIX ALGEBRA

Of these rules, those of matrix addition, matrix subtraction, and differentiation of a matrix with respect to time (or some other ordinary variable) are of the kind which is naively called "self-evident." Matrix multiplication is peculiar, and merits further comment. Since the right-hand members of Eqs. (79) and (80) are usually different, "multiplication" of a pair of matrices \mathbf{p} and \mathbf{q} is really the common name of two distinct operations denoted by \mathbf{qp} and \mathbf{pq} . These symbols stand for matrices, the difference between which is a matrix ($\mathbf{pq}-\mathbf{qp}$), thus defined:

$$(\mathbf{pq}-\mathbf{qp})_{ik} = \sum_j (\mathbf{p}_{ij}\mathbf{q}_{jk} - \mathbf{q}_{ij}\mathbf{p}_{jk}), \quad (81)$$

which in general does not vanish. This is usually expressed by saying that "matrix multiplication is noncommutative." It is, however, associative. We arrive at the same matrix if we form the product \mathbf{qp} (call it \mathbf{r}) and then the product \mathbf{pr} , as if we form the product \mathbf{pq} (call it \mathbf{s}) and then

the product \mathbf{sp} ; the symbol \mathbf{pqp} is applied to the result.

Immediately after Eq. (71), I gave various theorems about the systems of vibrations forming Fourier expansions. These can now be repeated for the vibration systems conforming to matrix algebra, in the same words but with the new meanings for the words. Let \mathbf{q} stand for a system represented by a matrix with a given spectrum of frequencies; then \mathbf{q}^2 is another matrix with the same spectrum. So are all the other positive-integer powers of \mathbf{q} ; \mathbf{q}^3 is the matrix product of \mathbf{q} and \mathbf{q}^2 , \mathbf{q}^4 that of \mathbf{q} and \mathbf{q}^3 , and so on upwards. So is any sum, finite or infinite, of positive-integer powers of \mathbf{q} multiplied by constants. So is any function of \mathbf{q} expressible as such a sum, and this is a very wide range of functions indeed. Let \mathbf{p} stand for another matrix with the same spectrum of frequencies as \mathbf{q} . Then \mathbf{qp} and \mathbf{pq} are matrices with the same frequencies, and so are \mathbf{qqp} and \mathbf{ppq} and \mathbf{qpq} , and all products of the form $\mathbf{p}^a\mathbf{q}^b\mathbf{p}^c\mathbf{q}^d\mathbf{p}^e\cdots$ where a, b, c, d, e, \cdots stand for positive integers, and all sums finite or infinite of such products multiplied by constants. We must be careful never to regroup the factors in such a product. If a quantum-mechanical argument leads us to the expression \mathbf{pqp} (for example) we must not cede to the temptation of rewriting it as $\mathbf{p}^2\mathbf{q}$. The first of these symbols represents the matrix-product $\mathbf{p}(\mathbf{qp})$ and the second represents $\mathbf{p}(\mathbf{pq})$; and since \mathbf{qp} and \mathbf{pq} are in general not the same, neither are \mathbf{pqp} and $\mathbf{p}^2\mathbf{q}$ the same.

"Matrix multiplication" is thus a peculiar process, which has proved to be highly expedient in quantum mechanics though it was originally defined for its own sake. "Differentiation of one matrix with respect to another" is likewise a peculiar process so defined as to be expedient in quantum mechanics. More precisely, there is a pair of processes, which Born defines as differentiation of the first kind and of the second kind respectively. With a function of the form $\mathbf{q}^a - a$ being a positive integer—the definitions coincide and the result looks familiar:

$$d(\mathbf{q}^a)/d\mathbf{q} = a\mathbf{q}^{a-1}. \quad (82)$$

This is the only case of which this book will treat. I will mention, however, that with a function such as \mathbf{pq}^2 the definitions differ, one giving $2\mathbf{pq}$ and the other $\mathbf{qp}+\mathbf{pq}$ for the derivative with re-

spect to q . If the function is rounded off by adding q^2p , the two definitions concur in giving $2(qp + p^2q)$ for $d(pq^2 + q^2p)/dq$. In quantum mechanics the energy function is always rounded off in this manner.²⁰ A product $p^a q^b$ does not appear except as a member of a sequence $(p^a q^b + p^{a-1} q^{b+1} p + p^{a-2} q^b p^2 + \dots + q^b p^a)$, which may be written $\sum_{s=0}^a p^{a-s} q^b p^s$. The derivatives of this sequence with respect to p and q are $a \sum_{s=0}^{a-1} p^{a-1-s} q^b p^s$ and $b \sum_{s=0}^{b-1} q^{b-1-s} p^a q^s$.

In certain of the equations which follow a matrix will be affirmed to be "constant in time"; what can this mean? Of all the members or "elements" forming such an ensemble as appears in Eq. (78), each involves a time factor, which is $\exp(2\pi i \nu_{rs} t)$ for the element (rs). The only elements which are independent of time are the ones for which the vanishing of ν_{rs} reduces this factor to unity. It was postulated at the start that the frequencies ν_{rs} should be the differences $(W_r - W_s)/h$ between the members of a set of term frequencies $W_1/h, W_2/h, \dots$. Thus the elements for which $r = s$ are independent of time. These are the elements lying along the "principal diagonal" of the matrix which starts at the left-hand upper corner. A matrix for which all the elements but these are zero is called a "diagonal matrix" and is independent of time.²¹ If in the course of a reasoning we arrive at the conclusion that a certain matrix is diagonal, that is tantamount to concluding that it is constant in time.

The quantities W_i , with their subscripts doubled, may themselves be regarded as the nonvanishing elements of a diagonal matrix \mathbf{W} thus defined:

$$\mathbf{W} = \begin{vmatrix} W_{11} & 0 & 0 & 0 & \dots \\ 0 & W_{22} & 0 & 0 & \dots \\ 0 & 0 & W_{33} & 0 & \dots \\ \dots & \dots & \dots & W_{44} & \dots \end{vmatrix}. \quad (83)$$

This enables us to make a striking assertion about any matrix, \mathbf{p} or \mathbf{q} for instance, having the fre-

quencies $\nu_{rs} = (W_{rr} - W_{ss})/h$. The derivative $\dot{\mathbf{p}}$ is the matrix of which the rs element is $2\pi i \nu_{rs} p_{rs}$. By applying the rules of matrix-multiplication which have just been formulated, and writing out the products \mathbf{Wp} and \mathbf{pW} and \mathbf{Wq} and \mathbf{qW} in the fashion illustrated in Eq. (78), the reader will find that:

$$\begin{aligned} \dot{\mathbf{p}} &= (2\pi i/h)(\mathbf{Wp} - \mathbf{pW}); \\ \dot{\mathbf{q}} &= (2\pi i/h)(\mathbf{Wq} - \mathbf{qW}), \end{aligned} \quad (84)$$

an interesting result, bringing out vividly the non-commutative quality of the process called "matrix multiplication," and largely responsible for the peculiar fitness of this type of algebra to atomic physics.

A special case of a diagonal matrix is that in which all the nonvanishing elements have a common (scalar) value c ; a still more special case is that in which this common value is unity, and the matrix is named the "unit matrix" and denoted by \mathbf{E} . Multiplication of any other matrix, say \mathbf{a} , by \mathbf{E} is commutative, and the product is \mathbf{a} itself. The diagonal matrix of which all the nonvanishing elements are equal to a scalar c is denoted (obviously) by $c\mathbf{E}$; multiplication of \mathbf{a} by $c\mathbf{E}$ is commutative, and the product is $c\mathbf{a}$.

DESIGN OF CLASSICAL AND OF QUANTUM-MECHANICAL ATOM MODELS

We are now ready to apply the methods of quantum mechanics to particular atom models. In keeping with the scope and size of this book, I will speak almost entirely of the simplest case of all: the molecule model based on the concept of the linear harmonic oscillator, and employed (as a first approximation) in interpreting the spectra of diatomic molecules. Of course it falls far short of the most general case, and it lacks some of the features which appear in the model of the hydrogen atom and produce the spectacular numerical agreements of theory with experiment which that particular model affords. Nevertheless it will serve as a good illustration of the basic ideas; and where it is too specialized, I can occasionally insert the more general theorems to which a less simplified model would lead.

What is an atom model, in quantum mechanics? To answer this, I begin by recalling what it is in classical theory, and in the earlier forms of

²⁰ See, however, N. H. McCoy, Proc. Nat. Acad. Sci. 18, 674-676 (1932).

²¹ If the quantities W_i are so chosen that some coincide with others, e.g., $W_i = W_j$, then a matrix may be constant in time even though its element (ij) does not vanish. Such cases (the so-called "degenerate cases") are extremely important in quantum mechanics, but we shall not reach any in this book.

quantum theory. Essentially, *it is an energy function, stated together with certain dynamical equations.* To the eye of the body the linear harmonic oscillator is a heavy weight attached to the end of a coiled spring, travelling to and fro along a definite path. To the eye of the mind, a linear harmonic oscillator is an energy function composed of a kinetic energy term $(\frac{1}{2})m\dot{x}^2$ and a potential energy term $(\frac{1}{2})k^2x^2$, with a momentum p equal to $m\dot{x}$ and certain equations for p and x .

This function combined with these equations is the oscillator, insofar as it does service as an atom model. I write them out in the "Hamiltonian" form, in which q denotes the coordinate x , and the momentum p is used in preference to \dot{q} as one of the independent variables. The energy, when expressed as function of p and q , is known as the "Hamiltonian function" $H(p, q)$. An atom model consists of some particular choice of a function $H(p, q)$, combined with the two "canonical equations" of Hamilton:

$$\partial H/\partial q = -dp/dt; \quad \partial H/\partial p = dq/dt. \quad (85)$$

From these equations we derive:

$$dH/dt = (\partial H/\partial q)(dq/dt) + (\partial H/\partial p)(dp/dt) = 0. \quad (86)$$

The energy function is constant in time. This may seem to exclude cases in which an atom is influenced by another atom or by an external field, but usually one adopts the artifice of regarding the two atoms or the atom-plus-field as a single system for which the energy is constant. (Otherwise one may make H a function $H(p, q, t)$ in which t appears explicitly.)

I exhibit these postulates in the general form appropriate to a system of f degrees of freedom:

$$H(p_1, \dots, p_f; q_1, \dots, q_f); \quad \partial H/\partial q_i = -p_i; \\ \partial H/\partial p_i = \dot{q}_i; \quad i = 1, 2, 3, \dots, f \quad (87)$$

and again in the special form belonging to the linear harmonic oscillator:

$$H = p^2/2m + k^2q^2/2; \quad dp/dt = -k^2q; \\ dq/dt = p/m. \quad (88)$$

From these equations one readily derives Eq. (72), and all the consequence which flow classically from that equation.

In designing an atom model, the procedure of classical mechanics consists in choosing some particular Hamiltonian function $H(p, q)$ and employing Eqs. (87). It leads to a spectrum of frequencies which are linear combinations of integer multiples of f fundamental frequencies, as well as to other conclusions which are not in accord with observation.

The procedure of quantum mechanics looks exactly the same on the surface. Variables p and q are introduced, and a function $H(p, q)$ is chosen, often looking precisely like the function selected in the classical theory for the same atom model; and Eqs. (87) are used. But now p and q are matrices \mathbf{p} and \mathbf{q} of the type (78); H is a matrix function \mathbf{H} ; every product of \mathbf{p} and \mathbf{q} which appears in \mathbf{H} is a matrix product; and the derivatives are obtained by matrix differentiation. The outward and visible form has subsisted, but the meaning beneath the symbols is transformed.

QUANTUM CONDITION IN MATRIX ALGEBRA, AND ITS CONSEQUENCES

Keeping to the case of the harmonic oscillator, we will now envisage Eqs. (88) as matrix equations.

In the left-hand members of the canonical equations stand the time derivatives, for which peculiar equivalent expressions have already been derived (Eqs. 84). In their right-hand members stand the derivatives $-\partial\mathbf{H}/\partial\mathbf{q}$ and $+\partial\mathbf{H}/\partial\mathbf{p}$, which in our special case are $-k^2\mathbf{q}$ and \mathbf{p}/m . For these also I proceed to derive peculiar equivalents, by a method which will seem highly artificial, but is evolved from a general theorem which shall be stated afterwards. We have:

$$\mathbf{Hq} - \mathbf{qH} = (\mathbf{p}^2/2m)\mathbf{q} + (k^2\mathbf{q}^2/2)\mathbf{q} \\ - \mathbf{q}(\mathbf{p}^2/2m) - \mathbf{q}(k^2\mathbf{q}^2/2) \\ = (1/2m)(\mathbf{p}^2\mathbf{q} - \mathbf{qp}^2) \\ = (1/2m)[\mathbf{p}(\mathbf{pq} - \mathbf{qp}) + (\mathbf{pq} - \mathbf{qp})\mathbf{p}] \quad (89a)$$

and similarly:

$$\mathbf{Hp} - \mathbf{pH} = (k^2/2)[-\mathbf{q}(\mathbf{pq} - \mathbf{qp}) - (\mathbf{pq} - \mathbf{qp})\mathbf{q}]. \quad (89b)$$

Let us now make the following assumption about the matrix $(\mathbf{pq} - \mathbf{qp})$:

$$\mathbf{pq} - \mathbf{qp} = \begin{vmatrix} h/2\pi i & 0 & 0 & \cdots \\ 0 & h/2\pi i & 0 & \cdots \\ 0 & 0 & h/2\pi i & \cdots \\ 0 & 0 & 0 & h/2\pi i \cdots \\ \cdot & \cdot & \cdot & \cdot \end{vmatrix}. \quad (90a)$$

This is the "quantum condition" of quantum mechanics, expressed in the way suitable to matrix algebra. I give alternative ways of expressing it. The next is the most succinct:

$$\mathbf{pq} - \mathbf{qp} = (h/2\pi i)\mathbf{E}, \quad (90b)$$

while the third will acquaint the reader with the practice of using the symbol δ_{rs} to denote a quantity which is unity when $r = s$ and zero when $r \neq s$:

$$(\mathbf{pq} - \mathbf{qp})_{rs} = (h/2\pi i)\delta_{rs}. \quad (90c)$$

Making this assumption, we convert the right-hand members of Eqs. (89) into $(\mathbf{p}/m)(h/2\pi i)$ and $-h^2\mathbf{q}(h/2\pi i)$ respectively. Substituting into Eqs. (88) and (84), we get:

$$\mathbf{Wp} - \mathbf{pW} = \mathbf{Hp} - \mathbf{pH}; \quad \mathbf{Wq} - \mathbf{qW} = \mathbf{Hq} - \mathbf{qH}, \quad (91)$$

which equations suggest at once that the energy matrix is identical with the matrix of the quantities \mathbf{W} . This is indeed the case: Eqs. (91) require that the energy matrix should be a diagonal matrix, and that its elements along the principal diagonal should differ only by a common arbitrary constant from the corresponding elements along the principal diagonal of \mathbf{W} . The reader has probably been taking this for granted all along, but it is a corollary of the quantum-condition (90), and therefore recedes to the rank of a theorem if Eq. (90) be taken as fundamental.

The general theorem mentioned above, of which Eq. (89) is a special case, is the following:

$$(\mathbf{fq} - \mathbf{qf}) = (d\mathbf{f}/d\mathbf{p})(\mathbf{pq} - \mathbf{qp});$$

$$(\mathbf{pf} - \mathbf{fp}) = (d\mathbf{f}/d\mathbf{q})(\mathbf{pq} - \mathbf{qp}). \quad (92)$$

With the "first kind" of differentiation it is true for any matrix function \mathbf{f} of \mathbf{p} and \mathbf{q} ; with the second kind of differentiation it is true for "rounded-off" functions such as those specified after Eq. (82).²² We could indeed accept Eqs.

²² The argument is, that if Eq. (92) is true for any two functions \mathbf{f} and \mathbf{g} , it is true also for $\mathbf{f} + \mathbf{g}$ and \mathbf{fg} . For

(92) as the definitions of the two processes of differentiation of a matrix function of \mathbf{p} and \mathbf{q} with respect to these two matrix variables. There would be nothing illogical about such a procedure; its disadvantage is, that it veils the partial analogy between these processes and the familiar kind of differentiation, which was illustrated in Eq. (82); but since the analogy is imperfect, there is something to be said for disregarding it in the definition of these new processes.

Thus the identity of \mathbf{H} and \mathbf{W} is not by any means restricted to the particular form of \mathbf{H} which was introduced in Eq. (88). If we use a definition of "differentiation" which is either given by Eq. (92) or else leads straight to Eq. (92), then any function \mathbf{H} which conforms to the canonical equations (it may be necessary to say in addition, that \mathbf{H} must be rounded off in the way above described) is substantially the same as \mathbf{W} .

The situation is now as follows:

We have laid aside the idea of coordinate \mathbf{q} and momentum \mathbf{p} as functions of time which can be written as sets of vibrations having frequencies $s\nu_0$ which are integer multiples of a fundamental frequency ν_0 ;

We have replaced it by the idea of coordinate \mathbf{q} and momentum \mathbf{p} as sets of vibrations having frequencies ν_{ij} which are the differences between the members of a set of terms W_{ii}/h ;

We have written these sets as matrices;

We have defined "addition" and "multiplication" of matrices in certain distinctive ways, with the result that such an expression as $[(1/2m)\mathbf{p}^2 + (h^2/2)\mathbf{q}^2]$ and all other expressions which are written like algebraic functions have definite ascertainable meanings and are matrix functions of \mathbf{p} and \mathbf{q} , being themselves matrices with the same spectrum of frequencies ν_{ij} as \mathbf{p} and \mathbf{q} possess;

We have defined "differentiation" of matrices in certain distinctive ways, such that such expressions as $d\mathbf{G}/dt$, $\partial\mathbf{G}/\partial\mathbf{p}$, $\partial\mathbf{G}/\partial\mathbf{q}$ — \mathbf{G} standing for \mathbf{p} or \mathbf{q} or any matrix function of \mathbf{p} and \mathbf{q} —have definite ascertainable meanings and are

$\mathbf{f} + \mathbf{g}$ the assertion is obviously true. For \mathbf{fg} we get:

$$\mathbf{fgq} - \mathbf{qgf} = \mathbf{f}(\mathbf{gq} - \mathbf{qg}) + (\mathbf{fq} - \mathbf{qf})\mathbf{g}$$

$$= [\mathbf{f}(d\mathbf{g}/d\mathbf{p}) + (d\mathbf{f}/d\mathbf{p})\mathbf{g}](\mathbf{pq} - \mathbf{qp}).$$

With the first kind of differentiation the quantity in square brackets is equal to $d(\mathbf{fg})/d\mathbf{p}$. With the second kind it is true if the stated condition is observed. Now (92) is true for $\mathbf{f} = \mathbf{q}$ and $\mathbf{g} = \mathbf{p}$; whence the theorem.

matrices with the same spectrum of frequencies as p and q :

We have introduced the quantum condition (90); and now it turns out, that the diagonal matrix constituted by the quantities W_{ii} is identical with a matrix function of p and q which obeys the canonical equations—that is to say, the matrix equivalents of the canonical equations—and which therefore there is full justification for calling the “energy” of the system characterized by these quantities W_{ii}/h and their difference frequencies ν_{ij} .

This strongly suggests that matrix algebra with the foregoing rules combined with the quantum condition is the proper, or at any rate a proper, technique for dealing with problems of atomic physics.

We require, however, much more than the foregoing result. First of all we require the actual values of the quantities W_{ii} . Can the technique of matrix algebra, combined with the quantum condition and with a suitable form for the Hamiltonian function $H(p, q)$, supply us with these?

In principle, it can; but the task is extraordinarily difficult, except in certain extremely simple cases. The technique of determining these quantities which was discovered by Schroedinger, and is known specifically as that of wave mechanics, is far more manageable; and it turns out that by an extension of that same technique, the elements of the matrices p and q and of matrix functions of p and q may be determined. I will close this booklet with an exposition of that method; but before we take it up, it will be interesting to confirm for ourselves that in the case of the linear harmonic oscillator with its Hamiltonian function previously given, the rules of matrix algebra and the quantum condition suffice to fix the quantities W_{ii} in entire agreement with Schroedinger's result which I quoted earlier.

Writing out the expression for any element on the principal diagonal of the matrix H , we have:

$$\begin{aligned} H_{ii} &= (1/2m)(p^2)_{ii} + (k^2/2)(q^2)_{ii} \\ &= (1/2m) \sum_j p_{ij} p_{ji} + (k^2/2) \sum_j q_{ij} q_{ji}. \end{aligned} \quad (93)$$

Replace each element of the matrix p by the corresponding element of the matrix $m\dot{q}$, so that

p_{ij} becomes $2\pi i \nu_{ij} q_{ij}$; write ν_0^2 for $(k^2/4\pi^2 m)$; we obtain:

$$H_{ii} = 2\pi^2 m \sum_j (-\nu_{ij} \nu_{ji} + \nu_0^2) q_{ij} q_{ji}. \quad (94)$$

Writing out the expression for any element of the principal diagonal of the matrix $(pq - qp)$, we find:

$$\begin{aligned} (pq - qp)_{ii} &= m(\dot{q}q - q\dot{q})_{ii} \\ &= 2\pi i m \sum_j (\nu_{ij} - \nu_{ji}) q_{ij} q_{ji}. \end{aligned} \quad (95)$$

This expression is to be equated to $(h/2\pi i)$,²³ and the result is to be used to evaluate H_{ii} . Before going further, we recall that according to the mode of construction of these matrices, ν_{ij} is equal to $-\nu_{ji}$ and q_{ji} is equal to q_{ij}^* , so that $q_{ij} q_{ji} = q_{ij} q_{ij}^* =$ squared amplitude of q_{ij} . Making these substitutions into Eqs. (93) and (95):

$$H_{ii} = 2\pi^2 m \sum_j (\nu_{ij}^2 + \nu_0^2) q_{ij} q_{ji}, \quad (96a)$$

$$-h/8\pi^2 m = \sum_j \nu_{ij} q_{ij} q_{ji}.$$

Now returning to Eq. (72), we envisage it as a matrix equation and derive from it a system of equations replacing Eq. (73):

$$4\pi^2 m \nu_{ij}^2 q_{ij} = k^2 q_{ij}. \quad (97)$$

From these it is inferred that for any element (ij) of the matrix q , either ν_{ij} is equal to $+\nu_0$ or to $-\nu_0$, or else the coefficient q_{ij} itself vanishes. Now think of all the elements belonging to a single row of the matrix, say the i th row. Not more than one of these may have the frequency $+\nu_0$, and not more than one may have the frequency $-\nu_0$; otherwise two or more of the quantities H_{ii} would have to be equal, a complexity which occurs in many cases, but not in that of the linear harmonic oscillator. We put $\nu_{i, i-1}$ equal to $+\nu_0$, and $\nu_{i, i+1}$ equal to $-\nu_0$; and we put q_{ij} equal to zero in all cases except $j = i \pm 1$. Now Eqs. (96) are reduced to the forms:

$$H_{ii} = 4\pi^2 m \nu_0^2 (q^2_{i, i-1} + q^2_{i, i+1}), \quad (98a)$$

$$-h/8\pi^2 m = \nu_0 (q^2_{i, i-1} - q^2_{i, i+1}), \quad (98b)$$

²³ It should perhaps have been emphasized earlier that when i is written as a factor it always stands for $(-1)^i$, and is never to be confused with subscript i .

and putting i equal consecutively to 1, 2, 3, \dots , we obtain:

$$H_{11} = \frac{1}{2}h\nu_0; H_{22} = (3/2)h\nu_0; \dots H_{jj} = (j + \frac{1}{2})h\nu_0 \quad (99)$$

which is the sequence of permitted energy values previously derived by Schroedinger's method.

The matrix technique thus proves itself capable of solving the first and fundamental problem of atomic theory,—the problem, that is to say, of devising plausible atom models with definite calculable systems of permitted energy values, which can be compared with experiment. In this respect it does not excel Schroedinger's technique of seeking the eigenvalues of a wave equation, and indeed is generally much harder to apply. However, in reviewing the foregoing process one notes a definite indication that the matrix method may be capable of something more. The coefficients q_{ij} , which were evaluated in the course of calculating the energy values H_{ii} , seem to be connected with the problem of radiation. For at the start, the concept of the matrix q was derived from the concept of the displacement of a particle; and according to classical theory, the displacement of an electrified particle is required to cause radiation, inasmuch as outflow of radiant energy results from acceleration of charges, and acceleration is always accompanied by displacement. Moreover in working through the problem of the harmonic oscillator we have been led, on the one hand to the inference that there is no radiation of frequency ν_{ij} except when $i-j = \pm 1$, and on the other hand to the inference that q_{ij} vanishes except when $i-j = \pm 1$; this strongly suggests that q_{ij} is some sort of a measure of the intensity of radiation of frequency ν_{ij} . To compute all these coefficients by the matrix technique would be a task of really extraordinary difficulty, except in this extremely simple case of the harmonic oscillator. It is therefore fortunate—as well as profoundly significant—that an extension of Schroedinger's technique permits us to compute them, and eventually to test the assertion that the matrix q contains the description of the light which an atom emits, much as the matrix H contains the description of the permitted energy values of that atom. To this task we now address ourselves.

CONCEPT AND EMPLOYMENT OF OPERATORS IN QUANTUM MECHANICS

We return to the wave equation (47) applied to an endless uniform beam, which I repeat with both sides multiplied by $(h/2\pi i)$:

$$(h/2\pi i)^2 \partial^2 \Psi / \partial x^2 = p^2 \Psi. \quad (101)$$

In mathematics and in mathematical physics, it is sometimes possible to make real progress by altering the phrasing of theorems and equations, and letting one's imagination wander in paths suggested by the altered wordings. I now introduce some ways of phrasing Eq. (101).

Let Ψ_0 denote any solution of that equation. To multiply Ψ_0 by p twice leads to the same result as to differentiate Ψ_0 twice with respect to x and multiply each time by $(h/2\pi i)$. Introducing the word "operation": the operations of multiplying by p , and of differentiating with respect to x and multiplying the result by $(h/2\pi i)$, are equivalent when applied to Ψ_0 . Introducing the word "operator": when applied to Ψ_0 , the operator p and the operator $(h/2\pi i)(\partial/\partial x)$ are equivalent. Rearranging the sentence: the solutions of Eq. (101), which is to say, the functions which we employ in the treatment of the uniform beam in field-free space, are the functions to which the application of operator p produces the same result as the application of operator $(h/2\pi i)\partial/\partial x$.

Now write down the Hamiltonian function for a system composed of a particle in a field of force which does not vary with time:

$$H(p, q) = \text{kinetic energy} + \text{potential energy} \\ = [(1/2m)(p_x^2 + p_y^2 + p_z^2) + V(x, y, z)]. \quad (102)$$

Write the symbol Ψ immediately after the quantity in square brackets, as though H were to be multiplied into Ψ . Following the hint of the preceding paragraph, replace p_x by $(h/2\pi i)(\partial/\partial x)$, and p_y by $(h/2\pi i)(\partial/\partial y)$, and p_z by $(h/2\pi i)(\partial/\partial z)$. Leave (x, y, z) unchanged, so that $V(x, y, z)\Psi$ continues to mean the product of V by Ψ . We now have the so-called "Hamiltonian operator"—hereafter to be denoted by H —operating on whatever is symbolized by Ψ :

$$\left[-\frac{\hbar^2}{8\pi^2 m} \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) + V \right] \Psi. \quad (103)$$

The reader will recognize two of the terms which appear in Schroedinger's wave equation (48). The equation itself appears if we let ourselves be guided by the classical conclusion that $H(p, q)$ is unvarying in time (Eq. 86). Multiply Ψ by a constant W ; or, to adopt the new phraseology, operate on Ψ by a constant W , the boldface italic letter implying that we are to think of the constant as an operator, though the operation which it performs is pure multiplication. Equate the result to $H\Psi$:

$$H\Psi = W\Psi. \quad (104)$$

Here is Schroedinger's equation again. It will be recalled that there are certain values of the constant W , for which there are distinctive solutions of Eq. (104). The result of applying the Hamiltonian operator to any one of these, say Ψ_{ii} , is the same as the result of multiplying Ψ_{ii} by the corresponding value W_{ii} of the constant. Rephrasing this so as to bring it into the form of a general and fundamental theorem:

Given an atom model, which is to say, a Hamiltonian function H . Convert it into the corresponding Hamiltonian operator H by writing $(h/2\pi i)(\partial/\partial q_1), \dots (h/2\pi i)(\partial/\partial q_r)$ for $p_1, \dots p_r$. There are certain distinctive functions (eigenfunctions) Ψ_{ii} , and corresponding to these there are constants (eigenvalues) W_{ii} of the dimensions of energy, such that when H is applied to Ψ_{ii} the result is the same as when W_{ii} is multiplied into Ψ_{ii} . The eigenvalues are the permitted energy values of the atom models, and the eigenfunctions are the Ψ -functions previously mentioned. This is a new way of setting up wave equations.²⁴

The connection between the method of wave mechanics and the method of matrix algebra can be made in the following fashion.

Form the eigenvalues W_{ii} into a diagonal

matrix, the one which appears in Eq. (83); and the eigenfunctions Ψ_{ii} into a diagonal matrix, of which the element (ii) is Ψ_{ii} . The product of these two is also a diagonal matrix, its element (ii) being the product $W_{ii}\Psi_{ii}$. Availing ourselves of the symbol δ previously employed (Eq. 90c) we may describe it fully by writing $\delta_{ij}W_{ii}\Psi_{jj}$ as its typical element. We may regard it as the ensemble of all the permitted values of the right-hand member of Eq. (104); and the operator H acting on Ψ may be conceived to produce this diagonal matrix.

If we are to do this, we must learn to think of $H\Psi$ as a matrix. Not only this, but also the result of operating on Ψ by x , and the result of operating on Ψ by p_x —these two must be considered as matrices, and so likewise must be the result of applying to Ψ any operator which is a function of coordinates and of the differential operators which replace momenta according to the foregoing rule. But how conceive of $x\Psi$ as a matrix, or of $(h/2\pi i)(\partial\Psi/\partial x)$ as a matrix, when x has been introduced as an ordinary scalar variable and Ψ thus far has denoted something which might be any of the eigenfunctions Ψ_{ii} ?

It helps to lean as much as possible on the analogy afforded by the Fourier expansion, where a function denoted by a single symbol is represented by a sequence of functions forming a complete orthogonal set, each multiplied by a constant evaluated in a certain way which the orthogonality of the set makes feasible. The eigenfunctions of any problem form a complete orthogonal set (a fundamental theorem, which there will not be space to prove in this book) and thus it is equally feasible to expand or represent any function in terms of them. But we wish to have not a single sequence of terms, but a two-dimensional pattern forming a matrix, to represent the result obtained when an operator acts on Ψ .

Recall now that in dealing with the thing denoted by Ψ , we have always dealt with cases where Ψ stood for one or another of the eigenfunctions Ψ_{ii} . In cases where it could not stand for one of these (as, for instance, when W in Eq. (48) or Eq. (104) has some value not comprised among the eigenvalues) nothing has been said about Ψ , and indeed such cases have been disregarded altogether. The symbol Ψ

²⁴ Certain difficulties are likely to occur if the coordinates p, q are not Cartesian, or if there are "constraints" (in the technical sense of analytical mechanics). Thus with non-Cartesian coordinates, the Hamiltonian may contain such a term as (for instance) pq , and this prescription by itself would not tell us whether to write the corresponding term in the wave equation as $(h/2\pi i) \partial(\Psi q)/\partial q$ or as $(h/2\pi i) q \partial\Psi/\partial q$ (and indeed there would be still other possibilities). However this trouble may be avoided by using Cartesians, and the other is not likely to occur in an atom model.

refers to the totality of the eigenfunctions Ψ_i , the symbol $H\Psi$ refers to any or all of the functions $H\Psi_i$; the symbol $x\Psi$ refers to any or all of the functions $x\Psi_i$; the symbol $p_x\Psi$ refers to any or all of the functions $(h/2\pi i)(\partial\Psi_i/\partial x)$. In general: let f stand for any operator constructed according to the foregoing rule out of a function of coordinates and momenta; then $f\Psi$ refers to any or all of the functions $f\Psi_i$. Henceforth I will use (for convenience of printing) ordinary italics for an operator acting on one of the functions Ψ_i , reserving boldface italics for operator symbols printed by themselves or in front of the general symbol Ψ .

These functions $f\Psi_i$ are now to be expanded in terms of the orthogonal set Ψ_i , even as earlier we expanded functions in terms of the orthogonal set of exponentials of Eq. (65). For definiteness I take first the simplest case of all, the operator x or q_1 . For the i th of the functions $x\Psi_i$, we write:

$$x\Psi_i = a_{1i}\Psi_1 + a_{2i}\Psi_2 + a_{3i}\Psi_3 + \dots = \sum_j a_{ji}\Psi_j, \quad (105)$$

where the symbols a_{ji} stand for constants presently to be evaluated.^{24a}

I repeat the definition of orthogonality from Eq. (66), specializing it a little:²⁵

$$\int \Psi_j^* \Psi_i dx = \delta_{ij}. \quad (106)$$

The integration is taken over the interval in which the set of functions is orthogonal, which varies from case to case; sometimes it is from $-\infty$ to $+\infty$, sometimes the variable x varies cyclically and the interval is one complete cycle. The use of x as sole variable of integration implies an atom model of a single degree of freedom, having one coordinate x and one momentum p_x . In general there will be several, say r , degrees of freedom, and the functions Ψ_i

^{24a} The indices in such equations are sometimes written as here, sometimes so that

$$x\Psi_i = \sum_j a_{ij}\Psi_j,$$

a prolific source of confusion.

²⁵ In the earlier definition the integral was stated to be different from zero when, and only when, $i=j$. To give it the specific value unity in all these cases is perfectly feasible, because each of the functions Ψ_i involves an arbitrary factor which can be adjusted to bring out this result. This adjustment is called "normalization."

will be functions of (q_1, \dots, q_r) and the integration will be with respect to all these variables. Instead of dx , the r -dimensional volume-element (in the special case of Cartesian coordinates, it is $dq_1 \dots dq_r$) will then appear in the integral: I write dq for it, and state the property of orthogonality thus:

$$\int \Psi_i^* \Psi_j dq = \delta_{ij}. \quad (107)$$

Now multiply both sides of (105) by Ψ_j^* , and integrate over the interval in question. We get:

$$\int \Psi_j^* x \Psi_i dq = a_{ji}. \quad (108)$$

Repeat the process for every value of j and for all the functions $x\Psi_i$. The coefficients a now form an ensemble which may be written after the fashion of a matrix. This is the matrix which we will always associate with the symbol $x\Psi$ which we have been calling "the result of operating on Ψ by x ." The symbol $x\Psi$ shall have no other meaning to us. The symbol x without the shadowy and undefinable Ψ shall have no other meaning to us; *the operator is the matrix*. Operator x is the matrix of which the typical (ji) element is $\int \Psi_j^* x \Psi_i dq$. Operator p_x is the matrix of which the typical element is $(h/2\pi i) \int \Psi_j^* (\partial\Psi_i/\partial x) dq$. In general: if f stands for any operator which has been obtained in the foregoing way, by taking some function of the p 's and q 's and substituting $(h/2\pi i)(\partial/\partial q_i)$ for p_i :

$$\text{Operator } f: \text{ matrix } \int \Psi_j^* f \Psi_i dq \quad (109)$$

and here $f\Psi_i$ is the function obtained by operating with f upon the function Ψ_i .

It is easy to say that the ensemble of the quantities $\int \Psi_j^* f \Psi_i dq$ shall be written as a matrix, but there would be nothing gained by doing so unless such ensembles could be proved to conform to the rules of matrix algebra. I will give the necessary proof for the peculiar rule of matrix multiplication. The product fg of two operators f and g is the operator which when applied to such a function as (for instance) Ψ_i produces the same function as is otherwise produced when first g is applied to Ψ_i ; and then f is applied to $g\Psi_i$. We go through the latter

process. First we obtain the function $g\Psi_i$, and expand it, getting:

$$g\Psi_i = g_{1i}\Psi_1 + g_{2i}\Psi_2 + \cdots = \sum_k g_{ki}\Psi_k. \quad (110)$$

Here the coefficients g_{ki} are constants evaluated as explained above. Apply to this the operator f ; we get:

$$f \cdot g\Psi_i = g_{1i}f\Psi_1 + g_{2i}f\Psi_2 + \cdots = \sum_k g_{ki}f\Psi_k. \quad (111)$$

Expand each of the functions $f\Psi_k$; we get:

$$f\Psi_k = \sum_j f_{jk}\Psi_j \quad (112)$$

the coefficients f_{jk} being constants. Substituting this into Eq. (111) and regrouping the terms:

$$f \cdot g\Psi_i = \sum_k f_{ik}g_{k1}\Psi_1 + \sum_k f_{ik}g_{k2}\Psi_2 + \cdots, \quad (113)$$

so that the coefficient of Ψ_j in the representation of $f \cdot g\Psi_i$ is $\sum_k f_{ik}g_{kj}$. But this is precisely the (ij) element of the matrix formed by multiplying f into g according to the rule of matrix multiplication; which is what was to be demonstrated. The other rules of matrix algebra are readily shown to hold for these ensembles.

In Eq. (111), the operator f is applied to a linear combination of the eigenfunctions. Earlier (before Eq. 105) I may have seemed to say that an operator could be applied only to individual eigenfunctions. It is difficult to avoid misapprehensions in these matters, and I was making every effort at that point to avoid giving the idea that Ψ is some function or other which is to be expanded in terms of the eigenfunctions. Such is not the case; Ψ in that situation has no meaning by itself; all that is meant by $f\Psi$ is the matrix already defined. Operators however may be applied to linear combinations of the eigenfunctions, under such conditions as occur in Eq. (111).

It is easy to see that the matrix of the operator H is diagonal, and moreover is precisely the diagonal matrix (83) of the eigenvalues W_{ii} . Owing to Eq. (104), the result of applying H to Ψ_i is simply Ψ_i itself, multiplied by the constant W_{ii} . The expansion of any member of a complete orthogonal set of functions in terms of the set is simply that member itself. Thus:

$$H\Psi_i = W_{ii}\Psi_i; \quad H_{ij} = \delta_{ij}W_{ii}. \quad (114)$$

The theorem thus follows automatically from our choice of the eigenfunctions of Eq. (104)—which may be called simply the eigenfunctions of H —as the complete orthogonal set to be used in the representations.²⁶

This theorem was earlier obtained by the process of matrix mechanics, culminating in Eq. (91). There it was obtained by assuming the so-called "quantum condition" of Eq. (90a). The equivalent of this condition must have entered somewhere into the present chain of reasoning, and we shall not have a hard task to locate it. Let us derive the matrices for the operators pq and qp , and form their difference. Applying this to an eigenfunction Ψ_i , we get:

$$\begin{aligned} (pq - qp)\Psi_i &= -\frac{\hbar}{2\pi i} \left[\frac{\partial}{\partial q}(q\Psi_i) - q\frac{\partial}{\partial q}\Psi_i \right] \\ &= \left(\frac{\hbar}{2\pi i} \right) \Psi_i. \end{aligned} \quad (115)$$

Thus $(pq - qp)$ is a diagonal matrix; moreover, it is that diagonal matrix for which all the elements on the principal diagonal have the common value $(\hbar/2\pi i)$; it is the matrix $(\hbar/2\pi i)\mathbf{E}$. Thus, when we substituted $(\hbar/2\pi i)(\partial/\partial q_i)$ for p_i in the operators, we were introducing the quantum condition.

Now it is essential to restore, to each element of these matrices (and of all the other matrices with which there is ever occasion to deal) the factor dependent on time which was introduced at the beginning long ago and subsequently was omitted for ease of printing and ease of reading the formulae. Schroedinger's equation, it will be recalled, is not the fundamental wave equation. The underlying equation (of which Eq. (44) is a special form) is one which involves derivatives with respect to time as well as derivatives with respect to coordinates. Certain of its solutions are of the form $\Psi_i \exp(2\pi i W_{ii}/\hbar)t$, where Ψ_i stands for any eigenfunction of the operator H , and W_{ii}

²⁶ Had any other complete orthogonal set been chosen, Eqs. (105) to (113) would have been valid in terms of the new set, but the matrix of H might not and usually would not have been diagonal. In the case of the so-called "rotator" (an atom model patterned after a rigid body, employed in molecular theory) it is expedient to use the set of eigenfunctions of the angular momentum operator. It turns out that with respect to this set the matrix of the energy operator H is diagonal, as well as that of the angular-momentum operator itself.

for the corresponding value of the energy of the system, that is to say, for the corresponding eigenvalue. Any operator, say f , is to be associated with a matrix of which the typical element is not fully given by Eq. (109), but is rather to be obtained from Eq. (109) by supplying each Ψ with its appropriate time factor:

Operator f :

$$\int \Psi_i^* f W_i dq \exp [(2\pi i/\hbar)(W_{ii} - W_{ii})t]. \quad (116)$$

The elements for which $i = j$ retain the character of constants, but in all the others the coefficients $\int \Psi_i^* f \Psi_i dq$ figure as amplitudes of vibrations having the frequencies $(W_{ii} - W_{ii})/\hbar$.

RADIATION FROM ATOMS INTERPRETED BY CLASSICAL AND BY QUANTUM-MECHANICAL THEORY

We make one more, and the last, of those frank and outright borrowings from classical theory which are essential to the success of quantum mechanics, and prevent us from regarding it as a self-contained and independent doctrine.

According to classical theory, an electron describing any sort of an orbit radiates energy in waves, which are light-waves. The frequencies of the light are those which appear in the Fourier expansion of the orbital motion, as I have already told at length. The polarization of the light is determined by the shape of the orbit. A to-and-fro vibration along a line—the only case which there will be room to treat—is a special kind of orbital motion, which produces plane-polarized waves with the electric vector parallel to the oscillation. If the vibration is perfectly sinusoidal the light also is of a single frequency, perfectly monochromatic. If the motion is not sinusoidal it may be resolved by the Fourier process into sinusoidal components, the frequencies of which then appear in the waves.

Much of the foregoing would follow from any plausible theory of light. The electromagnetic theory goes further, and predicts that the rate of radiation of energy R , from an electron describing sinusoidal vibrations of amplitude x_0 and frequency ν_0 , is proportional to $\nu_0^4 x_0^2$. It goes further yet, and prescribes the constant of proportionality:

$$R = (1/3c^3)(2\pi\nu_0)^4(e x_0)^2. \quad (117)$$

However this last paragraph contains an internal contradiction. If the vibrating electron is radiating energy, its amplitude cannot be constant. Its vibrations are damped; and the Fourier expansion of a damped vibration comprises frequencies extending over a continuous range on both sides of ν_0 , which should be present in the outflowing waves, and reveal themselves as a broadening of the spectrum line; while in Eq. (117), ν_0 ceases to be unique and x_0 becomes a variable. Superficially, the contradiction does not seem serious. Only a very small fraction of the energy of the vibration is radiated in each cycle, so that the amplitude changes only in a very small proportion from one cycle to the next, and Eq. (117) is not hindered from being a very close approximation. Moreover spectrum-lines do have appreciable breadths, some part of which could easily be ascribed to a steady decline in the amplitudes of oscillators. Fundamentally, though, the difficulty is so grave as to be fatal to the classical theory. The case of the linear oscillator is too specialized to furnish a good illustration, but that of an electron moving in the inverse-square central field of a positively-charged nucleus will serve. According to classical theory such an electron should describe a spiral terminating at the nucleus, and in doing so it should emit a "continuous" spectrum comprising every frequency.

This was the great objection to Rutherford's atom model of electrons circulating around a nucleus. Bohr, in order to save the atom model, introduced the remarkable assumption that there are certain orbits in which electrons can revolve indefinitely without losing energy by radiation, and that light is emitted in photons when an electron passes by a sudden and unvisualizable "transition" from one of these permitted orbits to another. Here was the genesis of the doctrine of stationary states, of permitted energy values, and of the emission of spectrum lines having frequencies equal to $1/\hbar$ times the differences between permitted energy values. It would be fascinating to trace the evolution of quantum mechanics from this beginning; but we must proceed directly to, I will not say the final, but the contemporary stage of the theory of radiation.

There is not space to treat a more general case than that of the linear oscillator in which the oscillating particle is an electron; but we can at least generalize Eq. (117) to some extent before departing from the classical theory. The sinusoidal vibration assumed in Eq. (117) is obtained by assuming a potential-energy function *const.* x^2 and ignoring the damping attendant on radiation. Assuming some other potential-energy function (and continue to ignore radiation) we obtain either a non-periodic motion (a case not to be treated here) or a motion which could be expanded by the Fourier process after the fashion of Eq. (63):

$$s(t) = \sum C_s \exp(2\pi i s \nu_0 t). \quad (C_s = C_{-s}^*). \quad (118)$$

According to classical theory, the outflowing waves could be resolved into wave trains having the various frequencies $\nu_0, 2\nu_0, 3\nu_0, \dots$; and the energy R_s flowing away per unit time in waves of frequency $s\nu_0$ would be given by the formula:

$$R_s = (1/3c^3)(2\pi s \nu_0)^4 e^2 C_s C_s^*. \quad (119)$$

Now by the procedure of quantum mechanics, we consider x as an operator \hat{x} and as the matrix which is associated with this operator. The (ij) element of this matrix, I recall, is $a_{ij} \exp(2\pi i \nu_{ij} t)$; the amplitudes a_{ij} are given by Eq. (108), the frequencies ν_{ij} are the quantities $(W_{ii} - W_{jj})/h$.

In quantum mechanics it is assumed that to each of these frequencies corresponds a rate of outflow of energy R_{ij} given by the similar formula:

$$R_{ij} = (1/3c^3)(2\pi \nu_{ij})^4 e^2 a_{ij} a_{ij}^*. \quad (120)$$

When using the wave picture of light of frequency ν_{ij} , we are to imagine a wave train of this frequency carrying energy steadily away from the atom at the rate R_{ij} , given by Eq. (120). When using the corpuscular picture of light, we are to imagine photons of energy $h\nu_{ij}$ emerging at such a rate that the average number per second is $R_{ij}/h\nu_{ij}$. This latter phrasing however is likely to lead us into the same difficulty as we encountered in trying to interpret $\Psi\Psi^*$ for a single atom. It is therefore desirable to say, either that Eq. (120) refers to a very great number of (non-interacting) atoms all together (so that $NR_{ij}/h\nu_{ij}$ will stand for the number of photons emitted from N atoms in unit time), or

else that R_{ij} stands for the probability that a photon shall be emitted by an atom in a given interval of time of unit length. The second of these statements may be interpreted to include the first, and therefore is superior.

Now it is a well-known fact of experience that the relative intensities of the various spectrum lines emitted from a gaseous assemblage of atoms depends on its condition; for instance, on its temperature and on the electrical current which may be traversing it. The factors a_{ij} must therefore be variable; and on investigation, one finds that the eigenfunctions Ψ_i which occur in these factors involve coefficients which can be adjusted to fit various conditions. On the other hand, it must not be supposed that the factor $(2\pi)^4 e^2 / 3c^3$ which appears in Eq. (120) is essentially unverifiable, and could as well be replaced by an arbitrary coefficient. It can be and has been successfully tested by experiments on scattering of high frequency light (x-rays) by atoms. It would however take too long to enter into these matters, and I must content myself with a statement of the consequences of the fact that with particular atom models, many of the matrix elements a_{ij} are equal to zero.

Referring to Eq. (108), one sees that if for any pair of values of i and j the integral $\int \Psi_i^* x \Psi_j dx$ is zero (the integration being taken, as always, from $-\infty$ to $+\infty$) the rate of radiation of energy R_{ij} corresponding to the frequency ν_{ij} is also zero. The corresponding line must then be missing from the spectrum of an atom which is adequately represented by an atom model having these eigenfunctions. Now in actual spectra, one evaluates the energy values W_{ii} and the corresponding "term frequencies" W_{ii}/h , by analysis and classification of the lines which are observed; and then one generally finds that a great number of the differences $(W_{ii} - W_{jj})/h$ correspond to vacant places in the spectrum—the expected lines are absent. If a line $(W_{ii} - W_{jj})/h$ is customarily missing from a spectrum, it is said that the line is "forbidden," and that the states of energy values W_{ii} and W_{jj} "do not combine with one another."

In the special case of the linear harmonic oscillator, with the potential-energy function $V(x) = \frac{1}{2}k^2x^2$, the consequences of the vanishing of certain factors a_{ij} are in full accord with what the

classical theory leads us to wish. The eigenfunctions for this case were given in Eqs. (52) and (55), from which one sees that the functions Ψ_i with odd-integer values of i are odd, while those with even-integer values of i are even. Thus the integral $\int \Psi_i^* x \Psi_j dx$ vanishes when i and j are both even or both odd, and it follows that no energy is radiated with any of the frequencies $2\nu_0, 4\nu_0, 6\nu_0, \dots$. If i is even and j is odd, the conclusion is not so immediate; but it follows from the general properties of the Hermite polynomials²⁷ that the integral vanishes unless i and j differ by one unit, so that no radiation occurs with frequencies $3\nu_0, 5\nu_0, 7\nu_0, \dots$. No frequency but ν_0 itself can appear in the outflowing radiation. This is just the result which is needed in order not to conflict with the classical theory; for in classical mechanics a linear oscillator with the potential-energy function $\frac{1}{2}k^2x^2$ has a sinusoidal motion with only one frequency, and only that one frequency can appear in the outflowing waves. In cases of linear oscillators with other potential-energy functions, the integrals $\int \Psi_i^* x \Psi_j dx$ do not in general vanish, and spectrum-lines of all the frequencies ν_{ij} are indicated. Likewise it is satisfactory that a departure from the particular form $\frac{1}{2}k^2x^2$ of the potential-energy function entails by both theories the appearance of new frequencies—in the one case the overtones $s\nu_0$, in the other the various frequencies ν_{ij} .

The too simple case of the linear oscillator can take us no further, which is a pity, as there are important theorems for which it cannot be invoked as an illustration. Indeed there is no possibility of giving an adequate idea of the scope and the range of quantum mechanics and of its verifications, without at the very least taking the nucleus-and-revolving-electron model of the hydrogen atom, applying to it the relativistic wave equation, making the allowance for the electron spin, and working out the effects of an externally-applied field, electric or magnetic. So great an enterprise transcends by far the scope of this article, and I must confine myself to stating the theorems without attempting to prove them.

The reader may imagine, for convenience,

²⁷ Courant-Hilbert, *Methoden der mathematischen Physik*, p. 76.

some three-dimensional atom model: say the nucleus and revolving electron of hydrogen and ionized helium, or the first-approximation model for sodium, in which last a single electron is supposed to move in an electric field designed to simulate the average resultant of the fields of the nucleus and of the other electrons of the atom. The Hamiltonian operator will (in general) involve three coordinates q_1, q_2, q_3 , which for ease of visualization may be conceived as Cartesian coordinates x, y, z referred to a frame fixed in the atom with origin at the nucleus thereof. The eigenfunctions will in general be functions of all three. Let Ψ_i, Ψ_j denote any pair of eigenfunctions; W_i, W_j the corresponding eigenvalues; $\nu_{ij} = (W_j - W_i)/h$ the corresponding frequency. (Strictly, the subscripts should be triple, but in these closing paragraphs it is not worth while to make the change.)

The verifications of quantum mechanics include the three following types:

(1) *Tests by comparison of the eigenvalues of the Hamiltonian operator with the observed energy values of the stationary states of the atom*, these latter deduced chiefly from spectra but partly also from electron-impact experiments. The agreements are amazingly good in the cases of hydrogen and ionized helium, if there is no external field at all and if there is a strong applied electric field; and there are also excellent agreements in the cases of various kinds of atoms in applied magnetic fields of various strengths.

(2) *Tests by missing lines*. Many pairs of eigenfunctions satisfy the following equations:

$$\int \Psi_i^* x \Psi_j dq = \int \Psi_i^* y \Psi_j dq = \int \Psi_i^* z \Psi_j dq = 0. \quad (121)$$

According to the theory there should be no radiation of the frequency ν_{ij} associated with such a pair. It is in fact almost a universal rule that such lines are missing from spectra. Occasional exceptions can be ascribed to the fact that atoms often pass through one another's fields, so that a Hamiltonian operator devised without regard to such interactions is not continuously correct. In this connection it is important that when an external electric field is applied to a radiating gas, many formerly-missing lines make their appearance; and on

introducing the proper term for this field into the Hamiltonian operator, it is found that the pairs of eigenfunctions corresponding to these lines have ceased to satisfy Eq. (121).

(3) *Tests by polarization.* Many pairs of eigenfunctions satisfy the following equations when the coordinates x, y, z are properly chosen:

$$\int \Psi_j^* x \Psi_i dq \neq 0;$$

$$\int \Psi_j^* y \Psi_i dq = \int \Psi_j^* z \Psi_i dq = 0. \quad (122)$$

According to the theory, the radiation of the frequency ν_{ij} associated with such a pair should be plane polarized, with the electric vector parallel to the x -direction. This is not always verifiable, however; for the coordinate frame (x, y, z) is fixed in the atom, and in a radiating gas the various atoms will be oriented quite at random unless there is some orienting agency acting from without. An externally-applied field, electric or magnetic, is such an agency. When the term describing such a field is incorporated into the Hamiltonian operator of an atom model, and the x -direction is chosen to coincide with the field direction, certain pairs of eigenfunctions satisfy Eq. (122), and the corresponding spectrum-lines are found experimentally to be plane polarized with electric vector parallel to the field. With electric fields there are also lines plane polarized at right angles to the x -direction, and with magnetic fields there are circularly polarized lines in the light radiated along the x -direction; all of which facts may be deduced from the values of the integrals appearing in Eqs. (121) and (122), when the corresponding eigenfunctions are inserted.

Quantum mechanics extends far beyond these limited cases of which I have spoken, and its verifications are more numerous and diversified by far than those which I have mentioned. The calculation of spectrum frequencies is not its

only function, and the observation and measurement of spectrum lines are not its only tests. Quantum mechanics is involved in the deflection and scattering of free rapidly-moving electrons and positive ions in gases; in the deflection of free rapidly-moving neutral atoms in non-uniform magnetic fields and in non-uniform electric fields; in the transfer of energy between electrons and atoms; in the deflection and scattering of light by atoms and electrons; in the polarization of gaseous, liquid and solid aggregations of atoms by electric and magnetic fields, which is evinced by measurements of dielectric constants and of susceptibility; in chemical reactions. Quantum mechanics is also the basis of the contemporary statistical theories of radiation, electricity and matter, from which are explained such diverse phenomena of nature as the distribution-law of black-body radiation, the specific heats of solids and of gases, and the laws of the conduction of electricity through metals and the escape of electrons from metals.²⁸ It would perhaps be shorter to list the divisions of physics, into which the quantum theory has as yet made no inroads and scored no successes! It seems audacious to have attempted to make even the most elementary of introductions to so vast a subject, in so small a space as forty-six pages. But every division of science is now so vast that if such attempts were not to be made, every book would be of encyclopedic length and every reader would be confined to his own specialty. I have therefore no hesitation in admitting that this article touches on only a small part of quantum mechanics; that in the untouched parts there are many important principles, many striking theorems, many brilliant successes of theory in the interpretation of experiments; and also that these successes fall far short of being universal.

²⁸ See for instance my earlier article, *Rev. Mod. Phys.* 1, 90 (1929); or Dr. Rabinowitsch's translation thereof, *Elementaire Einführung in die physikalische Statistik*, Hirzel, 1932.