Volume 2 *January, 1930* *Number 1*

REVIEWS
OF
MODERN PHYSICS

THE GENERAL PRINCIPLES OF QUANTUM MECHANICS. PART II

BY E. C. KEMBLE AND E. L. HILL*
HARVARD UNIVERSITY

SECTION 5. THE INTERACTION OF MATTER AND RADIATION

1. GENERAL INTRODUCTION

AS THE primary source of experimental evidence regarding the structure of the atom is the spectroscope, much of the mathematical structure developed in the preceding sections must hang in the air until we have constructed a quantum theory of the emission, absorption, and scattering of radiation by atoms. The desired theory should have a formal similarity to the elementary wave mechanics of Part I and should degenerate in the limit when h is made negligibly small into the classical electromagnetic theory of the emission and absorption of light. It must take into account the forces acting on the electrons due to the external radiation field, the modification of the radiation field due to the motion of the system, and finally, the damping influence of "radiation resistance." To be completely satisfactory the theory should also take into account the effect of the finite velocity of electromagnetic waves on the interaction of the different electrons in the atomic system and should satisfy the demands of the theory of relativity. The newly formulated Heisenberg-Pauli field theory goes a long way toward the solution of this great problem[1] but it would be quite impossible to introduce the reader to the mysteries of this theory with the mathematical tools developed in Part I of this report. Progress toward carrying out the above program has been made by successive approximations and we here content ourselves with reporting the initial steps in the investigation.

In the usual classical discussion of absorption, dispersion, and scattering, the forces acting on the electrons are somewhat arbitrarily divided into three classes. First, there are the forces internal to the atom, whose operation determines its unperturbed state. According to our present ideas of

---

* National Research Fellow.

[1] W. Heisenberg and W. Pauli, Zeits. f. Physik **56**, 1 (1929).

atomic structure these forces are to be identified with the electrostatic forces which act between electron and nucleus and between electron and electron. Second, there are the forces due to the external radiation field, and third, the retarding force on each electron due to its own radiation.[2] The elementary quantum theory of Part I takes cognizance only of the first class of forces. It can be extended without difficulty to include forces of the second class, but the forces of the third class, difficult to deal with even classically, are particularly intractable from the standpoint of wave mechanics. Radiation resistance forces are small, however, in the case of an atom near to its normal state and by neglecting them we may still hope to obtain a description of the early stages of the absorption process. The problem of dispersion can also be treated without considering radiation resistance if means are at hand for calculating the electromagnetic field due to a given electronic motion.

As a first step toward the formulation of an approximate quantum theory of absorption and dispersion a wave equation analogous to Eqs. (I,18) and (I, 38) must be set up to describe the behavior of the electrons in an arbitrary external electromagnetic field.

In view of the success of the strictly corpuscular theory of the Compton effect, it might seem that any attempt to account for the interaction of matter and radiation in terms of a continuous electromagnetic field is foredoomed to failure. We have already seen, however, that the field theory of diffraction can be interpreted as a statistical description of the scattering of photons and electrons by apertures. It is such a statistical description of the interaction of matter and radiation that we seek. Moreover, the continuous potential energy function V of the Schrödinger Eqs. (I, 18) and (I, 38) is in practice (Cf. Kepler problem, Section 2, 6) of electrostatic origin and if we admit such a continuous field to our theory we can hardly stop short of admitting the whole machinery of Maxwell. The Heisenberg-Pauli field theory provides a scheme for the modification of this theory, but we shall provisionally content ourselves with an attempt to graft the electromagnetic field of Maxwell unchanged to the matter waves of Schrödinger.

2. The Wave Equation for Systems of Electrons in an External Radiation Field

The principle of least action in the Newtonian dynamics is restricted to cases in which the energy is conserved.[3] Hence the simple method of setting up the wave equation sketched in Part I, Section 1 cannot be used for a system of electrons in a variable external field. Schrödinger himself, however, based his "derivation" of the wave equation (I, 18) on the Hamilton-Jacobi partial differential equation which is not subject to the above restriction.[4] His point of view permits us to set up and justify a generalization of Eqs. (I, 18) and (I, 38) appropriate to the problem in hand.

We have already shown that in the case of a constant electrostatic field the wave equation [(18) or (38), Part I] is obtained from the classical

---

[2] With these we may include the damping effects of collisions which play such an important part in the production of absorption in gases.

[3] That is, cases in which there is a Hamiltonian function $H(q, p)$ which does not contain the time explicitly.

[4] E. Schrödinger, Ann. d. Physik (4) **79**, 361 and 489 (1926).

Hamiltonian function $H(q, p)$ if we first form the operator $H(q, -\partial/\partial q)$ by the substitution $p_k \rightarrow -(h/2\pi i)\, \partial/\partial q_k$ and then use it to set up the relation

$$H\left(q_k, -\frac{\partial}{\partial q_k}\right)\psi = \frac{h}{2\pi i}\frac{\partial \psi}{\partial t}. \tag{1}$$

The parallelism between this second order first degree equation for ψ and the first order second degree Hamilton-Jacobi equation for the action function

$$H\left(q_k, \frac{\partial W}{\partial q_k}\right) = -\frac{\partial W}{\partial t} \tag{2}$$

W is obvious. A complete solution of the latter equation is a function of the coordinates, say $q_1, q_2, \cdots q_{3n}$, of the time, and of $3\,n$ constants of integration $\alpha_1, \cdots, \alpha_{3n}$, which represents a moving family of surfaces in configuration space $(q_1, \cdots, q_{3n})$ related to the possible mechanical orbits of the classical dynamics as the wave-fronts of geometrical optics are related to light rays. As wave fronts are surfaces of constant phase it is therefore not suprising to find that the classical action function is simply related to the phase of the ψ waves. If we change the independent variable in Eq. (I, 38) by means of the substitution

$$\psi = e^{-2\pi i W'/h} \tag{3}$$

we obtain the differential equation,

$$\sum_k \frac{1}{2\mu_k}\left(\frac{\partial W'}{\partial x_k}\right)^2 + V + \frac{\partial W'}{\partial t} = \frac{h}{4\pi i}\sum_k \frac{1}{\mu_k}\frac{\partial^2 W'}{\partial x_k^2}. \tag{4}$$

If the second derivatives on the right hand side of Eq. (4) are neglected, it becomes identical with the Hamilton-Jacobi equation (2) for the problem in hand. This approximation is equivalent to setting Planck's constant h equal to zero, and is valid in the region for which the classical mechanics is valid. Thus the classical action function W is an approximation for $-(h/2\pi i)\log\psi$. This fact may be used as the basis for a new proof of the basic proposition that suitable wave packet solutions of Eq. (I, 38) obey the laws of classical mechanics and this gives a justification of Eq. (I, 38) independent of that given in Section 1, Part I.[4a]

Turning now to the problem of the variable external field we see that a rational method of setting up the new wave equation is to construct the corresponding Hamiltonian function in Cartesian coordinates, convert it into an operator as before and finally use the operator as in Eq. (1) above.

Let $A_x^{(j)}$, $A_y^{(j)}$, $A_z^{(j)}$, denote the components of the vector potential evaluated at the point x_j, y_j, z_j where the jth electron is located. Let the algebraic value of the electronic charge be e. Then the classical Hamiltonian for a system of r electrons is[5]

[4a] For a direct and comprehensive proof of this proposition see also Sect. 6, 5 and Sect. 7, 1.

[5] Cf. M. Born, "Atommechanik" Sec. 34, p. 239. Many authors, including Born, use e for the absolute value of the electronic charge and so reverse the sign of the terms linear in e and A both in the Hamiltonian and the wave equation.

$$H(q,p)=\sum_{j=1}^{\tau}\frac{1}{2\mu}\left\{\left(p_{x_j}-\frac{e}{c}A_x^{(j)}\right)^2+\left(p_{y_j}-\frac{e}{c}A_y^{(j)}\right)^2+\left(p_{z_j}-\frac{e}{c}A_z^{(j)}\right)^2\right\}+V(x_1,\cdots z_\tau). \tag{5}$$

Making the substitution $p_k\rightarrow-(h/2\pi i)\partial/\partial q_k$ we obtain as the desired wave equation

$$\sum_{j=1}^{\tau}\left\{\left(\frac{\partial}{\partial x_j}+\frac{2\pi e i}{hc}A_x^{(j)}\right)^2+\left(\frac{\partial}{\partial y_j}+\frac{2\pi e i}{hc}A_y^{(j)}\right)^2+\left(\frac{\partial}{\partial z_j}+\frac{2\pi e i}{hc}A_z^{(j)}\right)^2\right\}\psi-\frac{8\pi^2\mu}{h^2}V\psi-\frac{4\pi\mu i}{h}\frac{\partial\psi}{\partial t}=0. \tag{6}$$

If we make the substitution $\psi=e^{-2\pi i W'/h}$, multiply by h^2, and allow h to approach zero, Eq. (6) yields the Hamilton-Jacobi equation for W' as in the case of a steady electrostatic field. From this fact we infer that Eq. (6) is in satisfactory agreement with classical mechanics in the limit when h is zero.

Expansion of the squared operators throws the differential equation into the form

$$\sum_{j=1}^{\tau}\left[\nabla_j^2\psi+\frac{4\pi e i}{hc}\left(A_x^{(j)}\frac{\partial\psi}{\partial x_j}+A_y^{(j)}\frac{\partial\psi}{\partial y_j}+A_z^{(j)}\frac{\partial\psi}{\partial z_j}\right)+\frac{2\pi e i}{hc}\left(\frac{\partial A_x^{(j)}}{\partial x_j}+\frac{\partial A_y^{(j)}}{\partial y_j}+\frac{\partial A_z^{(j)}}{\partial z_j}\right)\psi\right]-\frac{8\pi^2\mu}{h^2}V'\psi-\frac{4\pi\mu i}{h}\frac{\partial\psi}{\partial t}=0 \tag{7}$$

where[6] $V'=V+(e^2/2\mu c^2)\sum_{j=1}^{\tau}|A^{(j)}|^2$.

Referring to Eq. (I, 102) and to the discussion of Eq. (I, 149) on p. 207 Part I, we observe that to retain our fundamental physical interpretation of $\psi\,\psi^*dv$ it is necessary that

$$\mu\frac{\partial}{\partial t}(\psi\psi^*)=-\operatorname{div}F, \tag{8}$$

where F is a function which vanishes at infinity with ψ and which plays the rôle of mass current density. Using the modified wave equation (7) for a single particle (three dimensional case) we find that Eq. (8) is satisfied with[7]

$$F=\frac{h}{4\pi i}[\psi\text{ grad }\psi^*-\psi^*\text{ grad }\psi]-\frac{e}{c}A\psi\psi^*. \tag{9}$$

---

[6] The correction term $(e^2/2\mu c^2)\sum_{j=1}^{\tau}|A^{(j)}|^2$ is usually negligible.

[7] In Eq. (I, 102) which defines the vector current density, the sign of the right hand member should be reversed.

This is in agreement with the fact that classically the momentum p of a particle in a field with vector potential A is not μv, but $\mu v + eA/c$. Hence the operator associated with μv should be

$$-\frac{h}{2\pi i}\operatorname{grad}-\frac{e}{c}A.$$

From Eqs. (8) and (I, 78) we may deduce the validity of Eq. (I, 171) for the modified operator $H(q, -\partial/\partial q)$ and hence the Hermitian character of the modified matrix $H(n, m)$. Moreover, the whole discussion carries over without difficulty to the case of a system of τ particles.[8]

3. The Perturbation of an Atomic System by a Radiation Field: Absorption

Consider next the application of the wave equation (7) to the problem of the perturbation of an atomic system by an external radiation field.[9] The general solution ψ of (7) can be described in terms of the unperturbed characteristic functions by the aid of an expansion of the form

$$\psi = \sum_{n,m} c_{nm} u_{nm}(x) e^{2\pi i E_{nm} t/h} + \int \sum_{m} c_m(E) u_m(x, E) e^{2\pi i E t/h} dE. \tag{10}$$

(Here wave functions carry a double subscript to indicate degeneracy. In practice a more complicated notation may be necessary.) The coefficients c_{nm} and $c_m(\mathrm{E})$ become functions of the time during the period of application of the radiation field but settle down to a new set of permanent values different from the initial set when the field is removed. Our first problem is to correlate this change in the c's with the experimental absorption of light and to use the theory to derive theoretical values of Einstein's "transition probabilities." Due to the variation of the c's with time, we shall also find that the wave function executes oscillations which contain the frequencies characteristic of the radiation field, the natural frequencies of the atom, and combinations of the two. Our second problem is to use these three types of frequency to account for dispersion, normal scattering, emission, and Raman scattering.

As previously explained (Sect. 4, 3), Born interpets the ψ functions as descriptions of the behaviour of ensembles of atomic systems subjected to identical external influences. He assumes that the energy of an arbitrary system taken from the ensemble must in every case have an unique value identical with one of the familiar energy levels which play such an important part in the Bohr theory. Finally in the case of a non-degenerate system with a

---

[8] A wave equation for a single electron in a variable electromagnetic field was first derived by Gordon, Zeits. f. Physik **40**, 117 (1926), and independently by Klein, Zeits. f. Physik **41**, 407 (1927) in connection with their relativistic generalizations of the Schrödinger theory. Although the Gordon-Klein theory does not satisfy the requirements of Eq. (8) and hence involves a new definition of number density to take the place of $\psi\psi^*$, Eq. (6) in three dimensions can be derived from it by a suitable limiting process.

[9] The reader is referred to the article by K. L. Wolf and K. F. Herzfeld, "Absorption and Dispersion," Handbuch der Physik, XX, J. Springer, Berlin, 1928, for a review of this subject from both the experimental and theoretical standpoints.

discrete spectrum only, he identifies the probability of the energy E_n with the corresponding product $c_nc_n^*$.[10] This interpretation is based on the theorem that in the absence of a continuous spectrum

$$\sum_n c_nc_n^* = 1.$$

(Cf. Part I, Eqs. (81, 137, 172)). The generalization required to take into account a continuous spectrum will be evident from Eq. (I, 81).

These assumptions are in comformity with the discussion of Part I, Sect. 1,8 and seem to the authors quite reasonable.

A complete analysis of Born's assumptions cannot be given here, but the following suggestions may be helpful to the reader:

(a) Born supposes that an exact measurement of E for a single system is conceivable and would give a definite result. In support of this assumption we observe that photoelectric measurements with different wave-lengths indicate that we may determine the energies of photons by spectrum analysis, attributing to each photon a definite energy corresponding to the place at which it impinges on the photographic plate in an ideal spectroscopic experiment. In the same way a magnetic deflection experiment yields an energy spectrum for free electrons with a definite energy for each particle.

(b) In Part I, Sect. 1, 8 we have shown that if we perform a Fourier analysis of the wave function for a free particle into terms harmonic in the independent variables x, y, z the amplitudes give the probabilities of the conjugate dynamical variables p_x, p_y, p_z. Furthermore, from the wave standpoint a spectrometer is a mechanical device for analyzing a beam of radiation into a sum of terms which are harmonic in the time. The probability of each elementary energy range for an arbitrary photon is then fixed by the intensity of the corresponding component in the Fourier analysis with respect to time. Hence in the case of a system whose Hamiltonian function does not contain the time explicitly and in which an expansion of the type

$$\psi = \sum_n c_nu_n(x)\, e^{2\pi iE_nt/h} \tag{a}$$

is possible, a plausible extrapolation from the above special cases suggests that we identify $c_nc_n^*u_n(x)u_n^*(x)d\tau$ with the probability that the system has the energy E_n and also a configuration in the element $d\tau$ of configuration space. If the functions u_n are normalized, integration over all of configuration space gives $c_nc_n^*$ as the probability of the energy E_n independent of the configuration. This is perhaps sufficient justification for Born's assumption in the case under consideration.

(c) On the other hand, in the case of a system whose Hamiltonian does contain the time explicitly so that an expansion of the type (a) is valid only if the c_n's vary with the time, it is evident that Eq. (a) does not describe an exact Fourier analysis of the wave function. Moreover, energy and time are conjugate dynamical variables subject to the Heisenberg uncertainty principle, $\delta E\delta t \sim h$, and so the measurement of either the energy of a system or its energy distribution function must take an infinitely long time. Thus a rigorous non-arbitrary meaning cannot be attached to the phrase "instantaneous probability of the energy E_n." In practice, however, we do not expose photographic plates for an infinitely long time in spectroscopic measurements, and in the case of a fluctuating light source we understand by the frequency distribution at a given time the distribution obtained on a photographic plate exposed for a short time before and after the instant in question. This distribution depends on the duration of exposure as well as the time at which it is started, but is sufficiently definite for practical purposes. This practical procedure of exposing a plate for a finite time, say $t_0 \leqq t \leqq t_1$ is equivalent to the mathematical process of replacing a given function $f(t)$ by a second function $\phi(t)$ defined by

---

[10] This statement assumes that the expansion is made in terms of normalized wave functions.

$$\phi(t) = f(t),\ \text{if}\ t_0 \leqq t \leqq t_1;\ \phi(t) = 0,\ \text{if}\ \begin{matrix} t < t_0 \\ t > t_1, \end{matrix}$$

and performing a Fourier analysis on $\phi(t)$. For every pair of values of t_1-t_0 and $\bar{t} = (t_1+t_0)/2$ we get a definite distribution function which has a certain claim toward the title "distribution function for $t=\bar{t}$." Applying this idea to the case of an expansion of the type (a) where the c's vary slowly with time, we see that the instantaneous energy distribution function will necessarily be continuous but will have sharp maxima and minima near the energies E_n of the unperturbed system and will, in fact, be an approximation to the discontinuous spectrum postulated by Born's hypothesis. It can also be proved that if the c's are suddenly stabilized at any given instant by removing the perturbations which cause the c's to vary and if we determine by Fourier analysis the energy distribution function for all future time, we shall get exactly Born's result. In other words, Born's identification of $c_nc_n^*$ at the instant t with the instantaneous probability of the energy E_n is mathematically sound if we define the latter quantity as the probability of E_n, for all future time, *which would be created by removal of the perturbation at the instant in question.*

(d) It should be emphasized that the adoption of Born's statistical interpretation means that the wave functions have meaning only in connection with ensembles of atoms. It seems clear that if a wave function did exist for an individual atom, it would still be impossible to measure it, although in the case of an ensemble of similar atoms the observation of $\psi\psi^*$ is at least conceivable. If we adopt Bridgman's operational point of view we must then refuse to attribute objective reality to the wave functions except in connection with ensembles of identical systems. This view becomes even more inescapable when one considers the relation between the observed orbits of large scale bodies or of electrons and the associated wave packets. The wave packets must in all cases increase steadily in volume with time just as a blob of radiation emerging from a small aperture increases in volume with time. If, however, we observe that at any instant a particle is in the neighborhood of some particular point this observation modifies our expectation for the future and makes it necessary to discard our original wave-packet and substitute, as a basis for prediction, a new one based on our observation. This substitution of a new wave function for the old makes sense only if we adhere strictly to the view that wave functions are simply tools for the description of the statistical behavior of ensembles of identical mechanical systems.

The reader is referred to a recent paper of Lewis and Mayer, Proc. Nat. Acad. Sci. **15**, 127 (1929) for other very interesting considerations in this line.

On this basis the change in the value of $c_nc_n^*$ produced by any given perturbation measures the fraction of the whole number of systems in the ensemble which pass into the corresponding energy level as a result of the perturbation. We proceed to outline a scheme for computing this change in the case of a radiation field.

For simplicity let us assume that the radiation consists of a system of plane waves moving in the direction of the x-axis and polarized with the electric vector parallel to the y-axis. In conformity with the perturbation theory of Section 4, 3 we assume a finite wave train described for convenience by a complex Fourier integral. Let the vector potential and electric force be

$$\begin{aligned} A_y &= -\alpha f(t-x/c) = +\alpha c\int_{-\infty}^{+\infty}\frac{\phi_y(\nu)}{2\pi i\nu}e^{2\pi i\nu(t-x/c)}d\nu, \\ E_y &= -\frac{1}{c}\,\frac{\partial A_y}{\partial t} = -\alpha\int_{-\infty}^{+\infty}\phi_y(\nu)e^{2\pi i\nu(t-x/c)}d\nu. \end{aligned} \tag{11}$$

Here α is a real parameter which determines the amplitude of the wave system, while the complex function $\phi_y(\nu)$ fixes the wave form and may be subjected to the condition

$$\int_{-\infty}^{+\infty} \phi_y \phi_y^* d\nu = 1. \tag{12}$$

In the case of a radiation field the divergence of the vector potential vanishes. Using this fact and multiplying Eq. (7) by $-h^2/8\pi^2\mu$ we reduce it to the form (neglecting the term $(e^2/2\mu c^2)\sum_j |A^{(j)}|^2$),

$$\left[H_0 + \frac{h}{2\pi i}\,\frac{e}{\mu c}\sum_{j=1}^{\tau}\left(A_x^{(j)}\frac{\partial}{\partial x_j} + A_y^{(j)}\frac{\partial}{\partial y_j} + A_z^{(j)}\frac{\partial}{\partial z_j}\right)\right]\psi = \frac{h}{2\pi i}\,\frac{\partial \psi}{\partial t}, \tag{13}$$

where H_0 is the unperturbed Hamiltonian operator. Introducing the first of Eqs. (11) and identifying the parameter λ of Section 4, 3 with α we obtain the following expression for the perturbing operator $F(t)$:

$$F = -\frac{h}{2\pi i}\,\frac{e}{\mu c}\left[\sum_{j=1}^{\tau} f(t - x_j/c)\frac{\partial}{\partial y_j}\right]. \tag{14}$$

The first step in carrying through the perturbation scheme of Section 4, 3 is to evaluate the matrix elements

$$F_{mn}(t) = \int u_n^* F u_m dv.$$

If the wave-lengths of the radiation field are all large compared with the atomic dimensions we may regard A as constant over the atom and may replace $f(t-x_j/c)$ by the value of $f(t-x/c)$ at the nucleus of the atom in computing $F_{mn}(t)$. If the origin is at the nucleus

$$F_{mn}(t) = \frac{e}{\mu c} f(t) \int_{\infty} u_n^* \sum_{j=1}^{\tau}\left(-\frac{h}{2\pi i}\,\frac{\partial}{\partial y_j}\right) u_m dv. \tag{15}$$

$\sum_j(-h/2\pi i)\partial/\partial y_j$ will be recognized as the operator associated with the y component of the linear momentum (Cf. Section 3, 2 pp. 192-3). The integral in the right hand member is therefore the matrix element of $p_y(m,n)$ of the matrix defined by the expansion

$$p_y u_m = \sum_n p_y(m,n) u_n. \tag{16}$$

$p_y(m,n)$ in turn is the amplitude of the corresponding element $p_y(m,n/t)$ of the matrix function of t associated with p_y in the theory of Heisenberg, Born, and Jordan. The latter matrix is defined by an expansion similar to (16) with u_m and u_n replaced by ψ_m and ψ_n respectively. Thus

$$p_y(m,n)e^{(2\pi i/h)(E_m-E_n)t} = p_y(m,n/t) = \int_{\infty} \psi_n^* \sum_{j=1}^{\tau}\left(-\frac{h}{2\pi i}\,\frac{\partial}{\partial y_j}\right)\psi_m dv. \tag{17}$$

Classically the linear momentum is μ/e times the time derivative of the electric moment and a similar theorem holds for the corresponding matrices. Let P_y denote the y component of the electric moment, i.e., $\sum_j e_j y_j$ and let $P_y(m,n/t)$ and $P_y(m,n)$ denote elements of the matrices of P_y based on the functions ψ and u respectively. Then

$$(e/\mu)p_y(m,n/t)=\dot{P}_y(m,n/t)=(2\pi i/h)(E_m-E_n)P_y(m,n/t) \qquad (18a)$$

and consequently

$$(e/\mu)p_y(m,n)=(2\pi i/h)(E_m-E_n)P_y(m,n). \qquad (18b)$$

With the aid of Eq. (18b) the expression for $F_{mn}(t)$ can be thrown into the form

$$F_{mn}(t)=\frac{2\pi i}{hc}(E_m-E_n)P_y(m,n)f(t). \qquad (19)$$

The first order perturbations of the wave functions are now determined from Eq. (I, 179) which, with p set equal to unity, should read[11]

$$\overset{(1)}{f_{mn}}(T)=(2\pi i/h)\int_0^T F_{mn}(t')e^{2\pi i\nu_{mn}t'}dt' \qquad (20)$$

where ν_{mn} is defined by

$$h\nu_{mn}=E_m-E_n=-h\nu_{nm}.$$

By combining Eqs. (19) and (20) and introducing the complex Fourier coefficient ϕ_y of Eq. (11) we obtain[12]

$$\overset{(1)}{f_{mn}}(T)=-\left(\frac{2\pi i}{h}\right)P_y(mn)\phi_y(\nu_{nm}), \qquad (21)$$

and finally in first approximation

$$c_n=c_n{}^0-\alpha\sum_m c_m{}^0\left(\frac{2\pi i}{h}\right)P_y(mn)\phi_y(\nu_{nm})+\cdots. \qquad (22)$$

---

[11] Unfortunately the factor $e^{2\pi i\nu_{kn}t'}$ was accidentally omitted in preparing the manuscript of Part I.

[12] To elucidate this result we find first by substituting from Eqs. (11) and (19) for $f(t)$ and $F_{mn}(t)$ respectively into Eq. (20),

$$\overset{(1)}{f_{mn}}(T)=-\left(\frac{4\pi^2}{hc}\right)\nu_{mn}P_y(mn)\int_0^T f(t')e^{2\pi i\nu_{mn}t'}dt'.$$

But from the definition of the Fourier Integral and Eq. (11),

$$\frac{c\phi_y(\nu_{nm})}{2\pi i\nu_{nm}}=\int_{-\infty}^{+\infty} f(t')e^{2\pi i\nu_{mn}t'}dt'.$$

The replacement of the integral with respect to t' over the region 0 to T by an integral from $-\infty$ to $+\infty$ is valid because we have assumed explicitly $f(t')\equiv 0$ unless $0\leqq t'\leqq T$.

The same result is obtained somewhat more directly, but less rigorously, if one treats the product E_yP_y as a perturbing potential function depending explicitly on the time.

In case the radition field is not plane polarized, Eq. (22) is to be replaced by

$$c_n=c_n{}^0-\alpha\sum_m c_m{}^0\left(\frac{2\pi i}{h}\right)\left(P(mn)\phi(\nu_{nm})\right)+\cdots \tag{23}$$

where $P(mn)$ and $\phi(\nu_{nm})$ are vectors derived from the vector electric moment and vector electric force respectively.[13]

It is convenient at this stage of the argument to introduce the quantity Φ_{mn} defined by the equation

$$\Phi_{mn}=f_{mn}(T)f_{mn}{}^*(T). \tag{24}$$

In case all the atoms are initially in the same state m Eq. (I, 180) shows that Φ_{mn} measures the fraction of the atoms transferred from the state m to the state n by the perturbing field. Hence we may call Φ_{mn} a transition probability for jumps from the state m to the state n. Some analysis is required, however, before we can make the connection between Φ_{mn} and the corresponding transition probability as defined by Einstein. The derivation of the blackbody radiation formula by the latter author is based on the assumption that the rate of increase of the number of atoms, N_n, in the state n due to absorption in a natural radiation field is given by an expression of the form

$$\frac{dN_n}{d^t}=\sum_m(N_mB_{m\to n}-N_nB_{n\to m})u(\nu_{nm}) \tag{25}$$

where $u(\nu_{nm})$ is the monochromatic energy density of the field, and the B's are the probabilities of jumps from the state m to the state n and from n to m respectively, per atom per unit time when $u(\nu_{nm})$ is unity. It is evident that if the present theory is to harmonize with Einstein's it must be possible to derive Eq. (25) from Eq. (23). Since Einstein's theory ignores fluctuations from thermodynamic equilibrium we can identify N_n in Eq. (25) with the product of the number of atoms per unit volume, N, by the mean value of $c_nc_n{}^*$ obtained by averaging over all the atoms and also over an interval of time long enough to eliminate all high frequency terms.[14] $(1/N)dN_n/dt$ then becomes identical with the average "secular" rate of increase of $\overline{c_nc_n{}^*}$, and we have to prove that

$$\overline{\frac{d}{dt}\overline{c_nc_n{}^*}}=\frac{d}{dt}\overline{\overline{c_nc_n{}^*}}=\sum_m\left[\overline{\overline{c_mc_m{}^*}}B_{m\to n}-\overline{\overline{c_nc_n{}^*}}B_{n\to m}\right]u(\nu_{nm}). \tag{26}$$

---

[13] The large round brackets indicate a scalar product.

[14] We here treat each individual atom in any small volume element of a gas by the wave mechanics, neglecting all interactions. Each individual atom in the element then has its own ψ function correlated with an imaginary ensemble of similar independent atoms. Although the different atoms in the volume element are physically indistinguishable, a confusion of the group actually present in the element with the ensemble which defines the ψ function of any one of them is undesirable. Cf. the discussion on p. 15.

It is also permissible to restrict the proof to the case where the perturbing radiation field is "natural." This restriction is interpreted to mean that the field consists of many superposed short wave trains whose phases are distributed at random.

By Eq. (I, 180) there are terms of the form $c_l^0 c_n^{0*} f_{ln}(T) f_{ln}^*(T)$ in the expression for $c_n(T)c_n^*(T)$, but these average to zero for the special conditions under consideration and it is possible to prove[15] that Eq. (26) holds in first approximation with

$$B_{m\to n}u(\nu_{nm})=\frac{\Phi_{mn}}{T}=\frac{8\pi^3}{3h^2}\,|\,P(mn)\,|\,^2u(\nu_{nm}),\tag{27}$$

or

$$B_{m\to n}=B_{n\to m}=\frac{8\pi^3}{3h^2}\,|\,P(mn)\,|\,^2.\tag{28}$$

In virtue of Eq. (18) the above theoretical expression for the Einstein transition probability $B_{m\to n}$ can be expressed in the alternative form

$$B_{m\to n}=\frac{2\pi e^2}{3\mu^2}\,\frac{|\,p(m,n)\,|\,^2}{(E_m-E_n)^2}\tag{29}$$

provided that E_n is not equal to E_m.

The theory is applicable to degenerate systems if we interpret $B_{m\to n}$ as the transition probability from one elementary state to another and not from one energy level to another. For such systems it is desirable to introduce a more complete notation in which each of the symbols m and n is replaced by a pair of symbols one of which designates the energy level while the other indicates the particular sub-state under consideration. $B_{m\to n}$ then goes over into $B(m, l\rightarrow n, k)$ with m and n designating the energy levels. Eq. (25) becomes

$$\frac{d}{dt}N_{nk}=\sum_{m,l}[N_{ml}B(m,l\rightarrow n,k)-N_{nk}B(n,k\rightarrow m,l)]u(\nu_{nm}).\tag{30}$$

Summing both sides over all values of k (i.e., over all sub-states of the nth energy level) we obtain

$$\frac{dN_n}{dt}\equiv\frac{d}{dt}\left(\sum_k N_{nk}\right)=\sum_m[N_m\overline{B}_{m\to n}-N_n\overline{B}_{n\to m}]u(\nu_{nm})\tag{31}$$

with

$$\left.\begin{aligned}\overline{B}_{m\to n}&=\sum_{l,k}\frac{N_{ml}}{N_m}B(m,l\rightarrow n,k)=\frac{1}{g_m}\sum_{l,k}B(m,l\rightarrow n,k)\\ \overline{B}_{n\to m}&=\sum_{l,k}\frac{N_{nk}}{N_n}B(n,k\rightarrow m,l)=\frac{1}{g_n}\sum_{l,k}B(n,k\rightarrow m,l).\end{aligned}\right\}\tag{32}$$

---

[15] M. Born, Zeits. f. Physik **40**, 167 (1926); J. C. Slater, Proc. Nat. Acad. Sci. **13**, 7 (1927).

Here g_m denotes the statistical weight of the energy level E_m, and is equal to the number of degenerate levels associated with symbol m, thus being equal to N_m/N_{ml}. Eq. (31) agrees with the Einstein form for degenerate systems, and Eqs. (32) are in conformity with the familiar relation

$$\overline{B}_{n\rightarrow m}=(g_m/g_n)\overline{B}_{m\rightarrow n}.$$

Eq. (19) shows that a radiation field has no tendency to produce transitions from one member of a degenerate group of states to another member of the same group.

Eq. (27) shows that to a first approximation the intensity of any absorption line is proportional to the square of the absolute value of the corresponding matrix element of the dipole moment of the atomic system. To get the absorption (and hence emisson) due to the quadrupole moment which is important in connection with the appearance of "forbidden" lines in nebular spectra[16] one must take into account the variation in the vector potential A over the volume of the atom, which was neglected in deriving Eq. (15).

Evaluation of the matrix elements $P(mn)$ or $p(mn)$ yields the familiar selection principles for the magnetic and azimuthal quantum numbers of atomic systems.[17] In the case of the two-body problem, for example, using the wave functions of Sect. 2, 6 we obtain

$$P_z(n',l',m';n'',l'',m'')=e\left[\int_0^\infty R_{n',l'}R_{n''l''}r^3dr\right]\left[\int_0^\pi \Theta_{l',m'}\Theta_{l''m''}\sin\theta\cos\theta d\theta\right]$$
$$\times\left[\int_0^{2\pi}\Phi_{m'}\overset{*}{\Phi}_{m''}d\phi\right]=e\left[\int_0^\infty R_{n'l'}R_{n''l''}r^3dr\right]\delta_{l',l''\pm 1}\,\delta_{m',m''}, \tag{33}$$

with similar expressions for the matrix components of P_y and P_x.

The theory can be extended to include the continuous part of the absorption spectrum as well as the discontinuous part. To this end the perturbation theory of Section 4, 3 must be completed by the inclusion of the continuous characteristic functions in the expansion of the perturbed function [Eq. (68)]. For details the reader is referred to the original papers of Wentzel and others.[18]

4. EMISSION, DISPERSION, AND SCATTERING

Einstein's derivation of the black body radiation formula is based on the assumption that the mean rate of emission per unit volume of the frequency ν_{mn} associated with the quantum jump $m\rightarrow n$ in a gas in thermal equilibrium is given by an expression of the form

$$S(\nu_{mn})=N_mA_{m\rightarrow n}h\nu_{mn} \tag{34}$$

where N_m is the average number of atoms or molecules per unit volume in the state m, and $A_{m\rightarrow n}$ is a constant dependent only on the nature of the gas.

---

[16] Cf. James H. Bartlett, Phys. Rev. **34**, 1247 (1929).

[17] For details see L. Brillouin, Journ. de Physique **8**, 74 (1927).

[18] G. Wentzel, Zeits. f. Physik **40**, 574 (1926). J. R. Oppenheimer, Zeits. f. Physik **41**, 268 (1927); Phys. Rev. **31**, 66 (1928). P. A. M. Dirac, Proc. Roy. Soc. **A113**, 621 (1927).

The coefficients $A_{m\to n}$ are called "transition probabilities for spontaneous emission" and are related to the corresponding absorption probabilities by the equation

$$A_{m\to n}=(8\pi h\nu_{mn}{}^3/c^3)B_{m\to n}. \tag{35}$$

Combining Einstein's well-grounded and widely accepted theory with the wave-mechanics formula for $B_{m\to n}$ developed in the preceding article we arrive at an indirect method for computing $S(\nu_{mn})$, and hence the intensities of emission lines in gaseous spectra, applicable to all cases in which Eq. (34) is valid and values of N_m are obtainable. In the case of purely thermal radiation where the departure from thermodynamic equilibrium is small, the "population numbers" N_m can be estimated from the Maxwell-Boltzman law. In other cases these numbers are uncertain, but it is reasonable to assume in the case of an electric discharge that the populations of states whose energies are nearly the same are proportional to the statistical weights, as they are for thermal equilibrium. This assumption seems to be in accordance with the experimental facts. At the worst we can test the theory under conditions in which the population numbers are unknown by measuring and computing the relative intensities of the different spectrum lines originating in a common upper energy level.

A more direct method of computing emission intensities is highly desirable, however, and in the last analysis necessary. Eq. (34) asserts in effect the incoherence of the radiation fields of neighboring atoms and is therefore inapplicable when the electromagnetic waves sent out by neighboring atoms have definite phase relationships. Consider, for example, the metallic reflection of resonance radiation observed by Wood at the surface of a resonance bulb of mercury vapor.[19] The regularity of the reflection and the polarization of the reflected light show clearly that in this case the emission of the surface atoms is coherent so that amplitudes add, rather than intensities. By lowering the pressure, however, there is a continuous transition between this regular reflection and the emission of diffuse incoherent fluorescence light by the irradiated vapor. Similarly in the case of ordinary dispersion we have to do with elementary wavelets emitted by adjacent atoms with definitely related phases although the frequency emitted is a frequency of forced vibration which does not happen to coincide with any of the transition frequencies of the atoms.

As a first step toward the development of a theory of emission applicable to coherent as well as incoherent conditions we observe that since the ψ functions give a statistical description of the atomic motions while the electric and magnetic vectors give a corresponding statistical description of the emitted streams of photons, it is to be expected *a priori* that $\boldsymbol{E}$ and $\boldsymbol{H}$ for the emitted radiation shall be determined by the ψ functions just as in the classical theory $\boldsymbol{E}$ and $\boldsymbol{H}$ were determined by the motion of the electrical charges. Schrödinger's original suggestion was that the charge density ρ to be used in

---

[19] R. W. Wood, Phil. Mag. (6) **18,** 187 (1909); (6) **23,** 689 (1912); R. W Wood and M. Kimura, Phil. Mag. (6) **32,** 329 (1916).

Maxwell's equations should be identified, in the case of a single electron, with $e\psi\psi^*$. From our present point of view this means that ρ is to be identified with the statistical average charge per unit volume at the point in question. If the vector current density for a single electron is identified with

$$\frac{eh}{4\pi\mu i}[\psi \operatorname{grad} \psi^* - \psi^* \operatorname{grad} \psi] \tag{36}$$

[Cf. Part I, Section 3, 2] the law of continuity is satisfied, as well as the law of the conservation of electricity. In the case of a system of many electrons $\rho(x, y, z)$ and the current density are to be obtained by summing up the contributions of all the individual charges. The formulas are cumbersome but need not be introduced explicitly here since in practice the radiation field of a microscopic emitter is sufficiently determined by its dipole moment. In conformity with the above definition of ρ one may compute the statistical mean value of the electric moment $\boldsymbol{P}$ from the formula

$$\begin{aligned} \overline{\boldsymbol{P}}(t) &= \iiint_{\infty} \boldsymbol{r}\rho(x,y,z)\,dx\,dy\,dz \\ &= \int_{\text{all coordinate space}} \sum_{j=1}^{n} e\boldsymbol{r}_j\psi\psi^*\,d\tau \end{aligned} \tag{37}$$

where $\boldsymbol{r}$ is the radius vector from the origin to the point (x, y, z).

Expanding ψ in terms of the characteristic functions and neglecting the continuous spectrum as usual, we obtain

$$\overline{\boldsymbol{P}}(t) = \sum_{n}\sum_{m} c_n^* c_m \boldsymbol{P}(m, n/t) = \sum_{n}\sum_{m} c_n^* c_m \boldsymbol{P}(m, n) e^{2\pi i\nu_{mn}t}, \tag{38}$$

where $\boldsymbol{P}(m, n/t)$ and $\boldsymbol{P}(m,n)$ are elements of the vector matrices of $\boldsymbol{P}$ based on the functions $\psi(x, t)$ and $u(x)$ respectively. (Cf. p. 8.) The matrix $\boldsymbol{P}(m, n/t)$ is Hermitian so that each term in the above sum is the complex conjugate of the term obtained by interchanging the values of m and n. Hence $\overline{\boldsymbol{P}}(t)$ is real, as it should be. Eq. (38) shows that the frequencies of the electric moment as defined are identical with those given by the Bohr frequency condition and hence gives initial assurance that we are on the right track.

To determine the mean rate of emission of energy per atom in an ensemble with wave function ψ we insert $\overline{\boldsymbol{P}}(t)$ for the electric moment in the classical electrodynamic formula

$$\overline{I} = \frac{2}{3c^3}\overline{\left|\frac{d^2\overline{\boldsymbol{P}}}{dt^2}\right|^2} = \frac{2}{3c^3}\overline{[(\ddot{\overline{P}}_x)^2 + (\ddot{\overline{P}}_y)^2 + (\ddot{\overline{P}}_z)^2]} \tag{39}$$

Using the series of Eq. (38) and reducing, we obtain

$$\overline{I} = \frac{4}{3c^3}\sum_{\nu_{mn}}\sum_{>0} |c_n|^2 |c_m|^2 (2\pi\nu_{mn})^4 |\boldsymbol{P}(mn)|^2.$$

Defining $A_{m\to n}$ by Eqs. (35) and (28), this becomes

$$\bar{I}=\sum_{\nu_{mn}>0}\sum |c_n|^2|c_m|^2A_{m\to n}h\nu_{mn}. \tag{40}$$

Thus we find the emitted stream of energy made up of a sum of terms each having one of the frequencies of the actual spectrum and proportional to the corresponding transition probability $A_{m\to n}$. The theory is in harmony with the classical electromagnetic theory in the limiting case of a macroscopic system whose configuration can be sharply defined by a suitable wave packet since in this case $\bar{P}(t)$ reduces to the classical electric moment.

When, however, we apply our equation to the emission of radiation by a gas we strike a snag. Strictly speaking, we should treat the entire sample of gas as a single electro-mechanical system and compute the emission from a single function describing an ensemble of such systems. As this method of attack is too difficult, we content ourselves with a short cut suggested by classical considerations. Let us assume that the radiation from any element of volume can be obtained by compounding the electric moments of a set of independent atoms each of which is a member of a separate ensemble with a separate ψ function. The electric moment of the volume element *dxdydz* is then the sum of the values of $\bar{P}(t)$ for all atoms in the element. If the phases of the $\bar{P}$'s are haphazard, the intensities of emission add and the mean rate of emission of frequency ν_{mn} is[20]

$$\sum_{\text{All atoms}} |c_n|^2|c_m|^2A_{m\to n}h\nu_{mn}.$$

Let N denote the number of atoms per unit volume. Then the rate of emission of the frequency ν_{mn} per unit volume is

$$S(\nu_{mn})=N\overline{|c_m|^2|c_n|^2}A_{m\to n}h\nu_{mn}, \tag{41}$$

where the bar denotes an average over the atoms in the element. Clearly $N\overline{c_m}^2$ may be identified with the number of atoms per unit volume in the state m. Were it not for the factor $|c_n|^2$ in the right hand member, Eq. (41) would be in satisfactory agreement with Eq. (34). This factor, however, makes the emission of energy of any frequency depend on the number of atoms in the corresponding lower energy level as well as on the number in the upper energy level, and is in violent conflict with Einstein's theory and with experience.

The difficulty here encountered is fundamental and apparently can be met only by a modification of our conception of the electromagnetic field. The necessity for such modification was foreshadowed in early papers on the matrix mechanics by Born and Jordan, and by Born, Heisenberg, and Jordan[21] who assumed from the beginning that the kinematics of the electric and

---

[20] In order to take into account the effect of collisions between atoms, the perturbations by the ever present radiation field, or the damping effect of radiation resistance, one must regard the c_m's and c_n's as functions of the time. In that case it would be necessary to average the product with respect to time as well as over all the atoms in the volume element.

[21] M. Born and P. Jordan, Zeits. f. Physik **34**, 858 (1925); M. Born, W. Heisenberg, and P. Jordan, Zeits. f. Physik **35**, 557 (1926).

magnetic vectors must be the same as that of the coordinates and momenta of matter particles. To be specific, they assumed the validity of Maxwell's equations for the electromagnetic field but re-interpreted the dependent variables E and H in these equations as matrix functions of the coordinates instead of ordinary vector point functions. From the point of view of this review, this means that they regard E and H as quantities which are only partially determinate like the p's and q's in the wave mechanics of a system of material particles.

The plausibility of this procedure is perhaps most evident in the case of a radiation field in an evacuated rectangular box with perfectly reflecting walls. It is well known[22] that in the case of such a box the classical electromagnetic theory permits the resolution of the most general radiation field into the sum of an infinite series of discrete simple harmonic normal modes of vibration. The instantaneous amplitude of one of these vibrations may be regarded as a normal coordinate of the system and the equations of motion may be thrown into Lagrangian or Hamiltonian form. If q_n is any normal coordinate and p_n is the conjugate momentum, the expressions for the electric force E and vector potential A are of the form

$$E = \sum_n q_n f_n(x,y,z) = \sum_n E_n ;$$

$$A = \sum_n p_n a_n f_n(x,y,z) = \sum_n A_n .$$

The analogy between the equations of motion of the normal coordinates of this electromagnetic field and the equations of motion of the coordinates of a system of particles suggests at once that we treat the former like the latter and set up a wave equation based on the Hamiltonian function similar to that used in the treatment of a system of particles. The Heisenberg uncertainty principle then applies to the conjugate variables p_n and q_n and hence in a slightly modified form to the corresponding values of E_n and A_n.

Lorentz[22] has further shown that classically the same normal coordinates for the radiation field can be used to describe an interaction between matter and radiation in a box with perfectly reflecting walls, and Dirac[23] has been notably successful in dealing with the problem of the emission and absorption of radiation on the basis of quantum mechanics applied to matter and electromagnetic field together, using the device of the perfectly reflecting box to reduce the equations of motion to Hamiltonian form. In this way he derives the above formulas for the transition probabilities $A_{m\rightarrow n}$ and $B_{m\rightarrow n}$, and a theory of dispersion in a single self-consistent theory. The authors will not attempt, however, to review here these most important, but somewhat abstruse, papers of Dirac, primarily because up to this point we have not developed the necessary mathematical machinery.

---

[22] Cf. H. A. Lorentz in the report of the Solvay Conference of 1911: "La Theorie du Rayonnement et les Quanta," p. 12, Paris, 1912. The discussion by Lorentz is not limited to the case where the box is completely evacuated.

[23] P. A. M. Dirac, Proc. Roy. Soc. **A114**, 243 and 710 (1927).

It will be worth while, however, to examine briefly a less radical procedure recently suggested by Frenkel[24] which is based on the early discussion of the electromagnetic field from the matrix point of view by Born and Jordan.[21] The fundamental idea is that since $\overline{P}(t)$ [Cf. Eqs. (37) and (38)] is the statistical mean value of the electric moment of the emitting system, the electric and magnetic forces computed from it by classical formulas may be expected to be statistical mean values of the actual electric and magnetic forces for individual systems in the ensemble. To determine the corresponding rate of emission we have to compute the statistical mean value of the Poynting vector and then integrate over all directions of emission. If we now use the classical formula

$$\mathfrak{S}=\frac{c}{4\pi}\boldsymbol{E}\times\boldsymbol{H} \tag{42}$$

for the computation of $\overline{\mathfrak{S}}$ from $\overline{\boldsymbol{E}}$ and $\overline{\boldsymbol{H}}$ we obtain the incorrect expression (40) for the mean rate of emission. On the other hand we may proceed as follows. First we use the classical electromagnetic formulas[25]

$$\boldsymbol{E}\left(t+\frac{r}{c}\right)=\frac{1}{c^2r}\left[\boldsymbol{r}\left(\boldsymbol{r}_0\cdot\frac{d^2\boldsymbol{P}}{dt^2}\right)-\frac{d^2\boldsymbol{P}}{dt^2}\right];\ \boldsymbol{H}\left(t+\frac{r}{c}\right)=\frac{1}{c^2r}\left[\boldsymbol{r}_0\times\frac{d^2\boldsymbol{P}}{dt^2}\right] \tag{43}$$

to compute from each matrix component $\boldsymbol{P}(m,\, n/t)$ of $\boldsymbol{P}$ corresponding matrix elements of $\boldsymbol{E}$ and $\boldsymbol{H}$. Although these elements are functions of time we indicate them by simple subscripts as $\boldsymbol{E}_{mn}$ and $\boldsymbol{H}_{mn}$ repectively. From the definitions of $\boldsymbol{P}$ and of matrix elements it follows that

$$\overline{\boldsymbol{E}}=\sum_n\ \sum_m c_n{}^*c_m\boldsymbol{E}_{mn}\ ;\quad \overline{\boldsymbol{H}}=\sum_n\ \sum_m c_n{}^*c_m\boldsymbol{H}_{mn} \tag{44}$$

in analogy with Eq. (38). We now compute the Poynting vector $\overline{\mathfrak{S}}$ from the modified formulas

$$\overline{\mathfrak{S}}=\sum_n\sum_m c_n{}^*c_m\mathfrak{S}_{mn} \tag{45}$$

and [26]

$$\mathfrak{S}_{mn}=\frac{c}{8\pi}\ \sum_k(\boldsymbol{E}_{kn}\times\boldsymbol{H}_{mk}-\boldsymbol{H}_{kn}\times\boldsymbol{E}_{mk})\,. \tag{45a}$$

Forming the time average of the statistical average $\overline{\mathfrak{S}}$, we find that

$$\overline{\overline{\mathfrak{S}}}=\sum_n|\,c_n\,|^2\mathfrak{S}_{nn}=\frac{c}{4\pi}\ \sum_n\sum_k(\,|\,c_n\,|^2+|\,c_k\,|^2)(\boldsymbol{E}_{kn}\times\boldsymbol{H}_{nk})\,. \tag{46}$$

---

[24] J. Frenkel, "Einführung in die Wellenmechanik," Berlin, 1929, Chap. III, §5. Frenkel's volume reached us after the completion of Part I of this review.

[25] Here $\boldsymbol{r}_0$ denotes an unit vector along the radius drawn from the atomic center to the point for which $\boldsymbol{E}$ and $\boldsymbol{H}$ are to be calculated.

[26] By definition $(\boldsymbol{E}_{kn}\times\boldsymbol{H}_{mk})^{(x)}=(E_{kn}{}^{(y)}H_{mk}{}^{(z)}-E_{kn}{}^{(z)}H_{mk}{}^{(y)})$ where $E_{kn}{}^{(y)}$ is the yth component of the vector matrix $\boldsymbol{E}_{kn}$, etc. Cf. also footnote 28.

At this point Frenkel proceeds to take the unjustified step of replacing $(|c_n|^2+|c_k|^2)$ by $2|c_n|^2$. With this arbitrary alteration one gets

$$\bar{I}=\sum_{\nu_{nk}>0}\sum |c_n|^2 A_{n\to k}h\nu_{nk} \tag{47}$$

for the mean rate of emission in agreement with Eq. (34).

The above method of computing $\bar{\mathfrak{S}}$ up to Eq. (47) is exactly what we should expect to employ if the vectors $\boldsymbol{E}$, $\boldsymbol{H}$, and $\mathfrak{S}$ for the field of the atom are, like the electric moment $\boldsymbol{P}$ itself, definite functions of the p's and q's of the atom. Then the classical formulas for $\boldsymbol{E}$, $\boldsymbol{H}$, and $\mathfrak{S}$ in terms of the p's and q's determine operators $O_E(x, y, z, t)$, $O_H(x, y, z, t)$, and $O_{\mathfrak{S}}(x, y, z, t)$ which may be used to compute $\bar{\boldsymbol{E}}$, $\bar{\boldsymbol{H}}$, and $\bar{\mathfrak{S}}$ from the ψ function of the emitting ensemble of atoms. Defining the desired mean values by means of the operators as in Part I, Sect. 3 and expanding ψ in terms of the characteristic functions we obtain

$$\begin{aligned}
\bar{\boldsymbol{E}}&=\int \psi^* O_E\psi d\tau=\sum_{n,m} c_n^* c_m \boldsymbol{E}_{mn}\\
\bar{\boldsymbol{H}}&=\int \psi^* O_H\psi d\tau=\sum_{n,m} c_n^* c_m \boldsymbol{H}_{mn}\\
\bar{\mathfrak{S}}&=\int \psi^* O_{\mathfrak{S}}\psi d\tau=\sum_{n,m} c_n^* c_m \mathfrak{S}_{mn}.
\end{aligned} \tag{48}$$

The matrix elements $\boldsymbol{E}_{mn}$, $\boldsymbol{H}_{mn}$ as defined by Eqs. (48) may readily be proved to be identical with the elements $\boldsymbol{E}_{mn}$, $\boldsymbol{H}_{mn}$ previously defined in terms of $\boldsymbol{P}(m, n/t)$ (at least to the approximation to which Eqs. (43) are valid classically).[27] It remains to show that $\mathfrak{S}_{mn}$ is properly defined by Eq. (45a). The desired proof is a corollary on a general theorem of fundamental importance for the matrix formulation of the quantum mechanics (cf. the next section).

Theorem: Let O_α and O_β denote two operators associated with the dynamical variables α and β. Let α_{mn} be the matrix element defined by the equivalent equations

---

[27] Classically in the case of a system of electrons of equal mass and charge

$$\frac{d^2P}{dt^2}=\frac{e}{\mu}\sum_j \dot{p}_j=-\frac{e}{\mu}\sum_j \text{grad}_j V$$

where $\text{grad}_j V$ denotes the gradient with respect to the coordinates of the j^{th} particle. We may then use $-(e/\mu)\sum_j \text{grad}_j V$ as an operator for $d^2\boldsymbol{P}/dt^2$ and by combination with Eqs. (43) work out the operators $\boldsymbol{O_E}$ and $\boldsymbol{O_H}$. With the aid of the vector equivalent of Eq. (18a) and a generalization for Eq. (I, 92) it is then possible to show that

$$\frac{d^2\bar{P}}{dt^2}=-\frac{e}{\mu}\overline{\sum_j \text{grad}_j V}$$

and thus establish the equivalence of the two definitions of the matrices for $\boldsymbol{E}_{mn}$ and $\boldsymbol{H}_{mn}$.

$$\alpha_{mn}=\int \psi_n{}^*O_\alpha\psi_m d\tau \tag{49}$$

$$O_\alpha\psi_m=\sum_n \alpha_{mn}\psi_n. \tag{50}$$

Let β_{mn} be defined in the same way. Finally, let O_γ denote the operator product $O_\alpha O_\beta$. Then

$$\gamma_{mn}=\sum_k \beta_{mk}\alpha_{kn}=\sum_k \alpha_{kn}\beta_{mk}. \tag{51}$$

The right hand member of Eq. (51) is by definition the element (mn) of the matrix product $\{\beta_{mn}\}\ \{\alpha_{mn}\}$ so that the theorem states that the matrix of the product of two operators is the product of the matrices of the operators taken separately, the two products being formed in inverse order.[28]

Proof: It is only necessary to expand $O_\beta\psi_n$ in accordance with Eq. (50) and then apply the operator O_α to each term in the expansion. Then

$$O_\gamma\psi_m=O_\alpha\sum_k \beta_{mk}\psi_k=\sum_k \beta_{mk}\sum_n \alpha_{kn}\psi_n=\sum_n\left(\sum_k \beta_{mk}\alpha_{kn}\right)\psi_n. \tag{52}$$

This proves the theorem subject to the condition that is legitimate to apply the operator O_α term by term to the series $\sum_k \beta_{mk}\psi_k$.

Returning now to the Poynting vector, we observe that in view of the classical formula (42) we may define $\mathbf{O}_{\mathfrak{S}}$ by an equation of the form

$$\mathbf{O}_{\mathfrak{S}}=\frac{c}{4\pi}\mathbf{O}_E\times\mathbf{O}_H. \tag{53}$$

Here we may identify $\mathbf{O}_E$ and $\mathbf{O}_H$ with the vector operators

$$\mathbf{O}_E=iO_X+jO_Y+kO_Z$$
$$\mathbf{O}_H=iO_\alpha+jO_\beta+kO_\gamma$$

where X, Y, Z, α, β, γ, are the components of $\mathbf{E}$ and $\mathbf{H}$ respectively. Expanding (53) in terms of the corresponding scalar components we find an

---

[28] It is well known that the commutative law of ordinary multiplication does not apply to matrix multiplication and obviously it cannot apply to the multiplication of such operators as q and $\partial/\partial q$. For a presentation of the elements of matrix algebra the reader is referred to Bocher's "Higher Algebra" or to any of the current introductions to the matrix mechanics. Cf. also p. 25 ahead.

The inversion of order between the operator products and matrix products is an unhappy peculiarity of the present presentation of the subject, traceable in the first instance to the convention regarding the order of the subscripts in the equations which define the matrices in terms of the operators ((I, 131) and (49) and (50) above). Due to our original unfortunate choice of the sign of the exponent of the time factor of ψ, it was necessary to arrange the subscripts as we have done in order to get the same matrix elements (Hermitian case only) as those used by Born, Heisenberg, and Jordan in the matrix mechanics. The reader should take warning from this horrible example of the results of unconventionality!

ambiguity regard ing the order of the factors in the individual terms. We resolve this ambiguity by using in each case the mean value of the two products in direct and reverse order. This method of defining $O_{\mathfrak{S}}$ gives the corresponding matrix Hermitian form, and Eq. (45a) now follows as a corollary.

This completes the justification of Eq. (46) and indicates that one can make a certain amount of progress by the comparatively simple "quantization of the field" used in Eqs. (43) and (45a). The result is still incorrect, to be sure, since Eq. (47) cannot be derived from Eq. (46) by any legitimate procedure, but Frenkel's method does represent an advance over that employed on p. 14.

Consider next the phenomena of dispersion and scattering. Classically we account for them as due to the emission of secondary wavelets by atoms and molecules executing forced vibrations under the influence of the primary light waves. The first step in the theory is to compute the electric moment of the atom from the equations of motion of the electrons treated as oscillators held to fixed positions of equilibrium by elastic forces. Eqs. (43) are then used to compute the electric forces of the wavelets due to the individual atoms, and these are combined with due regard to amplitude and phase to give the scattered radiation and the coherently emitted beam which produces dispersion and reflection. In the quantum theory we may follow the same general program, the first step being the computation of the perturbed value of the mean electric moment of an individual atom under the influence of a monochromatic radiation field.

Let the field consist of a sinusoidal train moving along the x-axis with its initial wave-front at $x=ct$. We assume that the waves are polarized with the electric vector parallel to the y-axis, and define the vector potential by the equation

$$\begin{aligned} A_y &= -\alpha f(t-x/c) = -\alpha \sin 2\pi\nu_0(t-x/c), \qquad && t-x/c>0\,; \\ &=0, && t-x/c<0. \end{aligned} \tag{54}$$

We may use the machinery of Section 4, 3 and Section 5, 3, neglecting as before the variation in the phase of the radiation field over the volume of the individual atom. Let ξ denote the x coordinate of the center of gravity of the atom or molecule under consideration. Replacing the fixed time T of Section 5, 3 by the variable time t of Eqs. (I, 178-179) we have in first approximation

$$c_m(t) = c_m{}^0 + \alpha \sum_k c_k{}^0 \overset{(1)}{f}_{km}(t) \tag{55}$$

where

$$\overset{(1)}{f}_{km}(t) = -\frac{4\pi^2}{hc}\nu_{km}P_y(km)\int_{t'=\xi/c}^{t'=t} f(t'-\xi/c)e^{2\pi i\nu_{km}t'}dt' \tag{56}$$

Expressing $f(t'-\xi/c)$ in terms of exponentials and carrying out the integration, the latter equation becomes

$$f_{km}(t)=\frac{\pi}{hc}\nu_{km}P_y(km)\left\{\frac{e^{2\pi i[(\nu_{km}+\nu_0)t-\nu_0\xi/c]}-e^{2\pi i\nu_{km}\xi/c}}{\nu_{km}+\nu_0}-\frac{e^{2\pi i[(\nu_{km}-\nu_0)t+\nu_0\xi/c]}-e^{2\pi i\nu_{km}\xi/c}}{\nu_{km}-\nu_0}\right\} \tag{57}$$

provided that the radiation frequency ν_0 is not equal to the characteristic frequency ν_{km}, a condition which we shall suppose satisfied. Using the perturbed values of the c's given by Eqs. (55) and (57) in the general expression (38) for the mean electric moment of the atom, multiplying out, neglecting terms in the square of the amplitude of the light wave, and reducing, we obtain

$$\begin{aligned}\bar{\boldsymbol{P}}(t)= \sum_{n,m} c_n{}^{0*}c_m{}^0 \Bigg\{ & \boldsymbol{P}(\dot{m}n)e^{2\pi i\nu_{mn}t} \\ & +\frac{\alpha\pi}{hc}\sum_k \boldsymbol{P}(kn)P_y(mk)\nu_{mk}\left[\frac{e^{2\pi i[(\nu_{mn}+\nu_0)t-\nu_0\xi/c]}-e^{2\pi i(\nu_{kn}t+\nu_{mk}\xi/c)}}{\nu_{mk}+\nu_0}\right. \\ & \left.-\frac{e^{2\pi i[(\nu_{mn}-\nu_0)t+\nu_0\xi/c]}-e^{2\pi i(\nu_{kn}t+\nu_{mk}\xi/c)}}{\nu_{mk}-\nu_0}\right] \\ & +\frac{\alpha\pi}{hc}\sum_k \boldsymbol{P}^*(km)P_y{}^*(nk)\nu_{nk}\left[\frac{e^{2\pi i[(\nu_{mn}-\nu_0)t+\nu_0\xi/c]}-e^{2\pi i(\nu_{mk}t+\nu_{kn}\xi/c)}}{\nu_{nk}+\nu_0}\right. \\ & \left.-\frac{e^{2\pi i[(\nu_{mn}+\nu_0)t-\nu_0\xi/c]}-e^{2\pi i(\nu_{mk}t+\nu_{kn}\xi/c)}}{\nu_{nk}-\nu_0}\right]\Bigg\}. \end{aligned} \tag{58}$$

The coefficients of the initial amplitude products $c_n{}^{0*}\ c_m{}^0$ make up what we may call the perturbed value of the electric moment matrix $\boldsymbol{P}(m, n/t)$ to be used in computing $\boldsymbol{E}_{mn}$ and $\boldsymbol{H}_{mn}$. The terms having the frequencies ν_{mk} and ν_{kn} indicate a slight modification of the emission intensities due to the incidence of the perturbing light wave. In practice they would be quickly damped out by the radiation resistance which is always present, though we cannot include it in our theory, and we therefore neglect them.

There are three sorts of terms in the matrix, viz., those having the characteristic frequencies ν_{mn} of the unperturbed atom and giving rise to spontaneous incoherent emission; those having the frequency of the exciting light and producing reflection, dispersion, and normal Rayleigh scattering; and finally those having combination frequencies and producing Raman scattering. The Compton modified type of scattering is not represented in the above formula, as it does not take into account the continuous spectrum of characteristic functions of the atom. Compton scattering is entirely analogous to Raman scattering except that it is accompanied by atomic transitions from discrete energy levels to levels in the continuous spectrum.[29]

---

[29] In other words, the energy given up by a photon which experiences Compton scattering produces ionization of the atom struck, while the energy given up by a photon which experiences Raman scattering merely produces excitation of the atom struck.

The diagonal terms obtained by setting m equal to n are responsible for normal scattering and dispersion. If $\overline{\boldsymbol{D}}(t)$ is used to denote the part of the mean electric moment coming from these diagonal terms, we have, after some juggling of terms,

$$\overline{\boldsymbol{D}}(t)=\frac{2\alpha\pi}{hc}\sum_n |c_n{}^0|^2 R\left\{\sum_k \nu_{nk}\boldsymbol{P}(kn)P_y(nk)\left[\frac{e^{2\pi i\nu_0\zeta}}{\nu_{nk}+\nu_0}-\frac{e^{-2\pi i\nu_0\zeta}}{\nu_{nk}-\nu_0}\right]\right\} \tag{59}$$

where the symbol R indicates that only the real part of the expression which follows it is to be considered, and where ζ stands for $(t-\xi/c)$.[30]

Although we have assumed a light wave polarized with the electric vector parallel to the y-axis, the above formula shows that the part of the induced electric moment having the frequency ν_0 is not rigorously parallel to the y-axis. In consequence the scattered light is not completely polarized. The high degree of polarization observed experimentally is accounted for by the fact that while $P_y(m,n)\,P_y(n,m)$ is equal to $|P_y(n,m)|^2$ and hence is essentially positive, such products as $P_x(m,n)\,P_y(n,m)$ may be either positive or negative, and will on the average tend to cancel each other when one carries out the summation indicated in Eq. (59).

The important y-component of $\overline{\boldsymbol{D}}$ which is alone responsible for ordinary dispersion is readily reduced to the form

$$\overline{\boldsymbol{D}}_y(t)=-\frac{4\alpha\pi\nu_0}{hc}\cos 2\pi i\nu_0\zeta\sum_n |c_n{}^0|^2\left\{\sum_k \nu_{nk}|P_y(nk)|^2\left(\frac{1}{\nu_{nk}{}^2-\nu_0{}^2}\right)\right\} \tag{60}$$

To get the electric moment per unit volume we average $\overline{\boldsymbol{D}}_y$ over all the atoms in a unit volume. Denoting the number of atoms per unit volume by N, we have

$$\begin{aligned} n^2-1&=4\pi\frac{\text{Electric moment per unit volume}}{\text{Primary electric force}}\\ &=\frac{8\pi N}{h}\sum_n |c_n{}^0|^2\sum_m \nu_{mn}|P_y(nm)|^2\left(\frac{1}{\nu_{nm}{}^2-\nu_0{}^2}\right)\\ &=\frac{8\pi N}{h}\sum\sum_{\nu_{nm}>0}\left[|c_n{}^0|^2-|c_m{}^0|^2\right]|P_y(nm)|^2\frac{\nu_{nm}}{\nu_{nm}{}^2-\nu_0{}^2}\cdot \end{aligned} \tag{61}$$

This is Kramer's dispersion formula. The appearance of $|c_m{}^0|^2$ with a negative sign shows that the atoms in the lower energy levels make contributions to dispersion opposite in sign to those in the upper energy levels.

Eq. (61) can be put in more conventional form by using our previous considerations (Section 5, 3) on degenerate systems. We replace the indices (n, m) by $(n, l; m, k)$ where l and k are the quantum numbers which determine the orientation of the atom in space; i.e., l and k

---

[30] In deriving the above formula one makes use of the fact that since the operator for $\boldsymbol{P}$ or for any one of its components is a real function of the coordinates, the matrix $\boldsymbol{P}(m,n)$ is Hermitian. The reader will observe that the matrix elements of $\boldsymbol{P}$ and P_y which appear explicitly in (59) are the unperturbed elements.

are the magnetic quantum numbers associated with the elementary states n and m. In the absence of any constant external fields we may write

$$\sum_{n,l} | c^0_{n,l} |^2 \sum_{m,k} \nu_{mn} | P_y(n,l;m,k) |^2 \left(\frac{1}{\nu^2_{nm} - \nu_0^2} \right)$$
$$= \frac{1}{3} \sum_{n,l} \frac{1}{g_n} | c_n{}^0 |^2 \sum_{m,k} \nu_{mn} | P(n,l;m,k) |^2 \left(\frac{1}{\nu_{nm}{}^2 - \nu_0{}^2} \right) \tag{62}$$

as the ν_{nm}'s do not depend on the quantum numbers l and k, and also as[31]

$$\sum_{l,k} | P_y(n,l;m,k) |^2 = \tfrac{1}{3} \sum_{l,k} | P(n,l;m,k) |^2 \tag{63}$$

by symmetry. But by Eqs. (28) and (32)

$$\frac{1}{g_n} \sum_{l,k} | P(n,l;m,k) |^2 = \frac{3h^2}{8\pi^3} B_{n\to m} = \frac{3c^3 h}{64\pi^4 \nu_{nm}{}^3} A_{n\to m}. \tag{64}$$

Combining Eqs. (61), (62), and (64) we get

$$n^2 - 1 = \frac{Nc^3}{8\pi^3} \sum_n | c_n{}^0 |^2 \sum_m \frac{A_{n\to m}}{\nu_{nm}{}^2(\nu_{nm}{}^2 - \nu_0{}^2)}$$
$$= \frac{Nc^3}{8\pi^3} \sum\sum_{\nu_{nm} > 0} [| c_n{}^0 |^2 - | c_m{}^0 |^2] \frac{A_{n\to m}}{\nu_{nm}{}^2(\nu_{nm}{}^2 - \nu_0{}^2)}. \tag{65}$$

The Rayleigh scattering for a gas can be computed from the induced electric moment per atom $\overline{\boldsymbol{D}}$ as in the classical theory by adding the intensities of the waves scattered by the different atoms. The justification for this procedure depends on the haphazard location of gas molecules, or on the fluctuations in gas density, according to one's point of view.

The Raman scattering comes from the non-diagonal terms in Eq. (58). Neglecting the terms involving only the characteristic frequencies of the atom, and the terms which produce normal scattering and dispersion, we may lump the others together and call the resultant the "Raman moment" $\boldsymbol{R}(t)$ of the atom. This gives

$$R_{mn}(t) = \frac{\alpha\pi}{hc} \left\{ \left[\sum_k \frac{P(kn)P_y(mk)\nu_{mk}}{\nu_{mk} + \nu_0} + \frac{P^*(km)P_y{}^*(nk)\nu_{kn}}{\nu_{nk} - \nu_0} \right] e^{2\pi i(\nu_0 + \nu_{mn})t} \right.$$
$$\left. - \left[\sum_k \frac{P(kn)P_y(mk)\nu_{mk}}{\nu_{mk} - \nu_0} + \frac{P^*(km)P_y{}^*(nk)\nu_{kn}}{\nu_{nk} + \nu_0} \right] e^{-2\pi i(\nu_0 - \nu_{mn})t} \right\}. \tag{66}$$

This is substantially the expression for a typical element of the Raman moment matrix of the atomic system as used elsewhere by the authors in discussing the Raman effect.[32]

If we compute the intensities of the various Raman lines by means of Eq. (66) we meet with the same difficulty as in dealing with spontaneous

---

[31] This formula is of considerable generality, and is intimately connected with the question of "spectroscopic stability." Eqs. (62) and (63) can be shown to be correct to a very high degree of approximation even in case one or more constant external fields are present. Cf. J. H. Van Vleck, Phys. Rev. **29**, 727 (1927); **30**, 31 (1927).

[32] E. L. Hill and E. C. Kemble, Proc. Nat. Acad. Sci. **15**, 387 (1929). In the equation used in this paper the anti-Stokes terms of the Raman moment (i.e. those with frequencies greater than that of the exciting light) were omitted, as were the frequency factors ν_{mk} and ν_{kn} in the numerators of the fractions in Eq. (66).

emission. The theory makes the intensities depend in a symmetrical way on the populations of the initial and final energy levels. To get an empirically correct formula one must first resolve the matrix $\{R_{mn}(t)\}$ into two parts $\{R_{mn}^{(+)}\}$ and $\{R_{mn}^{(-)}\}$ associated with the anti-Stokes frequencies $(\nu_0+\nu_{mn})$ and the Stokes frequencies $(\nu_0-\nu_{mn})$ respectively. The emission from $\{R_{mn}^{(+)}\}$ is then computed using the populations of the upper energy levels, and from $\{R_{mn}^{(-)}\}$ using the populations of the lower levels.

SECTION 6. MATRIX THEORY

I. MATRIX ALGEBRA AND THE CHARACTERISTIC VALUE PROBLEM

In Part I, Section 3, and in subsequent sections we have made use of certain quantities which we called "matrix elements" because they appear in sets, each of which can be arranged in a two-dimensional array, or "matrix." Each such set is derived from some "dynamical variable" α (a function of the coordinates and momenta of the problem) and from a complete normalized orthogonal set of functions of the q's, say $f_1, f_2, \cdots, f_n, \cdots$. In order to find the matrix[33] of α one must first set up a corresponding operator O_α (in simple cases by the substitution of $-(h/2\pi i)\,\partial/\partial q_k$ for p_k in the classical formula for α) and then expand the function $O_\alpha f_m$ in terms of the complete set of f's. The matrix element $\alpha(m, n)$ is then defined by the expansion

$$O_\alpha f_m = \sum_n \alpha(m,n) f_n, \tag{67}$$

or by the corresponding explicit formula

$$\alpha(m,n) = \int_\infty f_n^* O_\alpha f_m d\tau. \tag{68}$$

We usually identify the f's with the characteristic functions which describe monochromatic solutions of Schrödinger's wave equation. We assume for the present that this equation has no continuous spectrum.[34] The above notation for the matrix elements will be used when f_n is set equal to the space factor $u_n(q)$, but if the complete characteristic function

$$\psi_n(q,t) = u_n(q) e^{2\pi i E_n t/h}$$

is to be used for f_n the symbol $\alpha(m,n)$ will be replaced by $\alpha(m, n/t)$.

---

[33] The matrix of α is said to be a "representation" of the operator O_α. It may be proved that the operator and the matrix are equivalent in the sense that either one determines the other. This is true despite the fact that the form of the matrix depends on the particular set of wave functions which is used. The equivalence depends on the invariant properties of the elements $\alpha(m, n/t)$ under a linear substitution of the functions ψ_n, a complete study of which would lead us far afield into the theory of linear substitutions which constitutes part of the mathematical theory of groups. Cf. A. Landé, "Optik, Mechanik, und Wellenmechanik," Sect. VII, Handbuch der Physik XX, Berlin, 1928.

[34] Every problem involving a continuous spectrum may be approximated by means of a modified problem in which the wave function is artificially required to vanish on some hypersphere of large radius in configuration space.

The matrices based on the characteristic functions may be used to determine the mean values of various dynamical variables in the case of an ensemble of systems whose ψ function is analysed into a sum of orthogonal characteristic functions. Several examples of this use have been given in the preceding section. It is also possible with the aid of matrices to reduce to a purely algebraic form much of the computation involved in solving problems in quantum mechanics. We have, in fact, shown (Part I, Sect. 4, 1) that if the matrix of the Hamiltonian operator is known for a complete set of normalized orthogonal functions of the coordinates, the location of the correct energy levels may be reduced to the algebraic problem of solving an infinite series of simultaneous linear equations in an infinite number of unknowns. (Cf. Eq. (I, 135)). A further study of this problem forms a natural method of approach to the matrix mechanics of Heisenberg, Born, and Jordan.

It will be convenient to begin with a brief review of the elements of matrix algebra. For this purpose we introduce the single symbol[35] $\mathbf{A}$ or $\{A(m, n)\}$ for the matrix

$$\mathbf{A} \equiv \{A(m,n)\} \equiv \left\| \begin{array}{ll} A(11) & A(12)\cdots \\ A(21) & A(22)\cdots \\ \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots \\ \cdots\cdots & \cdots\cdots \end{array} \right\| .$$

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be equal if, and only if, corresponding elements are equal; *i. e.*, if

$$A(m,n) = B(m,n). \qquad m,n = 1,2,\cdots,\infty.$$

To add or subtract two matrices $\mathbf{A}$ and $\mathbf{B}$ we add or subtract corresponding elements.
Thus

$$(A \pm B)(m,n) = A(m,n) \pm B(m,n).$$

To multiply two matrices we use the rule[28]

$$(AB)(m,n) = \sum_k A(m,k)B(k,m). \tag{69}$$

On the basis of the above rule it can readily be shown that the associative law of ordinary algebra applies to the multiplication of matrices, although the commutative law does not. Thus in general

$$\mathbf{AB} \neq \mathbf{BA}.$$

The product of a matrix $\mathbf{A}$ and an ordinary number c is the matrix whose typical element is $cA(m, n)$.

---

[35] In the preceding sections we used bold face type to distinguish three-dimensional vectors from scalars. In the remainder of the article, however, we shall have no need for vectors and *shall reserve the bold face type notation for matrices.*

The unit matrix is defined by

$$I \equiv \{\delta_{mn}\} \equiv \begin{Vmatrix} 1 & 0 & 0 \cdots \\ 0 & 1 & 0 \cdots \\ 0 & 0 & 1 \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{Vmatrix}.$$

It follows at once that

$$AI = IA = A.$$

The inverse or reciprocal of a matrix A is defined, in case it exists, by the relation

$$AA^{-1} = A^{-1}A = I. \tag{70}$$

Positive and negative integral powers of A can now be defined by the rules

$$\begin{aligned} A^p &= AAAAAA \cdots \text{ to } p \text{ factors} \\ A^{-p} &= (A^{-1})^p \\ A^0 &= I. \end{aligned}$$

With the aid of these definitions we may finally establish the usual exponential rules

$$A^\mu A^\nu = A^{\mu+\nu}\ ; \quad (A^\mu)^\nu = A^{\mu\nu}$$

for all positive and negative integral powers of A.

Consider now the problem of determining the energy levels of an atomic system. Let a set of approximate wave functions $\psi_1{}^0, \psi_2{}^0, \cdots, \psi_n{}^0, \cdots$ be given, and let the corresponding matrix for the Hamiltonian operator be $\{H^0(m, n/t)\}$. From the Schrödinger point of view the problem is to replace the approximate wave functions by exact wave functions $\psi_1, \psi_2, \cdots, \psi_n, \cdots$ which satisfy the differential equation,

$$H\psi_n = E_n\psi_n = \frac{h}{2\pi i}\frac{\partial\psi_n}{\partial t}, \tag{71}$$

and to locate the corresponding characteristic values of E. If, however, the original set of ψ's forms a complete orthogonal set, any new set is expressible in terms of the old by means of a suitable linear expansion. Hence the replacement of the initial functions by the actual characteristic functions of the problem may be described as a *linear transformation* of the initial set of functions. This transformation has the form

$$\psi_m{}^0 \rightarrow \psi_m = \sum_k S(m, k/t)\psi_k{}^0. \tag{72}$$

the coefficients forming a matrix $\{S(m, n/t)\}$ which completely determines the transformation. If the new functions, like the old, form a normalized orthogonal set, such as can always be picked out from the complete manifold of the solutions of Eq. (71), the elements of the transformation matrix satisfy the relation[36]

$$\sum_k S(n,k/t)S^*(m,k/t)=\delta_{nm}. \tag{73}$$

Eq. (73) is a generalization of the relations between the direction cosines in a rotation of axes in three-dimensional space. A linear transformation which satisfies (73) is said to be an *unitary* transformation.[37]

If we introduce the symbol $\tilde{S}$ for the matrix obtained from S by interchanging rows and columns, we see that the condition (73) is equivalent to the matrix equation

$$\tilde{S}^*S=S\tilde{S}^*=I. \tag{74}$$

But by the definition of the inverse matrix S^{-1} given in Eq. (70) this is equivalent to writing

$$S^*=S^{-1}. \tag{75}$$

Turning now to the matrix of the Hamiltonian operator, we readily prove that the application of the transformation whose matrix is S to the ψ functions leads to a similar linear transformation of the elements of the matrix of H. Thus

$$\begin{aligned} H^0(m,n/t)\rightarrow H(m,n/i) &= \int \psi_n{}^* H\psi_m d\tau \\ &= \int\left[\sum_l S^*(n,l/t)(\psi_l{}^0)^*\right]H\left[\sum_k S(m,k/t)\psi_k{}^0\right]d\tau \\ &= \sum_{l,k} S^*(n,l/t)H^0(k,l/t)S(m,k/t). \end{aligned} \tag{76}$$

In the language of matrix algebra, Eq. (76) takes the form

$$H^0\rightarrow H=SH^0\tilde{S}^*=SH^0S^{-1}. \tag{77}$$

Equation (77) describes the transformation of the Hamiltonian matrix due to the substitution of any new complete orthogonal set of functions for the initial set. We wish to demand, however, that the new set shall be the actual characteristic functions of the problem in hand. In that case we have directly by Eq. (71) that

$$H(m,n/t)=\int \psi_n{}^* E_m\psi_m d\tau=E_m\delta_{nm}. \tag{78}$$

In other words, we require the transformation to carry the matrix of H over into what is called *diagonal* form, all terms dropping out except those

---

[36] Cf. section in fine print at top of p. 211, Part I; also Eq. (I, 182).

[37] This is a slight generalization of the *orthogonal* transformation adapted to the algebra of complex quantities. Cf. H. Weyl, "Gruppentheorie und Quantenmechanik," p. 16, Leipzig, 1928.

for which $m=n$. The diagonal terms are then the desired energy values. The problem of locating the exact energy levels and exact wave functions is thus reduced to the problem of finding a matrix $\boldsymbol{S}$ which will carry $\boldsymbol{H}$ over into diagonal form when applied according to the rule of Eq. (77). If we now equate the right hand member of Eq. (77) to the matrix $\{E_m\delta_{mn}\}$, and multiply both sides of the equation behind by the matrix $\boldsymbol{S}$, we obtain[38]

$$\boldsymbol{SH}^0 = \{E_m\delta_{mn}\}\boldsymbol{S}. \tag{79}$$

Equating the two members of this equation element by element, we obtain a double infinity of simultaneous equations with the elements of the matrix S and the E values as unknowns; i.e.,

$$\sum_k S(m,k/t)\,[H^0(k,n/t)-E_m\delta_{kn}]=0, \quad m,n=1,2,\cdots\infty. \tag{80}$$

If the initial set of approximate wave functions are of the form

$$\psi_m{}^0 = u_m{}^0 e^{2\pi i E_m{}^0 t/h} \tag{81}$$

where the u^0's are independent of the time, we can simplify the set of Eqs. (80). By Eq. (71) ψ_m is of the form $u_m e^{2\pi i E_m t/h}$. Hence Eq. (72) is equivalent to

$$u_m e^{2\pi i E_m t/h} = \sum_k S(m,k/t)\,u_k{}^0 e^{2\pi i E_k{}^0 t/h}.$$

It follows that

$$u_m = \sum_k S(m,k)\,u_k{}^0$$

where $S(m, k)$ is independent of t and

$$S(m,k/t)=S(m,k)\,e^{2\pi i(E_m-E_k{}^0)t/h}. \qquad (82$$

Similarly

$$H^0(m,k/t)=H^0(m,k)\,e^{2\pi i(E_m{}^0-E_k{}^0)t/h}. \tag{83}$$

By means of Eqs. (82) and (83), Eqs. (80) are readily reduced to the time-free form

$$\sum_k S(m,k)\,[H^0(kn)-E_m\delta_{kn}]=0, \quad m,n=1,2,\cdots,\infty. \tag{84}$$

The set of equations obtained from the above by picking out all cases in which m has a given value is identical with the set (I, 135) except for the detail of notation. As explained in Part I, the solution of these equations can be carried through in general only by the successive approximations of perturbation theory.

So far we have used matrices only to re-state, in a different language, conclusions already reached in Part I by formally different methods. We can go much further than this, however, with the aid of the theorem of page 18 which states that the matrix of the product of two operators in terms of

---

[38] Cf. Born, Heisenberg, and Jordan, Zeits. f. Physik **35**, 557 (1926), Eq. (20).

any given set of orthogonal functions is equal to the product in inverse order of the matrices of the two operators taken independently. Since the H operator is built up from the operators for the p's and q's, it follows at once that the H matrix can be built up from the matrices of the p's and q's. Similarly we can build up the matrix of the electric moment from the matrices of the p's and q's. Thus the solution of both the energy level problem and the intensity problem can be made to turn on the determination of the matrices for the p's and q's. In what follows we shall see that in some cases, at least, the matrices in question can be found and the whole problem solved without the specific introduction of the initial orthogonal function system $\psi_1{}^0, \psi_2{}^0, \cdots$.

2. THE PROPERTIES OF THE p AND q MATRICES

The matrices for the coordinates q_k and the momenta p_k for a Cartesian coordinate system are derived from the operators q_k and $-(h/2\pi i)\partial/\partial q_k$ respectively. In conformity with the last paragraph of Section 3, 4 of Part I we provisionally define the operator for the momentum conjugate to the *generalized* coordinate q_k to be $-(h/2\pi i)\partial/\partial q_k$.[38a] Hence for any conjugate variables q_k and p_k[28]

$$[\boldsymbol{p}_k\boldsymbol{q}_k-\boldsymbol{q}_k\boldsymbol{p}_k](m,n)=\int \psi_n{}^*\left[q_k\left(-\frac{h}{2\pi i}\frac{\partial}{\partial q_k}\right)+\frac{h}{2\pi i}\frac{\partial}{\partial q_k}q_k\right]\psi_m d\tau$$

$$=\int \psi_n{}^*\frac{h}{2\pi i}\psi_m d\tau=\frac{h}{2\pi i}\delta_{mn}. \tag{85}$$

This is equivalent to the matrix equation

$$\boldsymbol{p}_k\boldsymbol{q}_k-\boldsymbol{q}_k\boldsymbol{p}_k=(h/2\pi i)\boldsymbol{I}. \tag{86}$$

By a similar argument we can prove that

$$\boldsymbol{p}_k\boldsymbol{q}_l-\boldsymbol{q}_l\boldsymbol{p}_k=0, \quad l\neq k \tag{87}$$

and

$$\boldsymbol{p}_k\boldsymbol{p}_l-\boldsymbol{p}_l\boldsymbol{p}_k=\boldsymbol{q}_k\boldsymbol{q}_l-\boldsymbol{q}_l\boldsymbol{q}_k=0. \tag{88}$$

In other words, the matrices for coordinates do not "commute" with the matrices for the conjugate momenta, though they do commute with the matrices for the other momenta. Also any momentum matrix commutes with any other, and any coordinate matrix with any other.

The fundamental relation of Eq. (86) between the matrices of the coordinates and of the conjugate momenta is the quantum mechanical equivalent of the Wilson-Sommerfeld quantum conditions of the Bohr theory[39]

[38a] Cf. pp. 36 and 42. The differential operator does not apply however when the coordinates assume discrete values as for instance in the use of a component of angular momentum as a matrix. Cf. also Sect. 7, 2 below.

[39] M. Born and P. Jordan, Zeits. f. Physik **34,** 858 (1925). Cf. also L. Brillouin, Journ. de Physique **7,** 135 (1926); reprinted in English in the volume "Selected Papers on Wave

$$\int p_k dq_k = nh \quad n=0,1,2,\cdots$$

The Eqs. (86), (87), (88) are called the "commutation rules." Dirac introduces the Poisson bracket symbol $[y, x]$ for the matrix function $(2\pi i/h)(\mathbf{xy}-\mathbf{yx})$, and writes the commutation rules in the form[40]

$$[\mathbf{q}_k, \mathbf{p}_l] = \mathbf{I}\delta_{kl}, \quad [\mathbf{q}_k, \mathbf{q}_l] = [\mathbf{p}_k, \mathbf{p}_l] = 0. \tag{89}$$

Another fundamental characteristic of the matrices of the p's and q's in Cartesian coordinates is their *Hermitian character* (Cf. Part I, p. 201) which is a consequence of the fact that the corresponding operators are *self-adjoint*. An operator O_α is by definition self-adjoint with respect to a coordinate system $q_1, \ldots, q_{3n}$, if

$$vO_\alpha u - uO_\alpha{}^* v = \sum_{k=1}^{3n} \frac{\partial f_k}{\partial q_k}, \tag{90}$$

where u and v are arbitrary continuous functions of the coordinates $q_1, \cdots, q_{3n}$, and $f_1, \cdots, f_{3n}$ are functions of the same coordinates, any or all of which may vanish in special cases.[40a] (The right hand member of the above equation is simply the divergence of the $3n$-dimensional vector F whose components are $f_1, \cdots, f_{3n}$). If we now take the volume integral of both sides of Eq. (90) over a hyper-sphere G surrounded by a surface S, we can convert the right hand member into a surface integral over S. If the functions u and v vanish rapidly at infinity, the vector F approaches zero with corresponding rapidity, and the surface integral vanishes in the limit as the radius of the sphere increases indefinitely. Under these circumstances

$$\int_\infty (vO_\alpha u - uO_\alpha{}^* v) dq_1 \cdots dq_{3n} = 0. \tag{91}$$

If u and v are identified respectively with the wave function ψ and its conjugate ψ^* for a suitable problem in wave mechanics we have

$$\bar{\alpha} = \int_\infty \psi^* O_\alpha \psi d\tau = \int_\infty \psi O_\alpha{}^* \psi^* d\tau = \bar{\alpha}^*. \tag{92}$$

Thus the mean value of a quantity α with a self-adjoint operator O_α is necessarily real provided that the ψ function for the problem in hand vanishes

---

Mechanics" by de Broglie and Brillouin, London, 1928. It may be remarked incidentally that Eqs. (86) can only be satisfied if q_k and p_k are infinite matrices. Cf. e.g., H. Turnbull, Mathematical Gazette **14,** 12 (1928).

[40] Dirac's notation is dictated by the fact that in the limiting case of large quantum numbers the matrix quantity $(2\pi i/h)(\mathbf{xy}-\mathbf{yx})$ corresponds to the classical Poisson bracket $[y,x]$ just as the matrix $\mathbf{p}$ corresponds to the classical value of the momentum p. Cf. P. A. M. Dirac, Proc. Roy. Soc. **A109,** 642 (1925).

[40a] An alternative definition of the self-adjoint operator equivalent to Eq. (90) is given by Jordan [Zeits. f. Physik **40,** 809 (1927)]. Eq. (91) can also be used to define the self-adjoint property.

with sufficient rapidity at infinity.[41] The latter restriction is of greater importance from the mathematical than from the physical point of view.

We can also identify u and v^* respectively with two characteristic functions ψ_n and ψ_m belonging to the discrete spectrum of the problem in hand. Then

$$\int_\infty (\psi_m{}^* O_\alpha \psi_n - \psi_n O_\alpha{}^* \psi_m{}^*) d\tau = 0$$

or

$$\alpha(n, m/t) = \alpha^*(m, n/t), \tag{93}$$

which is the condition that the matrix of α shall be Hermitian.

In classical mechanics and electrodynamics complex values of the dynamical variables have no meaning, and in quantum mechanics we may reasonably exclude complex values of measurable quantities as meaningless. This is equivalent to requiring that the operators correlated with all measurable quantities shall be self-adjoint, and that the corresponding matrices shall all be Hermitian. We have already proved that these conditions are satisfied by the energy operator and energy matrix. They are also satisfied for all functions of the coordinates only, and for the components of linear and angular momentum, as the reader can prove for himself without difficulty. Hereafter we shall assume that only self-adjoint operators and Hermitian matrices are to be used in building up the theory.

We have now proved that in order to solve the characteristic value problem for a given Hamiltonian operator H we can adopt as a starting point any complete orthogonal function system spread out over the whole of the configuration space for the problem, develop from it matrices for the p's and q's, compound them to form a matrix for H, and then apply a transformation of the *canonical type*[42]

$$T_c : \left\{ \begin{matrix} \mathbf{p}_k{}^0 \rightarrow \mathbf{p}_k = \mathbf{S}\mathbf{p}_k{}^0\mathbf{S}^{-1} \\ \mathbf{q}_k{}^0 \rightarrow \mathbf{q}_k = \mathbf{S}\mathbf{q}_k{}^0\mathbf{S}^{-1} \\ \mathbf{H}^0 \rightarrow \mathbf{H} = \mathbf{S}\mathbf{H}^0\mathbf{S}^{-1} \end{matrix} \right\} \quad \mathbf{S}^{-1} = \tilde{\mathbf{S}}^* \tag{94}$$

which reduces the Hamiltonian matrix to diagonal form. The initial matrices $(\mathbf{p}_k{}^0, \mathbf{q}_k{}^0)$ must be Hermitian and must satisfy the commutation rules. Without appreciable loss of generality we may assume that they have the standard form

$$p^0(m, n/t) = p^0(m, n) e^{2\pi i (E_m{}^0 - E_n{}^0) t/h}. \tag{95}$$

---

[41] It seems probable that the existence of the integral which defines $\bar{\alpha}$ is a sufficient practical criterion for the validity of Eq. (92).

[42] The term "*canonical*" is applied to this type of transformation because it preserves the commutation rules and hence the quantum mechanical form of Hamilton's equations of motion (to be derived below). The connection between this form of canonical transformation and the form used in the classical Hamiltonian theory has been traced by Jordan. Cf. P. Jordan, Zeits. f. Physik **38**, 513 (1926).

For the rest, it is clear that a wide variety of forms must be possible in any given case, and one surmises that any choice consistent with the above restrictions is permissible. In other words, it would not be unreasonable to assume—and in fact it has been commonly assumed—that the problem of finding a set of Hermitian matrices for the p's and q's which obey the commutation rules and reduce $\boldsymbol{H}(\boldsymbol{p}, \boldsymbol{q})$ to diagonal form has no other solutions than those obtainable from the Schrödinger equation. In the case of a non-degenerate system this would mean that the solutions of the matrix problem are unique.

In practice the solution of problems in quantum mechanics by starting from an arbitrary initial set of Hermitian matrices which satisfy the commutation rules, and applying to them a canonical transformation which reduces $\boldsymbol{H}(\boldsymbol{p}, \boldsymbol{q})$ to diagonal form is confined to cases in which the Hamiltonian function can be approximated by that of a slightly different problem which has been exactly solved and whose matrices form the starting point of a genuine perturbation method calculation.

3. THE EQUATIONS OF MOTION IN THE MATRIX THEORY

The original formulation of the matrix mechanics of Born, Heisenberg, and Jordan was based on an attempt to use, so far as possible, the formal machinery of the classical dynamics, but with a reinterpretation of the nature and the meaning of the symbols introduced. They *assumed* that the coordinates and momenta of every dynamical system were to be represented by matrices, that the Hamiltonian function was a function of the coordinate and momentum matrices, and that the canonical equations of Hamilton were applicable to these matrices. On the basis of these assumptions they were able to prove that the energy matrix $\boldsymbol{H}(\boldsymbol{p}, \boldsymbol{q})$ was constant in time (law of the conservation of energy),[43,44] which they showed to be equivalent to the statement that the energy matrix was of diagonal form if the system was non-degenerate. Otherwise they found it necessary to *assume* explicitly the diagonal form of $\boldsymbol{H}$.

According to our present point of view, the matrices of p, q, and H can have a great variety of equally legitimate forms since the ψ functions can be described in terms of many different orthogonal function systems. The solution of the characteristic value problem is obtained, however, by reducing the matrices to the form which corresponds to an expansion in terms of a complete orthogonal system of characteristic functions. We may call this the normal matrix form. By definition $\boldsymbol{H}(\boldsymbol{p}, \boldsymbol{q})$ is constant in time and diagonal if the matrices of the p's and q's are in normal form. We may now reverse the argument of Born and Heisenberg to prove that normal matrices satisfy equations formally identical with the Hamiltonian equations of classical theory. As a first step we must define the operation of differentiation of a matrix function.

---

[43] Cf. M. Born and P. Jordan, Zeits. f. Physik **34,** 858 (1925).

[44] M. Born, W. Heisenberg, and P. Jordan, Zeits. f. Physik **35,** 557 (1926).

Born, Heisenberg, and Jordan give two definitions of the derivative of a matrix function with respect to a matrix argument.[44] We shall use here what they call "derivatives of the first kind," which are defined by the equation

$$\frac{df}{dx}=\lim_{a\to 0}\frac{f(x+\boldsymbol{a})-f(x)}{a} \tag{96}$$

where $\boldsymbol{a}$ is the diagonal matrix obtained by multiplying the unit matrix $\boldsymbol{I}$ by an ordinary number a. It is readily proved that on the basis of this definition all the ordinary rules of differentiation apply to the differentiation of matrix functions except that one must have due regard for the order in dealing with product functions. Thus

$$\frac{d}{dx}(FG)=F\frac{dG}{dx}+\frac{dF}{dx}G,$$

where the order of the products cannot be altered in either member unless it is reversed throughout the equation.

It follows from this definition that if $\boldsymbol{f}(\boldsymbol{p}, \boldsymbol{q})$ is any function of a set of matrices $(\boldsymbol{q}_k, \boldsymbol{p}_k)$ which satisfy the commutation rules[45]

$$\begin{aligned}\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{p}_k}&=(2\pi i/h)(\boldsymbol{f}\boldsymbol{q}_k-\boldsymbol{q}_k\boldsymbol{f})=[\boldsymbol{q}_k,\boldsymbol{f}]\\ \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{q}_k}&=(2\pi i/h)(\boldsymbol{p}_k\boldsymbol{f}-\boldsymbol{f}\boldsymbol{p}_k)=[\boldsymbol{f},\boldsymbol{p}_k].\end{aligned} \tag{97}$$

Differentiation of a matrix $\boldsymbol{f}$ with respect to an ordinary number a is defined by the rule

$$\left(\frac{df}{da}\right)(m,n)=\frac{d}{da}f(m,n). \tag{98}$$

Applying the first of equations (97) to the Hamiltonian matrix $\boldsymbol{H}(\boldsymbol{p}_k, \boldsymbol{q}_k)$ we have

$$\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{p}_k}=\frac{2\pi i}{h}(\boldsymbol{H}\boldsymbol{q}_k-\boldsymbol{q}_k\boldsymbol{H}). \tag{99}$$

If the $\boldsymbol{p}$'s and $\boldsymbol{q}$'s are normal matrices, $\boldsymbol{H}$ is a diagonal matrix whose nonvanishing elements are the energy levels E_1, E_2, $\cdots$. Hence

$$\left(\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{p}_k}\right)(m,n)=\sum_j \frac{2\pi i}{h}\{E_m\delta_{mj}q_k(j,n/t)-q_k(m,j/t)E_n\delta_{nj}\}$$

---

[45] The proof of this theorem rests on the observation that it is true if we identify $\boldsymbol{f}$ with any one of the basic matrices $\boldsymbol{q}_1,\boldsymbol{q}_2,\cdots,\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots$. It is easy to show that if the theorem is true for any pair of functions $\boldsymbol{F}$ and $\boldsymbol{G}$ it is true for their sum and for their product. If we restrict ourselves to matrix functions which can be expressed as the sums of products of integral powers of the $\boldsymbol{p}$'s and $\boldsymbol{q}$'s, the rest follows by induction.

$$= \frac{2\pi i}{h}(E_m - E_n)q_k(m, n/t) \tag{100}$$

$$= \frac{d}{dt}q_k(m, n/t).$$

It follows from Eq. (98) that

$$\frac{dq_k}{dt} = \frac{\partial H}{\partial p_k}. \tag{101}$$

The second Hamiltonian equation

$$\frac{dp_k}{dt} = -\frac{\partial H}{\partial q_k} \tag{102}$$

is proved in exactly the same way.

Once Eqs. (101) and (102) have been established for normal matrices there is no difficulty in showing that they remain valid for any set of matrices obtained from normal ones by a canonical transformation with a constant matrix S.[46]

The equations of motion (101) and (102) were of considerable importance in the initial development of the matrix theory, since their direct integration in a few special cases gave clean solutions of simple problems which could then be used as the basis for an attack on more difficult problems by perturbation methods.[47]

4. Example: The Linear Oscillator

Consider the case of the linear oscillator whose Hamiltonian function has the form

$$H(p,q) = p^2/2\mu + kq^2/2. \tag{103}$$

The Hamiltonian equations of motion reduce to

$$\dot{q} = p/\mu\ ; \quad \dot{p} = -kq. \tag{104}$$

---

[46] The proof rests on the theorem that if

$$p_k = Sp_k^0S^{-1}; \quad q_k = Sq_k^0S^{-1}$$

then

$$f(p_k, q_k) = Sf(p_k^0, q_k^0)S^{-1}$$

and on the additional fact that if S is a constant matrix

$$dp_k/dt = S(dp_k^0/dt)S^{-1}.$$

The transformed matrices obtained from normal matrices by means of a constant canonical transformation do not have the form of Eq. (95).

[47] For examples of papers dealing with problems on the matrix (or q-number) theory which use the equations of motion, the following references can be consulted. In the papers of Mensing and Oppenheimer a certain amount of perturbation theory is used. P. A. M. Dirac, Proc. Roy. Soc. **A110**, 561 (1926), W. Pauli, Jr., Zeits. f. Physik **36**, 336 (1926), L. Mensing, Zeits. f. Physik **36**, 814 (1926), J. R. Oppenheimer, Proc. Cam. Phil. Soc. **23**, 327 (1926), I. Tamm, Zeits, f. Physik **37**, 685 (1926), D. M. Dennison, Phys. Rev. **28**, 318 (1926).

Eliminating p and introducing the classical frequency $\nu_0[=(1/2\pi)\ (k/\mu)^{1/2}]$ we have

$$\ddot{q}=\{(2\pi i\nu_{mn})^2 q(m,n/t)\}=-(2\pi\nu_0)^2 q$$

or

$$[\nu_{mn}{}^2-\nu_0{}^2]q(m,n/t)=0. \tag{105}$$

This equation shows that $q(m, n/t)$ must vanish unless

$$\nu_{mn}=\frac{E_m-E_n}{h}=\pm\nu_0. \tag{106}$$

It follows that the energy levels, or diagonal elements of the H matrix, are arranged in one or more equidistant series, and that there are just two non-vanishing elements in each row and column. From the Hermitian character of the matrices we know, of course, that the non-vanishing elements are symmetrically arranged with respect to the diagonal. Let us now assume that there is only one set of equidistant energy values as in the Bohr theory. The order of the elements of a diagonal matrix is arbitrary, so that we can assume without loss of generality that it is the same as the order of the energies. The non-vanishing elements of the matrices for q and p are then arranged in two lines parallel and adjacent to the principal diagonal. It follows at once that all non-diagonal elements of $[q,p]$ vanish as required by the commutation rule. The diagonal elements of $[q,p]$ must have the common value unity. In other words

$$\frac{h}{8\pi^2\mu}=-\sum_k \nu_{nk}\,|\,q(n,k/t)\,|^2=\nu_0[\,|\,q(n,n+1/t)\,|^2-|\,q(n;n-1/t)\,|^2]. \tag{107}$$

Hence the squares of the absolute values of the matrix elements $q(n, n+1/t)$ form an arithmetical progression with the common difference $h/8\pi^2\mu\nu_0$. These terms are essentially positive so that n must have a minimum value which we shall set equal to zero. For $n=0$, Eq. (107) then reduces to

$$|\,q(0,1/t)\,|^2=h/8\pi^2\mu\nu_0$$

and we have finally

$$|\,q(n,n+1/t)\,|^2=(n+1)h/8\pi^2\mu\nu_0,\quad n=0,1,2,\cdots. \tag{108}$$

For the matrix element $q(n+1, n/t)$ this yields

$$q(n+1,n/t)=q^*(n,n+1/t)=\left[\frac{(n+1)h}{8\pi^2\mu\nu_0}\right]^{1/2}e^{i(2\pi\nu_0 t+\phi_n)}, \tag{109}$$

where ϕ_n is an arbitrary phase constant. The factors $e^{i\phi_n}$ may be lumped with the complex amplitudes c_n in determining mean values in the case of a particular ensemble of linear oscillators. [Cf. for example, Eq. (38).] In the simple problem under consideration the electric moment is proportional to the displacement q so that by combining Eq. (108) with Eqs.

(28) and (35) we obtain at once the transition probabilities with which to compute the intensity of the emitted radiation.

Having evaluated the $\boldsymbol{q}$ matrix, we can next determine $\boldsymbol{p}$ by Eq. (104) and $\boldsymbol{H}$ by Eq. (103). For the latter we find

$$H(n,n/t)=E_n=(n+\tfrac{1}{2})h\nu_0. \tag{110}$$

These conclusions are in agreement with the pure wave theory of Sect. 2, 5 of Part I, and we thus see that the two theories lead to equivalent results.[48]

5. Hamilton's Equation for Wave Packets

Due to the general relation between matrices and mean values, viz.,

$$\bar{\alpha} = \int \psi^* O_\alpha \psi d\tau = \sum_{n,m} c_n^* c_m \alpha(m,n/t)$$

it follows that all matrix equations such as the Hamiltonian Eqs. (101) and (102) are equally valid if we replace each member by the corresponding mean value for some definite ensemble of systems. Thus we have

$$\overline{\frac{d\boldsymbol{q}_k}{dt}}=\frac{d}{dt}\overline{\boldsymbol{q}_k}=\overline{\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{p}_k}}\,; \quad \overline{\frac{d\boldsymbol{p}_k}{dt}}=\frac{d}{dt}\overline{\boldsymbol{p}_k}=-\overline{\frac{\partial \boldsymbol{H}}{\partial \boldsymbol{q}_k}}\,. \tag{111}$$

If, in particular, we consider an ensemble represented by a wave packet in which each of the coordinates and momenta is well defined, the mean value of any function of the coordinates and momenta will closely approximate the same function of the mean values of the individual coordinates and momenta. The classical mechanics is the mechanics of just such wave packets and the classical coordinates and momenta are evidently to be identified with the mean values of the actual coordinates and momenta. Thus in the limiting case where the coordinates and momenta are very well defined, Eqs. (111) go over into the classical Hamiltonian equations. The proof is independent of the coordinate system employed, and hence the theorem shows that the definition of the momentum operator in generalized coordinates given on p. 29 is justified by agreement with the classical definition of canonical pairs of variables in the limiting case of a sharply defined wave packet. On the other hand the proof does depend on the assumption that the Hamiltonian function has only a discrete spectrum of characteristic wave functions. As previously explained, one can approximate the solution of any problem in quantum mechanics involving a continuous spectrum by means of a modified

---

[48] Cf. in this connection L. Brillouin, Journ. de Physique **7**, 135 (1926). Brillouin's conclusion that the matrix theory leads to different results if n is allowed to assume negative values seems to the authors to be due to an algebraic error in his final energy formula. On checking the steps we find that the energy and coordinate matrices for his alternative solution are given by (110) and (109) with $(n+\frac{1}{2})$ and $(n+1)$ replaced by $(-n+\frac{1}{2})$ and $(-n+1)$ respectively, with $n=0, -1, -2, \cdots$. This amounts simply to a different system of numbering for the rows and columns and leads to no new results.

problem having only a discrete spectrum. Thus the continuous spectrum problem is a limiting case to which the Hamiltonian equations must also apply. A more direct treatment of the continuous spectrum case is given, however, in Sect. 7, 2.

6. Matrices and Bohr's Correspondence Principle

The matrix theory of Heisenberg, Born, and Jordan was developed in the beginning from Bohr's famous correspondence principle for conditionally periodic motions. This principle postulates the asymptotic identity of the frequencies and intensities of spectrum lines as observed and as computed from classical theory in the limiting case of high quantum numbers. The quantum mechanics is, of course, in harmony with Bohr's principle, which we have already used in a somewhat modified form in setting up the Schrödinger equation in Part I. As the process of tracing out the assumed correspondence is of practical value as well as historical interest, we briefly describe it here.

A conditionally periodic motion is a motion which from the classical standpoint should emit a line spectrum. In other words, it is a motion of such a character that every function of the coordinates and momenta, including the all-important electric moment, can be analysed into a multiple Fourier series. Thus if $P^{(x)}$ denotes the x component of the electric moment

$$P^{(x)}{}_{\text{classical}} = \sum_{\tau_1,\cdots,\tau_f=-\infty}^{+\infty} P^{(x)}_{\tau_1,\cdots,\tau_f} e^{2\pi i[\tau_1\nu_1+\cdots+\tau_f\nu_f]t} \tag{112}$$

where the τ's are positive or negative integers and the ν's are a set of fundamental frequencies characteristic of the motion. $P^{(x)}{}_{\tau_1,\cdots,\tau_f}$, is the complex conjugate of $P^{(x)}{}_{-\tau_1,\cdots-\tau_f}$ so that $P^{(x)}$ is real. A similar expansion holds for each individual coordinate and momentum component. The simplest type of conditionally periodic motion, and the one of the greatest importance in the Bohr theory, is that which occurs when by a suitable choice of coordinates the action function W can be resolved into the sum of a series of terms W_1, W_2, etc., each of which depends on a single coordinate. The variables are then said to be *separable*, and the Wilson-Sommerfeld quantum conditions can be applied to each of the coordinates, or what amounts to the same thing, to each of the terms in W. Under these circumstances the frequencies ν_1, ν_2, ν_3, $\cdots$ in Eq. (112) are the frequencies associated with the various coordinates used to define the system. Furthermore, each "stationary state," as defined by the quantum conditions, is associated with a set of quantum numbers n_1, n_2, $\cdots$ correlated in a one to one manner with the coordinates and with the frequencies.

Bohr has shown that according to the fundamental assumptions of his theory the frequency emitted in any jump, say

$$E(n_1', n_2', \cdots, n_f') \rightarrow E(n_1'', n_2'', \cdots, n_f''), \tag{113}$$

is equal to the mean value of the frequency $|(\tau_1\nu_1+\tau_2\nu_2+\cdots+\tau_f\nu_f)|$ of a corresponding pair of terms in Eq. (112) evaluated by taking a suitable average over a continuous series of classically permissible, but unquantized, states between the initial and the final stationary states.[49] One of the corresponding terms is that for which $\tau_1 = n_1' - n_1''$, $\tau_2 = n_2' - n_2''$, etc., while the other is the conjugate term. In the limiting case of very high quantum numbers the classical frequencies ν_1, ν_2, ν_3, etc. have the same values for the initial and final states, and hence the quantum frequencies become actually identical with the corresponding classical combination frequencies.

In addition to this limiting agreement in the frequencies, which can be deduced from the quantum conditions and from his other quantum postulates, Bohr assumed a limiting agreement between the intensities of the terms in (112) and the intensities of the corresponding spectrum lines, and then used this assumption as a tool for the approximate computation of quantum intensities.

As a first step toward exhibiting the correspondence principle in terms of the quantum mechanics, we observe that since the classical action function W obeys a differential equation which is a limiting case of the differential equation for[50] $-(h/2\pi i)\log\psi$, it follows that the process of resolving ψ into the product of several factors is equivalent to resolving W into a sum of corresponding terms. Thus if the variables in the Schrödinger equation for ψ are separable in the sense of Section 2,6, Part I, it follows that the variables in the corresponding classical action function will be separable in the sense just defined.

Let us assume that this condition is fulfilled. Then in the wave mechanics there will be a differential equation for every coordinate, which can be thrown into Sturm-Liouville form, and each energy level will be characterized by a set of quantum numbers $n_1, n_2, \cdots, n_f$, one for each coordinate. We expect that just as in the Bohr theory the frequency and intensity of the radiation produced by the transition (113) will be asymptotically equal to the frequency and intensity of the corresponding term in Eq. (112).

Since the classical mechanics may be regarded as a limiting case of the quantum mechanics of wave packets when the coordinates and momenta of the latter are defined with precision, we can prove the point by writing out the wave mechanical equivalent of Eq. (112) for a well defined packet and actually passing to the limit. The necessary equation is evidently

$$\overline{P^{(x)}} = \sum_{n_1',\cdots,n_f'} \sum_{n_1'',\cdots,n_f''} c_{n_1',\cdots,n'_f} \overset{*}{c}_{n_1'',\cdots,n_f''} P^{(x)}(n_1', \cdots, n_f'; n_1'', \cdots, n_f'') e^{2\pi i \nu(n_1',\cdots,n_f';n_1'',\cdots,n_f'')t} \tag{114}$$

This is equivalent to Eq. (38), but the notation has been amplified to bring out the explicit relation of each state to the corresponding quantum numbers. With the aid of suitable approximations the double sum in the above equation must collapse into a single sum like the right hand member of Eq. (112).

---

[49] Cf. J. H. Van Vleck, "Quantum Principles and Line Spectra" (Eq. (15a)) for an explicit statement of this theorem, which is also implicitly contained in Bohr's Danish Academy papers.

[50] Cf. Sect. 5, 2.

To accomplish the reduction the following assumptions are necessary:[51]

a) The characteristic functions in the packet which have amplitudes differing appreciably from zero all have very short wave-lengths. In other words, we are in the region of high quantum numbers. Without such functions it is impossible to fix coordinates and momenta simultaneously with the desired precision. (Cf. Part I, Section 1, 8 and 9).

b) The number of characteristic functions in the packet having amplitudes differing appreciably from zero must be very large if the mutual interference between these functions is to make the packet very small.

c) The range of energy values with appreciable amplitude coefficients must be so narrow that in this range the energy can be treated as a linear function of all of the quantum numbers.

d) The matrix elements of $P^{(x)}$ corresponding to transitions between the different characteristic functions of the packet must depend primarily on the quantum number differences between the different states involved, being practically independent of the absolute values of the initial quantum numbers through the range of the packet. This hypothesis is easily verified in special cases, and the work of Eckart and of Van Vleck[52] on the asymptotic relation between the matrix elements $P^{(x)}(n_1', \cdots, n_f'; n_1'', \cdots, n_f'')$ and the Fourier coefficients in Eq. (112) affords a sufficiently general proof.

e) Finally we must suppose that the amplitudes $c_{n_1, \cdots, n_f}$ which define the packet vary quite slowly in absolute value with the quantum numbers and have phases which like the energy values are linear functions of the quantum numbers. Then if the corresponding indices n_1', n_1''; n_2', n_2''; etc. differ by small integral values

$$c_{n_1', \cdots, n_f'} \cong c_{n_1'', \cdots, n_f''} e^{i[(n_1'-n_1'')\epsilon_1+(n_2'-n_2'')\epsilon_2+\cdots]}.$$

We shall undertake no careful verification of the above assumptions but shall content ourselves with exhibiting the relation between matrix elements and classical Fourier components to which they lead. For a more complete treatment of the problem the reader is referred to the paper by Van Vleck cited in footnote 52.

Let

$$\tau_k = n_k' - n_k''; \quad k = 1, 2, 3, \cdots, f.$$

---

[51] The writers are indebted to Professor H. A. Kramers for assistance in setting up the following scheme of reduction.

[52] C. Eckart, Proc. Nat. Acad. Sci. **12**, 684 (1926). J. H. Van Vleck, Proc. Nat. Acad. Sci. **14**, 178 (1928). Strictly speaking, there is an arbitrariness in the values of the matrix elements due to the arbitrariness in the phases of the characteristic functions. Thus if we multiply each of the characteristic functions of the linear oscillator $\psi_1, \psi_2, \cdots, \psi_n, \cdots$ by a corresponding phase factor $e^{i\epsilon_n}$ where ϵ_n is arbitrary, we get a new set of normalized wave functions which is just as good as the first set. This transformation of the characteristic functions multiplies each of the matrix elements in turn by a phase factor of the form $e^{i(\epsilon_n-\epsilon_m)}$. Consequently the hypothesis *d*) holds only for a particular choice of the phase constants of the basic characteristic functions.

Then in view of the hypotheses stated above

$$P^{(x)}(n_1', \cdots, n_f'; n_1'', \cdots, n_f'')e^{2\pi i\nu(n';n'')} \rightarrow P^{(x)}(\tau_1, \cdots, \tau_f)e^{(2\pi i/h)(\sum_k \tau_k \partial E/\partial n_k)t}.$$

Thus Eq. (114) reduces to

$$\overline{P^{(x)}} = \sum_{\tau_1, \cdots, \tau_f} P^{(x)}(\tau_1, \cdots, \tau_f)e^{(2\pi i/h)(\sum_k \tau_k \partial E/\partial n_k)t} \sum_{n_1', \cdots, n_f'} c_{n_1', \cdots, n_f'} \overset{*}{c}_{n_1'-\tau_1, \cdots, n_f'-\tau_f}.$$

Because of the assumed normalization of the wave packet and assumption **e**) this reduces to the form

$$\overline{P^{(x)}} = \sum_{\tau_1, \cdots, \tau_f} P^{(x)}(\tau_1, \cdots, \tau_f)e^{i(\sum_k \tau_k \epsilon_k)}e^{(2\pi i/h)(\sum_k \tau_k \partial E/\partial n_k)t} \tag{115}$$

which is similar to that of Eq. (112). Now we have already shown in Part I, Section 2, 4 and 5 that there is an asymptotic agreement between the values of the energy differences predicted by the wave mechanics and by the Bohr theory in the case of a linear oscillator. The same agreement is found in the case of the hydrogen atom problem, and may be proved for the general case of any conditionally periodic motion from the work of Van Vleck.[52] But according to the Bohr theory the values of the classical frequencies $\nu_1, \nu_2, \cdots, \nu_k, \cdots$ are asymptotically equal to the corresponding coefficients $h^{-1}(\partial E/\partial n_k)$ in the exponents of Eq. (115). Thus the frequency factors of the terms in Eq. (115) are the same as the frequency factors in the corresponding terms of Eq. (112), provided only that the wave packet of Eq. (115) is adjusted to the initial conditions characteristic of the classical motion to which Eq. (112) refers. The agreement between the matrix coefficients

$$P^{(x)}(\tau_1, \cdots, \tau_f)e^{i(\sum_k \tau_k \epsilon_k)} = e^{i(\sum_k \tau_k \epsilon_k)} \lim \left[P^{(x)}(n_1', \cdots, n_f'; n_1'-\tau_1, \cdots, n_f'-\tau_f)\right]$$

of the one series and the Fourier amplitudes

$$P^{(x)}{}_{\tau_1, \cdots, \tau_f}$$

of the other, can be proved directly by the method used by Van Vleck, or can be inferred from the fact that the motion of the center of gravity of the wave packet in configuration space is identical with the motion of the representative point of the classical system in the same space. (Cf. Hamilton's equations of motion, Section 6, 5).

SECTION 7. TRANSFORMATION THEORY

1. Operators and Canonical Transformation

In the so-called "q-number theory" of Dirac the same non-commutative algebraic machinery was used as in the matrix theory of Born, Heisenberg, and Jordan, but the symbols which the latter authors use for matrices were given a less specific interpretation. Dirac called these symbols "q-numbers" to distinguish them from ordinary numbers or "c-numbers" which obey the usual commutation rule for algebraic multiplication. These q-numbers were shown by him to be capable of representation by matrices in special cases. The matrix representation, however, is subject to limitations; e.g., in case angles, or "angle variables", are used as coordinates no corresponding matrices satisfying the commutation rules (89) exist. Also if problems involving continuous spectra are to be dealt with, one must use continuous matrices or matrices of partly continuous character, and convergence difficulties cause trouble.[53] These difficulties led Born and Wiener[54] to introduce operators instead of matrices for the discussion of aperiodic motions, and later Schrödinger,[55] Dirac,[56] and Eckart[57] identified the q-numbers with the operators for the corresponding classical quantities as defined in the Schrödinger theory and as used in the preceding sections of this paper (except for the difference in sign noted in the final foot-note of Part I). According to the conventions adopted in this paper the operators associated with the coordinate q and the conjugate momentum p are*

$$O_q = [q\times] = O_q^* \; ; \; O_p = -(h/2\pi i)\partial/\partial q = -O_p^*. \tag{116}$$

where $[q\times]$ denotes the operation of multiplying by q. The starred symbols represent the operators used by Schrödinger, Eckart, and Dirac. To avoid a confusing inversion of the usual order of multiplication it will be convenient in this section to use these latter operators and to adopt Eckart's symbolism for them. Thus Q is to replace the symbol $[q\times]$ and P is to represent the operator

$$P = O_p^* = (h/2\pi i)\partial/\partial q. \tag{116a}$$

If A and B are any two operators and if $f(q)$ is any function of the q's, then the operators C and D which satisfy the identical relations

$$Cf = Af + Bf \; ; \; Df = A(Bf),$$

are called the *sum* of the operators A and B, and the *product* of A and B respectively. We indicate these relations by the operator equations

$$C = A + B; \quad D = AB.$$

From the second of these equations it is obvious that the operators P and Q satisfy the same commutation rules as the matrices. Thus, since

$$\frac{h}{2\pi i}\left[\frac{\partial}{\partial q_k}q_k\psi - q_k\frac{\partial\psi}{\partial q_k}\right] = \frac{h}{2\pi i}\psi$$

---

[53] The continuous and mixed types of matrix are sufficiently defined in the fine print on p. 201 of Part I. The difficulties encountered in their use are briefly discussed by Born in his Massachusetts Institute of Technology lectures "Problems of Atomic Dynamics," Cambridge, 1926, at the end of Lecture 19, and by J. v. Neuman in his article "Mathematische Begrundung der Quantenmechanik," Göttingen Nachrichten (1927), p. 1.

[54] M. Born and N. Wiener, Zeits. f. Physik **36,** 174 (1926).

[55] E. Schrödinger, Ann. der Physik **79,** 734 (1926).

[56] P. A. M. Dirac, Proc. Roy. Soc. **A112,** 661 (1926).

[57] C. Eckart, Phys. Rev. **28,** 711 (1926).

* These are the operators to be applied to ψ_q. Cf. p. 44.

for any function $\psi(q)$, we have the identical operator equation

$$P_kQ_k-Q_kP_k=h/2\pi i. \tag{117}$$

The other operator products are commutable and we can summarize the commutation rules by the equations,

$$P_kQ_l-Q_lP_k=(h/2\pi i)\delta_{kl}\ ;\quad P_kP_l-P_lP_k=Q_kQ_l-Q_lQ_k=0. \tag{117a}$$

In developing the general transformation theory of quantum dynamics, Eq. (117) may be taken as the definition of canonical operators in all ordinary cases where the q's assume continuous ranges of values.[57a] The operators P, Q clearly obey the associative law of multiplication and the commutative, distributive, and associative laws of addition.

The reciprocal or inverse operators of Q and P are defined by

$$Q^{-1}=\left[\frac{1}{q}\times\right]\ ;\quad P^{-1}=\frac{2\pi i}{h}\int dq \tag{118}$$

in accordance with the rule

$$QQ^{-1}=I=[1\times]\quad PP^{-1}=P^{-1}P=I=[1\times]. \tag{119}$$

With the aid of these conventions we can manipulate the P's and Q's as we do matrices and can deal with any function of the P's and Q's involving only integral powers of the former. The derivative of an operator function $f(X)$ with respect to an argument operator X can be defined in accordance with Eq. (96) by the rule

$$\frac{d}{dX}f(X)=\lim_{A\to 0}\left(\frac{f(X+A)-f(X)}{A}\right) \tag{120}$$

where A is the operator $[a\times]$. When applied to operators P_k, Q_k, which obey the commutation rules, this definition is equivalent to the equations

$$\left.\begin{aligned}\frac{\partial f(P,Q)}{\partial P_k}&=\frac{2\pi i}{h}[fQ_k-Q_kf]=[Q_k,f]\\ \frac{\partial f(P,Q)}{\partial Q_k}&=\frac{2\pi i}{h}[P_kf-fP_k]=[f,P_k]\end{aligned}\right\}. \tag{121}$$

There is a difficulty, however, in connection with the time derivatives needed for the formulation of Hamilton's equations of motion, since the operators P and Q are not functions of the time operator. Furthermore the Hamiltonian operator $H(P, Q)$ is not a function of the time operator, so that so far as the writers can see, the Hamiltonian equations can be applied to the operators P, Q, $H(P, Q)$ only as definitions of the operators $\dot{P}$ and $\dot{Q}$.[58] This

---

[57a] In case any one of the coordinates q_k assumes only discrete values, e.g. when the energy is used as a coordinate, the operator $(h/2\pi i)\partial/\partial q_k$ has no meaning when applied to $\psi(q)$. Then Eq. (117) fails as do the corresponding matrix relations (89).

[58] Eckart assumes the Hamiltonian equations to be valid for suitable operators and then defines the momentum operators P_k and the time derivatives $\dot{P}_k$ and $\dot{Q}_k$ so as to reduce the equations of motion to identities. Dirac does not discuss the question at all.

means that the equations of motion no longer give any hold on the characteristic value problem.

The operator calculus is useful, however, in dealing with coordinate transformations, and to this problem we now address ourselves. Let the Hamiltonian function of a system be given in Cartesian coordinates, and let us suppose that we wish to make a change of variables. The simplest case is that of an ordinary point transformation in which the old coordinates $x_1, \cdots, x_n$ are functions of the new coordinates $q_1, \cdots, q_n$ and are independent of the momenta. Thus let

$$\begin{aligned} x_1 &= f_1(q_1, \cdots, q_n) \\ x_2 &= f_2(q_1, \cdots, q_n) \qquad (122) \\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{aligned}$$

The corresponding operator equations are simply

$$\begin{aligned} X_1 &= f_1(Q_1, \cdots, Q_n) \\ X_2 &= f_2(Q_1, \cdots, Q_n) \qquad (123) \\ &\cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \end{aligned}$$

Let the momenta conjugate to the original coordinates be designated by the symbols $p_1^0, \cdots, p_n^0$. The corresponding operators are

$$P_k^0 = (h/2\pi i)\partial/\partial x_k. \qquad (124)$$

Similarly we denote the momenta conjugate to the q's by $p_1, \cdots, p_n$ and we *define* the corresponding operators by

$$P_k = (h/2\pi i)\partial/\partial q_k. \qquad (125)$$

It is a straightforward job of differentiation to compute the P_k^0's in terms of the P_k's by means of Eqs. (122). The result will be of the form

$$P_k^0 = \sum_l g_{kl}(Q)P_l = \phi(Q_1, \cdots, Q_n; P_1, \cdots, P_n). \qquad (126)$$

By substitution from formulas (123) and (125) we can derive an expression for the original Hamiltonian operator in terms of the P's and Q's. We thus accomplish the change of variables

$$H^0(P^0, X) = H^0[\phi(P,Q), f(Q)] = H^\dagger(P,Q). \qquad (127)$$

The process just described is exactly equivalent to an ordinary change of variables in a partial differential equation, so that the solutions of the original equation[59]

$$H^0(P^0, Q^0)\psi^*(x) = -(h/2\pi i)\partial\psi^*(x)/\partial t \qquad (128)$$

when expressed in terms of the new coordinates become solutions of the new differential equation

$$H^\dagger(P,Q)\psi^{*\dagger}(q) = -(h/2\pi i)\partial\psi^{*\dagger}(q)/\partial t \qquad (129)$$

---

[59] This is the conjugate of Eq. (1).

with

$$\psi^{*\dagger}(q)=\psi^{*}[f(q)].$$

This change of variables will in some cases simplify the problem of determining the energy levels and wave functions but it does not reduce the problem in the new coordinates to a form equivalent to that which it had in the original coordinates. This is evident when we note that by changing the variables in the original normalization equation we obtain

$$\int_{\infty}\psi\psi^{*}dx_1\cdots dx_n=\int_{\infty}\psi^{\dagger}\psi^{*\dagger}\Delta dq_1\cdots dq_n=1 \tag{130}$$

where Δ denotes the Jacobian or functional determinant of the system of functions $f_1, \cdots, f_n$ with respect to the arguments $q_1, \cdots, q_n$; i.e.,

$$\Delta=\frac{\partial(f_1, \cdots, f_n)}{\partial(q_1, \cdots, q_n)}. \tag{131}$$

Thus the solutions of Eq. (128) still have a character which relates them to the original Cartesian coordinate system. On this account we may designate them with a subscript x in accordance with the notation of footnote 51, Part I. The probability that the system has a configuration corresponding to the volume element $dx_1, \cdots, dx_n$ of Cartesian coordinate space is $\psi_x\psi_x^{*}dx_1\cdots dx_n$, but the probability that it has a configuration corresponding to the volume element $dq_1\cdots dq_n$ of the new non-Cartesian coordinate space is $\psi_x^{\dagger}\psi_x^{*\dagger}\Delta dq_1\cdots dq_n$. The functions $\psi_x^{\dagger}$ have only conditional orthogonality in the q-space, and there is a further serious difficulty about computing mean values of the p's which will be indicated below (see especially footnote 60).

To avoid these difficulties we introduce a new wave-function

$$\psi_q=\psi_x^{\dagger}\Delta^{1/2}. \tag{132}$$

Since

$$\psi_x\psi_x^{*}dx_1\cdots dx_n=\psi_q\psi_q^{*}dq_1\cdots dq_n, \tag{133}$$

it is evident that ψ_q is the probability amplitude for the coordinate system $q_1, \cdots, q_n$. This equation also shows that the orthogonality of the ψ_x functions in Cartesian coordinate space carries with it the orthogonality of the ψ_q in the q-space. Reference to Sect. 3, Part I will show that we can compute the mean value of the momentum p_k from the formula[60]

$$\overline{p_k}=\int_{\infty}\psi_q^{*}\left(-\frac{h}{2\pi i}\frac{\partial}{\partial q_k}\right)\psi_q dq_1\cdots dq_n=\int_{\infty}\psi_q \overline{P}_k\psi_q^{*}dq_1\cdots dq_n. \tag{134}$$

---

[60] $\overline{P}_k$ is defined by an equation analogous to Eq. (93) of Part I. Its evaluation depends on the determination of the corresponding probability amplitude which is to be carried through by Fourier analysis as in Sect. 1, 8 of Part I. In order to carry out this analysis it is clearly essential that we start with ψ_q.

The operator P_k is by Eqs. (125) and (90) clearly self-adjoint with respect to the q coordinate system, and hence the values of $\overline{p_k}$ given by Eq. (134) are necessarily real.

To get the differential equation for this modified wave function we have only to make the substitution of $\psi_q\Delta^{-1/2}$ for $\psi_x^\dagger$ in Eq. (129). We find

$$\Delta^{1/2}H^\dagger(P,Q)\Delta^{-1/2}\psi_q^* = -\left(\frac{h}{2\pi i}\right)\frac{\partial\psi_q^*}{\partial t}\,. \tag{135}$$

In other words the correct Hamiltonian operator to be used in connection with the coordinate system $q_1, \cdots, q_n$ is

$$H_q(P,Q) = \Delta^{1/2}H^\dagger(P,Q)\Delta^{-1/2}. \tag{136}$$

The matrix elements of H_q with respect to the characteristic functions of Eq. (135) are the same as the matrix elements of H^0 with respect to the corresponding characteristic functions of Eq. (128), and we may conclude that H_q is self-adjoint with respect to the q coordinate system.[61]

Both steps in the derivation of H_q can be regarded as examples of *canonical transformations*. These transformations are re-statements of the initial problem in terms of new variables which, like the original ones, are grouped in canonical pairs. In the general case these transformations involve a mixing of the q's and p's like the contact transformations of classical mechanics. Thus we may define a transformation as the setting up of a new problem by means of a substitution of the form

$$Q^0 \rightarrow F(P,Q)\ ;\quad P^0 \rightarrow G(P,Q). \tag{137}$$

This converts $H^0(P^0, Q^0)$ into $H(P, Q)$. In case Eqs. (137) are of such character that

$$P_kQ_k - Q_kP_k = \frac{h}{2\pi i}$$

if

$$P_k^0Q_k^0 - Q_k^0P_k^0 = \frac{h}{2\pi i},$$

we say that the transformation is canonical. In this case we may interpret P_k as $(h/2\pi i)\partial/\partial q_k$ in accordance with Eq. (125), and then use $H(P, Q)$ to set up a Schrödinger equation for $\psi(q)$.

There are two standard forms of canonical transformation used in wave mechanics. One of these is described by the substitutions

$$\begin{aligned} Q^0 &\rightarrow T^{-1}(P,Q)\ Q\ T(P,Q) \\ P^0 &\rightarrow T^{-1}(P,Q)\ P\ T(P,Q) \end{aligned} \tag{138}$$

in which $T(P, Q)$ is an arbitrary operator function and $T^{-1}(P, Q)$ is its reciprocal. This substitution has the property of carrying over any operator

---

[61] In carrying through the computation of H_q from H^0 the formulas of Podolsky (Note 51, Part I) or of Schrödinger (Note 50, Part I) can be used to advantage.

function expressible as a power series in the P^0's and Q^0's into a corresponding function of the new variables according to the simple rule

$$F(P^0,Q^0)\rightarrow T^{-1}(P,Q)F(P,Q)T(P,Q). \tag{139}$$

The converse substitution is obviously

$$Q\rightarrow T(P^0,Q^0)Q^0T^{-1}(P^0,Q^0)\ ; \qquad P\rightarrow T(P^0,Q^0)P^0T^{-1}(P^0,Q^0). \tag{140}$$

From the rule of Eq. (139) it follows that the transformation is canonical, and it seems probable that it is in fact the most general type of canonical transformation.[62]

The transformed Hamiltonian function has the form

$$H(P,Q)=T^{-1}(P,Q)H^0(P,Q)T(P,Q) \tag{141}$$

and the new Schrödinger equation obtained from the transformation is

$$T^{-1}(P,Q)H^0(P,Q)T(P,Q)\psi^*(q)=-\frac{h}{2\pi i}\,\frac{\partial\psi^*(q)}{\partial t}\cdot \tag{142}$$

This is equivalent to

$$H^0(P,Q)T(P,Q)\psi^*(q)=-\frac{h}{2\pi i}\frac{\partial}{\partial t}[T(P,Q)\psi^*(q)] \tag{143}$$

if $T(P, Q)$ is independent of the time. If we set

$$\psi^{*0}(q)=T(P,Q)\psi^*(q)\ ;\ \psi^*(q)=T^{-1}(P,Q)\psi^{*0}(q) \tag{144}$$

Eq. (143) becomes identical with the initial differential equation

$$H^0(P^0,Q^0)\psi^{*0}(q^0)=-(h/2\pi i)\partial\psi^{*0}(q^0)/\partial t \tag{145}$$

except for the change in the symbols for the independent variables. It follows that for every solution of Eq. (145) there is a corresponding solution of Eq. (142) given by Eq. (144). It is assumed by London[63] that if $\psi^0(q^0)$ satisfies the boundary conditions appropriate to Eq. (145), the corresponding solution of Eq. (142) will satisfy the boundary conditions appropriate to the latter equation, but although this hypothesis is verified in various special cases it needs further examination. If we accept this assumption we may conclude that since the characteristic functions of Eq. (145) are derivable from those of (142) the latter equation is in fact a restatement of the original problem in different language.

It is worthy of note that by the use of a suitable exponential form for the operator $T(P, Q)$ it is possible to pass from Cartesian coordinates to spherical coordinates or to angle variables in which the region of definition of the transformed functions $\psi(q)$ is quite different from the region of definition of the initial functions $\psi^0(q^0)$.

---

[62] Jordan (Zeits. f. Physik **37**, 383 (1926)) has given a proof that the corresponding type of matrix transformation is the most general type of canonical matrix transformation.

[63] F. London, Zeits. f. Physik **40**, 193 (1926).

As a simple example of this type of canonical transformation we may cite the second of the two steps in the process of changing from Cartesian coordinates to generalized coordinates described on p. 44. Here we have $T=\Delta^{-1/2}$ and the substitution is

$$Q\rightarrow\Delta^{1/2}(Q')Q'\Delta^{-1/2}(Q')=Q'$$
$$P\rightarrow\Delta^{1/2}(Q')P'\Delta^{-1/2}(Q'),$$

which leaves the coordinate system unchanged.

A second type of canonical transformation is derived from classical analogy. Let $S(Q, P^0)$ denote an arbitrary function of the operators Q and P^0. Let the variables Q and P be defined in terms of the operators Q^0 and P^0 by means of the equations

$$Q^0=\frac{\partial S(Q,P^0)}{\partial P^0}\,;\quad P=\frac{\partial S(Q,P^0)}{\partial Q}\,\cdot \tag{146}$$

If we solve these equations for Q^0 and P^0 in terms of Q and P and substitute for each of the original variables its equivalent in terms of the new ones, we obtain a new Hamiltonian function. Furthermore the operators P, Q satisfy the commutation rules if they are satisfied by P^0 and Q^0 so that the transformation is canonical. The proof of this last proposition is that by a suitable use of exponentials it is possible to convert the transformation (146) into the form of the substitution (138).[64] By the same method one can show that the Schrödinger equation based on the new Hamiltonian has the same characteristic values as the original equation (128).

The first of the two steps required in changing over from Cartesian to generalized coordinates gives a convenient example of a case in which $S(Q, P^0)$ has a simple form. Using the notation of Eqs. (122) and (123) we set

$$S=\sum_l f_l(Q_1,\cdots,Q_n)P_l^0. \tag{147}$$

Then

$$X_k=\frac{\partial S}{\partial P_k^0}=f_k(Q_1,\cdots,Q_n) \tag{148}$$

in accordance with Eq. (123). Similarly

$$P_k=\frac{\partial S}{\partial Q_k}=\sum_l\frac{\partial f_l}{\partial Q_k}P_l^0=\frac{h}{2\pi i}\sum_l\frac{\partial x_l}{\partial q_k}\frac{\partial}{\partial x_l}\,\cdot \tag{149}$$

This is the inverse of Eq. (126). It shows that P_k is computed by a rule equivalent to the partial differentiation in the discussion on p. 43 and thus insures that the canonical transformation derived from Eq. (147) is an ordinary change of variables.

---

[64] Cf. P. Jordan, Zeits. f. Physik **38**, 513 (1926). A. Landé, Handbuch der Physik **XX**, Kap. 8, p. 434. P. A. M. Dirac, Proc. Roy. Soc. **A113**, 621 (1927).

In Article 5 of Sect. 6 we have shown that the classical Hamiltonian equations can be derived from the Hamiltonian equations of the matrix theory in the limiting case of a wave packet whose configuration is defined with precision. This proof requires extension due to the fact that it does not rigorously cover problems in which there is a continuous spectrum of E values but more especially because matrices for a coordinate q_k and its conjugate momentum p_k satisfying the commutation rules do not exist if either one is "quantized" so that it takes on only discrete values.[65] This situation arises when one uses the action and angle variables so important in classical dynamics and in the Bohr theory.

It would be natural to seek a means of bridging the above gap with the aid of operator equations of motion paralleling the matrix equations of motion. This method gets nowhere, however, on account of the impossibility of differentiating the P, Q operators with respect to the time. The following procedure, however, leads to the goal.

Let a system of generalized coordinates $q_1, \cdots, q_n$ be given and let ψ denote the probability amplitude ψ_q which goes with the system. The correct Hamiltonian function H is then self-adjoint with respect to the coordinates $q_1, \cdots, q_n$. Let θ denote an angular coordinate such that ψ is periodic in θ with period 2π. Consider now the time derivative of the mean value of $e^{i\theta}$

$$\begin{aligned}\frac{d}{dt}\int_\infty \psi^* e^{i\theta}\psi dq_1\cdots dq_n &= \int_\infty [\dot{\psi}^* e^{i\theta}\psi + \psi^* e^{i\theta}\dot{\psi}]d\tau \\ &= \frac{2\pi i}{h}\int_\infty\left[\psi^* e^{i\theta}H\left(-\frac{h}{2\pi i}\frac{\partial}{\partial q}, Q\right)\psi - \psi e^{i\theta}H\left(\frac{h}{2\pi i}\frac{\partial}{\partial q}, Q\right)\psi^*\right]d\tau. \qquad (150)\end{aligned}$$

Denoting the second of these Hamiltonian operators by $H(P, Q)$ and making use of the fact that it is self-adjoint we obtain

$$\frac{d}{dt}\overline{e^{i\theta}} = \frac{2\pi i}{h}\int_\infty \psi[H(P,Q)e^{i\theta} - e^{i\theta}H(P,Q)]\psi^* dq_1\cdots dq_n. \qquad (151)$$

This expression can be thrown into a more convenient form by means of the theorem[66]

---

[65] Cf. P. Jordan, Zeits. f. Physik **44**, 1 (1927). The reader will readily convince himself by actual multiplication that in the hydrogen atom problem the matrices for the azimuthal angle ϕ and its conjugate momentum M_z as defined by Eq. (68) do not satisfy the commutation rule of Eq.(85). The failure of what seems at first glance to be a clean-cut general theorem is to be referred to the restrictions on the preliminary theorem of p. 8.

[66] To prove the theorem note first that

$$P_\theta e^{i\theta} = \frac{h}{2\pi i}\frac{\partial}{\partial\theta}e^{i\theta} = e^{i\theta}\left[\frac{h}{2\pi i}\frac{\partial}{\partial\theta} + \frac{h}{2\pi}\right].$$

Obviously the theorem is true for $F(P_\theta, Q) = F(Q)$ and one can prove that it is true for the sum and product of two functions if it is true for the functions taken separately. The extension to sufficiently general cases follows by induction as in footnote 45. Cf. also P. A. M. Dirac, Proc. Roy. Soc. **A111**, 281 (1926).

$$F(P_\theta,Q)e^{i\theta}=e^{i\theta}F\left(P_\theta+\frac{h}{2\pi},Q\right). \tag{152}$$

Identifying $F(P_\theta, Q)$ with $H(P, Q)$ we obtain

$$\frac{d}{dt}\overline{e^{i\theta}}=i\int_\infty \psi e^{i\theta}\left[\frac{H\left(P_\theta+\frac{h}{2\pi},Q\right)-H(P_\theta,Q)}{h/2\pi}\right]\psi^* dq_1\cdots dq_n. \tag{153}$$

As a final step one can define the mean velocity associated with the angle θ by the relation

$$\overline{\dot{\theta}}\,\overline{e^{i\theta}}=\frac{d}{dt}\overline{e^{i\theta}}.$$

Then passing to the limit of a sharply defined wave packet, the right hand member of Eq. (153) becomes equal to the product of the partial derivative of the classical Hamiltonian into the function $\overline{e^{i\theta}}$. Hence

$$\overline{\dot{\theta}}=\frac{\partial H(\overline{p},\overline{q})}{\partial\overline{p_\theta}}. \tag{154}$$

This takes care of the difficult type of Hamiltonian equation for a coordinate of the angle-variable type. The other Hamiltonian equations can be derived in a similar manner by the transformation of the time derivatives of p_k and q_k.

2. The Dirac-Jordan Transformation Theory

Various gaps in the preceding discussion are bridged and loose ends tied together by the Dirac-Jordan transformation theory.[67,68] Unfortunately our exposition of their point of view must be very much abridged. In the main we shall follow the fuller and more general treatment of Jordan who has cast the essence of the matter in the form of a few postulates.

As the earlier sections of this paper testify, the quantum mechanics has much to say about "probability amplitudes." These are functions of coordinates, or systems of coordinates, whose squared absolute values determine the absolute or relative probability of the corresponding configurations. The probability depends, of course, on the circumstances and the circumstances which go with such functions are frequently labeled in the symbol for the function. For example, in the case of a one-dimensional oscillator, each of the characteristic functions $\psi_n(x, t)$ is a probability amplitude for the positional coordinate x when the energy E is known to have the value E_n. The "circumstances" are indicated by the subscript n. In this case the independent variable x has a continuous range of values, and accord-

---

[67] P. A. M. Dirac, Proc. Roy. Soc. **A113**, 621 (1927).

[68] P. Jordan, Zeits. f. Physik **40**, 809 (1927); **44**, 1 (1927). These two papers will be referred to as Jordan I and Jordan II respectively.

ingly $|\psi_n(x, t)|^2 dx$ measures the absolute or relative probability[69] that x has some value in the range from x to $x+dx$. In other cases the circumstances are not to be described so simply. For example, the solutions of the Schrödinger equation in the form which contains the time derivative (I, 38) depend on the initial conditions of the problem, which are fixed by the way in which each system of the ensemble is prepared. Any mode of preparation of the system may be regarded, however, as a specification of certain generalized coordinates,[70] and, following Jordan, one may formally attach subscripts for these coordinates to the symbols for the probability amplitudes. Thus in the case of the one-dimensional problem one can label the wave function $\psi(x, t)$ as $\psi_\beta(x, t)$ to indicate that the initial conditions of the problem are the equivalent of the specification of a suitably chosen "coordinate" β.

In case a coordinate β takes on a discrete series of values instead of a continuous range, there is a finite probability for each possible value of β. The probability of this value is then given directly by the product of the probability amplitude into its conjugate without multiplying in a factor $d\beta$. As an example, consider the probability amplitude for the energy in the case of a one-dimensional oscillator. This is the system of coefficients c_n, or $c(E_n)$, obtained when one analyses $\psi(x, t)$ into a series of characteristic functions $\psi_n(x, t)$. The probability of each discrete energy value is $c_n c_n^*$ while the probability of any energy range in the continuous spectrum is $c(E)c^*(E)dE$.[71]

As examples of probability amplitudes in problems involving more dimensions we cite the characteristic functions of the hydrogen atom problem obtained by separating the variables in spherical coordinates. These give the probabilities of the positional coordinates when the component of angular momentum along the z-axis, the square of the total angular momentum, and the total energy are given predetermined values. The functions $\psi(t,x, y,z)$ and $Q(t, p_x, p_y, p_z)$ met with in Sect. 1, 8 of Part I are probability amplitudes for position and momentum respectively for definite but unspecified modes of preparing the system.

Jordan adopts the symbol $\Phi_{\alpha,p}(\beta', q')$ for the probability amplitude of the coordinate system $\beta(=\beta_1, \cdots, \beta_f)$ with respect to the coordinate system $q(=q_1 \cdots, q_f)$; that is, $\Phi_{\alpha,p}(\beta', q')$ is the probability amplitude for the coordinates $\beta_1, \cdots, \beta_f$ when the coordinates $q_1, \cdots q_f$ are given specified values. Here the argument symbols are primed in order to distinguish the numerical values β_k', q_k' which the coordinates take on, from the symbols β_k, q_k for the coordinates themselves. The subscript α stands for the set of momenta conjugate to the coordinates β, while p stands for the momenta conjugate to the q's.[72]

---

[69] The former in case $\psi_n(x, t)$ is normalized.

[70] Here "generalized coordinate" means a function of the Cartesian coordinates and momenta.

[71] Cf. Eq. (I, 85) and the last paragraph on p. 214 of Part I.

[72] The momenta are not uniquely determined by the corresponding coordinates and the probability amplitude varies from one momentum system to another although, as Jordan proves, $|\Phi_{\alpha,p}(\beta', q')|^2$ depends only on the β's and q's. Cf. Jordan II, p. 20.

In normalizing the probability amplitudes we have to sum over all discrete values of the arguments and integrate over all continuous ranges. This process we indicate by the symbol $\overline{\sum}$ which degenerates into a simple summation sign or simple integration sign in special cases. As an example of the use of the symbolism we may take Eq. (I, 90) which gives the expansion of an arbitrary function of x, y, z in terms of the space factors of the hydrogen atom problem. The left hand member is a possible probability amplitude for the positional coordinates and the c's constitute the corresponding probability amplitude for the energies, so that the equation can be written in the form

$$\Phi_{p_x,p_q}(x,q)=\overline{\sum_E}\,\Phi_{p_x,p_E}(x,E)\,\Phi_{p_E,p_q}(E,q) \tag{155}$$

where x stands for the x, y, z coordinate system, q stands for an unspecified set of coordinates whose values fix the function $f(x, y, z)$, and E stands for the coordinate system E, l, m which defines the set of orthogonal functions employed in the expansion. (The momentum conjugate to the energy itself may be identified with the time, but in this case we use the symbol p_E since we have used E for a system of coordinates and not for one coordinate only.) The phrase "system of coordinates" has several different meanings in the quantum mechanics. The simplest type of coordinate system is a set of f parameters $q_1, \cdots, q_f$ which are independent functions of the Cartesian positional coordinates $x_1, \cdots, x_f$. Canonical transformations, however, lead to the introduction of coordinates which are independent functions of both the Cartesian coordinates and their momenta. The transformation theory requires a further modification of this conception in that it treats the values of $q_1, \cdots, q_f$ at each instant of time as forming different coordinate systems. Thus the determination of the coordinates at $t=t_1$ from those at $t=t_0$ is regarded as a coordinate transformation.

In carrying out the normalization we need in some cases an imaginary and magical function $\delta(x)$ first introduced by Dirac.[67] This is defined as a function which vanishes when x is not equal to 0 and whose integral over a range including the point $x=0$ is equal to unity. Strictly speaking, no such function exists, but we use the symbol as a convention to avoid circumlocution. The justification lies in the Dirichlet integral formula

$$\lim_{a\to\infty}\int_{-\infty}^{+\infty} dxF(x)\int_{-a}^{+a} dy\; e^{2\pi ixy}=F(0) \tag{156}$$

If we define $\delta_a(x)$ by the formula

$$\delta_a(x)=\int_{-a}^{+a} dy\; e^{2\pi ixy} \tag{157}$$

we see that when a is very large the properties of $\delta_a(x)$ approach those of the imaginary function $\delta(x)$ although we may not actually allow a to become infinite due to the non-convergence of the integral. The existence of the function $\delta_a(x)$ makes it possible to carry through many calculations formally just as if there were a genuine function $\delta(x)$, without making errors.[73]

For use in many-dimensional problems, Jordan has extended this conception by introducing a function $\Delta(\beta'-\beta'')$ of the pair of points $\beta_1', \beta_2', \cdots, \beta_f'$ and $\beta_1'', \cdots, \beta_f''$ in β-space. This function is defined to have the properties

$$\Delta(\beta'-\beta'')=0 \quad \text{unless} \quad \beta_1'=\beta_1''; \ \beta_2'=\beta_2''; \ \cdots\cdots \beta_f'=\beta_f''.$$

$$\overline{\sum_{\beta'}}\Delta(\beta'-\beta'')=1. \tag{158}$$

In one dimension it reduces to the Weierstrassian symbol[74] $\delta_{\beta',\beta''}$ if the argument β takes on discrete values, and to the δ function of the preceding paragraph with $x=\beta'-\beta''$ if β assumes continuous values.

By the aid of this symbol one can state the normalization and orthogonality relations for probability amplitudes with both discrete and continuous values of the arguments in the single form

$$\overline{\sum_{q'}}\Phi_{\alpha,p}(\beta',q')\Phi^{*}_{\alpha,p}(\beta'',q')=\Delta(\beta'-\beta''). \tag{159}$$

With these preliminary ideas in mind, we proceed to state Jordan's fundamental assumptions in the following manner:

(I) *In the case of any two systems of coordinates* β, q *it is possible to define conjugate systems of momenta* α, p *such that there exists a probability amplitude*[75]

$$\Phi_{\alpha,p}(\beta',q')$$

having the property that if we perform the operation $\overline{\sum}_{\beta'}$ *on the product* $\Phi_{\alpha p}(\beta', q')$ $\Phi_{\alpha p}{}^*(\beta', q')$ *over any region in* β*-space we obtain the relative or absolute probability of the corresponding configurations of the system* (β) *for the given values of the coordinates* q.

(II) *If we have the probability amplitudes for two coordinate systems with respect to a third we can determine the probability of the two with respect to one another by the rule*

$$\Phi_{\alpha,P}(\beta',O')=\overline{\sum_{q'}}\Phi_{\alpha,p}(\beta',q')\Phi_{p,P}(q',Q'). \tag{160}$$

This rule is exemplified in Eq. (155) above and in Eqs. (42), (43), and (87) of Part I.

(III) *The probability amplitude of a system of coordinates* β *with respect to the conjugate momenta* α *is*[76]

---

[73] Cf. E. H. Kennard, Zeits. f. Physik **44**, 326 (1927).

[74] Cf. Sect. 2, 7 of Part I.

[75] The probability amplitude connected with such a pair of coordinate systems must, of course, be independent of the particular problem under discussion; i.e., it must depend only on the relations existing between the two systems of coordinates.

[76] This rule leads to the identification of $\Phi_{\alpha p}(\beta', q')$ with the wave function which we would have indicated by the symbol $\psi^*(\beta)$ in our old notation.

$$\Phi_{\alpha,-\beta}(\beta',\alpha')=\text{const}\times\exp\left[\left(\frac{2\pi i}{h}\right)\sum_{k=1}^{f}\beta_k'\alpha_k'\right]. \tag{161}$$

This assumption re-states for the more general case the conclusion reached in Sect. 1 for an individual particle, that the momenta corresponding to a Cartesian coordinate system play the part of wave-number components for the waves in coordinate space. It is almost, but not quite, equivalent to the assumption of Eq. (I, 123), since it is applicable in case either β or α takes on a discrete range of values, while Eq. (I, 123) loses its meaning if the coordinate q has only discrete values. It can be regarded as a convenient definition of the relation existing between a coordinate and its canonically conjugate momentum. The subscripts on the symbol $\Phi_{\alpha,-\beta}(\beta',\alpha')$ indicate that if the α's are momenta conjugate to the β's, conversely if we treat the α's as coordinates then the β's with their signs changed become the new momenta. This reciprocal relationship between coordinates and conjugate momenta is a characteristic of the classical theory.

(IV) *If $\Phi_{\alpha p}(\beta', q')$ is the probability amplitude for the coordinate system β with respect to the system q, the corresponding amplitudes for the system q with respect to the system β is $\Phi^*_{\alpha p}(\beta', q')$. Thus*

$$\Phi_{p,\alpha}(q',\beta')=\Phi^*_{\alpha,p}(\beta',q'). \tag{162}$$

For the case in which the q's are canonically conjugate to the β's this rule is merely a corollary on (III).

(V) *If two canonical systems of coordinates β, α and $\bar{\beta}$, $\bar{\alpha}$ have a common coordinate $\beta_1=\bar{\beta}_1$, the amplitude $\Phi_{\alpha\bar{\alpha}}(\beta',\bar{\beta}')$ contains the factor $\Delta(\beta_1'-\bar{\beta}_1')$. Thus*

$$\Phi_{\alpha,\bar{\alpha}}(\beta',\bar{\beta}')=\Delta(\beta_1'-\bar{\beta}_1')\psi(\beta',\bar{\beta}') \text{ if } \beta_1=\bar{\beta}_1. \tag{163}$$

This axiom is probably derivable from the others.

These probability amplitudes have certain of the properties of matrices, viz., they are functions of two sets of variables which are equal in number, and furthermore the mode of computing the value of any of these functions for a particular set of values of the variables from related probability amplitudes (Cf. Eq. (160)) is formally similar to the rule for the multiplication of matrices (Cf. Eq. (69)).[77] On this account Jordan introduces the symbol $\Phi_{\alpha p}{}^{\beta q}$ for the whole system of values $\Phi_{\alpha p}(\beta', q')$ and using matrix notation replaces Eq. (160) by

$$\Phi_{\alpha P}{}^{\beta Q}=\Phi_{\alpha p}{}^{\beta q}\Phi_{pP}{}^{qQ}. \tag{165}$$

Eq. (165) is related to (160) just as the ordinary matrix equation

$$\boldsymbol{C}=\boldsymbol{AB}$$

is related to the equation

$$C(m,n)=\sum_k A(m,k)B(k,n).$$

---

[77] The probability amplitudes differ, however, from the matrices of Sect. 6 in that their rows and columns refer in general to different things (β's and q's respectively) instead of to the same set of variables. Jordan calls them "matrices of the first kind."

Equation (162) now takes the form

$$\Phi_{p\alpha}{}^{q\beta}=\Phi_{\alpha p}{}^{\beta q*}. \tag{166}$$

From Eq. (165) we deduce that

$$\Phi_{\alpha\alpha}{}^{\beta\beta}\Phi_{\alpha p}{}^{\beta q}=\Phi_{\alpha p}{}^{\beta q} \tag{167}$$

which means that as $\Phi_{\alpha p}{}^{\beta q}$ is arbitrary, $\Phi_{\alpha\alpha}{}^{\beta\beta}$ must play the role of an unit matrix. Explicitly written out this becomes

$$\Phi_{\alpha,\alpha}(\beta',\beta'')=\Delta(\beta'-\beta'').$$

The probability that $\beta=\beta'$, when β is by hypothesis equal to β'', is zero unless $\beta'=\beta''$. Furthermore it follows that

$$\Phi_{\alpha p}{}^{\beta q}\Phi_{p\alpha}{}^{q\beta}=\Phi_{p\alpha}{}^{q\beta}\Phi_{\alpha p}{}^{\beta q}=1. \tag{168}$$

This proves the orthogonality and the completeness[78] of the set of probability amplitudes, for explicitly written out it yields

$$\overline{\sum_{q'}}\Phi_{\alpha,p}(\beta',q')\Phi_{\alpha,p}{}^{*}(\beta'',q')=\Delta(\beta'-\beta''), \tag{169}$$

and

$$\overline{\sum_{\beta'}}\Phi_{\alpha,p}(\beta',q')\Phi_{\alpha,p}{}^{*}(\beta',q'')=\Delta(q'-q''). \tag{170}$$

In addition to the probability amplitudes, the development of the theory requires the introduction of linear operators for the various dynamical variables. The differential operators used hitherto are unsatisfactory, however, since they cannot be applied to functions in which the independent variables take on discrete values. Hence Jordan invents integral operators to take their place.

The fundamental characteristic of the operators previously employed is indicated by the much-used equation

$$O_\alpha\left(q,\frac{\partial}{\partial q}\right)\psi_\alpha(q)=\alpha\psi_\alpha(q) \tag{171}$$

whose solution is to give a probability amplitude for the q's corresponding to a fixed value of α. In terms of our present notation we may require of the operator for a dynamical variable β_1 that it shall have the property

$$O_{\beta_1}\Phi_{p\alpha}(q',\beta'')=\beta_1''\Phi_{p\alpha}(q',\beta''). \tag{172}$$

An equally satisfactory operator is one which transforms a function of q'' into a function of q' according to the rule

$$O_{\beta_1}\Phi_{p\alpha}(q'',\beta'')=\beta_1''\Phi_{p\alpha}(q',\beta''). \tag{173}$$

---

[78] Cf. Courant-Hilbert, Chap. II, Article 1, 3 Eq. (9).

An integral operator with this characteristic is[79]

$$O_{\beta_1} \equiv {}_{q''}\beta_{1_{q'}} = \overline{\sum_{\beta'}}\ \overline{\sum_{q''}} \Phi_{\alpha p}{}^*(\beta', q')\beta_1{}'\Phi_{\alpha p}(\beta', q'') \cdots \tag{174}$$

as the reader can easily verify. (The notation ${}_{q''}\beta_{1_{q'}}$ is used to indicate that the operator carries a function of q'' into a function of q'.)

In special cases this operator reduces to the corresponding operator of Sect. 7, 1. Suppose, for example, that we identify β_1 in Eq. (174) with q_1. Then we obtain the operator for q_1 with respect to the probability amplitudes of the system of coordinates to which q_1 belongs. Eq. (174) reduces to

$$\left.\begin{aligned} {}_{q''}q_{1_{q'}} &= \overline{\sum_{q'''}}\ \overline{\sum_{q''}} \Phi_{pp}{}^*(q''', q')q_1{}'''\Phi_{pp}(q''', q'') \cdots \\ &= \overline{\sum_{q'''}}\ \overline{\sum_{q''}} \Delta(q'''-q')q_1{}'''\Delta(q'''-q'') \cdots \end{aligned}\right\} \tag{175}$$

When applied to any probability amplitude with argument q'', say $\Phi_{p\alpha}(q'', \beta')$ this changes q'' into q' and multiplies by q'. Hence it is equivalent to the operator Q of Sect. 7, 1.

Consider next the operator for one of the momenta p_1 with respect to the same coordinates

$$\begin{aligned} {}_{q''}p_{1_{q'}} &= \overline{\sum_{p'}}\ \overline{\sum_{q''}} \overset{*}{\Phi}_{-q,p}(p', q')p_1{}'\Phi_{-qp}(p', q'') \cdots \\ &= \overline{\sum_{p'}}\ \overline{\sum_{q''}} e^{(2\pi i/h)p_1{}'(q_1{}'-q_1{}'')}p_1{}'\Delta(q_2{}'-q_2{}'') \cdots \Delta(q_f{}'-q_f{}'') \cdots . \end{aligned} \tag{176}$$

Evidently the operator merely substitutes $q_k{}'$ for $q_k{}''$ for every value of the subscript k except $k=1$. In case q_1 and p_1 both vary continuously it can be proved from Dirichlet's integral that

$${}_{q''}p_{1q'}\Phi_{p,\alpha}(q'', \beta') = \text{const.} \times \left(\frac{h}{2\pi i}\right)\left[\frac{\partial\Phi_{p,\alpha}(q''; \beta')}{\partial q_1{}''}\right]_{q''=q'}. \tag{177}$$

In this case also we see that the integral operator is equivalent to the corresponding differential operator P of Sect. 7, 1.

Consider next the matrices of the type used in Sect. 6 formed from the operator ${}_{q''}\beta_{1_{q'}}$ and the probability amplitudes $\Phi_{p,\tau}(q'', \mu'')$ and $\Phi_{p,\tau}(q', \mu')$. In analogy with Eq. (68) we define the element (μ', μ'') of this matrix as*

$$\widetilde{B}_1(\mu', \mu'') = \overline{\sum_{q'}} \Phi_{p,\tau}{}^*(q', \mu')\,{}_{q''}\beta_{1q'}\Phi_{p,\tau}(q'', \mu''). \tag{178}$$

---

[79] We here define the operator by means of the probability amplitude where in other cases we have defined the probability amplitude in terms of the operator. This procedure is a consequence of the fact that probability amplitudes for canonically conjugate quantities are treated as fundamental in the present theory.

* The use of the wavy line in $\widetilde{B}_1(\mu', \mu'')$ does not refer to a transposition of the indices μ' and μ'' as in Sect. 6, 1 but is used merely to distinguish the matrices of Eqs. (178) and (179) from those of (180).

Hence

$$\tilde{B}_1(\mu',\mu'') = \overline{\sum_{\beta'}} \Phi_{\alpha\tau}{}^*(\beta',\mu')\beta_1{}'\Phi_{\alpha\tau}(\beta',\mu''). \tag{179}$$

The second of the above expressions for $\tilde{B}_1(\mu',\mu'')$ becomes the analog of Eq. (I, 93) if we identify μ' and μ''. Thus this very general type of matrix formula contains within itself both the matrix elements of Sect. 6 and the mean value formulas of Sect. 3.

Finally, if we form the matrix $B_1(q',q'')$ of ${}_{q''}\beta_{1_{q'}}$ with respect to the coordinate system q, p we obtain

$$B_1(q',q'') = \overline{\sum_{\beta'}} \Phi_{\alpha p}{}^*(\beta',q')\beta_1{}'\Phi_{\alpha p}(\beta',q''). \tag{180}$$

Comparing with Eq. (174) we see that

$${}_{q''}\beta_{1q'} = \overline{\sum_{q''}} B_1(q',q'') \cdots. \tag{181}$$

Thus the operator of β_1 with respect to the coordinate system q, p is equivalent to a linear transformation whose matrix is $B_1(q',q'')$.

Matrices of the type $\tilde{B}_1(\mu',\mu'')$ and $B_1(q',q'')$ are called "*matrices of the second kind.*" If we introduce the matrix

$$\beta_1{}^{\beta\beta} = \beta_1{}'\Delta(\beta'-\beta'') \tag{182}$$

we can express Eq. (180) in the form

$$B_1 = (\Phi_{\alpha p}{}^{\beta q})^{-1}\beta_1{}^{\beta\beta}(\Phi_{\alpha p}{}^{\beta q}). \tag{183}$$

We call $\beta_1{}^{\beta\beta}$ the matrix of β_1 with respect to a system of coordinates of which it is itself a member, while B_1 is the matrix of β_1 with respect to the coordinate system q, p. This means, for example, that the probability amplitude $\psi_{nlm}(x,' y, z)$ in the hydrogen atom problem can be interpreted as the matrix of a transformation which carries the matrix of x (or y or z) with respect to an x, y, z coordinate system over into the matrix of x with respect to the "coordinate system" n, l, m.[80] Thus (for $\beta_1 = x$) Eq. (183) goes over into the form

$$X(n',l',m';n'',l'',m'') = \int_\infty \overset{*}{\psi}_{n''l''m''}(x,y,z)\,x\,\psi_{n'l'm'}(x,y,z)\,dxdydz. \tag{184}$$

If the matrix B_1 in Eq. (183) is known, we can use the equation to determine the probability amplitude $\Phi_{\alpha p}{}^{\beta q}$. Multiplying both sides in front by $\Phi_{\alpha p}{}^{\beta q}$ and writing out the explicit expression for the matrix element we obtain

---

[80] This use of the quantum numbers as a "coordinate system" is due to Dirac (ref. 67). More correctly we should say that the coordinate system consists of the set of dynamical variables determined by the quantum numbers; e.g., the total angular momentum, etc.

$$\overline{\sum_{q''}}\Phi_{\alpha p}(\beta',q'')B_k(q'',q')=\overline{\sum_{\beta''}}\beta_k'\Delta(\beta'-\beta'')\Phi_{\alpha p}(\beta'',q')$$
$$=\beta_k'\Phi_{\alpha p}(\beta',q'). \tag{185}$$

If the coordinates q have discrete spectra only, Eq. (185) is a homogeneous linear equation with the elements of the matrix $\Phi_{\alpha p}{}^{\beta q}$ as unknowns. If we allow q'' to take on all possible values we obtain an infinite set of simultaneous linear equations of which Eq. (I, 135) may be regarded as a special case where β is identical with the energy.

On the other hand, the above equation can be reduced, under suitable circumstances, to differential form. As B_k is Hermitian,

$$\overline{\sum_{q''}}B_k(q',q'')\Phi_{p\alpha}(q'',\beta')=\beta_k'\Phi_{p\alpha}(q',\beta'). \tag{186}$$

Define ${}_{q'}B_{k_{q'}}$ as the operator

$$\overline{\sum_{q''}}\Delta(q'-q'')\ \overline{\sum_{q'}}B_k(q'',q')\cdots.$$

${}_{q'}B_{k_{q'}}$ is equivalent to ${}_{q''}\beta_{k_{q'}}$ or to $\overline{\sum_{q''}}B_k(q', q'')$ except that it converts a function of q' into a function of q'. This reduces Eq. (186) to the form

$$\{{}_{q'}B_{k_{q'}}-\beta_k'\}\Phi_{p\alpha}(q',\beta')=0. \tag{187}$$

Finally, if the q coordinate system is continuous, so that

$${}_{q'}p_{k_{q'}}=\frac{h}{2\pi i}\,\frac{\partial}{\partial q_k'};\quad {}_{q'}q_{k_{q'}}=[q_k'\times]$$

and if ${}_{q'}B_{k_{q'}}=g_k({}_{q'}q_{k_{q'}};\ {}_{q'}p_{k_{q'}})$, we have

$$\left\{g_k\left(q',\frac{h}{2\pi i}\,\frac{\partial}{\partial q'}\right)-\beta_k'\right\}\Phi_{p\alpha}(q',\beta')=0. \tag{188}$$

This is paralleled by

$$\left\{f_k\left(q',\ \frac{h}{2\pi i}\,\frac{\partial}{\partial q'}\right)+\frac{h}{2\pi i}\,\frac{\partial}{\partial \beta_k'}\right\}\Phi_{p\alpha}(q',\beta')=0 \tag{188'}$$

in case β_k has a continuous range of values. Here, $f_k(q,\ p)=\alpha_k$. Suitable specialization of Eq. (188′) leads to the differential equation for Q referred to on p. 173 of Part I. If β_k is the energy and $g_k(q,\ p)$ is the Hamiltonian, (188) is the Schrödinger equation without the time. In general it is identical with Eqs. (I, 124) and (171).

The wave functions in the second Shcrödinger equation are probability amplitudes associated with the transformation from a fixed coordinate system ξ to a variable coordinate system q_t. The problem of determining $\psi(q,\ t)$

from the initial conditions as defined by $\psi(q, 0)$ is the problem of passing from $\Phi(q_0, \xi)$ to $\Phi(q_t, \xi)$, which is solved by use of the formula

$$\Phi(q_t, \xi) = \overline{\sum_{q_0}} \Phi(q_t, q_0)\Phi(q_0, \xi)$$

if $\Phi(q_t, q_0)$ can be found. Kennard[73] has discussed the problem in detail and has proved that if $\Phi(q_t, \xi)$ is formed in this way it is then a solution of an equation which differs from Eq. (I, 38) only through a term which is of no physical importance.

Eqs. (185), (186), and (187) can be paralleled by corresponding equations in which the matrix of β_k (*i.e.*, B_k) is replaced by the matrix of the momentum α_k conjugate to β_k. At the same time we have to replace the factor β_k' on the right by ${}_{\beta'}\alpha_{k_{\beta'}}$ [Cf. Eq. (178)]. Then we have, for example

$$\{ {}_{q'}A_{k_{q'}} - {}_{\beta'}\alpha_{k_{\beta'}}{}^* \} \Phi_{p\alpha}(q', \beta') = 0. \tag{187a}$$

This equation, regarded as a functional equation for the probability amplitude $\Phi_{p\alpha}(q', \beta')$ has the same solutions as Eq. (187). The equivalence of the solutions of the two equations is an expression of the fact that β_k and α_k are canonically conjugate quantities. If we now, starting with an arbitrary set of canonically conjugate coordinates and momenta (q, p) having continuous spectra, form arbitrary operator functions[81]

$$\left.\begin{aligned} {}_{q'}B_{k_{q'}} &= g_k({}_{q'}q_{k_{q'}}\ ; {}_{q'}p_{k_{q'}}) \\ {}_{q'}A_{k_{q'}} &= f_k({}_{q'}q_{k_{q'}}\ ; {}_{q'}p_{k_{q'}}), \end{aligned}\right\} k = 1, \cdots, f \tag{189}$$

we can set up a corresponding pair of equations of the form of Eq. (187) and (187a).[82] If the spectra for the β's and α's can be defined such that these equations have a simultaneous solution, then the coordinate system (β, α) thus defined is canonical. This rather formidable method of testing a set of proposed transformation equations like (189) reduces to the one previously used if the β spectrum, like the p and q spectra is continuous. Then (187a) goes over into

$$\left\{ f_k\left(q', \frac{h}{2\pi i}\frac{\partial}{\partial q'}\right) + \frac{h}{2\pi i}\frac{\partial}{\partial \beta_k'} \right\} \Phi_{p\alpha}(q', \beta') = 0. \tag{188a}$$

and the condition for the identity of the solutions of (188) and (188a) becomes[83]

$$f_k g_k - g_k f_k \equiv 1. \tag{190}$$

A canonical transformation can also be set up by means of the matrix equations

$$\bar{q}_k = (\Phi_{p\alpha}{}^{q\beta})^{-1} q_k (\Phi_{p\alpha}{}^{q\beta})$$
$$\bar{p}_k = (\Phi_{p\alpha}{}^{q\beta})^{-1} p_k (\Phi_{p\alpha}{}^{q\beta})$$

---

[81] Cf. Eqs. (137).

[82] The operator ${}_{\beta'}\alpha_k{}^*{}_{\beta'}$ is defined by Eq. (176).

[83] Cf. Jordan I, §5.

where $\Phi_{p\alpha}{}^{q\beta}$ is an arbitrary function of the q's and β's satisfying the conditions (166) and (168). If the transformation is to preserve the Hermitian character of the Hamiltonian matrix (which is equivalent to keeping the Hamiltonian operator self-adjoint), it is necessary that $\Phi_{p\alpha}{}^{q\beta}$ be independent of an interchange in the numerical values of the q's and β's.

For further details and applications of this beautiful theory the reader is referred to the original papers of Dirac and Jordan.[84]

ERRATA

Edwin C. Kemble

Physical Review Supplement 2, October, 1929

Page 174, last line. The right hand member of the equation should be multiplied by dx.

Page 192, Eq. (97) should read

$$\iiint_\infty(\psi_1\psi_2^*+\psi_2\psi_1^*)dx\,dy\,dz=\sum\iiint_\infty(Q_1Q_2^*+Q_2Q_1^*)dp_xdp_ydp_z.$$

Eq. (98) is best derived by means of Eq. (91) of Part II.

Page 193, Eq. (102). The sign of the right hand member should be reversed.

Page 193, line 13 from bottom. Read (94) and (96) for (94) and (95)

Page 203, Eq. (137) should read $\sum_k |c_{kn}|^2=1$.

---

[84] Cf. also D. Hilbert, L. Nordheim, and J. V. Neuman, Math. Ann. **98**, 1 (1927).